Modeling and Optimization of Active Distribution Network Operation Based on Deep Learning

Liao, Wenlong

DOI (link to publication from Publisher):
10.54337/aau539422562

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

MODELING AND OPTIMIZATION OF ACTIVE DISTRIBUTION NETWORK OPERATION BASED ON DEEP LEARNING

BY
WENLONG LIAO

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

MODELING AND OPTIMIZATION OF ACTIVE DISTRIBUTION NETWORK OPERATION BASED ON DEEP LEARNING

by

Wenlong Liao

AALBORG UNIVERSITY
DENMARK

Dissertation submitted March 2023

.

Dissertation submitted: March 2023

PhD supervisor: Professor Birgitte Bak-Jensen,
Aalborg University, Denmark

Assistant PhD supervisor: Associate Professor Jayakrishnan Radhakrishna Pillai
Aalborg University, Denmark

PhD committee: Professor Huai Wang
Department of Energy
Aalborg University, Denmark

Associate Professor Peiyuan Chen
Department of Energy and Environment
Chalmers University of Technology, Sweden

Professor Sami Petteri Repo
Department of Electrical Engineering
Tampere University of Technology, Republic of Finland

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: AAU Energy

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-730-7

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2023

CV

Wenlong Liao was born in Fengshan Village, Rentian Town, Shanghang County, Fujian Province, China.

He received his B.S. degree in Electrical Engineering from the College of Information and Electrical Engineering, China Agricultural University, Beijing, China, in 2017. He received his M.S. degree in Electrical Engineering from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2020. He was a visiting student at the Energy Digitalization Laboratory, The University of Hong Kong, from Sept. 2022 to Feb. 2023.

Currently, he is a PhD candidate at Aalborg University. During his doctoral studies, he aims to bridge the gap between artificial intelligence technologies and power systems. His research topics mainly include data cleaning, point prediction, interval prediction, scenario generation, and risk-based decision-making.

ENGLISH SUMMARY

The stochastic and fluctuating behaviors of power loads and renewable energy sources (RESs) bring risks to the operation and planning of active distribution networks (ADNs), which become more sophisticated and face more uncertainties than ever before. In this context, traditional models have difficulty in fully meeting the analysis and control requirements of ADNs. This Ph.D. project focuses on using deep learning technologies to transform the massive collected data into knowledge, and provide deeper insights into the past, better understandings of the future, and practical suggestions on possible decisions for the economic and secure operation of ADNs. To achieve this objective, this project aims to model and optimize the power loads and RESs in ADNs from multiple perspectives at the data level, the algorithm level, and the decision-making level.

At the data level, an advanced framework called context encoder (CE) is developed to fill in missing values of wind power datasets accurately. The CE is composed of three parts: a discriminator, an encoder, and a decoder. To extract the complicated latent features automatically, the convolutional neural network (CNN) is adopted to construct three parts of the CE. Further, a reconstruction loss function and an adversarial loss function are combined to a hybrid loss function, which enables CE to generate reasonable missing values for incomplete samples. Simulation results show that the CE not only considers correlations between wind powers and other attributes, but also captures the contextual information of wind power generation curves, making it highly adaptable to the missing data imputation with fast ramps. Moreover, CE outperforms popular baselines considering multiple metrics, especially for datasets with high missing ratios.

At the algorithm level, a new hybrid model is proposed for the ultra-short-term point prediction of RESs. In particular, multiple adjacent RESs and weather conditions are modeled as graphs from a new perspective in graph domains. Besides, a graph convolutional network (GCN) is presented to capture complicated spatial features between nodes, and then a long short-term memory (LSTM) is adopted to depict temporal features. Simulation results demonstrate that the proposed hybrid model is capable of capturing temporal and spatial features of multiple adjacent RESs and weather conditions accurately. Moreover, the proposed hybrid model performs better than widely used baselines on real-world datasets.

In addition to the point prediction model, two mainstream methods are proposed to represent the uncertainty of power loads and RESs from different perspectives (i.e., scenario prediction and interval prediction). Firstly, a new generative network called pixel convolutional neural network (PixelCNN) is presented to represent the uncertainty of power loads by creating a group of potential power load scenarios. An optimization model is proposed to filter these power load scenarios with similar

probability distributions, temporal correlations, and shapes as real ones. The proposed PixelCNN does not involve the difficulty in the training process as the existing models. According to numerical simulations on a real-world dataset, the proposed PixelCNN outperforms other popular generative networks, when it comes to the scenario prediction of power loads.

Secondly, a new ultra-short-term interval prediction method is proposed to represent the uncertainty of wind powers based on a point prediction model and an improved bootstrap technique. After training a point prediction model, an improved bootstrap technique is created to narrow prediction intervals (PIs) and boost prediction interval coverage percentages effectively. Simulation results demonstrate that the improved bootstrap technique balances the width and coverage percentage of PIs well.

At the decision-making level, two mainstream risk-based methods are proposed to represent the uncertainty in the decision-making process from different perspectives. Firstly, a robust day-ahead optimal scheduling (RDOS) model of power loads and RESs in ADNs is proposed to minimize the operational cost of a day considering prediction errors. Specifically, point prediction values and PIs are combined to form new kinds of worst-case scenarios, which are used to represent the full range of possible scenarios. In this case, the deterministic day-ahead optimal scheduling (DDOS) model is generalized into the RDOS model considering the operational constraints in two worst-case scenarios. Theoretically, the RDOS model is solved to obtain a robust solution, which guarantees the secure operation of ADNs for all possible scenarios. Simulation results show that the DDOS model-based solution can only guarantee the security of ADNs for the majority of situations, since the DDOS model ignores prediction errors of power loads and RESs. In contrast, the RDOS model takes prediction errors into consideration, and ensures that operational constraints are always met in theory.

Finally, a stochastic day-ahead optimal scheduling (SDOS) model of power loads and RESs in ADNs is proposed to minimize the power loss and voltage deviation by managing various power devices, such as electric vehicle charging stations (EVCSs), transformers, RESs, and heat pumps (HPs), from a probabilistic perspective. Specifically, a number of statistical models are created to describe charging curves of battery electric vehicles (BEVs), which are aggregated in a charging station (i.e., EVCS). Then, the Gaussian distribution model and K-means are employed to represent the uncertainty of power loads and RESs by constructing classical stochastic scenarios. In this case, the DDOS model is generalized into the SDOS model by taking classical stochastic scenarios as inputs. Simulations are carried out on a modified IEEE 33-bus distribution network. The impacts of BEVs and HPs with and without demand response on the voltage and power loss of ADNs are discussed. In comparison to the DDOS model, the SDOS model produces a more complete solution that takes into account both deterministic outputs as well as the potential values and associated probabilities.

DANSK RESUME

Den stokastiske og fluktuerende adfærd hos vedvarende energikilder (RES) og elbelastninger medfører risici for driften og planlægningen af aktive distributionsnet (ADN), som bliver mere sofistikerede og står over for større usikkerhed end nogensinde før. I denne sammenhæng har traditionelle modeller svært ved fuldt ud at opfylde ADN'ernes analyse- og styringskrav. Dette ph.d.-projekt fokuserer på at anvende deep learning-teknologier til at omdanne den store mængde indsamlede data til viden, give dybere indsigt i fortiden og bedre forståelse for fremtiden, samt give praktiske forslag til mulige beslutninger for at sikre en økonomisk og sikker drift af ADN. For at nå dette mål sigter dette projekt mod at modellere og optimere ADN fra flere perspektiver, herunder data, algoritme og beslutningstagning.

På dataniveau foreslås anvendelse af en effektiv metode kaldet kontekstenkoder (CE) til præcist at rekonstruere manglende værdier i vindkraftdatasæt. CE består af tre dele: en diskriminator, en koder og en dekoder. For at finde de komplicerede latente egenskaber automatisk anvendes foldnings neurale netværk (convolution neural networks, CNN) til at konstruere de tre dele af CE. Endvidere danner et rekonstruktionsstab og et modstandsdygtigt tab en hybrid tabsfunktion til at generere rimelige manglende værdier for ufuldstændige prøver. Simuleringsresultater viser, at CE ikke kun tager hensyn til korrelationer mellem vindenergien og andre parametre, men også fanger kontekst til vind produktionskurven, hvilket gør den yderst velegnet til rekonstruktion af manglende data i hurtige dynamiske forløb. Desuden klarer CE sig bedre end populære gældende basale metoder, når der tages hensyn til flere målinger, især for datasæt med høje manglende procentdele.

På algoritmeniveau foreslås en ny hybridmodel til ultrakortfristet forudsigelse af vedvarende produktion. I særdeleshed modelleres flere nærtstående vedvarende energikilder og vejrforhold som knudepunkter i grafer ud fra nye perspektiver i det grafiske domæne. Desuden præsenteres et graffoldnings netværk (GCN) til at indfange komplicerede rumlige egenskaber mellem knudepunkterne, og derefter anvendes en lang korttidshukommelse (long short term memory, LSTM) til at skildre tidsmæssige egenskaber. Simuleringsresultater viser, at den foreslåede hybridmodel er i stand til nøjagtigt at indfange tids- og stedsmæssige egenskaber ved flere tilstødende vedvarende energikilder og vejrforhold, og at den klarer sig bedre end de almindeligt anvendte metoder på datasæt fra den virkelige verden.

Ud over forudsigelsesmodellen foreslås to metoder til at repræsentere usikkerheden ved de vedvarende energikilder og elbelastningerne set fra forskellige perspektiver (dvs. scenarieforudsigelse og intervalforudsigelse). Først præsenteres et nyt generativt netværk kaldet pixel foldnings neural network (PixelCNNN) til at repræsentere usikkerheden i forbindelse med elbelastninger ved at skabe en gruppe af potentielle elbelastningsscenarier. Der foreslås en optimeringsmodel til at finde en

gruppe af potentielle strømbelastningsscenarier med lignende sandsynlighedsfordelinger, tidsmæssige korrelationer og former som de virkelige scenarier. Den foreslåede PixelCNN er ikke så vanskelig i træningsprocessen som de eksisterende modeller. Ifølge numeriske simuleringer på et datasæt fra den virkelige verden slår det foreslåede PixelCNNN andre populære generative netværk, når det drejer sig om forudsigelse af scenarier for elbelastninger.

Dernæst foreslås en ny ultrakorttids intervalprædiktionsmetode til at repræsentere usikkerheden ved vindkraft baseret på en prædiktionsmodel og en forbedret bootstrap-teknik. Efter træning af prædiktionsmodellen oprettes en forbedret bootstrap-teknik til at indsnævre forudsigelsesintervaller og øge dækningsgraden effektivt. Simuleringsresultater viser, at den foreslåede forbedrede bootstrap-teknik afbalancerer bredden og dækningsgraden af forudsigelsesintervallerne godt.

På beslutningsniveauet foreslås to risikobaserede metoder til at repræsentere usikkerheden i beslutningsprocessen ud fra forskellige perspektiver. På den ene side foreslås en robust optimal skeduleringmodel for næste dag (robust day-ahead optimal scheduling, RDOS-model) for aktive net for at opnå optimale løsninger under hensyntagen til forudsigelsesfejl i forbindelse med vedvarende energikilder og elbelastninger. Derefter kombineres forudsigelsesværdierne og forudsigelsesintervallerne for at danne to worst-case-scenarier, som bruges til at repræsentere hele spektret af mulige scenarier. I dette tilfælde generaliseres den deterministiske DDOS-model (day-ahead optimal scheduling) til RDOS-modellen under hensyntagen til de operationelle begrænsninger i to worst-case-scenarier. Endelig løses RDOS-modellen for de aktive distributionsnet' for at opnå en robust løsning, som teoretisk set garanterer sikker drift af nettene for alle mulige scenarier. Simuleringsresultaterne viser, at den DDOS-modelbaserede løsning kun kan garantere de aktive distributionsnets sikkerhed i de fleste situationer, da DDOS-modellen ikke tager hensyn til forudsigelsesfejl i forbindelse med elbelastninger og vedvarende energikilder. RDOS-modellen tager derimod hensyn til forudsigelsesfejl i forudsigelsesmodellen og sikrer, at spændingen ved hvert knudepunkt falder inden for det teoretisk tilladte område.

Endvidere foreslås en stokastisk day-ahead optimal scheduling (SDOS) model af belastninger og vedvarende energikilder de aktive distributionsnet for at minimere effekttabet og spændingsafvigelsen ved at styre forskellige belastninger, såsom ladestationer til elektriske biler (EVCS'er), transformere, vedvarende energikiler og varmepumper (HP'er), ud fra et sandsynlighedsperspektiv. Specifikt oprettes en række statistiske modeller til at beskrive ladekurver for elbilerne (BEV'er), som er aggregeret til en ladestastion (dvs. EVCS). Derefter anvendes en Gaussisk distributionsmodel og K-meansr til at repræsentere usikkerheden af belastningerne og de vedvarende energikilder ved at konstruere klassiske stokastiske scenarier. I dette tilfælde generaliseres DDOS-modellen til SDOS-modellen ved at tage klassiske stokastiske scenarier som input. Simuleringer udføres på et modificeret

IEEE 33-bus distributionsnetværk. Virkningerne af BEV'er og HP'er med og uden fleksibelt elforbrug på spænding og effekttab i de aktive net diskuteres. I sammenligning med DDOS-modellen producerer SDOS-modellen en mere komplet løsning, der tager hensyn til både deterministiske output samt de potentielle værdier og tilhørende sandsynligheder.

ACKNOWLEDGEMENTS

The Ph.D. project is run from Oct. 2020 to Jun. 2023. I would like to appreciate the funders of this Ph.D. project, including the Otto Monsted Foundation in Denmark, Aalborg University, and the China Scholarship Council.

I must give my deepest thank you to my supervisors, including Prof. Dr. Birgitte Bak-Jensen and Assoc. Prof. Dr. Jayakrishnan Radhakrishna Pillai. Although they are usually very busy, they still timely and carefully revise my papers from the whole framework to every sentence. They often encourage me in my daily life, and share their experiences and new ideas in my research. In addition, they are very supportive of my decisions. In particular, they know that I want to continue my career in academia. They give me a lot of valuable advice. Moreover, I am honored to be a member of the Intelligent Energy Systems and Flexible Markets Group led by Assoc. Prof. Dr. Amjad Mogahaddam. Thanks to every member of the Group for their help and support.

I would like to thank Assistant Professor Yi Wang and other members of the Energy Digitalization Laboratory at The University of Hong Kong. He provides support in my study and life during my visit to the University of Hong Kong. Especially, he gives a lot of advice about the postdoc. He often discusses research directions, research ideas, and research insights with me. The research interests of each member of EDL revolve around artificial intelligence and big data technology, and their many insights often inspire me.

I would like to thank my previous supervisors during my undergraduate and master studies, including Professor Shouxiang Wang at Tianjin University and Associate Professor Dechang Yang at China Agricultural University. Without their support and advice, I would not have chosen to go abroad for my Ph.D. studies.

I would like to thank the evaluation committee for attending my Ph.D. defense. In addition, I would like to thank my colleagues for supporting my study and life at AAU Energy, Aalborg University.

Lastly, I want to thank my family, especially my mom, my dad, and my brother-in-law Xingyuan Li. Also, thanks to my neighbors, upperclassmen, and roommates: Zhe Yang, Kuangpu Liu, Sihan Liu, Xinxin Chen, Bin Zhang, Yaqi Li, Siyu Jin, Yun Yu, and Hanchi Zhang. In addition, I would like to thank my other far-away friends Wenqing Yang, Yining Chen, Xiangxiang Dong, Yusen Wang, Xiang Ren, Yuelong Wang for occasionally discussing with me.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADN	active distribution network
AE	automatic encoder
AI	artificial intelligence
AMI	advanced metering infrastructure
BEV	battery electric vehicle
Bi-LSTM	bi-directional long short-term memory
CB	capacitor bank
CE	context encoder
CGAN	conditional generative adversarial network
CNN	convolutional neural network
Conv2D	2-D convolutional
Conv2DTran	2-D transposed convolutional
CWC	coverage width criterion
DDOS	deterministic day-ahead optimal scheduling
DL	deep learning
DNN	deep neural network
EVCS	electric vehicle charging station
GA	genetic algorithm
GAN	generative adversarial network
GCN	graph convolutional network
GRU	gated recurrent unit
HP	heat pump
KNN	k-nearest neighbor
LightGBM	light gradient boosting machine
LSTM	long short-term memory
LUBE	lower upper bound estimation
MAE	mean absolute error

Maxpool	maximum pooling
MLP	multi-layer perceptron
MMAE	maximum absolute error
NICE	non-linear independent component estimation
NWP	numerical weather prediction
PDF	probability distribution function
PI	prediction interval
PICP	prediction interval coverage percentage
PINAW	prediction interval normalized average width
PINC	prediction interval nominal confidence
PixelCNN	pixel convolutional neural network
PV	photovoltaic
RDOS	robust day-ahead optimal scheduling
RES	renewable energy source
RF	random forest
RMSE	root mean squared error
RNN	recurrent neural network
SCADA	supervisory control and data acquisition
SDOS	stochastic day-ahead optimal scheduling
SVM	support vector machine
V2G	vehicle-to-grid
VAE	variational auto-encoder
WT	wind turbine

LIST OF FIGURES

LIST OF TABLES

PART I REPORT

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

The global energy crisis and environmental problems necessitate the development of renewable energy sources (RESs). However, the stochastic and fluctuating behaviors of the renewable generation bring risks to the operation and planning of active distribution networks (ADNs) [1], which become more sophisticated and face more uncertainties than ever before. In this context, traditional models have difficulty in fully meeting the analysis and control requirements of ADNs [2]. For example, since traditional statistical models have difficulties in capturing the spatial and temporal features of RESs, they have limited prediction accuracy (i.e., prediction errors are relatively large) [3], which may lead to the voltage or current constraints being crossed in ADNs.

As one of the artificial intelligence (AI) technologies, deep learning (DL) has demonstrated exceptional performance in various fields [4], since it can automatically learn latent features from high-dimensional data without making assumptions or simplifications to the physical model of the system. The strong learning and representing capabilities of DL technologies bring new opportunities to address gaps in existing models for analysis and control requirements of ADNs.

In this context, the general goal of the project is to model and optimize the power loads and RESs accurately and efficiently to improve the economy and security of ADNs, especially from the data-driven perspective. Accordingly, the project aims at using the DL to transfer the massive collected data into the knowledge, and provide deeper insights into the past, better understandings of the future, and practical advice on possible decisions for the safe and economic operation of ADNs. To achieve this objective, the project will be conducted at the data level, algorithm level, and decision-making level:

At the data level, the missing values in datasets (e.g., wind farm datasets) are filled in precisely. This is the basis of the algorithm level and the decision-making level.

At the algorithm level, the energy predictions (e.g., power loads and output powers of RESs) including the deterministic point prediction and uncertainty prediction (e.g., interval prediction and scenario prediction), are executed to keep energy supply and load demand in balance. It provides inputs and boundary conditions for the decision-making at the next level.

At the decision-making level, the risk-based models (e.g., robust programming and stochastic programming) are formulated to obtain the optimal day-ahead scheduling of power loads and RESs in ADNs.

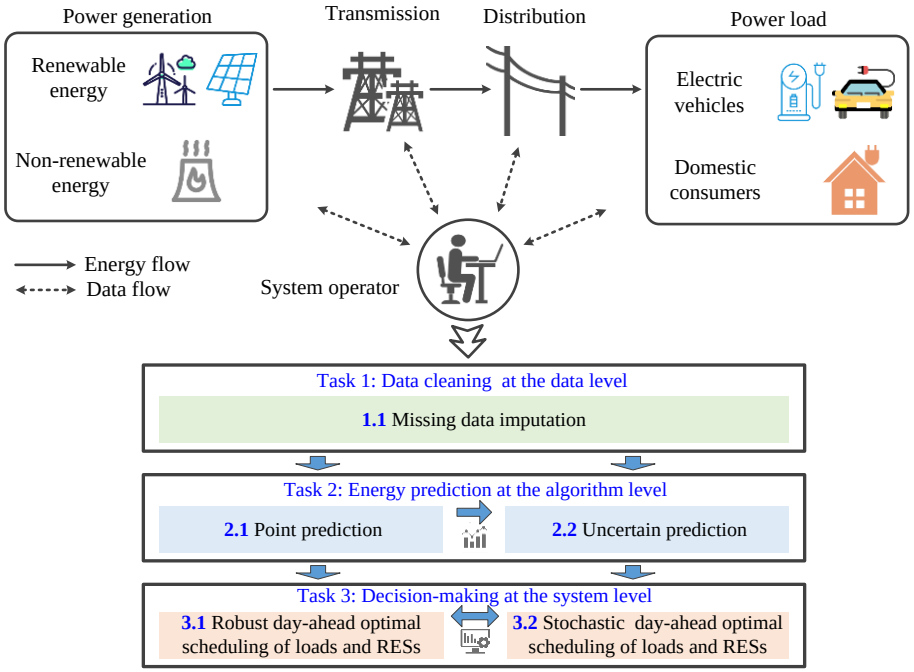


Figure 1-1 The project's structure, interrelationship, structure, and interdependencies.

The project is dedicated to addressing the above key themes, which consist of three tasks. The interrelationships among these tasks are shown in Figure 1-1. Each task consists of several sub-tasks, and each sub-task interacts with all the others. Specifically, task 1 belongs to data cleaning, which is a prerequisite for task 2.1. Task 2.2 is an extension of task 2.1 considering prediction errors. To obtain the optimal scheduling of power loads and RESs in ADNs considering risks, task 3.1 and task 3.2 should consider output powers and boundary conditions of power loads and RESs from Task 2.

1.2. LITERATURE REVIEW

1.2.1. LITERATURE REVIEW ON MISSING DATA IMPUTATION

The stochastic and fluctuating behaviors of power loads and RESs bring risks to the operation and planning of ADNs [5]. To ensure the safety and economy of ADNs, accurate point predictions of power loads and RESs should be conducted by using historical data from the supervisory control and data acquisition (SCADA) system and advanced metering infrastructure (AMI) to train an advanced model. However, given that the SCADA system and AMI are frequently disrupted by a variety of circumstances, such as cyber-attack, sensor failure, and communication congestion [6], such historical data may be incomplete. Therefore, missing data imputations of

power loads and RESs play a crucial part in the prediction tasks. The section mainly reviews the related works on missing data imputations for wind power, while missing data imputations for the photovoltaic (PV) power and power load can be treated in a similar way.

The traditional methods of missing data imputations fall into four main categories [7]: interpolation, parameter estimation, similarity-based method, and regression.

The piecewise constant interpolation, cubic interpolation, mimetic interpolation, and linear interpolation are the key components of the first category [8]. After building polynomials where parameters are acquired by utilizing the data surrounding the missing position, this kind of method fills in missing values easily. When the amount of missing data is relatively little, high accuracy can be achieved. Nevertheless, they neglect the correlation among various attributes, which limits the range of applications for these methods [9].

Interval estimation and point estimation make up the majority of the second category [10], [11]. Specifically, maximum likelihood estimation is a key procedure in this type of method. For instance, to achieve the purpose of missing data imputation, the expectation-maximization employs available data to estimate the missing values, and then uses complete data generated in the previous step to determine the model parameters in an iterative way [12]. However, the artificially assumed form of the probability distribution function (PDF) has a significant influence on the accuracy [13].

The third category mainly consists of the K-means, k-nearest neighbor (KNN), and mean imputation [14]. In particular, the mean imputation is an easy and commonly used method, which covers missing positions with the mean value of the non-missing part. However, it restricts attribute diversity and volatility, leading to low accuracies [15]. K-means imputes the missing values with the cluster centers of the classes to which the samples belong, while KNN estimates missing data points with the average values of the nearest multiple samples. The similarity-based models, like interpolations, also neglect the relationship among various attributes, which reduces their accuracy [16].

The fourth category mainly consists of multi-layer perceptrons (MLPs), random forests (RFs), and recurrent neural networks (RNNs) [17], [18]. For instance, an improved RF is presented to ensemble matrix transposition, linear interpolation, and matrix combination to impute the large-scale missing values of wind powers in [19]. The work in [20] generalizes the traditional RNN into a long short-term memory (LSTM) to predict the missing values of wind power curves. Regression-based methods have a wider application scope and a higher accuracy than interpolation-based methods, since they can fully exploit the correlation among attributes.

Recently, more and more publications have focused on the application of deep generative networks for missing data imputation of power loads and RESs [21]. The widely used methods mainly include the conditional generative adversarial network (CGAN) and automatic encoder (AE). To capture meaningful information from samples and minimize the damage, AE corrupts the inputs and then forces the decoder to restore them, but its corruption process is often quite localized and low-level [22]. The problem of unstable training in CGANs, such as diminished gradient, non-convergence, and mode collapse, remains a challenge in the previous works [23], [24], resulting in low accuracies of filling values.

In general, existing publications mainly include the following limitations: 1) Most of the publications have difficulties in capturing the contextual information of missing values and complex non-linear relationships among multiple attributes. 2) Many of the publications (e.g., similarity-based methods) have low accuracy of filling values.

1.2.2. LITERATURE REVIEW ON POINT PREDICTION

To preserve the balance between the energy supply and load demand, the development of point prediction models with high prediction accuracy for power loads and RESs is critical in the operation and planning of ADNs. In the following chapters, the risk-based decision-making is limited to the day-ahead scheduling of power loads and RESs in ADNs, so this project only focuses on the point prediction from the next 30 minutes into the next 24 hours, which belongs to the ultra-short-term prediction [25], [26]. Generally, the existing works of the ultra-short-term point prediction can be divided into three main categories: physical methods, statistical methods, and AI-based methods.

The physical methods usually require the surrounding environment (e.g., terrain, surface, and obstacle) and numerical weather prediction (NWP) data (e.g., temperature, wind speed, light intensity, pressure, and humidity from meteorological services) to model the relationship between inputs and outputs. Despite the advantages of the interpretability of the physical method, the NWP data and the numerous physical parameters impose a heavy computational burden [27]. Besides, the prediction accuracy of the physical method is sensitive to meteorological parameters. In other words, when a meteorological parameter is incorrect, it is possible for physical model mistakes to compound, reducing prediction accuracy significantly.

Popular statistical methods mainly consist of exponential smoothing models, auto-regressive integrated moving average models, gray models, auto-regressive models, and their hybrid models [28]. Statistical methods are less expensive than physical methods, since they do not require additional costly NWP simulations outside historical data. However, their prediction accuracies are limited, especially for the prediction task of RESs with strong stochastic and fluctuating behaviors [29]. The

reason is that statistical methods rely on historical data to predict future trends, and cannot take into account the correlations between outputs and weather conditions.

The traditional AI-based methods mainly consist of MLP, support vector machine (SVM), RF, and light gradient boosting machine (LightGBM) [30], [31]. These traditional AI-based methods are more cost-effective than physical methods, but they have trouble effectively depicting the spatio-temporal features from high-dimensional historical data. To address this issue, various deep neural networks (DNNs) have been developed recently. For example, the famous gated recurrent unit (GRU) and LSTM [32] have shown incredible performance in depicting the temporal feature from time series, significantly boosting the accuracy of ultra-short-term point predictions.

