2020

# Renewable Energy Integration in Distribution System with Artificial Intelligence

Yi Gu

# Renewable Energy Integration in Distribution System with Artificial Intelligence

---

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy


by

Yi Gu


August 2020

Advisor: David Wenzhong Gao

Author: Yi Gu
Title: Renewable Energy Integration in Distribution System with Artificial Intelligence
Advisor: David Wenzhong Gao
Degree Date: August 2020

# Abstract

With the increasing attention of renewable energy development in distribution power sys- tem, artificial intelligence (AI) can play an indispensiable role. In this thesis, a series of artificial intelligence based methods are studied and implemented to further enhance the performance of power system operation and control.

Due to the large volume of heterogeneous data provided by both the customer and the grid side, a big data visualization platform is built to feature out the hidden useful knowledge for smart grid (SG) operation, control and situation awareness. An open source cluster calculation framework with Apache Spark is used to discover big data hidden information. The data is transmitted with an Open System Interconnection (OSI) model to the data visualization platform with a high-speed communication architecture. Google Earth and Global Geographic Information System (GIS) are used to design the visualization platform and realize the results.

Based on the data visualization platform above, the external manifestation of the data is studied. In the following work, I try to understand the internal hidden information of the data. A short-term load forecasting approach is designed based on support vector regression (SVR) to provide a higher accuracy load forecasting for the network reconfiguration. The nonconvexity of three-phase balanced optimal power flow is relaxed to an optimal power flow (OPF) problem with the second-order cone program (SOCP). The alternating direction method of multipliers (ADMM) is used to compute the optimal power flow in distributed

manner. Considering the reality of distribution systems, a three-phase unbalanced distribtion system is built, which consists of the hourly operation scheduling at substation level and the minutes power flow operation at feeder level. The operaion cost of system with renewable generation is minimized at substation level. The stochastoc distribution model of renewable generation is simulated with a chance constraint, and the derived deterministic form is modeled with Gaussian Mixture Model (GMM) with genetic algorithm-based expectationmaximization (GAEM). The system cost is further reduced with OPF in real-time (RT) scheduling. The semidefinite programming (SDP) is used to relax the nonconvexity of the three-phase unbalanced distribution system into a convex problem, which helps to achieve the global optimal result. In the parallel manner, the ADMM is realizing getting the results in a short time.

Clouds have a big impact on solar energy forecasting. Firstly, a convolutional neural network based mathod is used to estimate the solar irradiance, Secondly, the regression results are collected to predict the renewable generation. After that, a novel approach is proposed to capture the Global horizontal irradiance (GHI) conveniently and accurately. Considering the nonstationary property of the GHI on cloudy days, the GHI capturing is cast as an image regression problem. In traditional approaches, the image regression problem is treated as two parts, feature extraction and regression, which are optimized separately and no interconnections. Considering the nonlinear regression capability, a convolutional neural network (CNN) based image regression approach is proposed to provide an End-to-End solution for the cloudy day GHI capturing problem in this paper. For data cleaning, the Gaussian mixture model with Bayesian inference is employed to detect and eliminate the anomaly data in a nonparametric manner. The purified data are used as input data for the proposed image regression approach. The numerical results demonstrate the feasibility and effectiveness of the proposed approach.

# Acknowledgements

First of all, I would like to extend my sincere gratitude to my supervisor, Dr. David Wenzhong Gao, for his instructive advice and useful suggestions on my thesis. I am deeply grateful of his help in the completion of this thesis. I learned a lot of knowledge from his excellent performance in teaching and research capabilities as well as responsibility to his students.

I also gratefully thank my formal advisor Dr. Jun Jason Zhang, for his profound knowledge of power systems and help.

I am also deeply indebted to all the other tutors and teachers for their direct and indirect help to me.

Special thanks should go to my friends and colleagues, especially Dr. Yingchen Zhang, Dr. Eduard Muljadi, Dr. Huaiguang Jiang at NREL, who have put considerable time and effort into their comments on the draft.

Finally, I am indebted to my parents for their continuous support and encouragement.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Artificial Intelligence (AI) was introduced in the mid 1950s, which has rapidly evolved in recent decades. It has been an essential tool, which has board applications in humans' daily life. Particularly useful in commercial and industrial applications. In the 1890, science fiction H.G.Wells started exploring the idea of robots and machine thinking as human. After that, in 1956, an AI application is realized to play checker and beat most of the humans. In 2011, Watson is introduced by IBM, which is used for much complicated techniques, such as deep learning. 2015 is a landmark for human research on AI, Google Clouds, AWS and others began to start research computer vision, natural language processing and other analysis tools. There are several methods used to support the AI technology. Machine learning, deep learning, Bayesian Network and Genetic Algorithm [1]. AI systems have been used across many areas of business and industry.