In fact, the input data of ultra-short-term point predictions for power loads and RESs should be viewed as a graph in a broad sense [33]. For example, adjacent PV systems are regarded as nodes, and the correlation between them is represented by the adjacency matrix. Further, the historical PV powers form feature vectors of nodes, which can be used to represent temporal features of PV powers [34]. However, these methods mentioned-above are defined in the Euclidean domain, and they cannot handle the graphs. Normally, these methods defined in the Euclidean domain chose to simplify graphs into Euclidean data by neglecting the adjacency matrix and only keeping feature vectors. This simplification operation is detrimental to boosting the accuracy of the ultra-short-term point prediction, because it loses the spatial features in adjacency matrices [35].

In general, existing publications mainly include the following limitations: 1) Most of the publications have difficulties in capturing spatio-temporal features from inputs (i.e., graphs) of ultra-short-term point predictions. 2) Most of the publications simplify inputs of ultra-short-term point predictions into Euclidean data by neglecting the adjacency matrix and only keeping feature vectors, which limits the prediction accuracy.

1.2.3. LITERATURE REVIEW ON UNCERTAINTY PREDICTION

The point predictions of power loads and RESs only provide deterministic values and overlook prediction errors induced by stochastic and fluctuating behaviors, which pose potential threats to the operation and planning of ADNs. Various methods have been proposed to model the uncertainty over the past few decades. Among them, interval prediction and scenario prediction are mainstream methods, and their solutions are usually used for risk-based decision-making models (e.g., stochastic programming and robust programming). Therefore, this section reviews the applications of the scenario prediction and interval prediction for power loads and RESs.

1) Literature review on interval prediction

To represent uncertainty (i.e., prediction errors), the interval prediction constructs prediction intervals (PIs) with a lower boundary and an upper boundary to cover real values. The widely used methods to construct PIs mainly consist of mean-variance estimation [36], ensemble Gaussian model [37], Bayesian model [38], Delta model [39], the lower upper bound estimation (LUBE) [40], and bootstrap techniques [41].

Concretely, the first four methods need to artificially assume and formulate a data distribution of prediction errors, which are affected by various factors, such as inputs, time horizons, and specific point prediction models. The probability distribution of prediction errors is often not easy to formulate accurately, which limits their application scope. The LUBE estimates the lower boundary and upper boundary by using a DNN with two-dimensional outputs, but it is still difficult to design a suitable loss function to update the weights of the DNN by using the gradient descent method [42].

The bootstrap technique constructs PIs by repeatedly resampling historical prediction errors without making assumptions about prediction error distributions. Due to its excellent performance and simple process, the bootstrap technique has been frequently utilized for the interval prediction of power loads and RESs [43].

Nonetheless, a significant shortcoming of the conventional bootstrap technique is that it has the same width of PIs for each prediction value. In ideal conditions, when prediction errors are small, the good PIs should be narrow. On the contrary, when prediction errors are large, the perfect PIs should be wide to ensure a high prediction interval coverage percentage (i.e., the probability that the PIs cover ture values) of PIs. In other words, the PIs of the conventional bootstrap technique are not flexible, since PIs with a fixed width are difficult to balance the width and coverage percentage.

In general, the existing publications mainly include the following limitations: 1) Most of the publications need to artificially assume and formulate a data distribution of prediction errors. 2) Most of the publications (e.g., conventional bootstrap technique) have difficulty in balancing the width and coverage percentage of PIs.

2) Literature review on scenario prediction

Scenario prediction is another commonly used way to represent the uncertainty of power loads and RESs by producing numerous possible scenarios [44]. Traditional methods of scenario prediction mainly consist of two categories: feature state-based methods and noise-based methods.

Noise-based methods produce numerous stochastic scenarios by combining deterministic point prediction values and random noises. For instance, prediction errors are considered to follow the Gaussian distribution in [45], and then different Gaussian noises and the point prediction value are summed to form a group of stochastic scenarios. In [46], a deep residual network is utilized to fit the variance and expectation, which are used to estimate prediction errors through the Monte Carlo method. To estimate prediction errors of power load which depend on both the point prediction model and inputs, a quantile regression technique is presented in [47]. The main drawback of noise-based methods is that the probability distribution of prediction errors has to be assumed and formulated artificially. In addition, the generated stochastic scenarios are hard to depict the diversity of different behavioral patterns, particularly if power loads and RESs include multiple behavioral patterns (e.g., heavy load and light loads).

Feature state-based methods need to formulate a physical model, bridging outputs (i.e., stochastic scenarios) and weather conditions (e.g., wind direction, temperature, wind speed, pressure, light intensity, and humidity). Then, a set of simulated weather conditions are input to the physical model to produce stochastic scenarios of power loads and RESs. For example, the work in [48] designs an empirical function to quantitatively produce temperature scenarios, which are used as inputs of a physical model for scenario predictions of power loads. To generate possible peak loads, the work in [49] proposes a polynomial to estimate the key factors of variations in load demands. Numerous load scenarios are created by encoding weather conditions to a generalized additive model in [50]. Generally, feature state-based methods outperform noise-based methods in terms of the diversity of generated scenarios. However, feature state-based methods fail to address the fundamental issue, because they shift the focus to how to get a good physical model and weather conditions.

To solve the aforementioned limitations of traditional methods, numerous deep generative networks have been developed in recent years for the scenario synthesis and scenario prediction of power loads and RESs [51], [52], [53]. For instance, various variants of generative adversarial networks (GANs) are adopted in [54], [55] to generate scenarios of wind and PV powers, while the works in [56], [57] apply non-linear independent component estimation (NICE) and variational auto-encoder (VAE) to model power load curves. Further, the GAN and NICE are generalized from scenario productions into scenario predictions for power loads and RESs in [58], [59]. Nevertheless, these generative networks involve either training problems or inaccurate loss functions, seriously affecting the quality of prediction scenarios. For example, the problem of unstable training in GANs is still a challenge [23], and the inference step of VAE is very intractable.

In general, existing publications mainly include the following limitations: 1) In most of the traditional methods, the probability distribution of prediction errors has to be

assumed and formulated artificially. 2) Most of the more recent generative networks involve either training problems or inaccurate loss functions, seriously affecting the quality of prediction scenarios.

1.2.4. LITERATURE REVIEW ON RISK-BASED OPTIMAL SCHEDULING

1) Literature review on stochastic day-ahead optimal scheduling of power loads and RESs in ADNs

Normally, the day-ahead scheduling of power loads and RESs in ADNs aims to find an optimal state of different power devices, such as transformers, RESs, battery electric vehicles (BEVs), and shunt capacitor banks (CBs), to minimize operational costs and maintain voltage amplitudes under a range of operational constraints. The traditional day-ahead optimal scheduling of power loads and RESs in ADNs is a deterministic model, in which the inputs (e.g., power loads and RESs) are deterministic point prediction values without considering prediction errors caused by stochastic and fluctuating behaviors of power loads and RESs [60]. In practical engineering, the prediction errors of point prediction values are unavoidable, which renders the security of ADN difficult to guarantee with deterministic model-based solutions [61].

Currently, the widely used method to estimate the risk caused by prediction errors mainly consists of interval programming, fuzzy programming, probabilistic programming, and robust programming [2]. Specifically, fuzzy programming employs the fuzzy set theory to represent the uncertainty, and yet its membership functions are very subjective (it relies on experience) [62]. Interval programming utilizes PIs to depict the uncertainty, but interval operations are very complex, and it is hard to find suitable intervals in some cases [63].

Stochastic programming models prediction errors from a probabilistic perspective. Stochastic programming not only provides expected values of voltages and losses like traditional deterministic models, but also describes their uncertainty by using probability distributions [64]. In light of this, an increasing number of publications focus on the application of stochastic programming to ADNs.

For instance, a day-ahead probabilistic scheduling model of RESs in ADNs considering prediction errors is presented in [65]. A stochastic programming model is designed to capture the uncertainty of RESs and minimize power losses of ADNs in [66]. To boost the operational performance of the power and heating systems, the work in [67] formulates a day-ahead stochastic scheduling scheme, which considers the prediction errors of wind powers and the flexibility of heat pumps (HPs). Similarly, the work in [68] proposes a stochastic probability model to minimize the costs of investments and operations, considering the uncertainty of power loads in

electric vehicle charging stations (EVCSs). However, previous publications mainly include two challenges that need to be addressed:

Firstly, few previous publications consider the impact of the simultaneous integration of RESs and flexible loads (e.g., HPs and BEVs) on ADNs. Most of the previous publications only model a part of them. Secondly, few previous publications represent BEVs and HPs with the demand response. Most of them just model BEVs and HPs without considering the demand response.

2) Literature review on robust day-ahead optimal scheduling of power loads and RESs in ADNs

As mentioned earlier, the inputs to the deterministic day-ahead optimal scheduling (DDOS) model of power loads and RESs in ADNs are point prediction values. This brings risks to the operation of ADNs, since the deterministic model ignores prediction errors of power loads and RESs. In addition to stochastic programming, robust programming is another popular way to depict the uncertainty of power loads and RESs by constructing worst-case scenarios [69]. The robust solution should guarantee that operational constraints are satisfied when ADNs operate in worst-case scenarios. The robust programming-based solutions can ensure the security of ADNs in all possible scenarios, as they have been tested in worst-case scenarios.

The key part of the robust day-ahead optimal scheduling (RDOS) of power loads and RESs in ADNs is how to construct worst-case scenarios. Conventionally, worst-case scenarios are represented by two extreme scenarios [70]: For the first extreme scenario, the power loads are taken as the maximum, and the output powers of RESs are taken as the minimum. For the second extreme scenario, the power loads are taken as the minimum, and the output powers of RESs are taken as the maximum. Although this traditional method is simple and effective, its solutions are too conservative. In other words, it pays a high price to secure ADNs.

Recently, numerous publications have been developed to balance economy and security by constructing suitable prediction errors, which are combined with the point prediction values to form new kinds of worst-case scenarios. For example, a Gaussian distribution model is used to produce prediction errors of wind powers and power loads in [71]. The work in [72] assumes prediction errors of wind powers follow the Laplace distribution, and groups of possible prediction errors are generated by using the Latin hypercube sampling method. In [73], [74], a Weibull distribution model is combined with a Rayleigh distribution to depict the variation of wind speeds. In the same way, the Monte Carlo method is adopted to produce prediction errors of PV powers by sampling the Beta distribution in [75]. Other popular distributions to capture prediction errors involve the Laplace distribution in [76], and the Cauchy distribution in [77]. However, prediction errors of power loads and RESs are affected by various factors (e.g., weather conditions and point

prediction models), making it difficult to formulate a uniform and accurate formula to describe real distributions.

In general, most of the publications are unable to represent prediction errors accurately. As a result, the constructed new kinds of worst-case scenarios are conservative, if real prediction errors are significantly smaller than generated prediction errors. On the contrary, new kinds of worst-case scenarios are too narrow to cover all possible scenarios, if real prediction errors are substantially larger than generated prediction errors. The good worst-case scenario constructions rely on the prediction error estimation, but it remains a great challenge to accurately estimate prediction error without assuming a probability density function.

1.3. RESEARCH MOTIVATION

1.3.1. RESEARCH QUESTIONS

As discussed in previous sections, although numerous methods have been developed to model power loads and RESs in ADNs, most of them have difficulty in capturing stochastic and fluctuating behaviors of power loads and RESs, which pose risks to the operation and planning of ADNs. This Ph.D. project aims to model and optimize the power loads and RESs in ADNs accurately and efficiently, especially from the data-driven perspective. Specifically, DL technologies are developed to transform the massive collected data into knowledge, and provide deeper insights into the past, better understandings of the future, and practical suggestions on possible decisions for the economy and secure operation of ADNs. To achieve this objective, the following questions are considered:

- **Q1:** How to design an advanced missing data imputation framework to achieve high imputation accuracy of filling values by capturing the contextual information of missing values and complex non-linear relationships among multiple attributes?

- **Q2:** How to capture spatio-temporal features from inputs (i.e., graphs) of the ultra-short-term point predictions instead of simplifying graphs into Euclidean data?

- **Q3:** How to accurately represent the uncertainties (i.e., prediction errors) of power loads and RESs without artificially assuming their probability distributions?

- **Q4:** How to formulate risk-based day-ahead optimal scheduling models of power loads and RESs in ADNs considering the uncertainties of RESs and loads?

1.3.2. RESEARCH TASKS

To solve these problems, this Ph.D. project is divided into multiple sub-tasks as follows:

- **T1: Develop an advanced framework to impute missing values.**

  To address Q1, a new framework is developed to depict the complicated non-linear correlation among multiple attributes. In addition, the contextual information of missing values is taken into account. In other words, the consistency of the missing values and other surrounding information is considered. It will be tested on datasets with common missing types, including missing completely at random and continuous missing.

- **T2: Develop an advanced model for the ultra-short-term point predictions.**

  To address Q2, a new point prediction model with high accuracy is proposed to depict the temporal features of the time series. Also, the spatial correlation between adjacent RESs is considered. The proposed point prediction model should outperform traditional methods for ultra-short-term predictions.

- **T3: Develop an interval prediction model and a scenario prediction model to represent the uncertainty, from two different perspectives.**

  To address Q3, two new models are proposed to represent the uncertainty of power loads and RESs from different perspectives. Firstly, a novel scenario prediction model is presented to depict the uncertainty by generating a set of possible scenarios. This model should not involve the same difficulty in the training process as the existing generative networks.

  Secondly, a new interval prediction model is designed to depict the prediction errors by using an upper boundary and a lower boundary. The constructed PIs should balance the width and coverage percentage well.

- **T4: Develop robust and stochastic day-ahead optimal scheduling models of power loads and RESs in ADNs, from two different perspectives.**

  To address Q4, two new models are developed to represent the uncertainty of power loads and RESs in the decision-making process from different perspectives. Firstly, an RDOS model of power loads and RESs in ADNs is formulated. In addition, it is also necessary to construct new kinds of worst-case scenarios as inputs. The difference between robust programming and deterministic programming is elaborated.

Secondly, a stochastic day-ahead optimal scheduling (SDOS) model of power loads and RESs in ADNs is formulated from a probabilistic perspective. The HPs and BEVs with and without demand response are considered. Lastly, the difference between stochastic programming and deterministic programming is elaborated.

1.4. PROJECT LIMITATIONS

This Ph.D. project project has a number of limitations:

- **L1:** It requires massive historical data to train the proposed DNNs before using them. Therefore, the proposed model is not applicable to ADNs without historical data. Moreover, the parameters of the proposed model may need to be fine-tuned when they are migrated to other datasets.

- **L2:** Due to the page limit, the impact of data outliers on the models is not discussed in this project. In addition, the dimensionality of each sample is assumed to be consistent.

- **L3:** This project assumes that the samples in the training set, validation set, and test sets have similar probability distributions, i.e., this project does not consider the effect of drastic climate change on the models.

- **L4:** The proposed point prediction model and interval prediction model are tested on datasets whose inputs only include historical data without considering NWP data, but the NWP data can be added to inputs in the extension work easily.

- **L5:** The control variables in RDOS and SDOS models of power loads and RESs in ADNs only consider the classical power devices, such as transformers, RESs, shunt CBs, HPs, and BEVs. Other power devices (e.g., energy storage, soft open points, and electrical boilers) are not considered. Note that the mathematical models of other power devices not discussed here are similar to those of classical power devices. Therefore, the proposed models can be easily generalized to ADNs with more power devices.

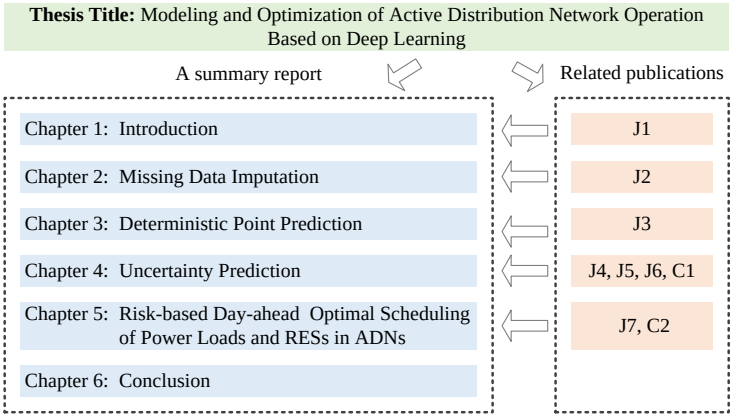


Figure 1-2 The relationships between chapters and related publications.

1.5. THESIS OUTLINE

This doctoral dissertation mainly includes two parts: a summary report and related publications. Specifically, the summary report describes the main outcomes of the previous publications and the relationship between them. As shown in Figure 1-2, the summary report includes six chapters.

- **Chapter 1: Introduction**

  In Chapter 1, the background is elaborated. Then, a comprehensive literature review is conducted. The research motivations and limitations of this project are clarified. A part of the work is related to [J1].

- **Chapter 2: Missing Data Imputation**

  Chapter 2 summarizes the main works in [J2]. Shortly, a framework called context encoder is proposed to impute the missing values of the wind power dataset. A hybrid loss function is designed to generate reasonable missing values by combining the adversarial loss function and reconstruction loss function. In addition, the proposed framework is tested on datasets with different missing types, including continuous missing and missing completely at random.

- **Chapter 3: Deterministic Point Prediction**

  Chapter 3 summarizes the main works in [J3]. Simply put, a hybrid model is designed for the ultra-short-term point prediction of RESs. The inputs of the ultra-short-term point predictions are modeled as graphs. A graph

convolutional network (GCN) and an LSTM model form a hybrid model to depict the temporal and spatial features from graphs. The proposed hybrid model is tested on different datasets.

- **Chapter 4: Uncertainty Prediction**

  Chapter 4 summarizes the main works in [J4], [J5], [J6], and [C1]. In simple terms, two new models are proposed to depict the uncertainty of power loads and RESs from different perspectives. Firstly, a new generative model named pixel convolutional neural network is proposed to depict the uncertainty of power loads by producing numerous possible scenarios. This generative model does not involve the same difficulty in the training process as the existing generative networks. Secondly, the conventional bootstrap technique is modified to depict the uncertainty by constructing PIs. The improved bootstrap technique balances the width and coverage percentage of PIs well. Both algorithms are tested on different real-world datasets.

- **Chapter 5: Risk-based Day Ahead Optimal Scheduling of Power Loads and RESs in ADNs**

  Chapter 5 summarizes the main works in [J7] and [C2]. In short, two new models are developed to represent the uncertainty of power loads and RESs in the decision-making process from different perspectives. Firstly, an RDOS model of power loads and RESs in ADNs is formulated. Then, the new kinds of worst-case scenarios are constructed by using the improved bootstrap technique. The difference between robust programming and deterministic programming is elaborated. Secondly, a SDOS model of power loads and RESs in ADNs is formulated. The HPs and BEVs with and without demand response are considered. The difference between stochastic programming and deterministic programming is elaborated. Lastly, the proposed models are tested on a modified IEEE 33-bus distribution network.

- **Chapter 6: Conclusion**

  In Chapter 6, the main conclusions and contributions are summarized. The possible future works are presented.

1.6. LIST OF PUBLICATIONS

This Ph.D. dissertation is related to the following 7 journal papers (J1-J7) and 2 conference papers (C1-C2). Most of the papers have been published, while a few papers are under review.

Journal Paper:

J1: W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, "A Review of Graph Neural Networks and Their Applications in Power Systems," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 345-360, Mar. 2022.

J2: W. Liao, B. Bak-Jensen, J. R. Pillai, D. Yang, and Y. Wang, "Data-driven Missing Data Imputation for Wind Farms Using Context Encoder," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 964-976, Jul. 2022.

J3: W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, "Short-Term Power Prediction for Renewable Energy Using Hybrid Graph Convolutional Network and Long Short-Term Memory Approach," *Electric Power Systems Research*, vol. 211, pp. 1-7, Oct. 2022 (Special Issue: XXII Power Systems Computation Conference. Note that all papers accepted for PSCC 2022 are published in Electric Power Systems Research.).

J4: W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, Y. Wang, and K. Liu, "Scenario Generations for Renewable Energy Sources and Loads Based on Implicit Maximum Likelihood Estimations," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1563-1575, Nov. 2022.

J5: W. Liao, L. Ge, B. Bak-Jensen, J. R. Pillai, and Z. Yang, "Scenario Prediction For Power Loads Using A Pixel Convolutional Neural Network and An Optimization Strategy," *Energy Reports*, vol. 8, pp. 6659-6671, Nov. 2022.

J6: W. Liao, S. Wang, B. Bak-Jensen, J. R. Pillai, Z. Yang and K. Liu, "Ultra-Short-Term Interval Prediction of Wind Power Based on Graph Neural Network and Improved Bootstrap Technique," *Journal of Modern Power Systems and Clean Energy*, E-pub ahead of print, doi: 10.35833/MPCE.2022.000632.

J7: W. Liao, S. Wang, B. Bak-Jensen, J. R. Pillai, and Z. Yang, "Bootstrap-Based Prediction Error Estimation for Robust Reactive Power Scheduling of Distribution Networks," *Journal of Modern Power Systems and Clean Energy*, Under Review.

Conference Paper:

C1: W. Liao, B. Bak-Jensen, J. R. Pillai, R. Zhu, and L. Song, "Data-Driven Scenarios Generation for Wind Power Profiles Using Implicit Maximum Likelihood Estimation," *in the 12th International Conference on Applied Energy(ICAE 2020)*, Dec. 2020, pp. 1-5.

C2: W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, Z. Li, and D. Yang, "Stochastic Day-ahead Optimal Scheduling of Active Distribution Networks with Renewable Energy Sources and Electric Vehicles," *in the 8th Asia Conference on Power and Electrical Engineering (ACPEE 2023)*, Apr. 2023, pp. 1-8 (Accepted).

CHAPTER 2. MISSING DATA IMPUTATION

Chapter 2 summarizes the main works in [J2].

J2: W. Liao, B. Bak-Jensen, J. R. Pillai, D. Yang, and Y. Wang, "Data-driven Missing Data Imputation for Wind Farms Using Context Encoder," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 964-976, Jul. 2022.

2.1. ABSTRACT AND KEY CONTRIBUTIONS

2.1.1. ABSTRACT

The operation and scheduling of wind power systems depend on high-quality datasets heavily. However, due to a variety of reasons, such as cyber-attack, sensor failure, and communication congestion, the datasets gathered by the SCADA system may contain missing values. To fill in the missing values of wind power datasets accurately, a data-driven approach named context encoder (CE) is proposed. The CE is composed of three parts: a discriminator, an encoder, and a decoder. To extract the complicated latent features automatically, the convolutional neural network (CNN) is adopted to construct three parts of the CE. Further, a reconstruction loss function and an adversarial loss function are combined to a hybrid loss function, which enables CE to generate reasonable missing values for incomplete samples. The simulation results show that the CE not only considers correlations between wind powers and other attributes, but also captures the contextual information of wind power generation curves, making it highly adaptable to the missing data imputation with fast ramps. Moreover, CE outperforms popular baselines considering mean absolute error, maximum absolute error, and root mean squared error in an integrated way, especially for datasets with high missing ratios.

2.1.2. KEY CONTRIBUTIONS

- A new framework is developed to impute the missing values of wind power datasets.
- The CNN is adopted to construct each part of the CE, so as to extract the complicated latent features of multiple attributes in wind power datasets automatically.
- A hybrid loss function is designed to ensure that CE generates reasonable missing values. Firstly, a reconstruction loss function is adopted to capture the

contextual information of missing values. Secondly, an adversarial loss function is adopted to make the filled values look more realistic and ensure that the probability distribution of the filled values is similar to the real one.

2.2. CONTEXT ENCODER FOR MISSING DATA IMPUTATION

2.2.1. THE FRAMEWORK OF THE CE

The overall architecture of the CE is presented in Figure 2-1, which includes three parts: discriminator, encoder, and decoder. Firstly, the inputs (i.e., wind powers and weather conditions) are reshaped into feature matrices, which are fed to the encoder and discriminator. The two-dimensional convolutional (Conv2D) layers and maximum pooling (Maxpool) layers are adopted to form the encoder, which projects feature matrices into the low-dimensional latent features. The decoder consists of several two-dimensional transposed convolutional (Conv2DTran) layers, and it is responsible for mapping latent features from the encoder to complete samples. As a detector, the discriminator consists of several convolutional, flatten, and dense layers. The discriminator is responsible for recognizing as many reconstructed samples as possible.

CNN is a famous feature extractor by performing complex convolutional operations. So far, CNN has been broadly applied to different fields (e.g., time series prediction, anomaly detection, and image generation), and has achieved outstanding achievements thanks to its strong feature extraction capability [78]. In view of this, the CNN is employed to form the three parts (i.e., the discriminator, encoder, and decoder) of the CE.

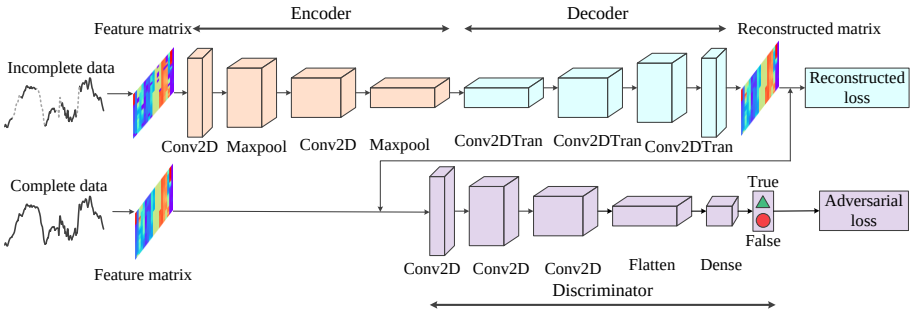


Figure 2-1 The overall architecture of the CE. Source: [J2].

2.2.2. THE STRUCTURE OF THE CE

The encoder is made up of a number of Maxpool layers and Conv2D layers. The main function of the Conv2D layer is to add a bias vector as inputs to the following layer after convolving the outputs of the previous layer. The Conv2D layer can be represented as follows:

$$Y_{\text{con}}^{i} = \sigma_{\text{con}}^{i}\left(X_{\text{con}}^{i} * W_{\text{con}}^{i} + B_{\text{con}}^{i}\right) \tag{2.1}$$

where X_{con}^{i} and Y_{con}^{i} are the inputs and outputs of the i^{th} Conv2D layer, respectively; $*$ is the convolutional operation; W_{con}^{i} and B_{con}^{i} are the weights and bias vectors of the i^{th} Conv2D layer, respectively; and $\sigma_{\text{con}}^{i}(\cdot)$ is the activation function of the i^{th} Conv2D layer [7].

To produce a low-dimensional representation of inputs from the previous Conv2D layer, the Maxpool layers, as shown in Figure 2-2, are employed to reduce their dimensionality:

$$Y_{\text{pool}}^{i} = \max_{j,k \in R}\left(Y_{j,k}^{i}\right) \tag{2.2}$$

where Y_{pool}^{i} and $Y_{j,k}^{i}$ are the outputs and inputs of the i^{th} Maxpool layer, respectively; and R is the maximum pooling area.