1. Health care: AI enables the doctors to understand the diseases at a deeper level, which aids to supply insights of the risks. And Internet of Things(IoT) and surgical robots have been used.

2. Agriculture: crop monitoring system by AI is set up for farmer to decide if they need to water, fertilizer and others. The complex biological problems are solved by mutation and crossover technology.

3. Finance: banks make credit decision automatically with models such as decision tree. Fraud is easier to spotted with algorithm by financial organization. One more general example, humans can make deposits by scanning the check with smart phone with AI.

4. Travel: Airline companies, hotels and car rentals use AI on demand forecasting and adjust the price automatically. And AI is also used to make the plans with routes, weather, customer loads and other variables.

5. Power System: An electric power system is a network of electrical components, which is used to supply, transmit and consume electric power. Power systems engineering is a brand of electrical engineering, which aims at power generation, transmission, distribution. Artificial Intelligence (AI) technology becomes popular for solving different problems in power systems like controlling, planning, scheduling and forecasting, etc. The techniques can used to solve difficult tasks faced by applications in present large power systems. The application of these techniques has been successful in many areas of power system engineering.

In recent decades, the increasing demand of world energy requires a rapid development of modern power systems with renewable energy. It is still facing many challenges in future development of hybrid power system. Due to advantages of green energy, such as economic benefits, almost no pollutions, less maintenance fee, a hybrid distributed power system is a top topic in recent power energy research. However, renewable energy is a highly weather depended resource. Because of that, a traditional distributed power system combined with renewable energy generation can be satisfactory to take the advantages and avoid disadvantages [2].

Traditional power system is dramatically changing with the rapid technology development. A large amount of massive heterogeneous data are provided and indispensable for the SG operation. The method on how to choose the useful information without spatial information is proposed as an open source cluster framework, which is based on Apache Spark and can be used to effectively collect, store and process data in parallel manner. The discovery knowledge is visualized with a Google Earth Platform, the operators can also monitor the power system based on that. The network reconfiguration is used to operate the power system with limited switches controlling, which aims to maintain voltage profiles, reduce system loss. An improved method on load prediction technology is used to compute the load deviations at two scheduled time points, which is ignored by traditional ways. The dynamic and efficient network reconfiguration method is simulated based on support vector regression (SVR). The short-term load forecasting approach is employed to minimize the error between real and forecasting demand load.

The distribution system markets are attracting more attention in recent research along with the rapid technology development of the power system. Considering the hybrid power system with high penetration of renewable energy, how to improve efficiency and reliability of the hybrid power system become a new challenge. Due to the economic benefits of renewable energy, a multi-timescale operational approach is scheduled on substation and feeder level. Different from traditional computation on distribution system, a three-phase unbalanced distribution system is used to minimize system loss. It should not be ignored with a desire of high accuracy on optimizing operation cost. And the corrective action is implemented at substation level. Reselling the redundant power to real-time power system should be helpful for operation cost optimization.

## 1.1 Current Research on Data Analysis and Knowledge Discovery

Big data representation and visualization are described in [3], which helps to avoid the problems, such as visual noise, large image perception, information loss and high rate of image change [4–10]. In this method, more than one view are collected for each representation display. However, how many views are needed for each representation display is still a problem. Recently the challenge of data analysis has risen to incorporation uncertainty into visual representation, a novel data analysis method of uncertainty is described in [11]. The data analysis platform will result an inaccuracy after process the data in the pipeline. [12] construct a visualization framework to analysis the relationship between transmission system capacity, scheduled power flows and actual network power flows. It includes the contour of bus, transmission line flow values, power flow values, but still limited on renewable energy power generation. For large-scale of electric power system, information visualization is necessary for research [13]. Several power system visualization tools are presented in [13], which realize 3D data analysis. Effective power system always require operators to analysis and interactive in a short time, presenting the data in a form for engineers to assess in an a quick manner is what [12] did. The platform can help operators to monitor and control the system, however, a visualization platform for large scale power system is still needed. data analysis is an established method and has been widely used because of the strong data management and knowledge display. [14] proposed some tools on how to visualize the data information of distribution system. A big amount of data analysis methods has been proposed in the recent decades, there are multi-direction model, large-scale transmission visualization model, while the researching on 3D distribution system visualization is still limited. In this chapter, I proposed a 3D data analysis platform for distribution system, the campus of University of Denver is used as the test bench, which helps to monitor the load for each building in real time. A lot of machine learning related approaches are imple-

mented in power system and related areas [15–26]. Similarly, optimization is a large topic, which contains both convex optimization and non-convex optimization approaches [27–34]. Building a smart grid (SG) with renewable energy generation is imperative, such as solar energy, wind power, hydropower, geothermal, bio energy, etc. [35–37]. Along with the rapid development of the modern power generation technology, the new inventions are dramatically improving the traditional power systems [37–40]. The large amount of smart sensors provide heterogeneous massive data, which are necessary to the smart grid operation and management. However, the useful knowledge always hides in the big amount of the data with less spatial information. In order to figure out the useful information, an open source cluster computing framework based on the Apache Spark is built to collect, save and process the big data in parallel manner. And then, the selected data are transmitted to the proposed visualization platform. For example, the way to improve the system efficiency and extend the service life to determine the optimal position of wind turbine [41]. According to the analysis of the market research, the service market grew from $3.2 billion in 2010 to $16.9 billion in 2015 [42].

The most important features are used to do the knowledge discovery are low value density, large volume and high velocity [41, 43–45].

1. Low value density: in smart grid abnormal data detection, most part of the collected data from operation are general data, the abnormal data only occupy very small parts [46].

2. Large volume: the various smart sensors with high speed sampling rate always produces a large volume of data, which can increase the TB level up to PB level [43].

3. High velocity: the rate speed of the smart sensors are increasing. For example, the synchrophasor measuring device, which measures from 20 samples/s to 1440 samples/s [46].

According to the characteristics, the proposed platform of big data is described to satisfy the requirements as follows. Four points are described here.

1. Efficient, resilient and distributed data storage skill for the collecting heterogeneous massive data.

2. Error tolerant, high speed, big data processing and analyzing in parallel manner.

3. Streaming processing platform for high speed in real-time data.

4. Hidden knowledge discovery based on machine learning to meet the requirements of different complex projects.

Hadoop is a widely used open-source software, which is released by Google file system in 2003. It is designed for big data processing and distributed storage in computer clusters [47, 48]. Compared with the traditional approach, MapReduce is the critical feature of Hadoop, it can divide the job into several smaller jobs and deal with them in distributed computer clustering. In each iteration, reloading the data from disk can cause a longer time consumption, especially for the iterative parameters optimization with machine learning [48]. Based on the new cluster computing framework, Apache Spark is designed to reduce the time-consumption on the clusters computation. The important advantages of Apache Spark are as follows [47]

1. Not frequently reload the data from disk for iterative calculations in machine learning.

2. Dividing the task into several smaller jobs and compute the results in distributed computer clusters.

3. Easy to use in many languages and different operation systems.

At present, Google Earth is widely used and can be operated in different operation systems such as Mac OS, Linux, Windows [49]. A Google Earth platform is used to achieve the data analysis with discovered knowledge.

The visualized platform is designed to meet the requirements of independent system operators (ISOs), customers and the utilities, which consists: 1) decision support, 2) forecasting assessment, 3) operation and control, 4) security and detection.

## 1.2 Current Research on Load Forecasting on SVR

Load forecasting is one of major field of research on power system operation. Many traditional ways have been tried to obtain good results. In [50], an early neural network is constructed to analysis the relationship between actual load and various features. However, the early model did not consider some reasons which may lead overfitting. Because of renewable integration, a high accuracy model of load forecasting has been more and more important for power planning and operation. A probabilistic electric load forecasting has been proposed in [51], the emerging technology has been considered during the load forecasting process. [52] introduces the definition of short-term load forecasting of power system. The current load forecasting methods are classified as new forecasting methods, traditional methods, intelligent methods and classic methods. It is reminded that not only the historic data is important, but also the right model selection. [53] proposed an approach of load forecasting with dynamic pricing for demand side management. Real time pricing and peak price are considered in details. There are plenty of methods considered in the development of load forecasting. In this chapter, mathematical and artificial intelligence computational model is a new trend on solving this problem, an SVR model is built with grid traversal to find out the best combination of hyper-parameters, which helps to reach the global minimal in our research.

Network reconfiguration is widely used to decrease system loss, control voltage stability [54–57]. The load prediction technology provides the load deviations between two time points. Considering the stochastic load deviation in distribution power system, it is important to set up an efficient and dynamic network reconfiguration approach [40, 58–62]. A support vector regression (SVR) is used to cooperate with the network reconfiguration

based on short-term load prediction method to minimize the system loss. Most parameters in SVR can be solved in a convex manner. Several parameters, defined as hyper-parameters, cannot be determined in a same time. The optimization of the parameters are really indispensable of SVR, and dramatically influence the performance and efficiency of the forecaster [63–66]. Due to this, a two-steps parameters optimization approach is proposed with grid traverse algorithm (GTA). According to the distributed computation frameworks in [67–70], the hyper-parameter optimization of SVR is described with MapReduce to optimize the parameters in parallel manner and reduce the calculation time consumption.