The fully Conv2DTran layer constitutes the decoder. The main function of the Conv2DTran layer is to add a bias vector as inputs of the next layer after transposing convolutional operations on the outputs of the previous layer. The Conv2DTran layer can be represented as follows:

$$Y_{\text{tran}}^{i} = \sigma_{\text{tran}}^{i}\left(X_{\text{tran}}^{i} *^{T} W_{\text{tran}}^{i} + B_{\text{tran}}^{i}\right) \tag{2.3}$$

where X_{tran}^{i} and Y_{tran}^{i} are the inputs and outputs of the i^{th} Conv2DTran layer, respectively; $*^{T}$ is the transposed convolutional operation; W_{tran}^{i} and B_{tran}^{i} are the weights and bias vectors of the i^{th} Conv2DTran layer, respectively; and $\sigma_{\text{tran}}^{i}(\cdot)$ is the activation function of the i^{th} Conv2DTran layer [7].

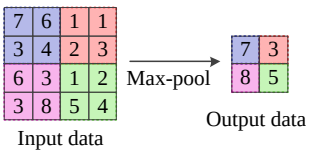


Figure 2-2 The maximum pooling operation in a straightforward example. Source: [J2].

The convolutional layer, dense layer, and flatten layer form the discriminator. In particular, flatten layers reshape the multi-dimensional features output from the Conv2D layer into one-dimensional features to be compatible with the input format of dense layers. The dense layer outputs a Boolean variable. The Boolean variable is 1, if the input of the discriminator is a complete sample. The Boolean variable is 0, if the input of the discriminator is a reconstructed sample from the encoder and decoder. The dense layer is represented as:

$$Y_{\text{dense}}^{i} = \sigma_{\text{dense}}^{i}\left(X_{\text{dense}}^{i} W_{\text{dense}}^{i} + B_{\text{dense}}^{i}\right) \tag{2.4}$$

where X_{dense}^{i} and Y_{dense}^{i} are inputs and outputs of the i^{th} dense layer, respectively; W_{dense}^{i} and B_{dense}^{i} are the weights and bias vectors of the i^{th} dense layer, respectively; and $\sigma_{\text{dense}}^{i}(\cdot)$ is the activation function of the i^{th} dense layer [7].

2.2.3. THE LOSS FUNCTION

The adversarial loss function and reconstruction loss function are combined into a hybrid loss function, which enables CE to generate reasonable missing values for incomplete samples. In particular, the reconstruction loss function is responsible for ensuring the consistency (i.e., contextual information) between missing values and the information around them. The reconstruction loss function is represented as:

$$L_{\text{rec}} = \left\| \left(X - F\left(X \odot (1-M)\right)\right) \odot M \right\|_2 \tag{2.5}$$

where X is the complete samples; $F(X\odot(1\text{-}M))$ is the sample reconstructed by the context encoder; L_{rec} is the reconstruction loss function; $\odot$ is the Hadamard product operation; $\| \cdot \|$ is the 2-norm; and M is the binary mask, representing the position of missing values. M=1 indicates a missing value, and M=0 indicates a complete value.

Further, the adversarial loss function not only makes the filled values look more realistic, but also ensures that the probability distribution of the filled values is similar to the true one. The adversarial loss function can be represented as follows:

$$L_{\text{adv}} = \max \mathrm{E}_{X}\left[\log\left(1 - D\left(F\left(X \odot (1-M)\right)\right)\right) + \log\left(D(X)\right)\right] \tag{2.6}$$

where L_{adv} is the adversarial loss function; E_X is the expected value of sample X; and $D(\cdot)$ is the output of the discriminator.

Lastly, the final loss function of CE is a mixture of the adversarial loss function and reconstruction loss function. The final loss function is represented as follows:

$$L_{\text{CE}} = \lambda L_{\text{rec}} + (1-\lambda) L_{\text{adv}} \tag{2.7}$$

where L_{CE} is the final loss function of the CE; and λ is a parameter to balance two sub-loss functions.

2.2.4. THE SPECIFIC STEPS

The missing types mainly include [79]: missing completely at random and continuous missing. The positions of the missing values are continuous for the latter, while the positions of the missing values are completely random for the former. To demonstrate the validity of the proposed CE on a dataset with missing completely at random and continuous missing types, the simulations are conducted by the following 4 steps.

Step 1: Datasets are imported. For example, a wind power dataset may include historical wind powers and weather conditions. Before taking these data as input, the widely used minimum-maximum normalization method is employed to normalize them to values with the same magnitude, ranging from 0 to 1.

Step 2: To ensure that the input data and CE are compatible, the 1-D time series are reshaped into a two-dimensional feature matrix [57]. Besides, the dataset is randomly separated into the training set, validation set, and test set.

Step 3: The parameters and structure of the CE are initialized. Then, the weights of the CE are updated by calculating the loss function and gradient. The weights are repeatedly iterated and updated until the maximum number of iterations allowed is achieved.

Step 4: Finally, the pre-trained CE is tested on the test set through several common metrics, including root mean squared error (RMSE), mean absolute error (MAE), and maximum absolute error (MMAE) [7]. The smaller these metrics are, the better the model performance.

2.3. CASE STUDY

2.3.1. SIMULATION SETTINGS

To test the performance of the proposed CE, a wind power dataset from [80] is adopted to conduct simulation and analysis. The simulation settings have been elaborated in [J2]. Due to the page limit, the next section only presents the simulation results on the dataset with missing values completely at random. The parameter discussion of the CE and simulation results on the dataset with continuous missing values are not shown here, but can be found in [J2].

2.3.2. SIMULATION RESULTS AND ANALYSIS

To compare the performance of the proposed CE and baselines (e.g., cubic interpolation, K-means, MLP in [18], KNN in [14], AE in [22], and CGAN in [24]), simulations are repeatedly conducted several times on the dataset with missing values completely at random. As a simple example, the missing ratio is set to 5%. The average filling errors are presented in Table 2-1 and Figure 2-3.

Although the fundamentals of K-means and KNN are straightforward and they are easy to use, their absolute errors are relatively large. For instance, the upper quartiles of K-means and KNN are greater than 0.5 p.u. and 0.18 p.u., respectively, whereas the CE's upper quartile is below 0.1 p.u.. Besides, the MMAE of CE is also far tinier than those of the K-means and KNN.

In terms of absolute errors, the lower, upper, and median quartiles of the CE are close to those of the CGAN, MLP, and AE, but slightly larger compared to those of cubic interpolation. Further, CE has a much smaller MMAE than cubic interpolation, MLP, and AE. In particular, the MMAE of cubic interpolation, MLP, AE, and CE are 0.995 p.u., 0.879 p.u., 0.801 p.u., and 0.567 p.u., respectively.

The AE and CE share most of their structures, but CE has an additional adversarial loss function and discriminator than AE. When absolute errors of CE, AE, and CGAN are compared, the adversarial loss function and discriminator are found to be useful in reducing maximum absolute errors. Furthermore, the loss function and framework of the CE outperform the traditional CGAN, since the MMAE, MAE, and RMSE of the CGAN are larger than those of the CE.

Table 2-1 The results on datasets with missing values completely at random. Source: [J2].

Methods	MAE (p.u.)	RMSE (p.u.)	MMAE (p.u.)
CGAN	0.0530	0.097	0.590
Cubic interpolation	0.018	0.077	0.995
KNN	0.075	0.135	1.011
MLP	0.048	0.100	0.879
K-means	0.204	0.275	0.975
CE	0.035	0.068	0.567
AE	0.030	0.082	0.801

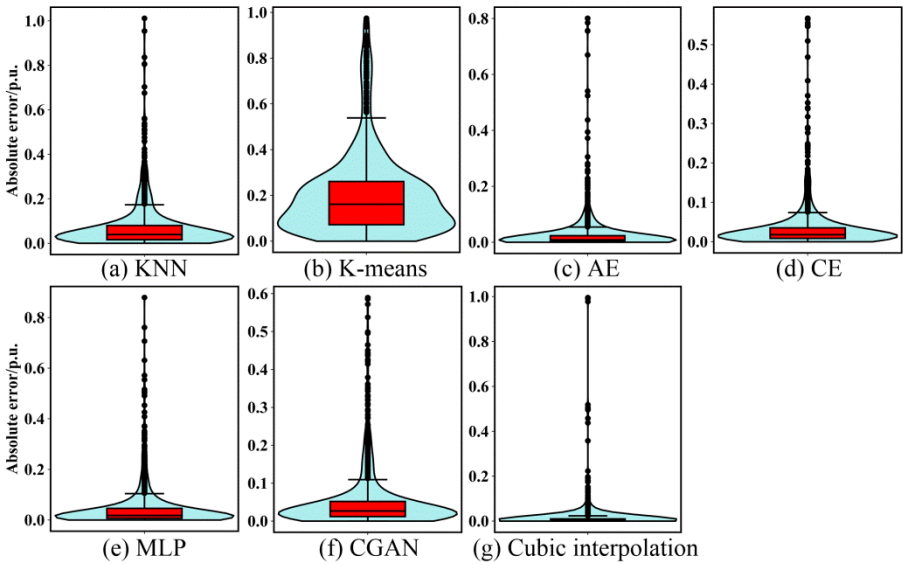


Figure 2-3 The visual analysis of a sample randomly selected from the test set. Source: [J2].

Additionally, a sample is chosen at random from the test set to compare how well various methods perform when filling in random missing values, as shown in Figure 2-4.

In terms of attributes with minor variations (e.g., density, pressure, temperature, wind direction, and wind speed), K-means and KNN perform significantly worse than other methods, because they belong to similarity-based methods, which neglect the temporal features of these attributes. Comparatively, cubic interpolation, CGAN, MLP, and CE achieve high accuracies by effectively exploiting the contextual information around missing values.

The CE outperforms the cubic interpolation for attributes (e.g., wind power) with significant variation. For instance, the wind power generation curve exists a clear peak at 22:00, as displayed in Figure 2-4(e). The filling accuracy of cubic interpolation is very limited in this case. The reason is that the cubic interpolation only uses the contextual information to fill in missing values, while the information at early and late times can be very different due to the large fluctuations. In contrast, CE not only considers correlations between wind powers and other attributes, but also captures temporal features of wind power generation curves, making it highly adaptable to the missing data imputation with fast ramps.

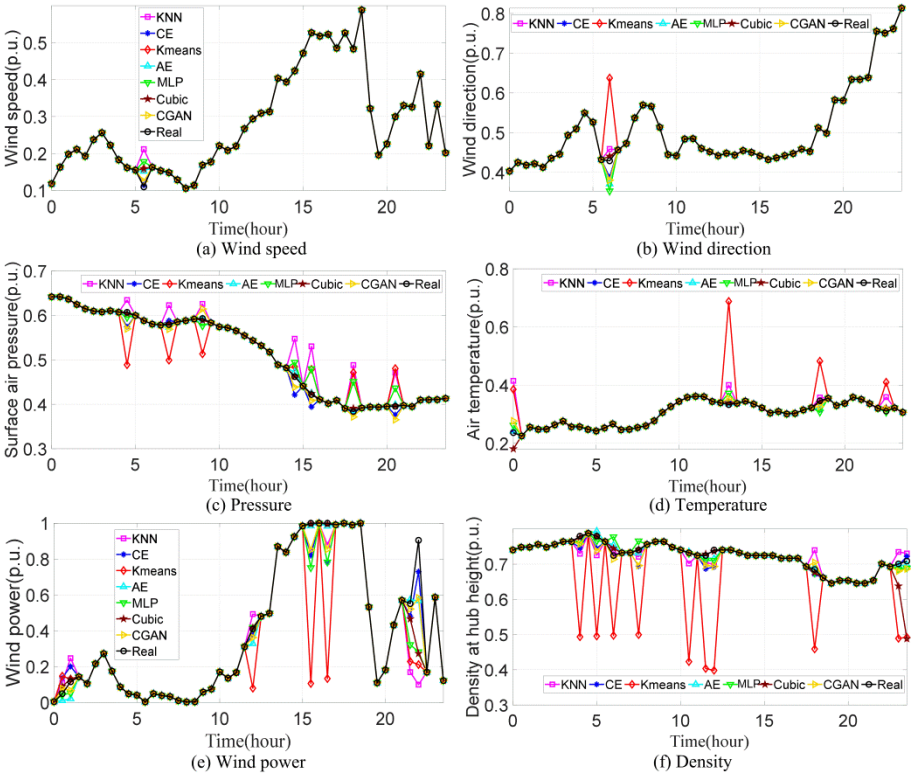


Figure 2-4 The visual analysis of a sample with missing values. Source: [J2].

To explore the performance of the CE on datasets with different missing ratios, the missing ratio in each sample ranges from 10% to 50%. The positions of missing values are randomly initialized. The CE and baselines are trained and tested several times, and the average filling errors are presented as shown in Table 2-2.

The RMSE, MAE, and MMAE of the CGAN, K-means, MLP, and KNN are always greater than those of the CE for different missing ratios, indicating that CE has superior filling accuracy. Specifically, simulation results on datasets with different missing ratios show that the RMSE, MMAE, and MAE of CE are tinier than 0.085 p.u., 0.652 p.u., and 0.045 p.u., respectively.

Table 2-2 The results on datasets with various missing ratios. Source: [J2].

Methods	Missing ratio is 10%			Missing ratio is 20%			Missing ratio is 30%		
	MAE (p.u.)	MMAE (p.u.)	RMSE (p.u.)	MAE (p.u.)	MMAE (p.u.)	RMSE (p.u.)	MAE (p.u.)	MMAE (p.u.)	RMSE (p.u.)
CGAN	0.042	0.093	0.692	0.043	0.089	0.694	0.047	0.083	0.727
Cubic interpolation	0.018	0.074	0.991	0.021	0.144	8.053	0.021	0.126	8.053
KNN	0.080	0.130	1.068	0.085	0.143	1.337	0.094	0.147	1.257
MLP	0.047	0.104	0.954	0.046	0.098	0.941	0.046	0.096	0.860
K-means	0.191	0.265	0.939	0.189	0.260	0.956	0.198	0.269	0.968
CE	0.037	0.071	0.619	0.038	0.075	0.606	0.041	0.078	0.590
AE	0.031	0.083	0.904	0.030	0.078	0.900	0.033	0.085	1.019

Methods	Missing ratio is 40%			Missing ratio is 50%		
	MAE (p.u.)	MMAE (p.u.)	RMSE (p.u.)	MAE (p.u.)	MMAE (p.u.)	RMSE (p.u.)
CGAN	0.049	0.091	0.713	0.050	0.092	0.758
Cubic interpolation	0.029	0.226	9.965	0.033	0.253	13.031
KNN	0.100	0.153	1.388	0.107	0.166	1.448
MLP	0.047	0.096	0.851	0.047	0.097	0.974
K-means	0.198	0.268	0.960	0.194	0.266	0.963
CE	0.044	0.083	0.615	0.045	0.085	0.652
AE	0.038	0.093	0.980	0.040	0.098	0.962

MAE calculates the average magnitude of filled errors, regardless of their directions. Besides, MAE is a linear metric, in which all filling errors are equally weighted in the mean. In contrast, when calculating the RMSE, the weights depend on the error. The larger the error is, the larger the weight is. If large errors are particularly undesirable, then RMSE is more useful than MAE. In particular, although the MAE of the CE is slightly larger than those of cubic interpolation and AE, its RMSE is smaller than those of cubic interpolation and AE, which demonstrates that the maximum filling errors of CE are much tinier than those of cubic interpolation and AE. In other words, CE outperforms the cubic interpolation and AE in terms of controlling the maximum filling error.

The MMAE of the cubic interpolation increases rapidly with increasing missing ratios. This phenomenon shows that cubic interpolation is particularly sensitive to missing ratios, and it is only suitable for filling in datasets with a small number of missing values. The MMAE of the other methods increases gradually with the missing ratios, indicating that they are applicable to missing data imputation with high missing ratios.

2.4. CONCLUSION

Missing data imputation is the basis of point predictions. This chapter summarizes a new approach to boost the filling accuracy of missing data imputations. The following conclusions are reached by performing simulation and analysis.

K-means and KNN perform significantly worse than other methods, because they belong to similarity-based methods, which neglect the contextual information of different attributes. Comparatively, cubic interpolation, CGAN, MLP, and CE can achieve higher accuracy by effectively exploiting the contextual information around missing values.

The cubic interpolation is particularly sensitive to the percentage of missing values. When the missing ratio is large, the filling errors of the cubic interpolation significantly exceed the maximum allowed value.

The adversarial loss function and discriminator in the CE are found to be useful in reducing maximum absolute errors.

CE not only considers correlations between wind powers and other attributes, but also captures the contextual information of wind power generation curves, making it highly adaptable to the missing data imputation with fast ramps.

From the simulation results on the dataset with continuous missing values, the MMAE, MAE, and RMSE of the CE are tinier than those of other algorithms in the case of high missing ratios, which indicates that the CE outperforms baselines on the dataset with large-scale continuous missing values (the detail of this part can be found in [J2]).

In general, CE is a good choice for missing data imputations of wind powers considering MMAE, MAE, and RMSE, especially for datasets with high missing ratios.

CHAPTER 3. DETERMINISTIC POINT PREDICTION

Chapter 3 summarizes the main works in [J3].

J3: W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, "Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach," *Electric Power Systems Research*, vol. 211, pp. 1-7, Oct. 2022 (Special Issue: XXII Power Systems Computation Conference. Note that all papers accepted for PSCC 2022 are published in Electric Power Systems Research.).

3.1. ABSTRACT AND KEY CONTRIBUTIONS

3.1.1. ABSTRACT

The ultra-short-term point prediction with high accuracy is important to the day-ahead optimal scheduling of power loads and RESs in ADNs. However, due to the complex temporal and spatial correlations of multiple adjacent RESs and weather conditions, the ultra-short-term point prediction of RESs has traditionally been regarded as a complicated regression task. To depict the temporal and spatial features, a new hybrid model is proposed for the ultra-short-term point prediction of RESs by integrating a GCN and an LSTM. In particular, multiple adjacent RESs and weather conditions are modeled as nodes of graphs with correlation matrices and feature matrices. Besides, a GCN is presented to extract complicated spatial features between nodes, and then an LSTM is adopted to depict temporal features. Simulation results demonstrate that the proposed hybrid model is capable of extracting temporal and spatial features of multiple adjacent RESs and weather conditions accurately. Moreover, it performs better than widely used baselines on real-world datasets.

3.1.2. KEY CONTRIBUTIONS

- Multiple adjacent RESs and weather conditions are modeled as nodes of graphs with correlation matrices and feature matrices, from a new perspective in the graph domain.

- A new hybrid model is proposed for the ultra-short-term point prediction of RESs. In particular, a GCN is presented to extract complicated spatial features between nodes, and then an LSTM is adopted to depict temporal features.

- Numerical simulations are performed to verify that the hybrid model achieves state-of-the-art for the ultra-short-term point prediction of wind and PV powers.

3.2. A HYBRID MODEL FOR ULTRA-SHORT-TERM PREDICTION

3.2.1. PROBLEM DEFINITION

1) Definition of graphs

Normally, the ultra-short-term point prediction aims to predict the output powers of RESs given historical output powers and weather conditions (i.e., wind speed, light intensity, humidity, pressure, and temperature) [25].

In this chapter, the inputs are transformed into an undirected graph $G=(V,E)$ with nodes. In particular, the output powers of RESs are considered as real nodes, and the weather conditions are viewed as virtual nodes [34]. The real nodes and virtual nodes form all the nodes in the undirected graph.

In practical engineering, output powers of RESs and weather conditions are not simultaneously available sometimes. For instance, some datasets do not collect the surrounding weather conditions, and the inputs are composed of real nodes without virtual nodes.

Traditionally, the correlation between nodes is generally represented by an adjacency matrix. 1 indicates the presence of an edge between two nodes, and 0 indicates no edge. Obviously, it is difficult to depict the strength of the correlation between two nodes by using Boolean variables. Therefore, the correlation between nodes is represented by the Pearson correlation coefficient.

Until now, the inputs have been transformed into undirected graphs, which can be fed to the following hybrid model directly.

2) The specific architecture of the hybrid model

Figure 3-1 shows the specific architecture of the proposed hybrid model, consisting of a GCN and an LSTM [34]. Firstly, the GCN is utilized to extract spatial features between nodes from the feature matrix and the correlation matrix of graphs. Secondly, the LSTM is adopted to depict temporal features from the time series, which are output from the previous GCN. Finally, a dense layer is placed at the end of the LSTM to predict the output powers of RESs.

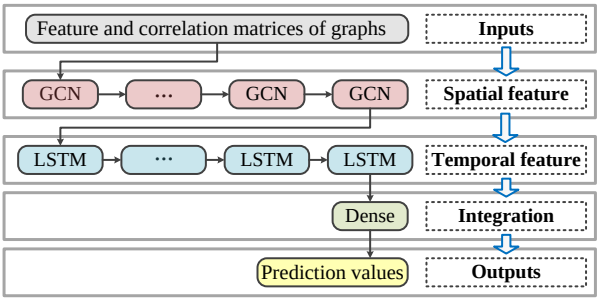


Figure 3-1 The framework of the proposed hybrid model. Source: [J3].

3.2.2. CAPTURING SPATIAL FEATURES WITH GCN

One of the main challenges in predicting the ultra-short-term power of RESs is capturing spatial features from inputs. The CNN has difficulty in effectively capturing the spatial information and making excellent use of the correlation matrix between nodes. Fortunately, to handle the tasks defined in the graph domain, the CNN has been expanded into the GCN defined in the graph domain [81], and the GCN has drawn increasing interest due to its potent capabilities.

The GCN and its variants are broadly separated into two categories: spectral-based GCN and spatial-based GCN [4]. It is challenging to determine which performs better. The spectral-based GCN is more popular than the spatial-based GCN due to the fact that it was developed earlier. As a simple example, the widely used spectral-based GCN is adopted to capture the spatial properties of inputs without losing generality.

The feature matrix X_{feature} and correlation matrix $C(t)$ of graphs are considered as inputs. The graph convolutional layers construct a filter in the Fourier domain, and then utilize its first-order polynomial in the Laplacian to capture the spatial information between nodes. A spectral-based GCN, as depicted in Figure 3-2, typically consists of several graph convolutional layers. Note that the GCN-derived time series with spatial features are regarded as inputs of the LSTM.

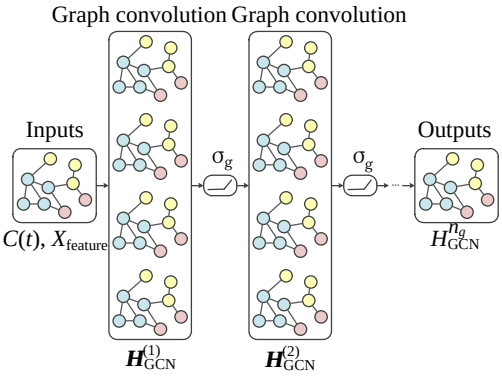


Figure 3-2 The framework of the GCN. Source: [J3].

3.2.3. CAPTURING TEMPORAL FEATURES WITH LSTM

Capturing temporal features is another crucial component of the ultra-short-term point prediction of RESs. Traditional DNNs, like MLP, are ineffective at modeling temporal features of time series. The more recent RNN is a kind of very promising method that excels at handling time series (e.g., audio). Given that the difficulty of vanishing gradient is part of the standard RNN, various advanced variants of RNNs have been proposed and have performed well in many fields [82]. In order to extract temporal features of time series from the final GCN layer, a more modern advanced variant named LSTM is adopted.

A cell state, an output gate, an input gate, and a forget gate make up the structure of a basic LSTM layer, which is depicted in Figure 3-3. The three gates alter the data flow out and in the cell state, which memorizes values throughout a range of time durations. Note that the LSTM-derived outputs with temporal features are regarded as inputs of the dense layer.

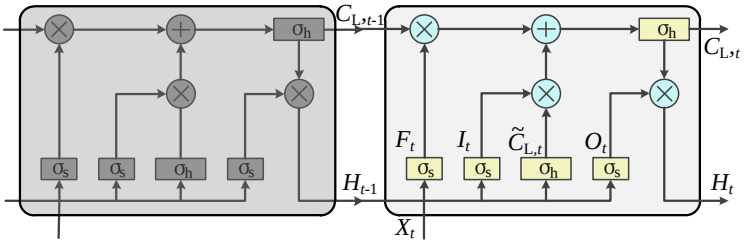


Figure 3-3 The framework of the LSTM. Source: [J3].

3.2.4. ULTRA-SHORT-TERM PREDICTION WITH A HYBRID MODEL

In the end, a dense layer with 1 neuron at the top of the hybrid model predicts the output power based on outputs from the previous LSTM layer. Normally, the activation function is the sigmoid function.

3.2.5. THE SPECIFIC STEPS

To test the model performance, the simulations are conducted by the following 4 steps.

Step 1: Datasets are imported. For example, a wind power dataset may include historical wind powers and weather conditions. Before taking these data as input, the widely used minimum-maximum normalization method is employed to normalize them to values with the same magnitude, ranging from 0 to 1.

Step 2: The historical output powers and weather conditions from time t-h to time h are used to construct graphs with a correlation matrix and historical features, which are considered as inputs of the proposed hybrid model. The output data is the output power of the i^{th} RES at time t+k.

Step 3: The dataset is randomly separated into the training set, validation set, and test set. The parameters and structure of the proposed hybrid model are initialized. Then, the weights of the proposed hybrid model are updated by calculating the loss function and gradient. The weights are repeatedly iterated and updated until the maximum number of iterations allowed is achieved.

Step 4: Finally, the pre-trained hybrid model is tested on the test set through several common metrics, such as MAE and RMSE.

3.3. CASE STUDY

3.3.1. SIMULATION SETTINGS

Simulation and analysis are conducted on two real-world datasets from [80], [83]. The simulation settings are elaborated in [J3]. Due to the page limit, the next sections only analyze the performance of the proposed hybrid model for ultra-short-term predictions of wind and PV powers. The parameter discussion of the proposed hybrid model is not shown here, but can be found in [J3].

3.3.2. RESULT AND DISCUSSION

The proposed hybrid model and baselines (e.g., the GCN in [81], CNN in [29], LSTM in [32], MLP in [31], and a hybrid model of CNN and LSTM) are conducted

30 times independently. Then, Table 3-1 and Table 3-2 display the average simulation results of the wind power dataset and PV power dataset, respectively.

1) The ability to capture temporal and spatial features

A portion of DNNs (e.g., LSTM and the proposed hybrid model) which aim to depict temporal features of power generation curves, typically outperform the conventional MLP. For instance, for the 2-hour wind power prediction, the RMSEs of the LSTM and the proposed hybrid model are decreased by about 26.13% and 33.71%, respectively, compared to the MLP [34]. Similarly, the MAEs are lowered by about 25.39% and 36.27%, respectively. The reason for this is that the non-stationary and complex power generation curves of RESs are challenging for the conventional MLP to handle. Besides, the prediction accuracy of GCN is not the best, because it neglects temporal features of power generation curves, and only takes spatial features into consideration.