## 1.3 Current Research on multi-timescale distributed power system

With advanced development of power generation technology, distribution system markets attracts more and more attention in recent research [71–73]. Considering the output of high penetration renewable energy generation, distribution power system faces a big challenge. Many approaches are used to improve the net reliability. In [71], a day-ahead market energy auction is built for distribution system operation. In [74,75], a multi-timescale power system operation approach is proposed for renewable energy generation. In [76], a storage-based operation approach is used to decrease the operation cost of a distribution power system with renewable energies. In [72, 77], the optimal operation methods are proposed with economic dispatch in distribution system with renewable energy. However, all these studies ignored the system loss, which would significantly influence the operation cost. A multi-timescale operation approach is proposed to reduce the operation cost. A three-phase unbalanced OPF is built to reduce the system loss.

In [78], the stochastic programming optimization (SPO) provides many benefits to distribution systems. The distribution of the forecasting errors is given to generate the error models, which are formulated as a chance-constraint for SPO in distribution systems. It is

assumed that there are more than one renewable generators operating on certain feeders. The aggregated error distribution usually includes multiple Gaussian models, which can be modeled as the Gaussian mixture model (GMM) [79]. The variant Gaussian models indicates the uncertainties of the renewable energies. The genetic algorithm-based expectation maximization (GAEM) is applied for GMM to determine the amount of Gaussian models automatically [80]. According to the this, the chance-constraint of the renewable generation is formulated into a deterministic form for further reducing the computation burden [78]. In [81, 82], the variability of the renewable generation is operated by hourly demand response with day-ahead scheduling, which is scheduled without considering the stochastic net load deviation within an hour. In [83], a single line model instead of three-phase system is used to compute the system loss, which ignores the three-phase unbalanced configuration in distribution power systems. Because of the nonconvexity of the three-phase unbalanced OPF in distribution system, the heuristic based approach is proposed in [84]. However, this method hardly avoids to fall into the local minimums with a high computation time-consumption for a large-scale distribution systems [85]. To avoid the issues above, an inequality constraint is proposed based on semidefinite programming (SDP) to relax the three-phase unbalanced OPF problem. After that, the proposed formulation for system loss is solved in parallel manner with alternating direction methods of multipliers (ADMM). It helps to further reduce the total operation cost hourly.

## 1.4 Current Research on image regression with convolutional neural network

Global horizontal irradiance (GHI) ia a critical index which we use to predict the output power of PV generators. In [86], time series forecasting model is set up for short-term load forecasting. The logical procedures for model developing, which use the autocorrelation function helps to obtain a better performance. A novel approach for daily peak load fore-

casting is proposed in [87], the result shows that the daily peak load can be forecast in a high accuracy with artificial neural network. ARMA (autoregressive moving-average) model is studied for short-term load forecasting of power system in [88]. The Box-Jenkins transfer function is used as a important tool to increase the forecasting accuracy. However, the time series forecasting model always need continuous data, in this chapter, a deep learning model is built to overcome the non-serial data. [89] set up a probability approach for long-term load forecasting, the probability distribution of weekly peak electricity was predicted in his method. A deep learning framework is proposed for short-term load forecasting, a deep belief model is used to do the hourly load forecasting in [90].

## 1.5 Organization of thesis and Main Contributions

In Chapter 1, the literature review of the current research around renewable energy distribution power systems are described, which provides background study of the renewable energy. In Chapter 2, The useful hidden data is transmitted into a visualization platform. The result provides the audience a convenient way to get the information. In Chapter 3, network reconfiguration is used to operate the power system with an improved load forecasting technology based on SVR. In Chapter 4, a multi-timescale operational approach is used on a two level distribution power system, which consists of the substation and the feeder level. The numerical results are obtained from simulation on IEEE 123-bus, IEEE-8500, University of Denver campus distribution power system and Feeder J1. Feeder J1 is [91] selected for analysis because 1.7 MW of clients owned PV generators exists on the feeder. The system locates in the northeast of US, which serves 1300 commercial, small industrial and residential customers. In Chapter 5, a novel approach is proposed to capture the GHI conveniently and accurately. Considering the nonlinear regression capability, a convolutional neural network (CNN) based image regression approach is proposed to provide an End-to-End solution for the cloudy day GHI capturing problem. After that, a novel ensemble learning model is constructed with different machine learning and deep learning

model. An efficient way is introduced to construct the renewable energy forecasting model. In Chapter 6, the conclusion and future work is summarized.

The **main contributions** of this thesis are:

1. data analysis platforms are constructed for operators to assess the load power and renewable energy in convenience. The visualization platforms provide the operators a clear understanding of the information in a short time. The platform can be used to visualize a large volumes of data, which helps the customers to make decision more efficiently with less time consumption. Depends on current research, it is necessary for operators to work on this platform.

2. For a distribution system operator, a multi-timescale approach is proposed based on a three-phase unbalanced distribution system. With the day-ahead dispatch of the substation level, the objective is provided to optimize the system cost with renewable energy. The chance constraint is built based on GMM and GAEM. At the feeder level, the objective is minimizing the system loss. An OPF problem is formulated for the three-phase unbalanced system, calculate the results in parallel with ADMM. In the proposed multi-timescales model, the feeder scheduling is described with higher time resolution and update frequencies, which can be obtained near RT calculation.

3. A big data processing approach is used in renewable energy area for solar irradiation capturing, which contains big data cleaning and deep learning based image regression. Compared with the traditional solar irradiation capturing approach, the CNN based approach is cheap, fast, accurate, and convenient to be extended for large-area monitoring.

4. Based on the Dirichlet process, variational inference, and Bayesian theory, a Gaussian mixture model with Bayesian inference approach is employed. It determines the mixture components automatically.

5. Based on the CNN architecture for classification, a new regression CNN architecture is designed for image regression problem. The input of the proposed approach is the image set and the output is the continuous variable set, which can be extended for multiple regression problems. In future study, the related problems, for example, the large-area PV output forecasting, can be studied based on our research.

6. A new ensemble learning model is built for energy forecasting, compare with transitional ensemble learning model. The proposed model aims to figure out the most efficient combination of the machine learning models.

# Chapter 2

# Data analysis and data computation platform of smart grid operation

## 2.1 Introduction

Recently, with widely spread smart sensors, multidimensional data such as voltage, current, wind direction, wind speed, and solar irradiance can be collected in highly simpling rate over large area. The collected large-volume data contains a large amount of knowledge and requires a bunch of new method to process and discover. For example, a modern hybrid power system usually contains high penetration of wind and solar, which is very difficult to identify its real system operation states with traditional methods. And traditional methods are also not suitable for large-volume and high-speed data process. Therefore, considering advanced data-processing and data-driven technologies, machine learning based approaches are proposed to discover the hidden information for power system operation and control. In addition, data visualization is a significant indispensable step for advanced data-processing and data-driven technologies, which displays rich information vividly for both data preprocessing and result analysis.

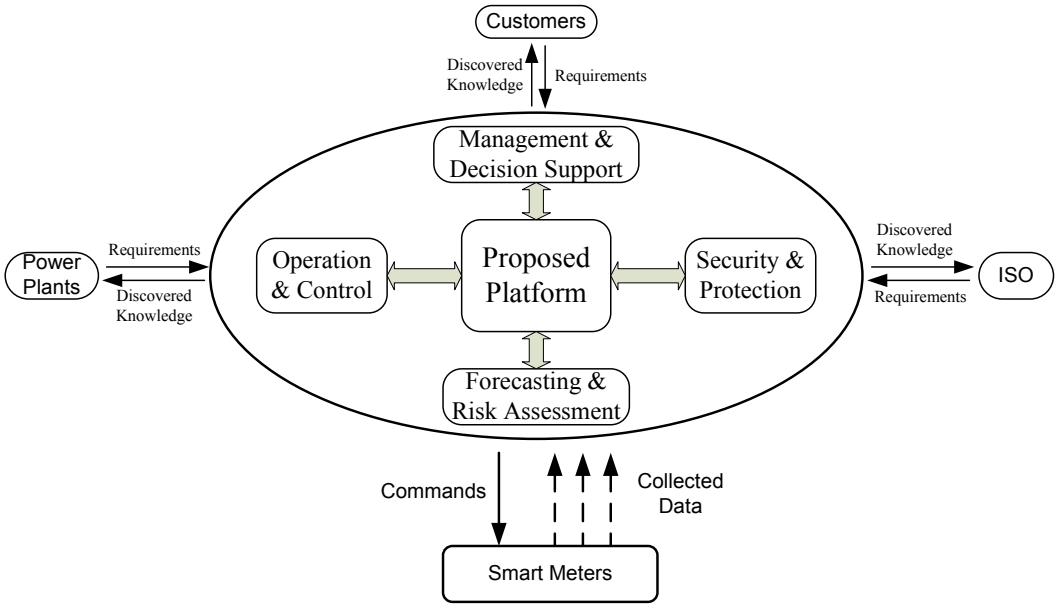



Figure 2.1: Main architecture of the proposed big data platform.