Table 3-1 The mean results of various methods for the wind power dataset. Source: [J3].

The time horizon of point predictions	Metrics	CNN	A hybrid of CNN&LSTM	A hybrid of GCN&LSTM	MLP	GCN	LSTM
Time horizon is 1 hour	MAE(MW)	0.99	1.02	0.82	1.11	0.90	1.03
	RMSE(MW)	1.47	1.47	1.20	1.59	1.27	1.49
Time horizon is 2 hours	MAE(MW)	1.63	1.42	1.23	1.93	1.45	1.44
	RMSE(MW)	2.32	1.91	1.75	2.64	1.93	1.95
Time horizon is 3 hours	MAE(MW)	2.22	1.90	1.80	2.35	1.95	1.91
	RMSE(MW)	3.07	2.64	2.44	3.04	2.64	2.67
Time horizon is 4 hours	MAE(MW)	2.62	2.47	2.18	2.76	2.36	2.50
	RMSE(MW)	3.54	3.33	2.93	3.61	3.13	3.38
Time horizon is 5 hours	MAE(MW)	3.15	3.06	2.69	3.16	2.74	3.07
	RMSE(MW)	4.30	4.03	3.52	3.94	3.63	3.99

Table 3-2 The mean results of various methods for the PV power dataset. Source: [J3].

The time horizon of point predictions	Metrics	CNN	A hybrid of CNN&LSTM	A hybrid of GCN&LSTM	MLP	GCN	LSTM
Time horizon is 1 hour	MAE(MW)	0.94	0.86	0.76	0.94	0.88	0.87
	RMSE(MW)	1.84	1.72	1.57	1.97	1.81	1.73
Time horizon is 2 hours	MAE(MW)	1.05	1.04	0.97	1.07	1.04	0.98
	RMSE(MW)	2.08	2.14	1.99	2.25	2.08	2.07
Time horizon is 3 hours	MAE(MW)	1.16	1.11	1.07	1.17	1.12	1.12
	RMSE(MW)	2.39	2.28	2.21	2.38	2.36	2.31
Time horizon is 4 hours	MAE(MW)	1.18	1.16	1.11	1.25	1.14	1.20
	RMSE(MW)	2.48	2.34	2.29	2.60	2.33	2.45
Time horizon is 5 hours	MAE(MW)	1.22	1.21	1.18	1.24	1.18	1.18
	RMSE(MW)	2.48	2.43	2.38	2.55	2.42	2.49

2) The analysis of the ablation study

Obviously, the proposed hybrid model based on spatial features and temporal features has a smaller RMSE and MAE than those of the GCN and LSTM based on single features, suggesting that temporal and spatial features of power generation curves can be captured by the proposed hybrid model accurately. For instance, for the wind power prediction with a 2-hour time horizon, the MAE of the proposed hybrid model is decreased by roughly 15.17% than the GCN that just takes into account spatial features, while the RMSE is lowered by 9.32%. Similarly, for the PV power prediction with a 1-hour time horizon, the RMSE and MAE of the proposed model are reduced by about 9.25% and 12.64%, respectively, compared to the LSTM, which solely account for temporal features [34].

3) The comparative analysis between GCN and CNN

After comparing the performance of the proposed hybrid model with another hybrid model of LSTM and CNN, it is found that the former performs better than the latter. The reason is that CNN cannot handle graphs (i.e., the inputs of ultra-short-term predictions), and it has to reduce the graphs to feature matrices by discarding the correlation matrix, which impairs the prediction accuracy.

Generally, the proposed hybrid model achieves the best prediction results across all evaluation metrics and various time horizons (time horizon ranges from 1 hour to 5 hours) of point predictions, demonstrating the superiority and adaptability of the proposed hybrid model for the ultra-short-term prediction of RESs considering the spatial and temporal features.

3.4. CONCLUSION

To boost the prediction accuracy of the ultra-short-term prediction for RESs, the inputs are modeled as graphs, and then a spectral-based GCN and an LSTM are combined to form a hybrid model, which captures temporal and spatial features from graphs. Simulations are conducted on two datasets.

The results show that the proposed hybrid model achieves the best prediction results across all evaluation metrics (e.g., MAE and RMSE) and various time horizons of point predictions (time horizon ranges from 1 hour to 5 hours) for the ultra-short-term prediction of RESs by considering the spatial and temporal features.

CHAPTER 4. UNCERTAINTY PREDICTION

Scenario prediction and interval prediction are the mainstream methods to depict the uncertainty of power loads and RESs, from two different perspectives. Chapter 4 consists of two parts: 1) A summary of the scenario prediction in [J4], [J5], and [C1]. 2) A summary of the interval prediction in [J6].

J4: **W. Liao**, B. Bak-Jensen, J. R. Pillai, Z. Yang, Y. Wang, and K. Liu, "Scenario Generations for Renewable Energy Sources and Loads Based on Implicit Maximum Likelihood Estimations," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1563-1575, Nov. 2022.

J5: **W. Liao**, L. Ge, B. Bak-Jensen, J. R. Pillai, and Z. Yang, "Scenario prediction for power loads using a pixel convolutional neural network and an optimization strategy," *Energy Reports*, vol. 8, pp. 6659-6671, Nov. 2022.

J6: **W. Liao**, S. Wang, B. Bak-Jensen, J. R. Pillai, Z. Yang and K. Liu, "Ultra-Short-Term Interval Prediction of Wind Power Based on Graph Neural Network and Improved Bootstrap Technique," *Journal of Modern Power Systems and Clean Energy*, E-pub ahead of print, doi: 10.35833/MPCE.2022.000632.

C1: **W. Liao**, B. Bak-Jensen, J. R. Pillai, R. Zhu, and L. Song, "Data-Driven Scenarios Generation for Wind Power Profiles Using Implicit Maximum Likelihood Estimation," *in the 12th International Conference on Applied Energy(ICAE 2020)*, Dec. 2020, pp. 1-5.

4.1. SCENARIO PREDICTION

4.1.1. ABSTRACT AND KEY CONTRIBUTIONS

1) Abstract

Reliable and accurate power load prediction plays a vital role in the operation of ADNs. However, due to the stochastic and fluctuating behaviors of power loads, prediction errors of deterministic point predictions cannot be avoided, posing risks to the operation of ADNs. The scenario prediction is a broadly applied technique to depict stochastic and fluctuating behaviors of power loads by creating a group of potential power load scenarios, allowing system operators to take into account the uncertainty of power loads. To represent the uncertainty of power loads, a new generative model called pixel convolutional neural network (PixelCNN) is

developed for stochastic scenario generations of power loads. Then, a constrained optimization model is proposed to filter a group of potential power load scenarios with similar probability distributions, temporal correlations, and shapes as real ones. Simulation results show that the proposed PixelCNN performs better than other popular generative models, when it comes to the scenario prediction of power loads.

2) Key Contributions

The key contributions are summarized as follows.

- The PixelCNN is presented for the stochastic scenario generation of power loads. The suitable structure of the scenario generator and loss function are designed.

- A constrained optimization model is formulated to identify a set of potential power load scenarios based on both deterministic point predictions and probabilistic scenarios derived from a pre-trained PixelCNN model. With just a few minor parameter adjustments, this optimization model is capable of generating power load scenarios for various time horizons.

- Through numerical simulations, it has been demonstrated that the proposed PixelCNN exhibits superior performance compared to other widely used generative networks, when it comes to the scenario prediction of power loads.

4.1.2. THE PIXELCNN FOR STOCHASTIC SCENARIO GENERATION

1) The framework of the PixelCNN

Figure 4-1 displays the conceptual structure of the PixelCNN. Specifically, by leveraging historical power load curves, a scenario generator is trained to map random noises (e.g., Gaussian noises) into stochastic scenarios that closely resemble real-world patterns.

Typically, the loss function is defined as the difference between generated and real stochastic scenarios [51]. The PixelCNN is free from assuming probability distributions of prediction errors and power load curves. After unsupervised learning, the PixelCNN produces numerous realistic stochastic scenarios of power loads [52], which are narrowed down to a set of appropriate scenarios by an optimization model.

2) The fundamental principle of a scenario generator

Normally, each stochastic scenario is a time series, denoted as $X=\{x_1,x_2,...,x_n\}$. n is the number of data points (i.e., power loads). The following formula is a representation of the joint probability distribution of various points:

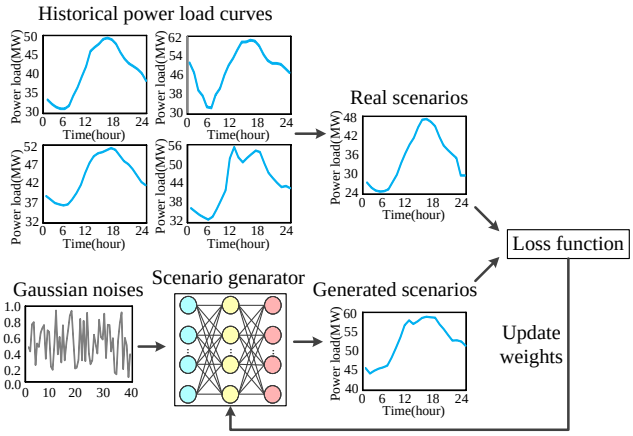


Figure 4-1 The framework of the PixelCNN. Source: [J5].

$$p(X) = p\left(x_1, x_2, \ldots, x_n\right) \tag{4.1}$$

where $p(X)$ denotes the joint probability distribution.

Moreover, the chain rule can be employed to factorize the $p(X)$ into a new form, i.e., the product of several probability distributions:

$$p(X) = p(x_1)p(x_2)\ldots p(x_n) \tag{4.2}$$

where $p(x_n)$ denotes the probability distribution of the n^{th} power load.

The works in [57], [84] demonstrated that power loads are strongly time-dependent. In other words, past, present, and future power loads are interacting with each other. With this in mind, the i^{th} point can be estimated by using the first i-1 points. The stochastic scenario generation of power loads is performed point by point [52]. These are the precise steps:

Step 1: The first conditional probability distribution $p(x_1,x_2)=p(x_1)p(x_2|x_1)$ is used to obtain the second point x_2 by using the first point x_1 as the input data.

Step 2: The second conditional probability distribution $p(x_1,x_2,x_3)=p(x_1)p(x_2|x_1)p(x_3|x_1,x_2)$ is used to obtain the third point x_3 by using the first two points x_1 and x_2 as the input data.

Step 3: Similarly, the $(n\text{-}1)^{\text{th}}$ conditional probability distribution is used to obtain the n^{th} point x_n by using the first n-1 points $x_1,x_2,\ldots,x_{n-1}$ as the input data [52]. As a result, the conditional probability distributions can be utilized to represent the joint probability distribution of various points:

$$p\left(x_1, x_2, \ldots, x_n\right) = \prod_{i=2}^{n} p\left(x_i \left| x_1, \ldots x_{i-1}\right.\right) \tag{4.3}$$

where Π denotes the multiplication operation.

The scenario generator is created by maximizing the likelihood of samples in the training set after constructing the joint probability distribution of various points. A DNN is employed to substitute conditional probability distributions, since it is often challenging to accurately and comprehensively represent the conditional probability distributions of power loads by using mathematical formulas.

3) The structure of a scenario generator

As mentioned in previous chapters, CNN has found broad applications across various domains, and has achieved outstanding achievements thanks to its strong feature extraction capability [78]. In view of this, CNN is adopted to construct the scenario generator.

However, the standard CNN is not a direct substitute for the conditional probability distribution. The reason is that the conditional probability distribution takes the first $i-1$ points as inputs to predict the i^{th} point, whereas the standard convolutional layer employs all data points around the i^{th} point to predict the i^{th} point, as depicted in Figure 4-2(a).

To allow the standard CNN to replace the conditional probability distribution, a mask matrix is utilized to obscure the unnecessary data behind the i^{th} point, as presented in Figure 4-2(b). In the mask matrix, the values in the first i-1 positions are ones, and the values in other positions are zeros. Then, the standard Conv2D layer can be generalized into the masked Conv2D layer:

$$Y_{\text{con}}^{i} = \sigma_{\text{con}}^{i}\left(\left(X_{\text{con}}^{i} \odot M_{\text{S}}\right) * W_{\text{con}}^{i} + B_{\text{con}}^{i}\right) \tag{4.4}$$

where M_{S} is the mask matrix; $\odot$ is the Hadamard product operation; X_{con}^{i} and Y_{con}^{i} are the inputs and outputs of the i^{th} masked Conv2D layer, respectively; * is the convolutional operation; W_{con}^{i} and B_{con}^{i} are the weights and bias vectors of the i^{th} masked Conv2D layer, respectively; and $\sigma_{\text{con}}^{i}(\cdot)$ is the activation function of the i^{th} masked Conv2D layer.

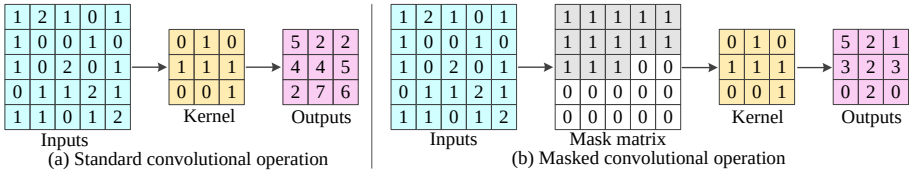


Figure 4-2 The standard and masked convolutional operations. Source: [J5].

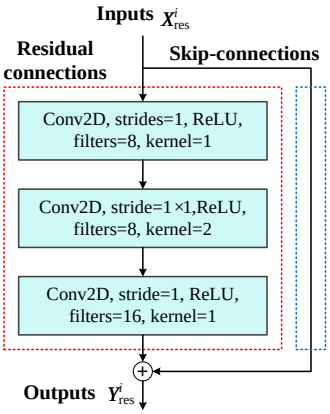


Figure 4-3 The residual block. Source: [J5].

Typically, the masked convolutional layer serves as the input layer of the scenario generator [52], with its outputs being received by the middle layer. Previous publications have demonstrated that adding more convolutional layers can boost the feature learning ability of the model [85]. Nevertheless, an excessive number of convolutional layers could result in problems with over-fitting and degradation of the model performance. This occurs because including too many middle layers leads to a reduction in the amount of original information.

To boost the feature learning ability and address the over-fitting issue, the residual blocks (i.e., a widely used technique in computer vision [85]) are employed to construct the middle layer of the scenario generator. As depicted in Figure 4-3, a residual block consists of two components: skip connection and residual connection. In particular, the skip connection focuses on maintaining the original information by adding the input directly to the output, while the residual connection aims to capture the underlying features through several convolutional layers. In this instance, network degradation can be prevented, since the information in the latter layer is richer than that in the earlier layer. The formula of the residual block is as follows [52].

$$Y^i_{\text{res}} = X^i_{\text{res}} + F_{\text{res}}\left(X^i_{\text{res}}\right) \tag{4.5}$$

where Y_{res}^{i} and X_{res}^{i} are inputs and outputs of the i^{th} residual block, respectively; and $F_{\text{res}}(\cdot)$ denotes several convolutional layers.

To sum it up, the input layer and output layer of the scenario generator are masked Conv2D layer and standard Conv2D layer, respectively. Multiple residual blocks make up the middle layer. After constructing the scenario generator, the one-hot coding is utilized to represent stochastic scenarios by encoding power load curves into binary numbers [52], since the one-hot coding-based loss function is easier to train as compared to the conventional real encoding-based loss function [86]. Then, the categorical cross-entropy loss function is used to update the weights of the scenario generator.

4.1.3. A CONSTRAINED OPTIMIZATION MODEL

It is clear that the various stochastic scenarios generated by the scenario generator are disordered. In other words, there is no correlation between generated stochastic scenarios and deterministic point prediction values. To address this issue, a constrained optimization model is formulated to select specific scenarios, which are close to historical power loads $X_{\text{hist}}=(x_{t-h},...x_{t-1},x_t)$ and point prediction values $X_{\text{pred}}=(x_{t+1},x_{t+2},...,x_{t+m})$. m is the prediction time horizon, and h is the time horizon of the historical data.

The general framework of the constrained optimization model is shown in Figure 4-4. Firstly, the scenario generator of the pre-trained PixelCNN is fed with Gaussian noises to produce numerous disordered stochastic scenarios. Secondly, constraints and the objective function of the constrained optimization model are formulated given inputs, which include synthetic stochastic scenarios $S=(X_1^{'},X_2^{'},\cdots X_N^{'})$, historical power load X_{hist}, and point prediction values X_{pred}. Finally, a set of potential power load scenarios is obtained by solving the constrained optimization model. N denotes the number of synthetic stochastic scenarios.

Further, each stochastic scenario $X_i^{'}$ ranging from time t-h to time h+m can be separated into two parts: the first part $X_{\text{first}}^{'}=(x_{t-h}^{'},\cdots x_{t-1}^{'},x_t^{'})$; and the second part $X_{\text{second}}^{'}=(x_{t+1}^{'},x_{t+2}^{'},\cdots x_{t+m}^{'})$. They can be represented as follows [52].

$$X_i^{'}=\left[X_{\text{first}}^{'},X_{\text{second}}^{'}\right] \tag{4.6}$$

$$X_{\text{first}}^{'}=(x_{t-h}^{'},\cdots x_{t-1}^{'},x_t^{'});\ X_{\text{second}}^{'}=(x_{t+1}^{'},x_{t+2}^{'},\cdots x_{t+m}^{'}) \tag{4.7}$$

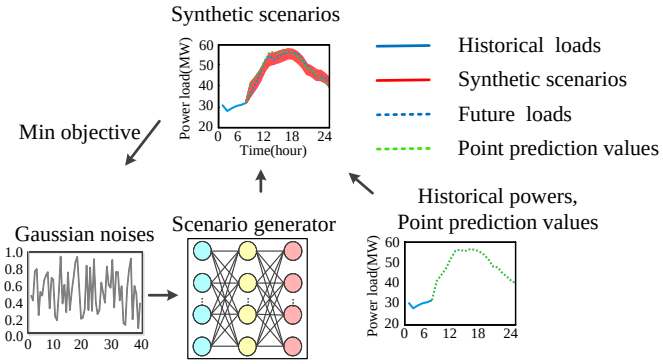


Figure 4-4 The overall framework of the optimization model. Source: [J5].

It is clear that the various stochastic scenarios of power loads generated by the scenario generator are disordered. To search out a set of stochastic scenarios related to historical power loads and point prediction values, the following two requirements need to be met [59].

Requirement 1: The first part X'_{first} of the stochastic scenario and the historical power load X_{hist} should be close to each other.

Requirement 2: The second part X'_{second} should center on the point prediction value X_{pred}.

To filter the specified stochastic scenarios satisfying these two requirements from numerous unordered stochastic scenarios, a constrained optimization model is formulated [52]:

$$\begin{aligned} &\min_{z} \left\| X_{\text{hist}} - X'_{\text{first}} \right\|_2 \\ &s.t. \quad z \in Z \\ &L(\alpha) \le X'_{\text{second}} \le U(\alpha) \end{aligned} \tag{4.8}$$

where α is a parameter to balance the width and coverage percentage of PIs; and $L(\alpha)$ and $U(\alpha)$ are the lower and upper boundaries, respectively.

PIs can be created by using a wide variety of techniques, such as mean-variance estimation [36], ensemble Gaussian model [37], Bayesian model [38], Delta model [39], LUBE [40], and bootstrap techniques [41]. Given that each technique has unique benefits and traits, it is difficult to say which technique is the best. To make it easier to compare baselines with the proposed method, the PIs in [59], [87] are used as a simple example [52]:

$$U(\alpha)=\max\left(X_{\text{pred}}\right)\times\alpha+X_{\text{pred}},L(\alpha)=-\max\left(X_{\text{pred}}\right)\times\alpha+X_{\text{pred}} \tag{4.9}$$

4.1.4. THE SPECIFIC STEPS

To demonstrate the validity of the proposed PixelCNN and constrained optimization model, the simulations are conducted by the following 4 steps.

Step 1: Datasets are imported. For example, a power load dataset may include historical power loads and weather conditions. Before taking these data as input, the widely used minimum-maximum normalization method is employed to normalize them to values with the same magnitude, ranging from 0 to 1. The dataset is randomly separated into the training set, validation set, and test set.

Step 2: The PixelCNN is trained by using the training set and validation set. The parameters and structure of the proposed PixelCNN are initialized. Then, the weights of the proposed PixelCNN are updated by calculating the loss function and gradient. The weights are repeatedly iterated and updated until the maximum number of iterations allowed is achieved. Numerous disordered stochastic scenarios are synthesized by encoding Gaussian noises to the PixelCNN.

Step 3: A point prediction model (e.g., LSTM) is trained by using the training set and validation set. The parameters and structure of the point prediction model are initialized (details of the parameters can be found in [34]). Then, the weights of the point prediction model are updated by calculating the loss function and gradient. The weights are repeatedly iterated and updated until the maximum number of iterations allowed is achieved. The point prediction values of the test set are obtained by using the pre-trained point prediction model.

Step 4: The optimization model is solved by using the genetic algorithm (GA) whose parameters can be found in [52]. Finally, the performance of the proposed PixelCNN and optimization model is measured on the test set through several common metrics, such as prediction interval coverage percentage (PICP) and prediction interval normalized average width (PINAW). Their definitions can be found in [52]. When other parameters are fixed, the larger the PICP is, the better the model performance is. On the contrary, the smaller the PINAW is, the better the model performance is.

4.1.5. CASE STUDY

1) Simulation settings

A real electricity consumption dataset from the University of Texas at Austin [88] is utilized to conduct simulations. The simulation settings have been elaborated in [J5]. Due to the page limit, the next section only presents the simulation results of

scenario prediction by solving the optimization model to select specific scenarios from numerous stochastic scenarios. The parameter discussion, training process, and scenario generation of the PixelCNN are not shown here, but can be found in [J5].

2) Results and analysis of scenario prediction

To demonstrate the effectiveness of the proposed PixelCNN and the constrained optimization model for scenario predictions of power loads, three samples are picked from the test set randomly. Next, the point prediction values of these three samples are predicted by an LSTM. After generating numerous scenarios by the PixelCNN, a set of specific scenarios are filtered by solving the constrained optimization model. For easy visualization, 100 prediction scenarios are randomly selected for each point prediction value and real power load, as depicted in Figure 4-5.

The box-plot reveals that there are significant discrepancies between real values and the point prediction values in data distributions. For instance, the predicted maximum value is substantially lower than the actual peak, when the prediction time horizon is 48 hours, as shown in Figure 4-5(d) to Figure 4-5(f). In this case, the voltage of nodes in ADNs may fall below the lower boundary, if the point prediction value is utilized to solve the solutions of the day-ahead optimal scheduling of power loads in ADNs. Relatively, by adjusting the parameter α, the prediction scenarios are able to encompass the whole range of actual values. Further, these prediction scenarios can be considered as inputs of risk-based decision-making models (e.g., the RDOS in ADNs) to obtain robust solutions,which ensure the security of ADNs.

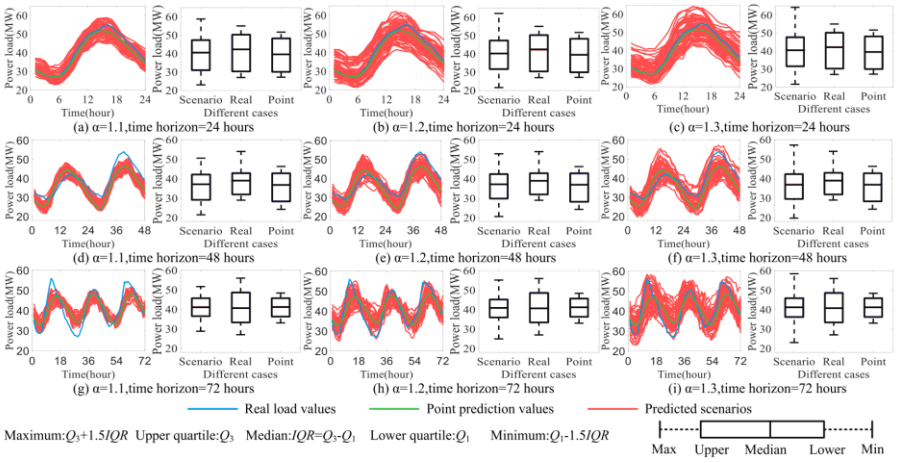


Figure 4-5 Three samples and their prediction results. Source: [J5].

By adjusting the parameter α, it is easy to balance reliability and sharpness. When the parameter α is increased, both PINAW and PICP will rise simultaneously. For instance, when the parameter α is 1.1 and the prediction time horizon is 48 hours, the prediction scenarios of power loads closely resemble point prediction values, whereas prediction scenarios may not cover the actual valley and peak. As the parameter α increases, although prediction scenarios are less concentrated, they have a higher probability of covering all real power loads. Generally speaking, parameter α can be flexibly adjusted flexibly to meet the requirements of the system operator [52].

Besides qualitative findings, the superiority of the proposed model is further confirmed by calculating evaluation metrics of the proposed model and baselines (e.g., VAE in [56], NICE in [57], and GAN in [59]) quantitatively. Each generative network is conducted 30 times, respectively. Then, Table 4-1 shows the average evaluation metrics (e.g., PINAWs and PICPs) of the test set.

The VAE performs better than the NICE, since it has both a smaller PINAW and a larger PICP than the NICE simultaneously. Although the VAE has the smallest PINAW, its PICP is also very limited. As opposed to other generative networks, the GAN exhibits the largest PICP among all models, but it comes at a cost because its PINAW is significantly greater than those of other models. The PINAW of the PixelCNN is not only much smaller than that of GAN, but also its PICP is significantly larger than that of VAE. That is to say, the equilibrium between PICP and PINAW is adequately maintained by the PixelCNN. For instance, when the parameter α is 1.1 and the prediction time horizon is 24 hours [52], comparing the GAN and PixelCNN to the VAE, their PICPs have increased by 28.84% and 26.82%, respectively, while their PINAWs are also elevated by 70.87% and 28.93%, respectively. Therefore, PixelCNN should be a better choice than the GAN. Practically speaking, the prediction scenarios of the VAE fall short of covering sufficient real power loads. It has difficulty in ensuring the secure operation of ADNs, since small PIs may not cover worst-case scenarios. The GAN exhibits a PINAW that is excessively large, leading to an increase in the reserve capacity of the ADN for optimal scheduling of power loads and RESs (i.e., a surplus of reserve capacity results in economic waste.). When considering the PINAW and PICP simultaneously, the PixelCNN performs better than well-known generative networks (e.g., GAN, NICE, and VAE). In other words, PixelCNN strikes a balance between security and economy.

The PICPs of various models decrease as the prediction time horizon increases, when the parameter α is fixed. For instance, when the parameter α is set to 1.2, the PICPs of the 72-hour and 48-hour time horizons are decreased by roughly 10.47% and 6.24%, respectively, relative to the 24-hour time horizon. This is due to the fact that point prediction models become less accurate as prediction time horizons are extended [52], which increases the uncertainty associated with power loads.

Table 4-1 The average results of each method. Source: [J5].

Parameters		PICP(%)				PINAW(MW)			
		NICE	VAE	GAN	PixelCNN	NICE	VAE	GAN	PixelCNN
Prediction time horizon is 24 hours	α=1.1	56.75%	58.33%	87.17%	85.15%	5.75	5.15	8.80	6.64
	α=1.15	85.69%	88.61%	96.94%	94.92%	9.85	9.06	12.84	10.52
	α=1.2	93.50%	96.25%	99.58%	99.37%	13.11	12.50	16.33	13.86
Prediction time horizon is 48 hours	α=1.1	50.13%	49.43%	79.65%	77.60%	5.46	4.88	8.69	5.84
	α=1.15	75.46%	77.99%	91.25%	88.68%	8.93	8.74	12.76	9.65
	α=1.2	88.95%	90.56%	97.08%	96.20%	12.75	12.44	16.37	13.41
Prediction time horizon is 72 hours	α=1.1	44.54%	44.58%	68.29%	66.05%	5.53	4.93	8.11	5.98
	α=1.15	53.87%	54.54%	86.11%	84.45%	8.16	7.20	12.24	8.75
	α=1.2	69.25%	70.19%	93.52%	90.06%	10.42	9.80	15.90	10.10

In practical engineering, the adjustment of parameter α is essential. When adjusting the parameter α, the decision maker has to take full account of PICP and PINAW, which are affected by various factors, such as the generative model, point prediction model, and prediction time horizon. As an illustration, the PixelCNN model produces prediction scenarios with a coverage probability of 96.20% for the actual values, when the parameter α is 1.2 and the prediction time horizon is 48 hours [52]. Further, the prediction scenario-based robust solution prevents the limitations from being crossed with a 96.20% probability. The system operators can increase the parameter α, if they desire more security, but doing so comes at a cost, because robust solutions become more conservative as the parameter α is raised. Before making a choice, it is recommended that system operators perform tests on PICP and PINAW for various parameters to determine their optimal values. Afterward, a reasonable parameter α can be chosen by balancing the economy and safety. In practical engineering, the parameter α can be determined based on the previous data from the adjacent days, since the consumption habits of users today are usually similar to those of adjacent days.

4.1.6. CONCLUSION

To depict the uncertainty of power loads, a new generative network called PixelCNN and a constrained optimization model are developed for stochastic scenario predictions of power loads. The following conclusions are reached by performing simulation and analysis.

The temporal features of power load curves with various time horizons can be accurately depicted by PixelCNN. The probability distribution of power loads can be captured by generative networks. Besides, the main shortcoming of PixelCNN is that it requires substantially more training time and inference time to produce stochastic scenarios than other well-known generative models. Fortunately, The training time and inference time of PixelCNN are few hours and several minutes,

respectively, which are within the acceptable timeframes for the day-ahead optimal scheduling of power loads and RESs in practical engineering (the detail of this part can be found in [J5]).

By adjusting the parameter α, it is easy to balance reliability and sharpness. When the parameter α is small, the prediction scenarios closely resemble point prediction values, whereas prediction scenarios have difficulty covering the actual valley and peak. As the parameter α increases, although prediction scenarios are less concentrated, they have a higher probability of covering all real power loads. Lastly, when considering the PINAW and PICP simultaneously, the PixelCNN performs better than well-known generative networks (e.g., NICE, GAN, and VAE). In other words, PixelCNN strikes a balance between security and economy.

4.2. INTERVAL PREDICTION

4.2.1. ABSTRACT AND KEY CONTRIBUTIONS

1) Abstract

The ultra-short-term prediction of wind power with high accuracy plays a key role in the operation of ADNs. However, due to the stochastic and fluctuating behaviors of wind powers, prediction errors of deterministic point predictions cannot be avoided, posing risks to the operation of ADNs. In view of this, a novel ultra-short-term interval prediction model is proposed to represent the uncertainty of wind powers. In particular, multiple adjacent wind farms and weather conditions are represented as nodes of graphs with correlation matrices and feature matrices. Besides, a GCN is presented to capture complicated spatial features between nodes, and then a bi-directional long short-term memory (Bi-LSTM) is adopted to depict temporal features. Next, an improved bootstrap technique is created to narrow PIs and boost prediction interval coverage percentages effectively. Simulation results demonstrate that the proposed point prediction model is capable of capturing temporal and spatial features from graphs accurately, and the proposed improved bootstrap technique performs better than widely used baselines in terms of prediction interval construction.

2) Key contributions

The key contributions are summarized as follows.

- The bidirectional learning is applied to boost the performance of the point prediction model. The uncertainty of wind powers is considered by constructing suitable PIs.

- The traditional bootstrap technique is improved to balance the width and coverage percentage of PIs flexibly and effectively. Additionally, it is free from the assumption of the particular distribution of prediction errors.

- Multiple adjacent wind farms and weather conditions are represented as nodes of graphs. Then, a GCN and a Bi-LSTM are combined to capture complicated spatial and temporal features from graphs.

4.2.2. THE POINT PREDICTION MODEL WITH BIDIRECTIONAL LEARNING

For sequence learning problems, bidirectional learning is a popular strategy to boost the performance of the classic LSTM, because the output information of time series prediction is a constantly correlated component, rather than the single result of the previous input information [57]. Instead of being trained to depict temporal features in the forward path only, bidirectional learning can help classic LSTM capture temporal information in both the forward and reverse paths. In a variety of sequence learning tasks, such as audio signal processing, the variant of LSTM with bidirectional learning called Bi-LSTM has been shown to outperform the classic LSTM [89]. As a result, the Bi-LSTM is employed to replace the classic LSTM of the hybrid model in Chapter 3.

Note that only the principle of Bi-LSTM is presented here. Additional parts of point prediction can be found in Chapter 3. After replacing the classic LSTM with the Bi-LSTM, the framework of the point prediction model can be generalized from Figure 3-1 into Figure 4-6.

Figure 4-7 shows the structure of a Bi-LSTM unit. Specifically, the complex relationship between the cell state vector at time t-1, the latent state vector at time t-1 (i.e., H_{t-1}), and the feature information at time t (i.e., $X_{L,t}$) is modeled by using the forward LSTM [90]. In the same way, the complex relationship between the cell state vector at time t+1, the latent state vector at time t+1 (i.e., H_{t+1}), and the feature information at time t (i.e., $X_{L,t}$) is modeled by using the backward LSTM. Finally, a mathematical operator $\sigma_{Bi}(\cdot)$, such as concatenation and summation, is utilized to combine the forward LSTM and the backward LSTM to obtain the final outputs.

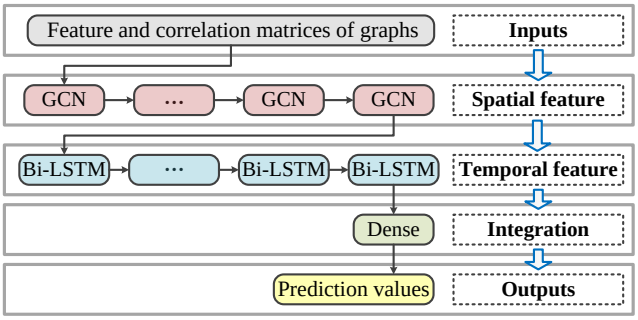


Figure 4-6 A framework of a point prediction model with bidirectional learning. Source: [J6].

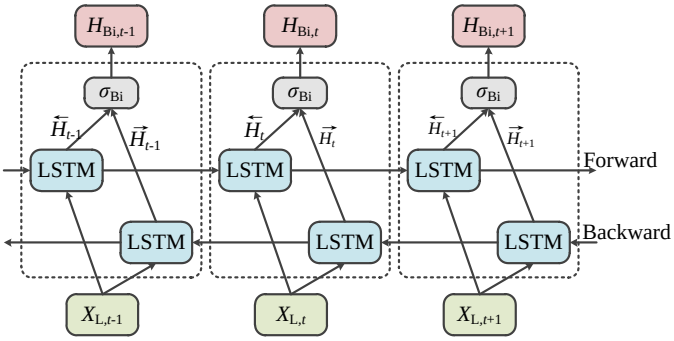


Figure 4-7 The structure of a Bi-LSTM unit. Source: [J6].

4.2.3. THE PREDICTION INTERVAL CONSTRUCTION

After training a point prediction model, an improved bootstrap technique is proposed to represent prediction errors by using upper and lower boundaries.

Although the conventional bootstrap technique is a commonly used and flexible way to represent uncertainty by using PIs, its PIs are too wide [91]. To address this gap, the conventional bootstrap technique is generalized into the improved bootstrap technique, which mainly consists of three steps, as shown in Figure 4-8.

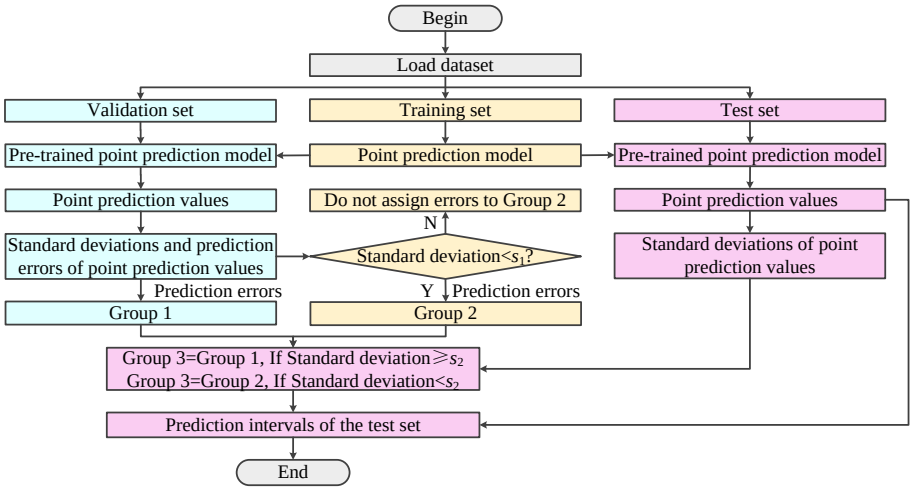


Figure 4-8 Flow chart of the improved bootstrap technique.

Step 1: A point prediction model is trained by using the training set. Then, the standard deviation and prediction error are computed for point prediction values of the validation set. All prediction errors are assigned to a group named group 1. In the meantime, the prediction error of point prediction values is also assigned to a group named group 2, if the corresponding standard deviation is smaller than the parameter s_1.

Step 2: The goal is to construct PIs for the point prediction value of the test set. Firstly, the pre-trained point prediction model is used to obtain the point prediction values of the test set. Then, the standard deviation of the point prediction value is calculated. If the standard deviation is larger than the parameter s_2, all prediction errors in the group 1 are assigned to a group named group 3. Otherwise, all prediction errors in the group 2 are assigned to the group 3. In theory, the parameter s_2 should be smaller than the parameter s_1. If the parameter s_2 is larger than the parameter s_1, the PIs are not wide enough to cover real values.

Step 3: The prediction errors at the percentile α are added to the point prediction values to obtain upper and lower boundaries given a prediction interval nominal confidence (PINC) after sorting prediction errors in the group 3 in descending order. For example, when the system operator wants to use the confidence intervals given 90% PINC, the prediction error at the 5% percentile and the point prediction value are summed to get the lower boundary, and the prediction error at the 95% percentile and the point prediction value are summed to get the upper boundary.

4.2.4. THE SPECIFIC STEPS

To demonstrate the validity of the proposed model for the ultra-short-term interval prediction of wind power, the simulations are conducted by the following 4 steps.

Step 1: Datasets are imported. For example, a wind power dataset may include historical wind powers and weather conditions. Before taking these data as input, the widely used minimum-maximum normalization method is employed to normalize them to values with the same magnitude, ranging from 0 to 1. The dataset is randomly separated into the training set, validation set, and test set.

Step 2: A point prediction model (e.g., a hybrid model consisting of a GCN and a Bi-LSTM.) is trained by using the training set. The parameters and structure of the point prediction model are initialized (details of the parameters can be found in [J6]). Then, the weights of the point prediction model are updated by calculating the loss function and gradient. The weights are repeatedly iterated and updated until the maximum number of iterations allowed is achieved. Next, the point prediction values of the validation set and test set are calculated by using the pre-trained point prediction model.

Step 3: The parameter s_1 and the parameter s_2 are initialized. Then, PIs of point prediction values are constructed by using the improved bootstrap technique given different PINCs.

Step 4: Finally, the model performance is measured on the test set through several common metrics, such as PICP and PINAW. To balance the conflicting metrics (i.e., PICP and PINAW), previous works usually use the coverage width criterion (CWC) to comprehensively evaluate the model performance [92], [93]. The smaller the CWC, the better the model performance.

4.2.5. CASE STUDY

1) Simulation settings

Simulations are conducted on two real-world datasets from [80]. The simulation settings have been elaborated in [J6]. Due to the page limit, the next section only discusses the model performance for ultra-short-term interval prediction of wind powers with a 1-hour prediction time horizon. The parameter discussion, comparative analysis of point prediction models, and prediction accuracy at different prediction time horizons are not shown here, but they can be found in [J6].

2) Comparative analysis of prediction interval construction

The construction of narrow and reliable PIs is the goal of the ultra-short-term interval prediction of wind power. The PINC is initialized from 90% to 99%, since high PINCs are more practically significant compared to low PINCs.

The widely used ensemble Gaussian model in [37] and the conventional bootstrap technique in [41] are viewed as baselines for prediction interval constructions. The ultra-short-term interval prediction of wind power with a 1-hour prediction time horizon is considered as a simple example. After using a point prediction model to get point prediction values, the improved bootstrap technique and baselines are used to construct PIs given different PINCs, respectively. Then, different metrics are calculated for various PIs, as shown in Tables 4-2 to Tables 4-5.

Compared with conventional and improved bootstrap techniques, the width of the PIs generated by the ensemble Gaussian model is the narrowest (i.e., the smallest PINAW), but its PICP is much lower than the expected value (i.e., the PINC). The reason may be that the distribution of prediction errors is not the Gaussian distribution. In other words, the PIs generated by conventional and improved bootstrap techniques are more reliable, because their PICPs are close to the PINC.

Table 4-2 The average prediction results of the test set in Spring. Source: [J6].

Method	PINC(%)	The 1st dataset			The 2nd dataset		
		PICP(%)	PINAW(p.u.)	CWC(p.u.)	PICP(%)	PINAW(p.u.)	CWC(p.u.)
Ensemble	90%	42.1%	0.09	1.074	58.8%	0.208	1.196
Gaussian	95%	49.2%	0.107	1.164	68.7%	0.248	1.176
model	99%	59.3%	0.141	1.167	78.9%	0.327	1.22
Conventional	90%	84.3%	0.300	0.700	80.4%	0.309	0.808
bootstrap	95%	91.1%	0.424	0.939	87.1%	0.404	1.004
technique	99%	99.1%	0.847	0.847	93.8%	0.772	1.774
Improved	90%	84.4%	0.274	0.637	81.1%	0.283	0.725
bootstrap	95%	91.1%	0.371	0.822	87.0%	0.372	0.926
technique	99%	98.7%	0.627	1.262	96.1%	0.633	1.364

Table 4-3 The average prediction results of the test set in Summer. Source: [J6].

Method	PINC(%)	The 1st dataset			The 2nd dataset		
		PICP(%)	PINAW(p.u.)	CWC(p.u.)	PICP(%)	PINAW(p.u.)	CWC(p.u.)
Ensemble	90%	60.3%	0.024	0.129	60.9%	0.163	0.862
Gaussian	95%	64.2%	0.028	0.161	68.6%	0.195	0.929
model	99%	69.7%	0.037	0.199	78.9%	0.257	0.959
Conventional	90%	92.7%	0.168	0.168	92.1%	0.347	0.347
bootstrap	95%	97.0%	0.248	0.248	96.8%	0.466	0.466
technique	99%	100.0%	0.553	0.553	99.7%	0.858	0.858
Improved	90%	92.3%	0.109	0.109	91.9%	0.323	0.323
bootstrap	95%	97.6%	0.160	0.16	97.2%	0.445	0.445
technique	99%	99.7%	0.314	0.314	99.2%	0.716	0.716

Table 4-4 The average prediction results of the test set in Autumn. Source: [J6].

Method	PINC(%)	The 1st dataset			The 2nd dataset		
		PICP(%)	PINAW(p.u.)	CWC(p.u.)	PICP(%)	PINAW(p.u.)	CWC(p.u.)
Ensemble	90%	57.8%	0.127	0.761	51.7%	0.173	1.348
Gaussian	95%	64.4%	0.152	0.853	60.8%	0.207	1.352
model	99%	73.7%	0.200	0.907	76.6%	0.272	1.104
Conventional	90%	85.7%	0.215	0.482	82.5%	0.314	0.771
bootstrap	95%	91.9%	0.345	0.748	92.6%	0.444	0.945
technique	99%	98.3%	0.637	1.296	98.1%	0.741	1.518
Improved	90%	84.2%	0.164	0.383	82.9%	0.262	0.635
bootstrap	95%	92.6%	0.268	0.571	93.0%	0.391	0.823
technique	99%	98.3%	0.457	0.93	98.1%	0.585	1.199

Table 4-5 The average prediction results of the test set in Winter. Source: [J6].

Method	PINC(%)	The 1st dataset			The 2nd dataset		
		PICP(%)	PINAW(p.u.)	CWC(p.u.)	PICP(%)	PINAW(p.u.)	CWC(p.u.)
Ensemble	90%	47.8%	0.100	0.921	59.5%	0.115	0.646
Gaussian	95%	54.3%	0.119	1.031	67.8%	0.138	0.675
model	99%	63.8%	0.157	1.069	77.9%	0.184	0.712
Conventional	90%	81.8%	0.267	0.669	90.0%	0.28	0.56
bootstrap	95%	91.5%	0.368	0.806	94.9%	0.425	0.852
technique	99%	99.2%	0.822	0.822	98.8%	0.805	1.617
Improved	90%	82.4%	0.236	0.579	89.5%	0.234	0.474
bootstrap	95%	91.5%	0.316	0.692	94.2%	0.331	0.676
technique	99%	99.2%	0.612	0.612	99.0%	0.562	0.562

It should be noted that in some cases (e.g., the autumn of the 1st dataset with 99% PINC), the CWCs of bootstrap techniques are larger than those of the ensemble Gaussian method, but this does not imply that the bootstrap techniques are inferior to the ensemble Gaussian model. The reason is that the PICP of the ensemble Gaussian model is much lower than the PINC, while the PICPs of bootstrap techniques are close to the PINC, which ensures the secure operation of ADNs with the expected probability.

Additionally, when the PINAW and PICP of the conventional and improved bootstrap techniques are compared, it can be seen that while they have quite similar PICPs, the conventional bootstrap technique has a larger PINAW than that of the improved bootstrap technique, indicating that the improved bootstrap can greatly reduce the width of the PIs with almost no reduction in PICPs. For instance, for the winter of the 1st dataset with 99% PINC, the PICP of both the conventional and improved bootstrap approaches is 99.2%, but the improved one resulted in a 25.54% decrease in PINAW compared to the conventional one.

The conventional bootstrap technique has a smaller CWC than the improved one in the spring of the 1st dataset with 99% PINC. The main reason is that the definition of CWC given in previous works [92], [93] is inappropriate in some extreme cases. The improved bootstrap technique has a PICP of 98.7%, which is a little under the PINC (i.e., 99%). The CWC increases significantly as a result of the penalty coefficient in the definition. In fact, the PINAW of the improved bootstrap technique is decreased by 25.97%, while its PICP is just 0.4% lower than the conventional one. In exchange for tighter PIs, a slight reduction in the prediction interval coverage percentage is worthwhile. In brief, the improved bootstrap technique still outperforms the conventional one in these cases.

To compare conventional and improved bootstrap techniques visually, two samples are randomly selected from the first and second datasets, and then they are visualized as shown in Figure 4-9. The prediction time horizon is 1 hour.

Obviously, for large volatile regions (e.g., valley, peak, and steep ramps), the PIs of the improved bootstrap technique are close to those of the conventional one, since the improved bootstrap technique focuses on maintaining the width of the PIs in large volatile regions, and only narrowing the PIs in small volatile regions. For instance, the improved bootstrap technique successfully narrows the PIs in elliptically enclosed regions, in which the volatility is small, whereas PIs constructed by the conventional bootstrap technique are overly wide.

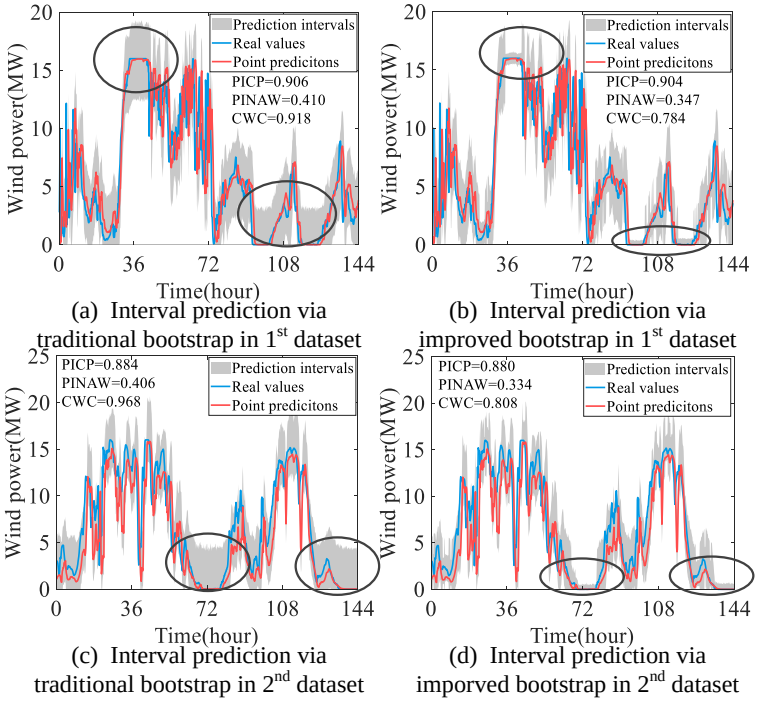


Figure 4-9 Two samples selected from the first and second datasets randomly. Source: [J6].

4.2.6. CONCLUSION

A novel ultra-short-term interval prediction model is proposed to depict the uncertainty of wind powers. The following conclusions are reached by performing simulation and analysis.

The improved bootstrap technique can greatly reduce the width of the PIs with almost no reduction in PICPs. Compared with the ensemble Gaussian model and conventional bootstrap technique, the improved bootstrap technique is the best choice to construct PIs for the ultra-short-term interval prediction of wind powers.

The temporal-spatial features of the wind power generation curve can be accurately captured by the proposed point prediction model. As a result, it achieves the highest prediction accuracy when compared to the popular baselines. Besides, the bidirectional learning helps to boost the accuracy of the ultra-short-term interval prediction of wind power. Furthermore, the proposed model always outperforms the popular baselines for the ultra-short-term interval prediction of wind power, regardless of how the prediction time horizon (it ranges from 30 minutes to 2 hours) is altered (the detail of this part can be found in [J6]).

CHAPTER 5. RISK-BASED DAY-AHEAD OPTIMAL SCHEDULING OF POWER LOADS AND RESS IN ADNS

Robust programming and stochastic programming are the mainstream methods for risk-based decision-making models considering the uncertainty of power loads and RESs, from two different perspectives. Chapter 5 consists of two parts: 1) A summary of the RDOS model of power loads and RESs in ADNs [J7]. 2) A summary of the SDOS of power loads and RESs in ADNs [C2].

J7: W. Liao, S. Wang, B. Bak-Jensen, J. R. Pillai, and Z. Yang, "Bootstrap-Based Prediction Error Estimation for Robust Reactive Power Scheduling of Distribution Networks," *Journal of Modern Power Systems and Clean Energy*, Under Review.

C2: W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, Z. Li, and D. Yang, "Stochastic Day-ahead Optimal Scheduling of Active Distribution Networks with Renewable Energy Sources and Electric Vehicles," *in the 8th Asia Conference on Power and Electrical Engineering (ACPEE 2023)*, Apr. 2023, pp. 1-8 (Accepted).

5.1. ROBUST DAY-AHEAD OPTIMAL SCHEDULING OF POWER LOADS AND RESS IN ADNS

5.1.1. ABSTRACT AND KEY CONTRIBUTIONS

1) Abstract

Due to the stochastic and fluctuating behaviors of power loads and RESs, prediction errors of deterministic point prediction models cannot be avoided, posing risks to the operation of ADNs. To obtain the day-ahead optimal scheduling scheme of power loads and RESs in ADNs considering risks, a robust programming model is proposed. Firstly, a DDOS model of power loads and RESs in ADNs is formulated. Then, an improved bootstrap technique is employed to represent prediction errors of power loads and RESs by constructing PIs. Next, prediction errors and point prediction values are combined to form new kinds of worst-case scenarios, which are used to encompass the entire range of possible scenarios. In this case, the DDOS model is generalized into the RDOS model considering the operational constraints in worst-case scenarios. Lastly, the RDOS model is solved to get a robust solution, which guarantees the secure operation of ADNs for all possible scenarios theoretically. Simulation results show that the improved bootstrap technique performs better than widely used baselines in terms of worst-case scenario

constructions. Additionally, the improved bootstrap technique-based RROS model also strikes a good balance between security and economy.

2) Key contributions

The key contributions are summarized as follows.

- Point prediction values and prediction errors are combined to form new kinds of worst-case scenarios, which are used to encompass the entire range of possible scenarios.

- An RDOS model of power loads and RESs in ADNs is formulated to completely take into consideration the prediction errors, arising from the stochastic and fluctuating behaviors of power loads and RESs.

- The DDOS and RDOS models are contrasted in terms of security and economics.

5.1.2. DETERMINISTIC DAY-AHEAD OPTIMAL SCHEDULING MODEL

1) Objective function

The DDOS model focuses on decreasing the power loss and keeping the voltage amplitude within the allowable range by managing various power devices. Without loss of generality, the objective function of the DDOS model is specified to minimize the operational cost of a day [94] by managing classical power devices, such as shunt CBs, transformers, and RESs. For example, the operational cost includes the energy cost, change cost of transformers, and change cost of shunt CBs. Note that the mathematical models of other power devices (e.g., energy storage, soft open points, and electrical boilers) not discussed here are similar to those of classical power devices. Therefore, the proposed models can be easily generalized to ADNs with more power devices.

2) Constraints

The DDOS model should satisfy a large number of operational constraints [60], such as power flow constraints, current constraints, voltage constraints, tap constraints of transformers, and capacity constraints of RESs [95]. ADNs may also include constraints for possible load activation and reactive power consumption of loads, which will be discussed in the extension work.