## 2.2 Architecture of the Platform

### 2.2.1 Main Architecture

The proposed platform in Fig. 2.1 has four major applications which consists of 1) forecasting and risk assessment, 2) decision support and management, 3) operation and control, 4) security and protection. The independent system operators (ISOs) could ask the useful knowledge from the proposed platform, respectively.

### 2.2.2 Detailed Architecture

As shown in Fig 2.2, the major architecture of knowledge discovery platform contains several parts such as Data collection, resilient distribution dataset, Spark master, GIS visualization, and Applications.

1. **Resilient Distributed Dataset**: the major task is to store and process data with supervising of Spark master. On one hand, the work nodes from 1 to $n$ are working in parallel. On the other hand, the distributed datasets receive original dataset and

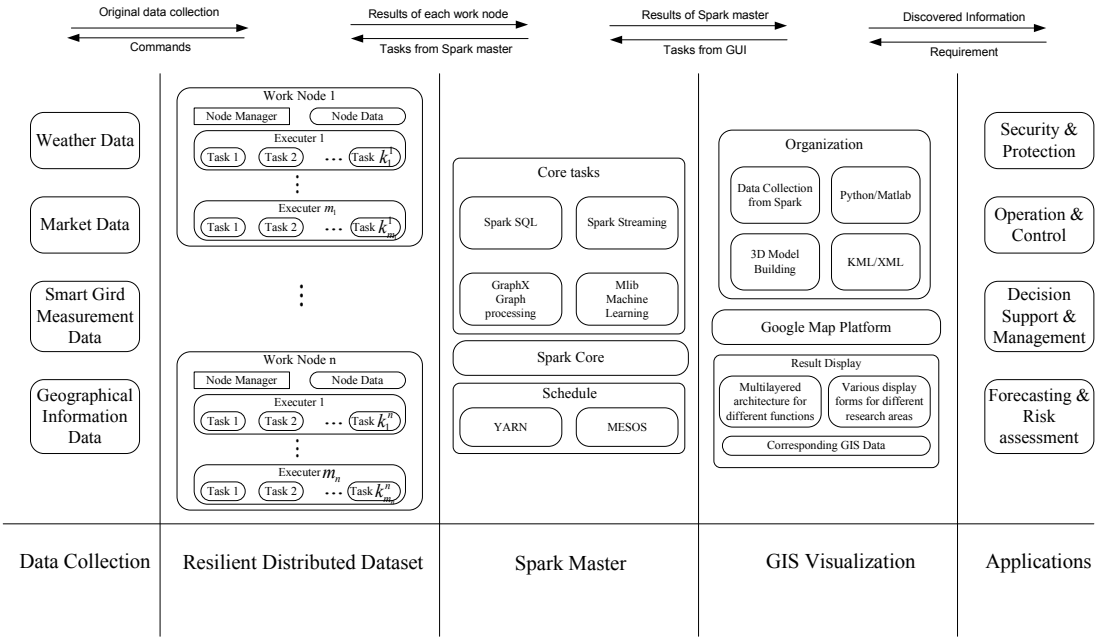



Figure 2.2: Detail architecture of the proposed big data platform.

store the datasets at distributed data centers. It can reduce the communication burden to Spark master (It is an advanced big data process engine, which has been widely used in both industrial and academic research). In addition, the critical data can be compressed and stored in distributed to increase the resilient of databases [46].

2. **Spark master**: the major task is to collect different types of data in natural world. In this part, the smart sensors provide the original data from natural world and the smart sensors can receive commands from the Spark master about measurement time periods.

3. **GIS visualization**: the data flow of the discovered knowledge is obtained from Spark master, Python and Matlab are used to generate the KML files for visualizing data in various forms. Based on different kinds of scenarios, multi-layered architecture is applied, and different kinds of display forms are used for different research areas such as operation, control, security, and management.