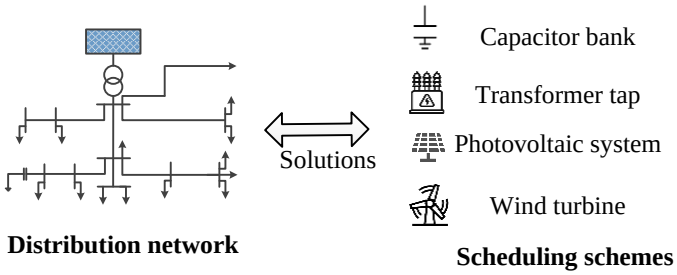


Figure 5-1 The framework of the DDOS model. Source: [J7].

To sum up, Figure 5-1 shows the schematic architecture of the DDOS model. The inputs of the DDOS model are point prediction values of power loads and active power output of RESs. The outputs of the DDOS model include voltages and power loss, which can be used to obtain the operational cost or check constraints. Given a set of operational constraints, the decision variables include the reactive powers supplied by RESs, the reactive powers supplied by shunt CBs, and the tap position of transformers at each time hour. In the extension work, the active power from RESs as well as power loads may be decision variables to optimize the operation of ADNs.

5.1.3. WORST-CASE SCENARIO CONSTRUCTION

Conventionally, worst-case scenarios are represented by two extreme scenarios [70]: For the first extreme scenario, the power loads are taken as the maximum, and the output powers of RESs are taken as the minimum. For the second extreme scenario, the power loads are taken as the minimum, and the output powers of RESs are taken as the maximum. Although this traditional method is simple and effective, its solutions are too conservative. In other words, it pays a high price to secure ADNs.

To balance economy and security, the improved bootstrap technique in Chapter 4 is used to produce PIs of point prediction values. Then, new kinds of worst-case scenarios are constructed by combining point prediction errors and PIs of prediction errors. Typically, the worst-case scenarios include the following two combinations between power loads and output powers of RESs [71].

Case 1: The point prediction values of output powers of RESs are larger than the real values, and the point prediction values of power loads are tinier than the real values. In this worst-case scenario, the voltages of nodes may exceed the lower boundary. This worst-case scenario can be formulated as:

$$\begin{cases} P_{\text{Case1},i,t} = P_{i,t} + \Delta P_{\text{UP},i,t} & i = 1,2,\cdots m \\ Q_{\text{Case1},i,t} = Q_{i,t} + \Delta Q_{\text{UP},i,t} & i = 1,2,\cdots m \end{cases} \tag{5.1}$$

$$P_{\text{Case1RES},i,t} = P_{\text{RES},i,t} + \Delta P_{\text{Low RES},i,t} \qquad i = 1,2,\cdots n_{\text{R}} \tag{5.2}$$

where m is the number of nodes; n_R is the number of RESs; $Q_{i,t}$ and $P_{i,t}$ are the reactive and active power loads (point prediction values) of the i^{th} node at time t, respectively; $\Delta Q_{Up,i,t}$ is the upper boundary of prediction errors for the reactive power load of the i^{th} node at time t; $P_{Case1,i,t}$ is the active power load of the i^{th} node at time t for the case 1(as shown in the upper boundary of the blue area in Figure 5-2); $Q_{Case1,i,t}$ is the reactive power load of the i^{th} node at time t for the case 1; $\Delta P_{Up,i,t}$ is the upper boundary of prediction errors for the active power load of the i^{th} node at time t; $\Delta P_{LowRES,i,t}$ is the lower boundary of prediction errors of active power supplied by the i^{th} RES at time t; and $P_{Case1RES,i,t}$ is the active power supplied by the i^{th} RES at time t for the case 1 (as shown in the lower boundary of the blue area in Figure 5-3).

Case 2: The point prediction values of output powers of RESs are tinier than the real values, and the point prediction values of power loads are larger than the real values. In this worst-case scenario, the voltages of nodes may exceed the upper boundary. This worst-case scenario can be formulated as:

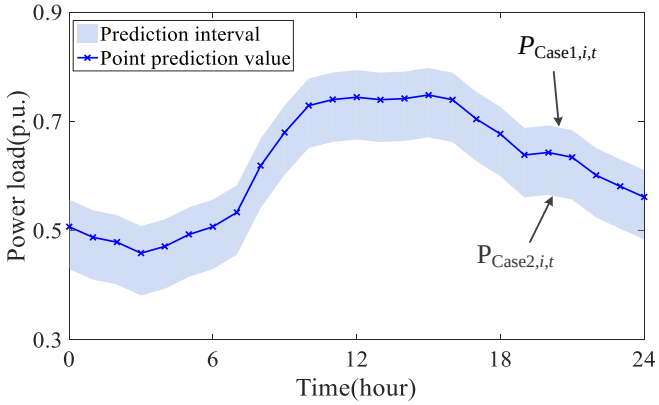


Figure 5-2 The lower boundary and upper boundary of power loads. Source: [J7].

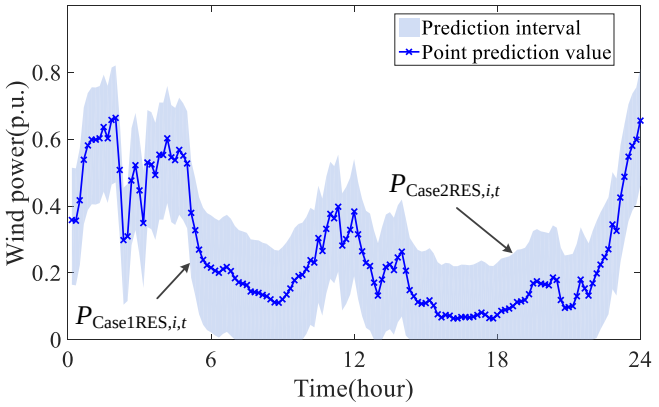


Figure 5-3 The lower boundary and upper boundary of wind powers. Source: [J7].

$$\begin{cases} P_{\text{Case 2},i,t} = P_{i,t} + \Delta P_{\text{Low},i,t} & i = 1,2,\cdots m \\ Q_{\text{Case 2},i,t} = Q_{i,t} + \Delta Q_{\text{Low},i,t} & i = 1,2,\cdots m \end{cases} \tag{5.3}$$

$$P_{\text{Case 2 RES},i,t} = P_{\text{RES},i,t} + \Delta P_{\text{Up RES},i,t} \qquad i = 1,2,\cdots n_{\text{R}} \tag{5.4}$$

where $\Delta Q_{\text{Low},i,t}$ is the lower boundary of prediction errors for the reactive power load of the i^{th} node at time t; $P_{\text{Case2},i,t}$ is the active power load of the i^{th} node at time t for the case 2 (as shown in the lower boundary of the blue area in Figure 5-2); $Q_{\text{Case2},i,t}$ is the reactive power load of the i^{th} node at time t for the case 2; $\Delta P_{\text{Low},i,t}$ is the lower boundary of prediction errors for the active power load of the i^{th} node at time t; $\Delta P_{\text{UpRES},i,t}$ is the upper boundary of prediction errors for active power supplied by the i^{th} RES at time t; and $P_{\text{Case2RES},i,t}$ is the active power supplied by the i^{th} RES at time t for the case 2 (as shown in the upper boundary of the blue area in Figure 5-3).

Note that $\Delta P_{\text{Up},i,t}$, $\Delta Q_{\text{Up},i,t}$, $\Delta P_{\text{LowRES},i,t}$, $\Delta P_{\text{Low},i,t}$, $\Delta Q_{\text{Low},i,t}$, and $\Delta P_{\text{UpRES},i,t}$ are constructed by using the improved bootstrap technique in Chapter 4.

5.1.4. ROBUST DAY-AHEAD OPTIMAL SCHEDULING MODEL

After using the improved bootstrap technique to produce two worst-case scenarios, the next step is to construct an RDOS model based on these two worst-case scenarios. Figure 5-4 presents the framework of the RDOS model.

In particular, the worst-case scenarios have a low probability of occurrence, and the point prediction value represents the most likely event. In other words, the point prediction value is the expected value of the future scenario. Therefore, the objective function of the DDOS model can also be treated as the objective function of the RDOS model (i.e., the RDOS model minimizes the operational cost with point prediction values as inputs), but the RDOS model also needs to take into account the constraints in two worst-case scenarios.

In summary, the RDOS model mainly consists of two parts. Firstly, the point prediction values are considered as inputs to obtain operational cost. Secondly, two worst-case scenarios are considered as inputs to ensure that the ADN meets operational constraints in worst-case scenarios.

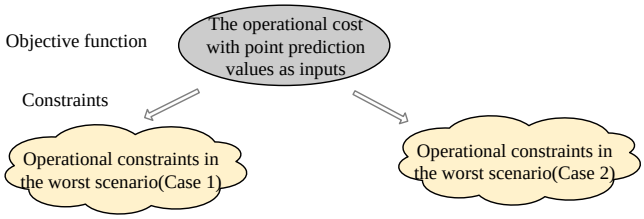


Figure 5-4 The framework of the RDOS model. Source: [J7].

So far, the RDOS model of ADNs has been formulated. By solving this model, robust solutions can be obtained, and they guarantee the secure operation of ADNs for all possible scenarios given a large enough PINC theoretically.

5.1.5. CASE STUDY

1) Simulation settings

Simulations and analysis are executed on the modified IEEE 33-bus distribution network, the simulation settings can be found in [J7]. Various power devices, such as PV systems, shunt CBs, transformers, and wind turbines (WTs), are integrated to the modified IEEE 33-bus distribution network, the topology of which is presented in Figure 5-5.

Due to the page limit, the next section mainly discusses the voltage differences between the DDOS model and the RDOS model, and compares the security and economy of the DDOS and RDOS models. The difference between the DDOS model-based solution and the RDOS model-based solution is not shown here, and it can be found in [J7].

2) Comparative analysis of voltages between DDOS and RDOS models

Here, the voltage differences between the DDOS model and the RDOS model are explained by performing simulations on two new kinds of worst-case scenarios mentioned before.

Case 1: Voltages of nodes exceed the lower boundary considering uncertainties of power loads and RESs.

Firstly, a specific sample with a large prediction error is chosen from the test set. Secondly, the DDOS model and RDOS model are solved to obtain solutions, as presented in Table 5-1. After performing the power flow analysis, the voltage of each node is obtained, as depicted in Figure 5-6.

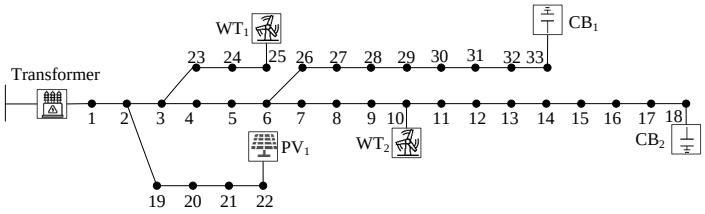


Figure 5-5 The framework of the modified IEEE 33-bus distribution network. Source: [J7].

Table 5-1 The tap position and reactive power output plans of DDOS and RDOS models in case 1. Source: [J7].

Models	Tap position	CB_1 (kvar)	CB_2 (kvar)	WT_1 (kvar)	WT_2 (kvar)	PV_1 (kvar)
DDOS	4×1.25%	500	200	299.17	295.69	113.83
RDOS	4×1.25%	500	500	299.52	294.37	230.58

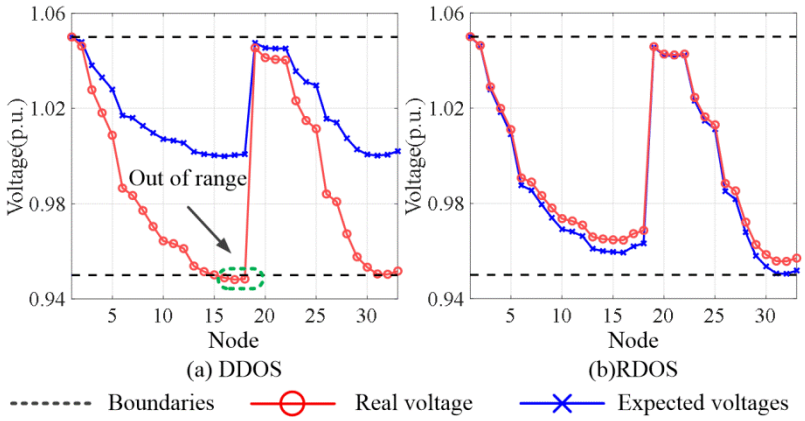


Figure 5-6 The voltage of each node in case 1. Source: [J7].

If the point prediction model is perfect without prediction errors, the DDOS model-based solution ensures that the voltage of each node falls within the allowable range, as presented by the blue line in Figure 5-6(a). Nevertheless, prediction errors of deterministic point prediction models cannot be avoided due to the stochastic and fluctuating behaviors of power loads and RESs. For this sample, the point prediction values of output powers of RESs are larger than the real values, and the point prediction values of power loads are smaller than the real values. As a result, reactive power outputs arranged by the DDOS model-based solution are not enough to maintain voltages, and the voltages at some nodes (e.g., the 16th node to the 18th node) are below the lower boundary, as presented by the red line in Figure 5-6(a).

In contrast, the RDOS model constructs PIs through the improved bootstrap technique that fully accounts for prediction errors of power loads and RESs, and then supplies enough reactive power to guarantee that the voltage at each node falls within the allowed range in the worst-case scenario, as shown by the red line in Figure 5-6(b).

Case 2: Voltages of nodes exceed the upper boundary considering uncertainties of power loads and RESs.

There could be instances that are the opposite of case 1. In particular, another specific sample with a large prediction error is chosen from the test set. Secondly, the DDOS model and RDOS model are solved to obtain solutions, as presented in Table 5-2. After performing the power flow analysis, the voltage of each node is obtained, as depicted in Figure 5-7.

As demonstrated by the blue line in Figure 5-7(a), the DDOS model-based solution makes an effort to guarantee the secure operation of ADNs for the deterministic point prediction scenario. However, for this specific sample, the point prediction values of output powers of RESs are tinier than the real values, and the point prediction values of power loads are larger than the real values. Accordingly, prediction errors of the point prediction model lead voltages at some nodes (e.g., the 20^{th} node to the 22^{nd} node) to exceed the upper boundary, as represented by the red line in Figure 5-7(a).

In contrast, the RDOS model takes prediction errors of the point prediction model into consideration. By reducing leading reactive powers and providing lagging reactive powers, the RDOS model-based solution ensures that the voltage at each node falls within the allowed range in the worst-case scenario, as shown by the red line in Figure 5-7(b).

Table 5-2 The tap position and reactive power output plans of DDOS and RDOS models in case 2. Source: [J7].

Models	Tap position	CB_1 (kvar)	CB_2 (kvar)	WT_1 (kvar)	WT_2 (kvar)	PV_1 (kvar)
DDOS	4×1.25%	500	100	89.46	124.66	97.95
RDOS	4×1.25%	500	200	20.74	17.37	-160.79

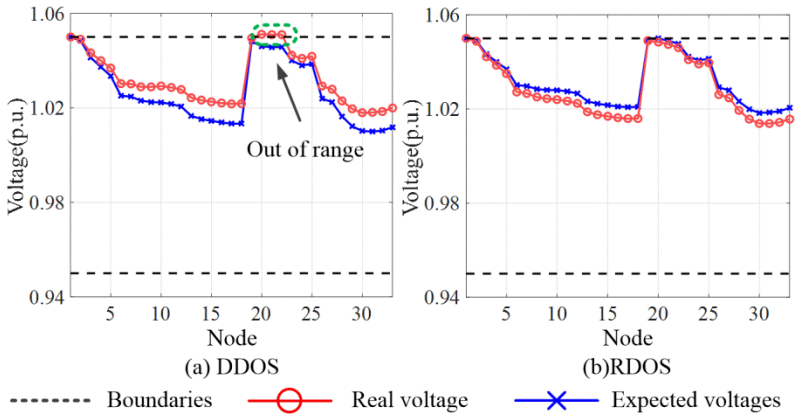


Figure 5-7 The voltage of each node in case 2. Source: [J7].

3) Comparative analysis of security and economy

Here, the security and economy of the DDOS and RDOS models are compared by using all samples in the test set.

In terms of the DDOS model, a point prediction model is applied to estimate the point prediction values of power loads and RESs. Then, the DDOS model is solved given these point prediction values as inputs.

In terms of RDOS models, two worst-case scenarios are constructed by the conventional bootstrap technique, improved bootstrap technique, ensemble Gaussian model, and the traditional method in [70], respectively. In particular, there is no need to estimate point prediction values in the traditional method in [70], which simply uses the maximum and minimum values to construct two worst-case scenarios: In the first worst-case scenario, the power loads are taken as the maximum, and the output powers of RESs are taken as the minimum. In the second worst-case scenario, the power loads are taken as the minimum, and the output powers of RESs are taken as the maximum. The conventional bootstrap technique, improved bootstrap technique, and ensemble Gaussian model require point prediction values to estimate prediction errors. Then, worst-case scenarios are produced by combining prediction errors and point prediction values given the specific PINC.

The GA is used to solve these RDOS models after creating worst-case scenarios. The average operational cost and percentage of unsafe scenarios of the test set are displayed in Table 5-3.

When there are substantial prediction errors, the operational constraints cannot be satisfied by the DDOS model, and the DDOS model-based solution can only guarantee the security of ADNs for the majority of situations (the proportion of risky situations is 12.86%), since the DDOS model ignores prediction errors of power loads and RESs.

The security of ADNs cannot always be guaranteed, even if the ensemble Gaussian model-based RDOS model and the conventional bootstrap technique-based RDOS model both somewhat reduce the fraction of unsafe scenarios. A relatively small proportion of insecure scenarios are produced by the improved bootstrap technique-based RDOS model. For instance, there are 0% insecure scenarios, when the PINC is 99%. This demonstrates how an improved bootstrap technique-based RDOS model guarantees that ADNs always meet the operational constraints.

Table 5-3 The average operational cost and percentage of unsafe scenarios of the test set. Source: [J7].

Models	PINC (%)	Average percentage of insecure scenarios (%)	Average operational costs
DDOS	None	12.86%	465.49 ¥(i.e., 66.26$)
The traditional method-based RDOS model [70]	None	0.00%	525.51 ¥(i.e., 74.81$)
The ensemble Gaussian model-based RDOS	90%	8.21%	477.15 ¥(i.e., 67.92$)
The conventional bootstrap technique-based RDOS		3.75%	490.14 ¥(i.e., 69.77$)
The improved bootstrap technique-based RDOS		1.67%	494.31 ¥(i.e., 70.37$)
The ensemble Gaussian model-based RDOS	99%	4.98%	483.10 ¥(i.e., 68.77$)
The conventional bootstrap technique-based RDOS		0.14%	503.56 ¥(i.e., 71.68$)
The improved bootstrap technique-based RDOS		0.00%	506.98 ¥(i.e., 72.17$)

Although the operational costs of the conventional bootstrap technique-based RDOS model and ensemble Gaussian model-based RDOS model are marginally cheaper than those of the proposed model, they are unable to guarantee the security of ADNs, even when the PINC is 99%. On the contrary, the traditional model-based RDOS model in [70] has zero percent insecure scenarios, but it has the highest operational costs (i.e., over-conservatism).

Additionally, the improved bootstrap technique-based RDOS model can be seen as a good compromise strategy that strikes a good balance between economy and security. For instance, when the PINC is 99%, compared to the DDOS model, the mean operational cost of the improved bootstrap technique-based RDOS model is only 8.91% higher, but the proportion of insecure situations is decreased by 12.86%. In brief, the improved bootstrap technique-based RDOS model is safer than the conventional bootstrap technique-based RDOS model, ensemble Gaussian model-based RDOS model, and DDOS model, while it is more economical than the traditional model-based RDOS model in [70].

5.1.6. CONCLUSION

To obtain the optimal solutions considering risks, an RDOS model of power loads and RESs in ADNs is proposed. The following conclusions are reached by performing simulation and analysis.

The DDOS model-based solution can only guarantee the security of ADNs for the majority of situations, since the DDOS model ignores prediction errors of power loads and RESs. In contrast, the improved bootstrap technique-based RDOS model takes prediction errors into consideration, and ensures that operational constraints are always met in theory when a large enough PINC is given.

The DDOS model-based solutions are occasionally unworkable, since the production schedules of RESs exceed the maximum available reactive powers. Contrarily, the RDOS model-based solutions guarantee that production schedules of RESs never exceed the maximum available reactive powers (the detail of this part can be found in [J7]).

The improved bootstrap technique-based RDOS model can be seen as a good compromise strategy that strikes a good balance between economy and security. In particular, the improved bootstrap technique-based RDOS model is safer than the conventional bootstrap technique-based RDOS model, ensemble Gaussian model-based RDOS model, and DDOS model, while it is more economical than the traditional model-based RDOS model in [70].

5.2. STOCHASTIC DAY-AHEAD OPTIMAL SCHEDULING OF POWER LOADS AND RESS IN ADNS

5.2.1. ABSTRACT AND KEY CONTRIBUTIONS

1) Abstract

Due to the stochastic and fluctuating behaviors of power loads and RESs, prediction errors of deterministic point prediction models cannot be avoided, posing risks to the operation of ADNs. To obtain the day-ahead optimal scheduling scheme of power loads and RESs in ADNs considering risks, a stochastic programming model is proposed. Specifically, a DDOS model of ADNs is formulated to minimize the power loss and voltage deviation by managing various power devices, such as EVCS, transformers, RESs, and HPs. To depict EVCSs, a number of statistical models are created to describe the charging curves of BEVs with and without participating in the demand response. Then, the Gaussian distribution model and K-means are utilized to depict the uncertainty of power loads and RESs by constructing classical stochastic scenarios. In this case, the DDOS model is generalized into the SDOS model by taking classical stochastic scenarios as inputs. Simulation results demonstrate that the proposed SDOS model is capable of depicting the uncertainty of power losses and voltages accurately from a probabilistic perspective.

2) Key contributions

The key contributions are summarized as follows.

- Statistical models are created to describe the charging curves of BEVs with and without participating in the demand response. In addition, the impact of BEVs with and without demand response on voltages and power losses in ADNs is discussed.

- The mathematical model of HPs is formulated. Besides, the impact of HPs with and without demand response on voltages and power losses in ADNs is discussed.

- A SDOS model is formulated to completely take into consideration the prediction errors, arising from the stochastic and fluctuating behaviors of power loads and RESs. The difference between the DDOS model and SDOS model is demonstrated.

5.2.2. DETERMINISTIC DAY-AHEAD OPTIMAL SCHEDULING MODEL

1) Objective function

The DDOS model of power loads and RESs in ADNs focuses on decreasing the power loss and keeping the voltage amplitude within the allowable range by managing various power devices, such as EVCSs, transformers, RESs, and HPs. Without loss of generality, the objective function of the DDOS model is specified to achieve minimum power loss and voltage deviation. Next, the linear weighted sum approach is adopted to combine the multiple objectives into a single goal after normalizing the power loss and voltage deviation.

2) Constraints

Normally, the DDOS model of ADNs should satisfy a large number of operational constraints, such as power flow constraints, current constraints, voltage constraints, tap constraints of transformers, constraints of EVCSs, constraints of HPs, and capacity constraints of RESs. Except for the HPs, the other operational constraints have been formulated in [C2]. Therefore, only the model and control strategy of HPs are supplemented here.

HPs are usually considered as flexible loads, which are often used to regulate the indoor temperature. To ensure thermal comfort, the room temperature should be greater than a lower boundary and less than an upper boundary [96]:

$$T_{\text{Room,min}} \le T_{\text{Room},t} \le T_{\text{Room,max}} \tag{5.5}$$

where $T_{\text{Room,min}}$ and $T_{\text{Room,max}}$ are the lower boundary and upper boundary of the room temperature, respectively; and $T_{\text{Room},t}$ is the room temperature at time t.

Normally, the room temperature is influenced by the input thermal power of the radiator as well as by the ambient temperature. As a simple example, the room temperature at time t can be represented by an indoor thermal dynamics model [97]:

$$T_{\text{Room},t} = T_{\text{Room},t-1} + \left(\frac{T_{\text{Amb},t} - T_{\text{Room},t-1}}{1000 R_{\text{house}}} + P_{\text{Rad,t}} \right) \Delta t / C_{\text{house}} \tag{5.6}$$

$$0 \le P_{\text{Rad,t}} \le P_{\text{Rad,max}} \tag{5.7}$$

where Δt is the time horizon (e.g., 1 hour); T_{Amb} is the ambient temperature; C_{house} is the heat capacity; R_{house} is the thermal resistance of the house; $P_{\text{Rad,max}}$ is the maximum input thermal power of the radiator; and $P_{\text{Rad},t}$ is the input thermal power of the radiator at time t, which is a decision variable to be optimized.

Ideally, the input thermal powers of all HPs in 24 hours are optimized at the same time. However, this will cause the dimension of decision variables to be too long, and the model is difficult to be solved. For example, if the ADN includes 100 HPs, the dimension of decision variables is 2400.

To simplify the model, the input thermal powers of all the HPs are not optimized simultaneously. Alternatively, the input thermal power of each HP is optimized separately. For example, if the ADN includes 100 HPs, the GA is executed 100 times. In this case, each time the GA is executed, the dimensionality of the decision variables is 24.

To reduce power loss and voltage deviation, it aims to reduce the thermal power of the HP given the room temperature constraint. Therefore, when the thermal power of one HP is optimized, the objective function can be:

$$\min F_{\text{HP}} = \sum_{t=1}^{24} P_{\text{Rad},t} \tag{5.8}$$

Note that this is just one simple method to control HPs, and other models can be considered in future work. For example, reinforcement learning may be developed to solve this complex model with high-dimensional decision variables.

To sum up, when the input thermal power of one HP is optimized, the objective function is equation (5.8) subject to equation (5.5) to equation (5.7). After determining the input thermal power of each HP, the other decision variables (i.e.,

tap ratios of transformers, reactive power outputs of RESs, and reactive power outputs of EVCSs at each time hour) are optimized simultaneously. Specifically, the objective function aims to achieve minimum power loss and voltage deviation, subject to operational constraints (e.g., power flow constraints, current constraints, voltage constraints, tap constraints of transformers, constraints of EVCSs, and capacity constraints of RESs).

5.2.3. STOCHASTIC DAY-AHEAD OPTIMAL SCHEDULING MODEL

1) The simple framework of the SDOS model

Figure 5-8 compares the DDOS model and SDOS model in a straightforward manner.

In terms of the DDOS model, the inputs include point prediction values of power loads and RESs, and the outputs include deterministic estimated values of the power loss and voltage deviation.

In terms of the SDOS model, the inputs generalize from point prediction values into a set of stochastic scenarios, and the outputs also generalize from deterministic estimated values into probability distributions of the power loss and voltage deviation. The expected values (mean values) of voltage, current, and power loss are obtained after feeding stochastic scenarios to the SDOS model, and these expected values should satisfy their constraints.

2) Stochastic scenario generations

Here, it will explain how to use a K-means model and a Gaussian distribution model to expand point prediction values into stochastic scenarios.

With point prediction values of power loads and RESs as inputs, stochastic scenarios can be produced by using a variety of techniques, such as flow-based generative networks [57], generative moment matching networks [84], pixel convolutional neural networks [52], and Gaussian distribution models [98].