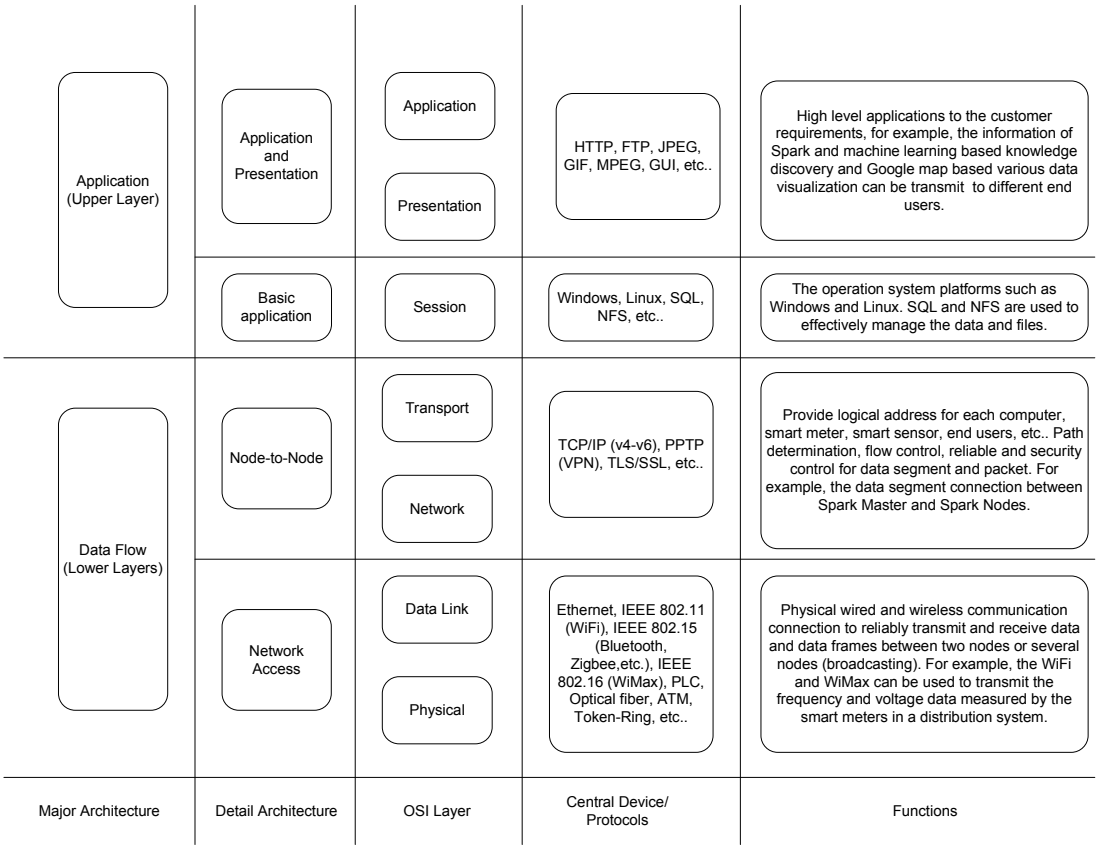



Figure 2.3: Architecture of the communication design.

### 2.2.3 Data Communication Architecture

The architecture of communication system in the proposed platform is described in Fig 2.3. The two major layers are: data flow in lower layers, application in upper layers. The standard Open System Interconnection (OSI) model is used to compare and explain the corresponding abilities and applications of each layer in the detailed architecture.

1. **Network Access**: The major task of this layer is to build a reliable physical wired or wireless communication connection among the work nodes for transmitting and receiving data. In the data collection part, as shown in Fig 2.2, the smart sensors are pervasively located. Both wired and wireless communication connections can be implemented according to the data volume and transmission speed.

2. **Node-to-Node**: The major task of this layer is to provide addresses for network access devices such as computers, smart sensors, servers, and end users, to send and receive data among the network access devices, determine the path of data flow, control the security and quality of service (QoS).

3. **Application and Presentation**: The major task of this layer is to implement the complex data structure on the data visualization platform.

## 2.3 The Spark Based Knowledge Discovery

Apche Spark based computer cluster is used to discover the hidden knowledge in the collected big data, which provides a resilient, fast, and effective parallel computation platform for many machine learning algorithms such as regression, classification, recognition, etc.. [92]. As shown in Fig 2.2, the Apache Spark contains: Spark core, Spark SQL, Spark Streaming, Spark GraphX, Spark Mlib.

### 2.3.1 Spark Core

The Spark Core contains a lot of basic functions such as task scheduling, memory management, fault recovery, etc. [47]. RDD is one of the most significant features of Spark core, which means that many distributed computer nodes with databases can be organized and manipulated in parallel. RDD saves the application states such as check points periodically, which means that the system and computation can be recovered quickly after some work nodes loss or failed. Considering the different requirements from different customers, multiple tasks can be computed in parallel with the high-effective platform. The big data processing results can be accessed in multiple end users, respectively.