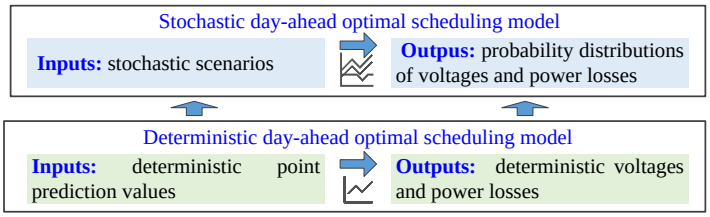


Figure 5-8 The relationship between DDOS and SDOS models. Source: [C2].

As a simple example, point prediction values are generalized into a variety of stochastic scenarios by combining point prediction values and prediction errors from the commonly used Gaussian distribution model, because prediction errors of power loads and RESs roughly follow Gaussian distributions, according to findings from earlier research [98].

After sampling the Gaussian distribution, point prediction values can be generalized into thousands of stochastic scenarios, which will pose a heavy computational burden when the SDOS model is solved [65]. To solve the SDOS model given limited computational resources, the K-means model in [2] is utilized to reduce numerous stochastic scenarios to hundreds of (e.g., 100) classical scenarios.

5.2.4. THE CHARGING CURVES OF BATTERY ELECTRIC VEHICLES

To depict the charging curves of EVCSs, a number of statistical models are created to describe the charging curves of BEVs with and without participating in the demand response. It should be noted that this is merely one of the modeling techniques [99], and the overall optimization is independent of the charging model that is actually used. BEVs are divided into the following four categories to make descriptions easier:

BEV_1: BEVs only charge and do not discharge. Charging behaviors only depend on user habits. BEVs do not participate in the demand response by shifting charging behavior to off-peak hours.

BEV_2: BEVs participate in the demand response. BEVs have not only charging behaviors, but also discharging behaviors, known as vehicle-to-grid (V2G). BEVs do not shift their charging behaviors to off-peak hours, and do not shift discharging behaviors to peak hours. The charging and discharging behaviors only depend on user habits.

BEV_3: BEVs only charge and do not discharge. BEVs participate in the demand response by shifting charging behaviors to off-peak hours.

BEV_4: BEVs have not only charging behaviors, but also discharging behaviors. BEVs participate in the demand response by shifting charging behaviors to off-peak hours, and shifting discharging behaviors to peak hours.

According to previous works in [100], [101], the statistical models are formulated to PDFs of BEVs, which are sampled by using the Monte Carlo method to obtain the expected charging curves of BEVs.

1) The charging curves of BEV_1 and BEV_2

As a simple example, Figure 5-9 presents the expected values of the charging curves of the BEV_1 and BEV_2.

Due to the fact that the majority of users commute to work in the morning, the charging power of the BEV_1 is low during this time. In the afternoon, the charging power progressively grows, because they are plugged into charging stations in the company for charging behaviors. The charging power of the BEV_1 slowly decreases from night until morning, since the majority of them are already charged fully.

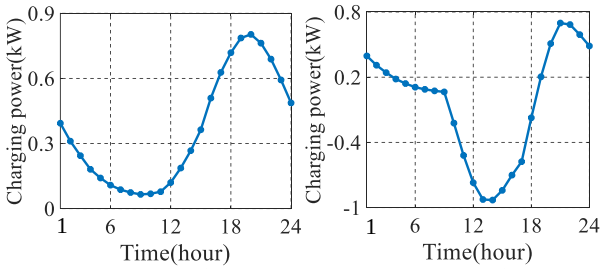
The charging curves of the BEV_1 and the BEV_2 look similar to each other, and they differ in that the BEV_2 has an additional discharging behavior. For instance, the majority of users arrive at the office before 10:00 and leave around 18:00. As a result, the BEV_2 may discharge during this window.

2) The charging curves of BEV_3 and BEV_4

The charging behaviors or discharging behaviors of the BEV_1 and BEV_2 only depend on user habits. They do not shift their charging time or discharging time. Here, the charging curves of the BEV_3 and BEV_4 are discussed. They can shift their charging time or discharging time.

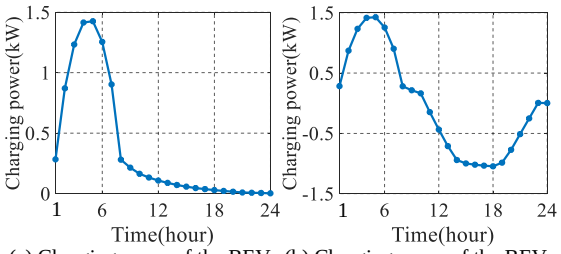
To describe the BEV_3 and BEV_4, the off-peak hours are assumed to begin at 1:00, and end at 8:00. The peak hours begin at 10:00 and end at 23:00. The proposed PDFs are sampled by using the Monte Carlo method to obtain expected values of the charging curves of the BEV_3 and BEV_4, as depicted in Figure 5-10.

It is clear that the charging behaviors of the BEV_3 and BEV_4 have been shifted to the off-peak hours, and the discharging behaviors of the BEV_4 have been shifted to the peak hours.



(a) Charging curve of the BEV_1 (b) Charging curve of the BEV_2

Figure 5-9 The charging curves of the BEV_1 and the BEV_2. Source: [C2].



(a) Charging curve of the BEV_3 (b) Charging curve of the BEV_4

Figure 5-10 The charging curves of the BEV_3 and the BEV_4. Source: [C2].

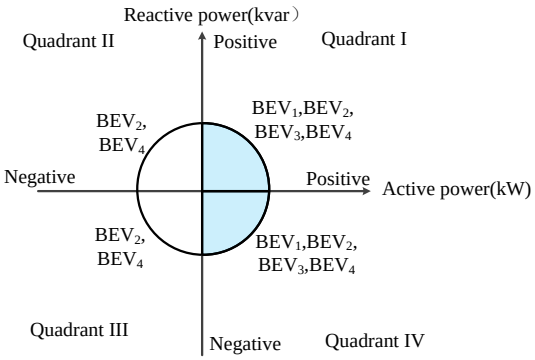


Figure 5-11 The capacity curve of BEVs. Source: [C2].

3) The reactive power supplied by BEVs.

According to the analysis and discussion of the previous work [102], the active power and reactive power supplied by BEVs are bounded by the capacity curve, as shown in Figure 5-11.

The BEV_1 and BEV_3 can supply lagging reactive powers and leading reactive powers. In addition, they cannot supply active powers, and can only consume active powers. Therefore, the BEV_1 and BEV_3 generally operate in the quadrant IV and quadrant I.

Relatively, BEV_2 and BEV_4 not only provide lagging reactive powers and leading reactive powers, but also produce active powers and consume active powers. Therefore, BEV_2 and BEV_4 generally operate in four quadrants.

In summary, the statistical models can be used to obtain the number of BEVs in each hour, and the active charging power of BEVs in each hour with and without demand response. Similarly, a large number of BEVs form an aggregator (i.e., EVCS) to simplify the day-ahead optimal scheduling model. Therefore, the reactive power outputs of EVCSs at each hour are decision variables to be optimized. After

obtaining the reactive power output of EVCSs, the reactive power output of each BEV at each hour can be determined according to the proportion of their reactive capacity to the total capacity (i.e., the reactive power of EVCSs).

Note that this is also one simple method (it aggregates multiple BEVs into one EVCS) to control BEVs, and other models can be considered in future work. For example, the charging power and reactive power of each BEV can be optimized individually, and then reinforcement learning may be developed to solve this complex model with high-dimensional decision variables.

5.2.5. CASE STUDY

1) Simulation settings

Simulations and analysis are conducted on the modified IEEE 33-bus distribution network, the parameters of which can be found in [C2]. Various power devices (e.g., EVCSs, transformers, RESs, and HPs) are integrated to the modified IEEE 33-bus distribution network, the topology of which is presented in Figure 5-12.

Due to the page limit, the next section mainly analyzes the impact of HPs on ADNs, and compares the DDOS model and SDOS model. The impact of BEVs on ADNs is not shown here, but it can be found in [C2].

2) The impact of HPs on ADNs

The 100 HPs are integrated into the 7th node, and 100 HPs are also integrated into the 16th node. All the HPs participate in the demand response. The maximum input thermal power of the radiator is 6 kW [103]. The lower boundary and upper boundary of the room temperature are 19℃ and 24 ℃, respectively. The heat capacity is 2166.4 kJ/℃. The thermal resistance of the house is 0.0067 ℃/W [103]. The day-ahead prediction values of ambient temperature are shown in Figure 5-13.

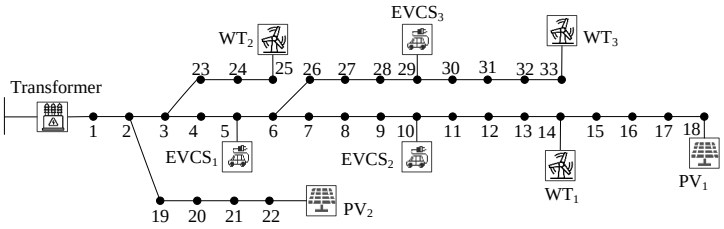


Figure 5-12 The topology of the modified IEEE 33-bus distribution network. Source: [C2].

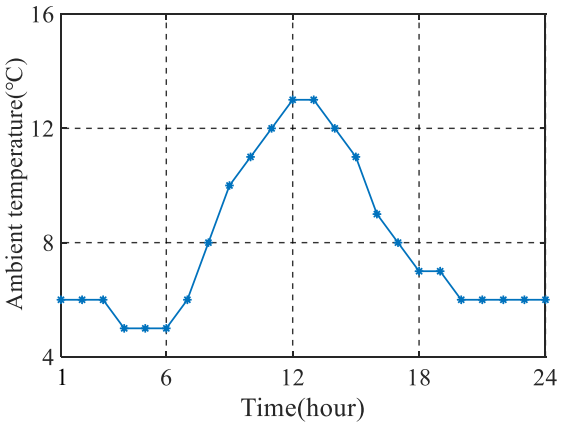


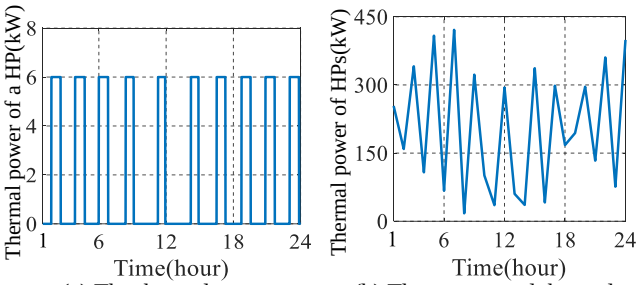
Figure 5-13 The day-ahead prediction values of ambient temperatures.

To analyze the impact of HPs on ADNs, simulations are performed in the following 3 cases, respectively.

Case 1: The ADN includes transformers, BEVs, and RESs. The EVCSs include 50% of the BEV_3 and 50% of the BEV_4. The HPs are not integrated into the 7^{th} node and 16^{th} node of the ADN.

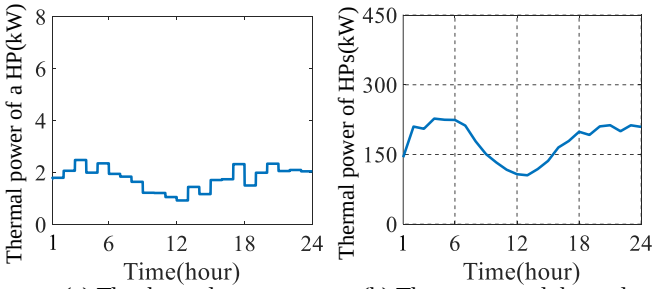
Case 2: The ADN includes transformers, BEVs, RESs, and HPs. The EVCSs include 50% of the BEV_3 and 50% of the BEV_4. The HPs do not participate in the demand response. Specifically, when the room temperature is below the lower boundary, the HP turns on, and the input thermal power of the radiator is the rated power. When the room temperature is greater than the upper boundary, the HP turns off, and the input thermal power of the radiator is 0. The initial temperature in each room is initialized from 19℃ to 24℃ randomly. The input thermal power of one HP is presented in Figure 5-14(a), and the aggregated thermal power of 100 HPs at the 7^{th} node is shown in Figure 5-14(b). The aggregated thermal power of 100 HPs at the 16^{th} node is also similar to Figure 5-14(b).

Case 3: The ADN includes transformers, BEVs, RESs, and HPs. The EVCSs include 50% of the BEV_3 and 50% of the BEV_4. The HPs participate in the demand response. The initial temperature in each room is initialized from 19℃ to 24℃ randomly. Specifically, the input thermal power of the radiator is the decision variable to be optimized. The room temperature should be maintained within the boundaries. The decision variables of each HP are optimized separately. After optimizing the input thermal power at each hour, the input thermal power of one HP is presented in Figure 5-15(a), and the aggregated thermal power of 100 HPs at the 7^{th} node is shown in Figure 5-15(b). Similarly, the aggregated thermal power of 100 HPs at the 16^{th} node also looks like Figure 5-15(b).



(a) The thermal power of a HP (b) The aggregated thermal power of 100 HPs

Figure 5-14 The thermal powers of 100 HPs at the 7th node without demand response.



(a) The thermal power of a HP (b) The aggregated thermal power of 100 HPs

Figure 5-15 The thermal powers of 100 HPs at the 7th node with demand response.

After solving the DDOS model, the power flow analysis is performed. Then, the room temperatures, power losses, and voltage deviations are obtained, as presented in Figure 5-16, Figure 5-17, Figure 5-18, and Table 5-4.

Whether it is case 2 or case 3, the room temperature is always within the boundary, as shown in Figure 5-16. The difference between case 2 and case 3 is the input thermal power of the radiator. Specifically, the conventional control strategy is used in case 2 without demand response, in which the thermal power of the radiator is switched between 0 kW (i.e., turn off) and 6 kW(i.e., turns on). Relatively, the HPs are always heating without shutting down in case 3, and the input thermal power of the radiator is optimized according to the objective function and constraints. Comparing Figure 5-14(b) and Figure 5-15(b), it is found that the input thermal power of the radiator is relatively flat after optimization, which plays a role in peak-shaving to some extent.

The integration of HPs in ADNs increases the power loss and voltage deviation. For example, compared to case 1, power losses increase by 35.6% and 28.7% in Case 2 and Case 3, respectively.

Further, the power loss and voltage deviation drop after HPs participate in the demand response. For example, compared with the case 2, the power loss and voltage deviation of the ADN in case 3 are reduced by 5.10% and 10.28%, respectively. The minimum voltage is raised from 0.957 p.u. in case 2 to 0.968 p.u. in case 3. The reason is that the peak of the load is shaved to some extent after the HPs participate in the demand response.

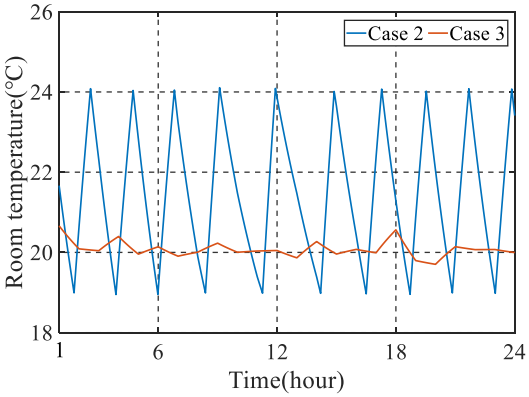


Figure 5-16 The dynamic room temperatures of the ADN with HPs.

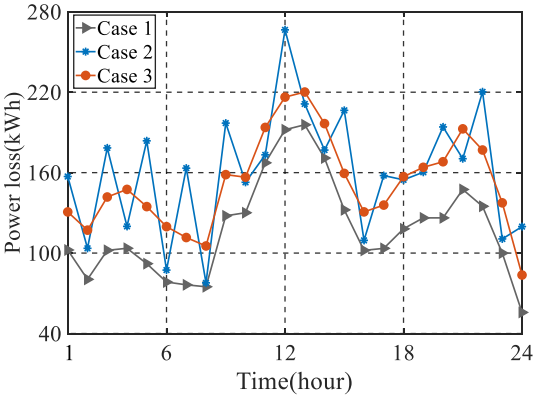


Figure 5-17 The dynamic power losses of the ADN with and without HPs.

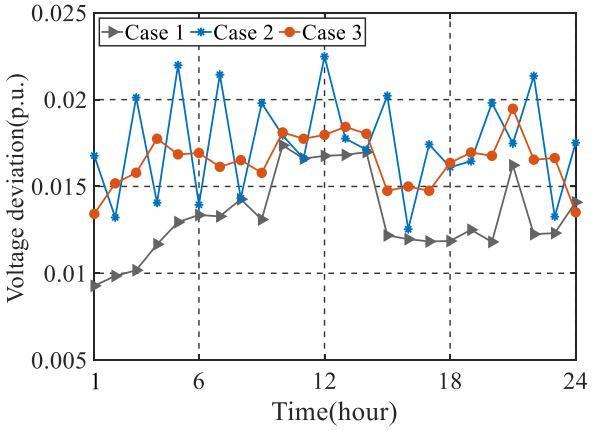


Figure 5-18 The dynamic voltage deviations of the ADN with and without HPs.

Table 5-4 The average voltage deviation and total power loss of the ADN with HPs.

Metrics	Case 1	Case 2	Case 3
Total power loss (kWh)	2840.32	3852.57	3655.92
Average voltage deviation (p.u.)	0.0133	0.0175	0.0157
The voltage range (p.u.)	[0.981,1.05]	[0.957,1.05]	[0.968,1.05]

3) The comparison analysis between DDOS and RSOS models

Here, the DDOS model and RDOS model are conducted to discuss their differences. In each EVCS, half of the BEVs belong to the third category (i.e., the BEV_3) and the other half of BEVs belong to the fourth category (i.e., the BEV_4). All HPs participate in the demand response. To generate stochastic scenarios by using the Gaussian distribution model, the standard deviation of EVCSs is set to 10% [98].

Figure 5-19 displays the voltage deviation and power loss of one day before and after the DDOS model is solved, and Figure 5-20 depicts the voltages at the peak hour.

After adjusting power devices (e.g., EVCSs, RESs, transformers, and HPs), both the voltage deviation and power loss greatly decrease, which should be attributable to the flexible operation of various power devices. Particularly, the voltage security is at risk before optimization, since voltages at part of the nodes (e.g., 14^{th} node to 18^{th} node) are below the lower boundary during the peak hour. After solving the DDOS model, the voltages at nodes are maintained within the boundary, as presented by the orange line in Figure 5-20.

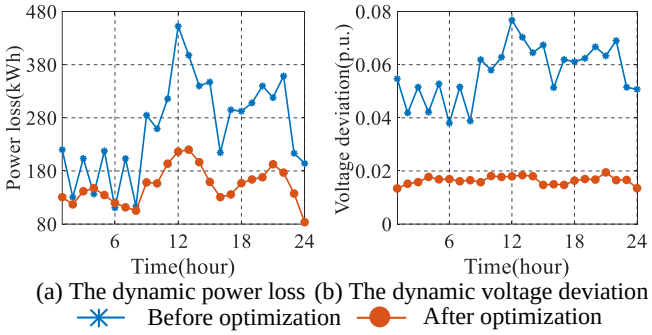


Figure 5-19 The voltage deviation and power loss of one day. Source: [C2].

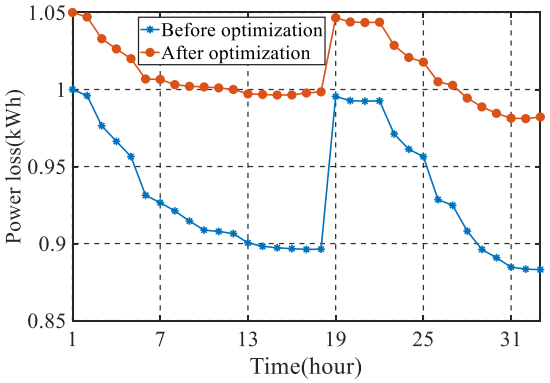


Figure 5-20 The voltages at the peak hour. Source: [C2].

It is obvious that the DDOS model can only produce deterministic outputs (e.g., the expected values of the voltage deviation and power loss), and it is unable to depict the uncertainties resulting from prediction errors of power loads and RESs.

To capture the uncertainty, the SDOS model is employed to generalize deterministic outputs into probability distributions. For instance, probability distributions of the voltage deviation, power loss, and voltages at the peak hour are shown in Figure 5-21, Figure 5-22, and Figure 5-23, respectively.

In comparison to the DDOS model, the SDOS model produces a more complete solution that takes into account both deterministic outputs as well as the potential values and associated probabilities.

To observe the power loss in 2-D space clearly, the PDFs of the power loss at the peak hour are displayed in Figure 5-24. Also, the voltage at the 17^{th} node is considered as a simple example, and the PDF of its voltage at the peak hour is depicted in Figure 5-25.

After performing the SDOS model, the expected value of the power loss is significantly decreased. Before optimization, the probability that the voltage at the 17th node is below the lower boundary is 100%, while after optimization, there is a 100% probability that the voltage is within the boundary.

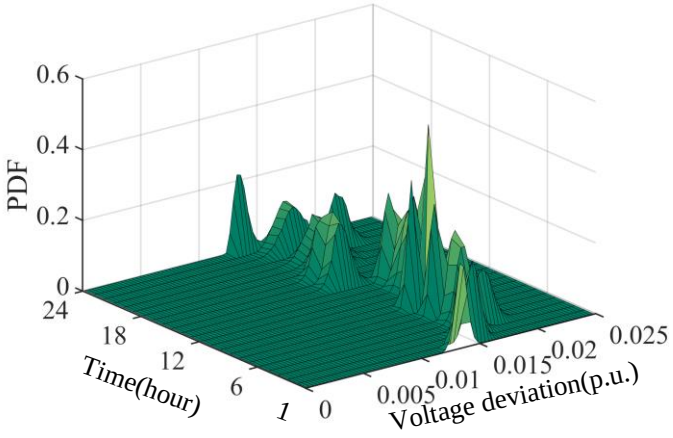


Figure 5-21 The probability density function of the voltages deviation. Source: [C2].

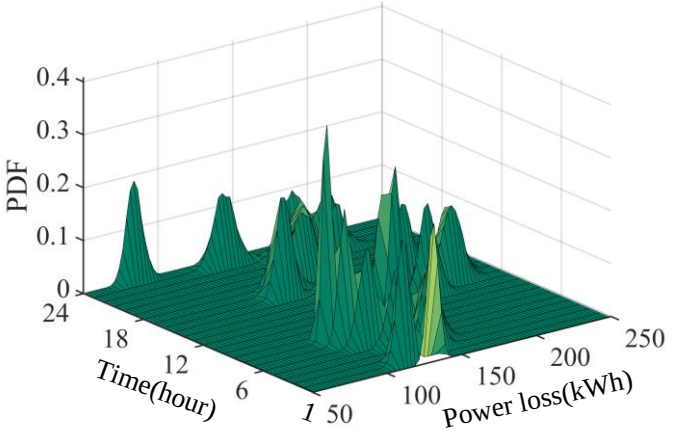


Figure 5-22 The probability density function of the power loss. Source: [C2].

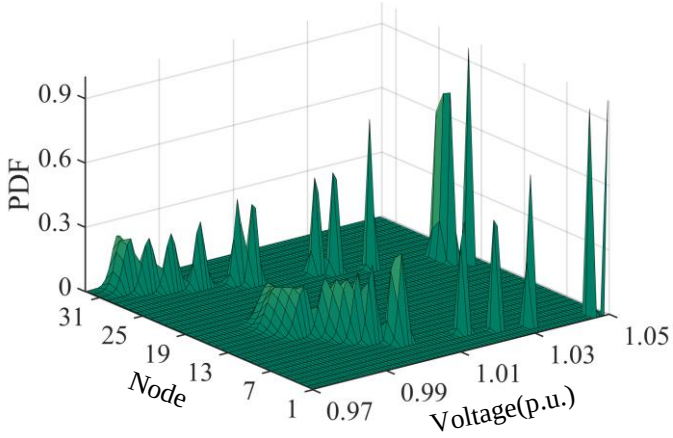


Figure 5-23 The probability density function of the voltage at the peak hour. Source: [C2].

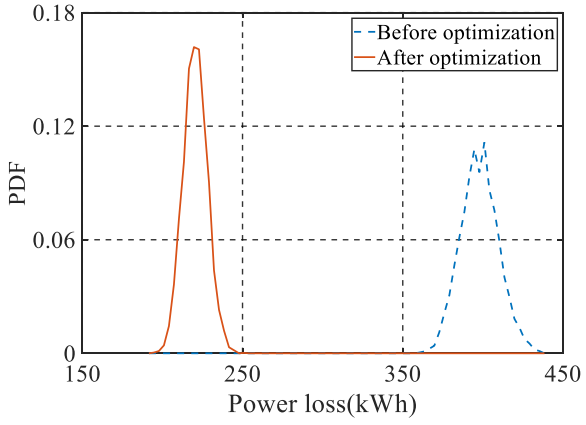


Figure 5-24 The probability density function of the power loss at the peak hour. Source: [C2].

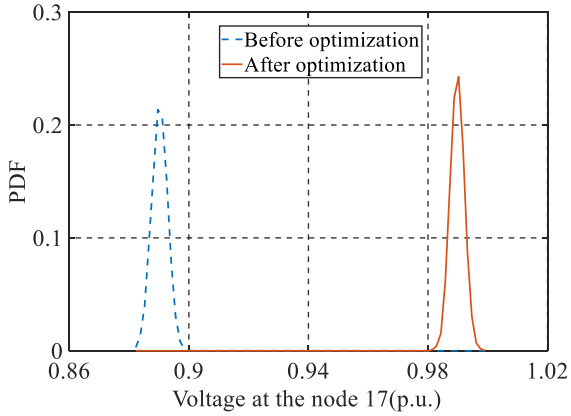


Figure 5-25 The probability density function of the voltage of the 17th node at the peak hour. Source: [C2].

5.2.6. CONCLUSION

To obtain the day-ahead optimal scheduling scheme of power loads and RESs in ADNs considering risks, a stochastic programming model is proposed. The following conclusions are reached by performing simulation and analysis.

The power loss and voltage deviation of the ADN increase after integrating the BEVs without discharging behaviors, since they only charge and do not discharge. The discharging behaviors of BEVs reduce the power loss and voltage deviation of the ADN. The time shifting of charging behaviors or discharging behaviors can reduce the voltage deviation and power loss, regardless of which kind of BEV (the detail of this part can be found in [C2]).

HPs with demand response can not only ensure that the indoor temperature is within the boundary, but also increase a smaller power loss and voltage deviation of ADNs than conventional HPs without demand response.

In comparison to the DDOS model, the SDOS model produces a more complete solution that takes into account both deterministic outputs as well as the potential values and associated probabilities.

CHAPTER 6. CONCLUSION

6.1. SUMMARY

This Ph.D. project focuses on using DL technologies to transform the massive collected data into knowledge, and provide deeper insights into the past, better understandings of the future, and practical suggestions on possible decisions for the safe and economic operation of ADNs. To achieve this objective, the project is conducted to model and optimize the power loads and RESs at the data level, algorithm level, and decision-making level.

In Chapter 1, the background is elaborated. Then, a comprehensive literature review, ranging from the data level to the decision-making level, is conducted. The research gaps, tasks, and limitations of this project are clarified.

In Chapter 2, an advanced framework called CE is proposed to achieve a high accuracy of missing data imputation. The key contributions mainly include two points: 1) The CNN is adopted to construct each part of the CE, so as to extract the complicated latent features of multiple attributes automatically. 2) A hybrid loss function is designed to generate reasonable missing values by combining the adversarial loss function and reconstruction loss function. The simulation results show that the CE not only considers correlations between wind powers and other attributes, but also captures contextual information of wind power generation curves, making it highly adaptable to the missing data imputation with fast ramps. Moreover, CE outperforms popular baselines considering root mean squared error, mean absolute error, and maximum absolute error in an integrated way, especially for datasets with high missing ratios.

In Chapter 3, a new hybrid model is developed to boost the prediction accuracy for the ultra-short-term point prediction of RESs. The key contributions mainly include two points: 1) The inputs of ultra-short-term point predictions are modeled as graphs, from a new perspective in the graph domain. 2) A new hybrid model is designed to depict spatio-temporal features from graphs. The simulation results show that the proposed hybrid model is capable of capturing temporal and spatial features of multiple adjacent RESs and weather conditions accurately. Moreover, it performs better than widely used baselines on real-world datasets.

In Chapter 4, two models are designed to depict the uncertainty of power loads and RESs from two different perspectives (i.e., scenario prediction and interval prediction). Firstly, a new generative network named PixelCNN and an optimization model are proposed to represent the uncertainty of power loads by creating numerous stochastic scenarios. The proposed method mainly solves two gaps existing in most of the publications: 1) The particular data distribution of prediction

errors is exempt from being artificially assumed and formulated. 2) The proposed method does not involve the difficulty in the training process as the existing models (e.g., GANs). Besides, the proposed PixelCNN performs better than other popular generative networks, when it comes to the scenario prediction of power loads.

Secondly, a new ultra-short-term interval prediction method is developed to represent the uncertainty of wind powers. The key contributions mainly include two points: 1) The bidirectional learning is adopted to boost the accuracy of the point prediction model. 2) An improved bootstrap technique is presented to construct PIs without assuming the probability distribution of prediction errors. Simulation results demonstrate that the proposed improved bootstrap technique balances the width and coverage percentage of PIs well.

In Chapter 5, two risk-based methods are proposed to represent the uncertainty in the decision-making process from two different perspectives. Firstly, an RDOS model of power loads and RESs in ADNs is developed to minimize the operational cost of a day considering prediction errors of power loads and RESs. The key contributions mainly include two points: 1) Point prediction values and prediction errors are combined to form new kinds of worst-case scenarios, which are used to encompass the entire range of possible scenarios. 2) The DDOS model and RDOS model are contrasted in terms of security and economy. Simulation results demonstrate that the proposed improved bootstrap technique-based worst-case scenarios perform better than widely used baselines. Additionally, the proposed RDOS model can be seen as a good compromise strategy that strikes a good balance between economy and security.

Secondly, a SDOS model of power loads and RESs in ADNs is proposed to achieve minimum power loss and voltage deviation considering risks from a probabilistic perspective. The key contributions mainly include three points: 1) The mathematical models of HPs and BEVs are formulated to depict their consumption behaviors. 2) The impacts of HPs and BEVs with and without demand response on voltages and power losses in ADNs are discussed. 3) The difference between the DDOS model and SDOS model is demonstrated. Simulations and analysis are executed on the modified IEEE 33-bus distribution network. In comparison to the DDOS model, the SDOS model produces a more complete solution that takes into account both deterministic outputs as well as the potential values and associated probabilities.

6.2. FUTURE WORK

This Ph.D. project proposes several approaches to solve the multiple problems of ADNs at different levels. Although the proposed approaches outperform the popular baselines, a number of limitations exist:

- Massive historical data is needed to train the proposed CE and hybrid model in Chapter 2 and Chapter 3 before using them. This is also a major limitation of DNNs. Therefore, the proposed model is not suitable for ADNs without massive historical data. Fortunately, with the development of the AMI and SCADA system, most of the ADNs are recording a huge amount of data. In future work, the modeling and optimization of ADNs can be discussed in small sample cases.

- This project assumes that the samples in the training set, validation set, and test set have similar probability distributions, i.e., this project does not consider the effect of drastic climate change on the models, which rarely happens in practical engineering. In the future, transfer learning may be adopted to solve the challenges that the training set, validation set, and test set have different data distributions.

- The proposed point prediction model in Chapter 3 and interval prediction model in Chapter 4 are tested on datasets whose inputs only include historical data without considering NWP data due to the unavailability of good datasets. The NWP data can be added to inputs in the extension work.

- The control variables in the DDOS model, RDOS model, and SDOS model only consider the classical power devices, such as transformers, RESs, shunt CBs, EVCSs, and HPs. Other power devices (e.g., energy storage, soft open points, and electrical boilers) have similar mathematical models to classical power devices, and they can be easily integrated into ADNs with more power devices in future work.

LITERATURE LIST

[1] C. Wan, W. Qian, C. Zhao, Y. Song, and G. Yang, "Probabilistic Forecasting Based Sizing and Control of Hybrid Energy Storage for Wind Power Smoothing," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 4, pp. 1841-1852, Oct. 2021.

[2] Q. Zhao, W. Liao, S. Wang, and J. R. Pillai, "Robust Voltage Control Considering Uncertainties of Renewable Energies and Loads via Improved Generative Adversarial Network," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1104-1114, Nov. 2020.

[3] H. Quan, A. Khosravi, D. Yang, and D. Srinivasan, "A Survey of Computational Intelligence Techniques for Wind Power Uncertainty Quantification in Smart Grids," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4582-4599, Nov. 2020.

[4] W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, "A Review of Graph Neural Networks and Their Applications in Power Systems," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 345-360, Mar. 2022.

[5] S. Paul, A. Sharma, and N. P. Padhy, "Risk Constrained Energy Efficient Optimal Operation of a Converter Governed AC/DC Hybrid Distribution Network With Distributed Energy Resources and Volt-VAR Controlling Devices," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 4263-4277, July. 2021.

[6] R. Razavi-Far, M. Farajzadeh-Zanjani, M. Saif, and S. Chakrabarti, "Correlation Clustering Imputation for Diagnosing Attacks and Faults With Missing Power Grid Data," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1453-1464, Mar. 2020.

[7] W. Liao, B. Bak-Jensen, J. R. Pillai, D. Yang, and Y. Wang, "Data-driven Missing Data Imputation for Wind Farms Using Context Encoder," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 4, pp. 964-976, July 2022.

[8] Q. Lin and J. Wang, "Vertically Correlated Echelon Model for the Interpolation of Missing Wind Speed Data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 804-812, Jul. 2014.

[9] Y. Hu, Y. Qiao, J. Liu, and H. Zhu, "Adaptive Confidence Boundary Modeling of Wind Turbine Power Curve Using SCADA Data and Its Application," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1330-1341, July 2019.

[10] Y. Wang, Y. Sun, Z. Wei, and G. Sun, "Parameters Estimation of Electromechanical Oscillation With Incomplete Measurement Information," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5016-5028, Sept. 2018.

[11] N. Eklund, "Using genetic algorithms to estimate confidence intervals for missing spatial data," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 36, no. 4, pp. 519-523, Jul. 2006.

[12] J. Liu, S. Kumar, and D. P. Palomar, "Parameter Estimation of Heavy-Tailed AR Model With Missing Data Via Stochastic EM," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2159-2172, 15 Apr., 2019.

[13] S. Zargar, M. Farsangi, and M. Zare, "Probabilistic multi-objective state estimation-based PMU placement in the presence of bad data and missing measurements," *IET Generation, Transmission & Distribution*, vol. 14, no. 15, pp. 3042-3051, Aug. 2020.

[14] P. Jonsson and C. Wohlin, "An evaluation of k-nearest neighbour imputation using Likert data," *in Proceedings of 10th International Symposium on Software Metrics*, Nov. 2004, pp. 108-118.

[15] J. Du, M. Hu, and W. Zhang, "Missing Data Problem in the Monitoring System: A Review," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 13984-13998, Dec. 2020.

[16] S. Tak, S. Woo, and H. Yeo, "Data-Driven Imputation Method for Traffic Data in Sectional Units of Road Links," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1762-1771, Jun. 2016.

[17] J. Yoon, W. Zame, and M. Schaar, "Estimating Missing Data in Temporal Data Streams Using Multi-Directional Recurrent Neural Networks," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1477-1490, May. 2019.

[18] Y. Mao and M. Jian, "Data completing of missing wind power data based on adaptive BP neural network," *in International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct. 2016, pp. 1-6.

[19] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu, and L. Zhu, "A missing power data filling method based on improved random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33-39, Dec. 2019.

[20] T. Li, J. Tang, F. Jiang, X. Xu, C. Li, J. Bai, and T. Ding, "Fill Missing Data for Wind Farms Using Long Short-Term Memory Based Recurrent Neural Network," *in IEEE 3rd International Electrical and Energy Conference (CIEEC)*, Sept. 2019, pp. 705-709.

[21] X. Liu and Z. Zhang, "A Two-Stage Deep Autoencoder-Based Missing Data Imputation Method for Wind Farm SCADA Data," *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10933-10945, May. 2021.

[22] S. Ryu, M. Kim, and H. Kim, "Denoising Autoencoder-Based Missing Value Imputation for Smart Meters," *IEEE Access*, vol. 8, pp. 40656-40666, Feb. 2020.

[23] Z. Yao and C. Zhao, "FIGAN: A Missing Industrial Data Imputation Method Customized for Soft Sensor Application," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 3712-3722, Oct. 2022.

[24] Z. Wang, D. Wang, Q. Duan, G. Sha, C. Ma, and C. Zhao, "Missing Load Situation Reconstruction Based on Generative Adversarial Networks," *in*

IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Jul. 2020, pp. 1528-1534.

[25] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1322-1330, Dec. 2015.

[26] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review" *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, Jul. 2016.

[27] E. Pelikan, K. Eben, J. Resler, P. Jurus, P. Krc, M. Brabec, T. Brabec, and P. Musilek, "Wind power forecasting by an empirical model using NWP outputs," *in 9th International Conference on Environment and Electrical Engineering*, May. 2010, pp. 45-48.

[28] S. Hu, Y. Xiang, H. Zhang, S. Xie, J. Li, C. Gu, W. Sun, and J. Liu, "Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction," *Applied. Energy*, vol. 293, pp. 1-18, Apr. 2021.

[29] Z. Si, Y. Yu, M. Yang, and P. Li, "Hybrid Solar Forecasting Method Using Satellite Visible Images and Modified Convolutional Neural Networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 1, pp. 5-16, Jan. 2021.

[30] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376-388, Oct. 2020.

[31] M. Madhiarasan and S. Deepa, "Comparative analysis on hidden neurons estimation in multi layer perceptron neural networks for wind speed forecasting," *Artificial Intelligence Review*, vol. 48, pp. 449–471, Dec. 2017.

[32] F. Shahid, A. Zameer, A. Mehmood, and M. Raja, "A novel wavenets long short term memory paradigm for wind power prediction," *Applied Energy*, vol. 269, pp. 1-11, Jul. 2020.

[33] M. Khodayar and J. Wang, "Spatio-Temporal Graph Deep Neural Network for Short-Term Wind Speed Forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 670-681, Apr. 2019.

[34] W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, and K. Liu, "Short-term power prediction for renewable energy using hybrid graph convolutional network and long short-term memory approach," *Electric Power Systems Research*, vol. 211, pp. 1-7, Oct. 2022.

[35] Q. Wu, H. Zheng, X. Guo, and G. Liu, "Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks" *Renewable Energy*, vol. 199, pp. 977-992, Nov. 2022.

[36] A. Khosravi, S. Nahavandi, D. Creighton, and J. Jaafar, "Wind farm power uncertainty quantification using a mean-variance estimation method," *in IEEE International Conference on Power System Technology (POWERCON)*, Oct. 2012, pp. 1-6.

[37] C. Sheng, H. Wang, X. Lu, Z. Zhang, W. Cui, and Y. Li, "Distributed Gaussian Granular Neural Networks Ensemble for Prediction Intervals Construction" *Complexity*, vol. 2019, pp. 1-17, Jul. 2019.

[38] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341-1356, Sept. 2011.

[39] A. Khosravi, S. Nahavandi, and D. Creighton,"Construction of optimal prediction intervals for load forecasting problems," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1496-1503, Aug. 2010.

[40] A. Saeed, C. Li, M. Danish, S. Rubaiee, G. Tang, Z. Gan, and A. Ahmed, "Hybrid Bidirectional LSTM Model for Short-Term Wind Speed Interval Prediction," *IEEE Access*, vol. 8, pp. 182283-182294, Sept. 2020.

[41] Y. Wen, D. ALHakeem, P. Mandal, S. Chakraborty, Y. Wu, T. Senjyu, S. Paudyal, and T. Tseng, "Performance Evaluation of Probabilistic Methods Based on Bootstrap and Quantile Regression to Quantify PV Power Point Forecast Uncertainty," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 4, pp. 1134-1144, Jun. 2019.

[42] Y. Li, X. Chen, C. Li, G. Tang, Z. Gan, and X. An, "A Hybrid Deep Interval Prediction Model for Wind Speed Forecasting," *IEEE Access*, vol. 9, pp. 7323-7335, Dec. 2021.

[43] K. Li, R. Wang, H. Lei, T. Zhang, Y. Liu, and X. Zheng, "Interval prediction of solar power using an Improved Bootstrap method," *Solar Energy*, vol.159, pp. 97-112, Jan. 2018.

[44] Z. Wang, W. Wang, C. Liu, and B. Wang, "Forecasted Scenarios of Regional Wind Farms Based on Regular Vine Copulas," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 77-85, Jan. 2020.

[45] M. Shepero, D. Meer, J. Munkhammar, and J. Widen, "Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data," *Applied Energy*, vol. 218, pp. 159-172, May. 2018.

[46] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-Term Load Forecasting With Deep Residual Networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, July 2019.

[47] Y. Wang, Q. Chen, N. Zhang, and Y. Wang, "Conditional Residual Modeling for Probabilistic Load Forecasting," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 7327-7330, Nov. 2018.

[48] J. Xie and T. Hong, "Temperature Scenario Generation for Probabilistic Load Forecasting," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1680-1687, May. 2018.

[49] P. McSharry, S. Bouwman, and G. Bloemhof, "Probabilistic forecasts of the magnitude and timing of peak electricity demand," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1166-1172, May. 2005.

[50] V. Dordonnat, A. Pichavant, and A. Pierrot, "GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression

models," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1005-1011, Jul. 2016.
[51] W. Liao, B. Bak-Jensen, J. R. Pillai, Z. Yang, Y. Wang, and K. Liu, "Scenario Generations for Renewable Energy Sources and Loads Based on Implicit Maximum Likelihood Estimations," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1563-1575, Nov. 2022.
[52] W. Liao, L. Ge, B. Bak-Jensen, J. R. Pillai, and Z. Yang, "Scenario prediction for power loads using a pixel convolutional neural network and an optimization strategy," *Energy Reports*, vol. 8, pp. 6659-6671, Nov. 2022.
[53] W. Liao, B. Bak-Jensen, J. R. Pillai, R. Zhu, and L. Song, "Data-Driven Scenarios Generation for Wind Power Profiles Using Implicit Maximum Likelihood Estimation," *in the 12th International Conference on Applied Energy(ICAE 2020)*, Dec. 2020, pp. 1-5.
[54] Y. Chen, Y. Wang, D. Kirschen and B. Zhang, "Model-Free Renewable Scenario Generation Using Generative Adversarial Networks," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3265-3275, May. 2018.
[55] Y. Li, J. Li, and Y. Wang, "Privacy-Preserving Spatiotemporal Scenario Generation of Renewable Energies: A Federated Deep Generative Learning Approach," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2310-2320, Apr. 2022.
[56] Z. Pan, J. Wang, W. Liao, H. Chen, D. Yuan, W. Zhu, X. Fang, and Z. Zhu, "Data-Driven EV Load Profiles Generation Using a Variational Auto-Encoder," *Energies*, vol. 12, no. 5, pp. 1-15, Feb. 2019.
[57] L. Ge, W. Liao, S. Wang, B. Bak-Jensen, and J. R. Pillai, "Modeling Daily Load Profiles of Distribution Network for Scenario Generation Using Flow-Based Generative Network," *IEEE Access*, vol. 8, pp. 77587-77597, Apr. 2020.
[58] L. Zhang and B. Zhang, "Scenario Forecasting of Residential Load Profiles," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 84-95, Jan. 2020.
[59] Y. Chen, X. Wang and B. Zhang, "An Unsupervised Deep Learning Approach for Scenario Forecasts," *in Power Systems Computation Conference (PSCC)*, Jun. 2018, pp. 1-7.
[60] W. Liao, J. Chen, Q. Liu, R. Zhu, L. Song, and Z. Yang, "Data-driven reactive power optimization for distribution networks using capsule networks," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 5, pp. 1274-1287, Sept. 2022.
[61] Z. Huang, Y. Zhang, and S. Xie, "Data-adaptive robust coordinated optimization of dynamic active and reactive power flow in active distribution networks," *Renewable Energy*, vol. 188, pp. 164-183, Apr. 2022.
[62] R. Liang, Y. Chen, and Y. Chen, "Volt/Var control in a distribution system by a fuzzy optimization approach," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 2, pp. 278-287, Feb. 2011.

[63] P. Zhang, W. Li, and S. Wang, "Reliability-oriented distribution network reconfiguration considering uncertainties of data by interval analysis," *International Journal of Electrical Power & Energy Systems*, vol. 34, no. 1, pp. 138-144, Jan. 2012.

[64] S. Cai, M. Zhang, Y. Xie, Q. Wu, X. Jin, and Z. Xiang, "Hybrid Stochastic-Robust Service Restoration for Wind Power Penetrated Distribution Systems Considering Subsequent Random Contingencies," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2859-2872, Jul. 2022.

[65] A. Samimi, "Probabilistic day-ahead simultaneous active/reactive power management in active distribution systems," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 6, pp. 1596-1607, Nov. 2019.

[66] T. Ding, Q. Yang, Y. Yang, C. Li, Z. Bie, and F. Blaabjerg, "A Data-Driven Stochastic Reactive Power Optimization Considering Uncertainties in Active Distribution Networks and Decomposition Method," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4994-5004, Sept. 2018

[67] M. Zhang, Q. Wu, J. Wen, B. Zhou, Q. Guan, J. Tan, Z. Lin, and F. Fang, "Day-ahead stochastic scheduling of integrated electricity and heat system considering reserve provision by large-scale heat pumps," *Applied Energy*, vol. 307, pp. 1-18, Feb. 2022.

[68] S. Wang, Z. Dong, F. Luo, K. Meng, and Y. Zhang, "Stochastic Collaborative Planning of Electric Vehicle Charging Stations and Power Distribution System," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 1, pp. 321-331, Jan. 2018.

[69] M. Yan, N. Zhang, X. Ai, M. Shahidehpour, C. Kang, and J. Wen, "Robust Two-Stage Regional-District Scheduling of Multi-carrier Energy Systems With a Large Penetration of Wind Power," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1227-1239, Jul. 2019.

[70] M. A. Mahmud, M. J. Hossain, and H. R. Pota, "Voltage Variation on Distribution Networks With Distributed Generation: Worst Case Scenario," *IEEE Systems Journal*, vol. 8, no. 4, pp. 1096-1103, Dec. 2014.

[71] Y. Wang, W. Wu, B. Zhang, Z. Li, and W. Zheng, "Robust voltage control model for active distribution network considering PVs and loads uncertainties," *in IEEE Power & Energy Society General Meeting*, Jul. 2015, pp. 1-5.

[72] M. Sun, C. Feng, E. Chartan, B. Hodge, and J. Zhang, "A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization", *Applied Energy*, vol. 238, pp. 1497-1505, Mar. 2019.

[73] H. Bidaoui, I. Abbassi, A. Bouardi, and A. Darcherif, "Wind Speed Data Analysis Using Weibull and Rayleigh Distribution Functions, Case Study: Five Cities Northern Morocco", *Procedia Manufacturing*, vol. 32, pp. 786-793, Oct. 2019.

[74] K. A. Singh, M. G. M. Khan, and M. R. Ahmed, "Wind Energy Resource Assessment for Cook Islands With Accurate Estimation of Weibull

Parameters Using Frequentist and Bayesian Methods," *IEEE Access*, vol. 10, pp. 25935-25953, Mar. 2022.

[75] A. AI-Sumaiti, M. Ahmed, S. Rivera, M. Moursi, M. Salama, and T. AIsumaiti, "Stochastic PV model for power system planning applications," *IET Renewable Power Generation*, vol. 13, no. 16, pp. 3168-3179, Dec. 2019.

[76] J. Wu, B. Zhang, H. Li, Z. Li, Y. Chen, and X. Miao, "Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 100-107, Feb. 2014.

[77] S. Xu, W. Wu, Y. Yang, B. Wang, and X. Wang "Analytical solution of stochastic real-time dispatch incorporating wind power uncertainty characterized by Cauchy distribution," *IET Renewable Power Generation*, vol. 15, no. 10, pp. 2286– 2301, Jul. 2021.

[78] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," *in International Conference on Communication and Signal Processing (ICCSP)*, Apr. 2017, pp. 588-592.

[79] S. Hosseini, C. Tang, and J. Jiang, "Calibration of a wind farm wind speed model with incomplete wind data," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 1, pp. 343-350, Jan. 2014.

[80] C. Draxl, A. Clifton, B. Hodge, J. McCaa, "The wind integration national dataset (WIND) toolkit," *Applied Energy*, vol. 151, pp. 355-366, Aug. 2015.

[81] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848-3858, Sept. 2020.

[82] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235-1270, Jul. 2019.

[83] National Renewable Energy Laboratory, "Solar integration national dataset toolkit," Nov. 2012, [Online]. Available: https://www.nrel.gov/grid/sind-toolkit.html.

[84] W. Liao, Y. Wang, Y. Wang, K. Powell, Q. Liu, and Z. Yang, "Scenario generation for cooling, heating, and power loads using generative moment matching networks," *CSEE Journal of Power and Energy Systems*, early access, May. 2022, doi: 10.17775/CSEEJPES.2021.00680.

[85] K. Zhang, M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu, "Residual Networks of Residual Networks: Multilevel Residual Networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1303-1314, Jun. 2018.

[86] D. Wang, L. Wu, X. Zhang, and X. Wu, "A novel hardware Trojan design based on one-hot code," *in 6th International Symposium on Digital Forensic and Security (ISDFS)*, Mar. 2018, pp. 1-5.

[87] W. Liao, Z. Yang, X. Chen, and Y. Li, "WindGMMN: Scenario Forecasting for Wind Power Using Generative Moment Matching Networks," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 5, pp. 843-850, Oct. 2022.

[88] K. Powell, A. Sriprasad, W. Cole, and T. Edgar, "Heating, cooling, and electrical load forecasting for a large-scale district energy system," *Energy*, vol. 74, no. 1, pp. 877-885, Sept. 2014.

[89] X. Zhang, S. Kuenzel, N. Colombo, and C. Watkins, "Hybrid Short-term Load Forecasting Method Based on Empirical Wavelet Transform and Bidirectional Long Short-term Memory Neural Networks," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 5, pp. 1216-1228, Sept. 2022.

[90] W. Liao, S. Wang, B. Bak-Jensen, J. R. Pillai, Z. Yang and K. Liu, "Ultra-Short-Term Interval Prediction of Wind Power Based on Graph Neural Network and Improved Bootstrap Technique," *Journal of Modern Power Systems and Clean Energy*, E-pub ahead of print, doi: 10.35833/MPCE.2022.000632.

[91] A. Khosravi, S. Nahavandi, D. Srinivasan, and R. Khosravi, "Constructing Optimal Prediction Intervals by Using Neural Networks and Bootstrap Method," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1810-1815, Aug. 2015.

[92] C. Li, G. Tang, X. Xue, A. Saeed, and X. Hu, "Short-Term Wind Speed Interval Prediction Based on Ensemble GRU Model," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1370-1380, Jul. 2020.

[93] C. Li, G. Tang, X. Xue, X. Chen, R. Wang, C. Zhang, "The short-term interval prediction of wind power using the deep learning model with gradient descend optimization," *Renewable Energy*, vol. 155, pp. 197-211, Aug. 2020.

[94] T. Ibrahim, T. Rubira, A. Rosso, M. Patel, S. Guggilam, and A. Mohamed, "Alternating Optimization Approach for Voltage-Secure Multi-Period Optimal Reactive Power Dispatch," *IEEE Transactions on Power Systems*, vol. 37, no. 5, pp. 3805-3816, Sept. 2022.

[95] S. Adhikari, F. Li, and H. Li, "P-Q and P-V Control of Photovoltaic Generators in Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2929-2941, Nov. 2015.

[96] D. Zhang, S. Li, M. Sun, and Z. Neill, "An Optimal and Learning-Based Demand Response and Home Energy Management System," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1790-1801, July 2016.

[97] B. Cui, C. Fan, J. Munk, N. Mao, F. Xiao, J. Dong, and T. Kuruganti, "A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses", *Applied Energy*, vol 236, pp. 101-116, 2019.

[98] M. Doostizadeh and H. Ghasemi, "Day-ahead scheduling of an active distribution network considering energy and reserve markets," *International*

Transactions on Electrical Energy Systems, vol. 23, no. 7, pp. 930-945, Oct. 2013.

[99] S. Cheng and Z. Li, "Multi-objective Distribution Network Reconfiguration Considering Electric Vehicle V2G Model," *Energy Procedia*, vol. 158, pp. 278-283, Feb. 2019.

[100] L. Tian, S. Shi, and Z. Jia, "A statistical model for charging power demand of electric vehicles," *Power System Technology*, vol. 34, no.11, pp. 126-130, Nov. 2010.

[101] S. Ge, L. Wang, H. Liu,L. Feng, L. Huang, and T. Zhu, "An optimization model of peak-valley price time-interval considering vehicle-to-grid," *Power System Technology*, vol. 37, no.8, pp. 2316-2321, Aug. 2013.

[102] M. Mazumder and S. Debbarma, "EV Charging Stations With a Provision of V2G and Voltage Support in a Distribution Network," *IEEE Systems Journal*, vol. 15, no. 1, pp. 662-671, Mar. 2021.

[103] Q. Cui, X. Bai and W. Dong, "Collaborative planning of distributed wind power generation and distribution network with large-scale heat pumps," *CSEE Journal of Power and Energy Systems*, vol. 5, no. 3, pp. 335-347, Sept. 2019.

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-730-7

AALBORG UNIVERSITY PRESS