### 2.3.2 Spark SQL

Spark SQL is an efficient package focusing on processing data, and supporting a lot of data forms [47]. Python is used as the programming language, which also supports a lot of data forms such as XLS, MAT, etc. [93].

### 2.3.3 Spark Streaming

Spark Streaming is a useful component for processing the live streaming of data [47]. Compared with Hadoop, which needs to frequently load and reload data from disk, Spark can save and store the data in memory, which dramatically increasing the data processing speed.

## 2.4 The Google Earth Based Data Visualization

Many open source visualization tools are used to support a simple platform for the researchers to integrate any kind of data with their geospatial products [49, 94]. Both of the two platforms (Google Earth and arcGIS) can be attached multiple layers with small editing. For Google Earth, the map and other information are displayed in KML format. Compared with arcGIS, Google Earth is more convenient and vivid for primer operations. Google Earth is considered as a better choice to build the power system model for DU campus.

## 2.5 Results

### 2.5.1 An Example of Demand Response

The proposed platform can be used to satisfy a variety of different requirements for different users. Fig. 2.4 illustrates the demand response with the temperatures and the unit price in the SG of DU campus. Specifically, demand response is an electrical power demand

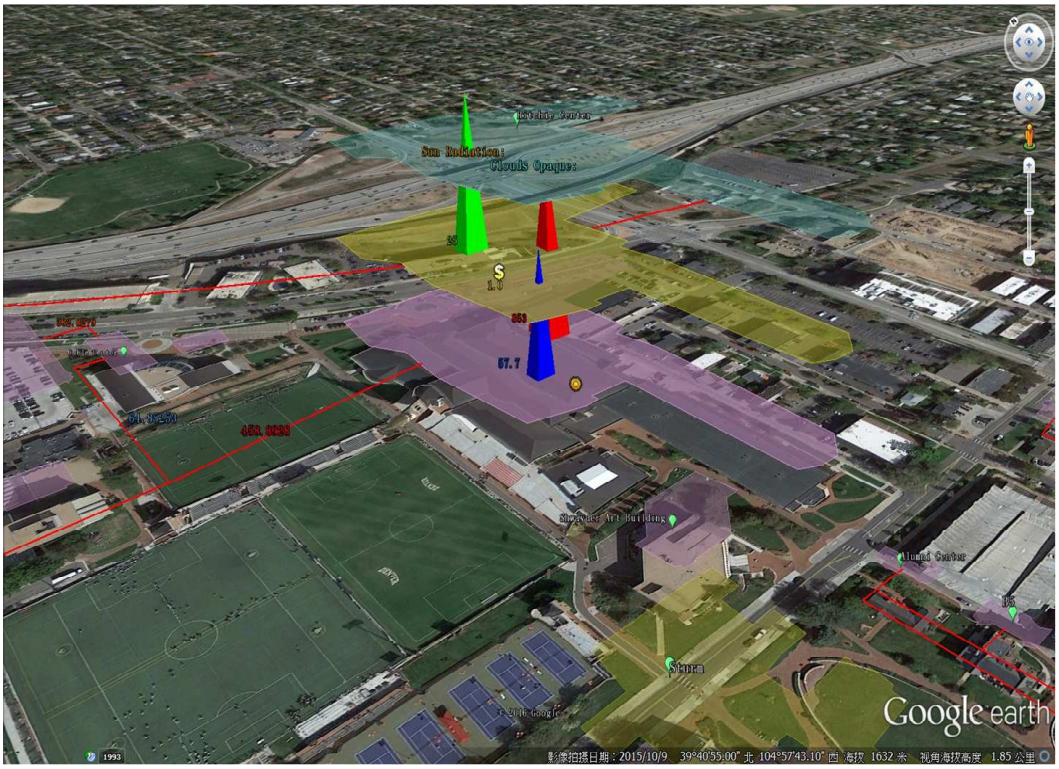



Figure 2.4: Real-time demand response in DU campus.

shift skill which allow users to shift their loads such as dishwasher and washing machine to avoid high price periods. Massive heterogeneous data such as temperature, electrical price, voltage, frequency, etc. are collected by the pervasively located smart sensors in real-time. Then the designed platform is used to compute the demand responses according to different time periods. The red and blue pyramids indicate the real-time temperature and electricity price, respectively. The green pyramid denotes the corresponding demand response in real-time.

### 2.5.2 A Description of Renewable Energy-PV

As shown in Fig. 2.5, this model illustrates the ratio between PV generation and total electrical power consumption. In the red circle, the three white bars denote the total power consumption of three areas in Ritchie Center building. The three red bars indicate the

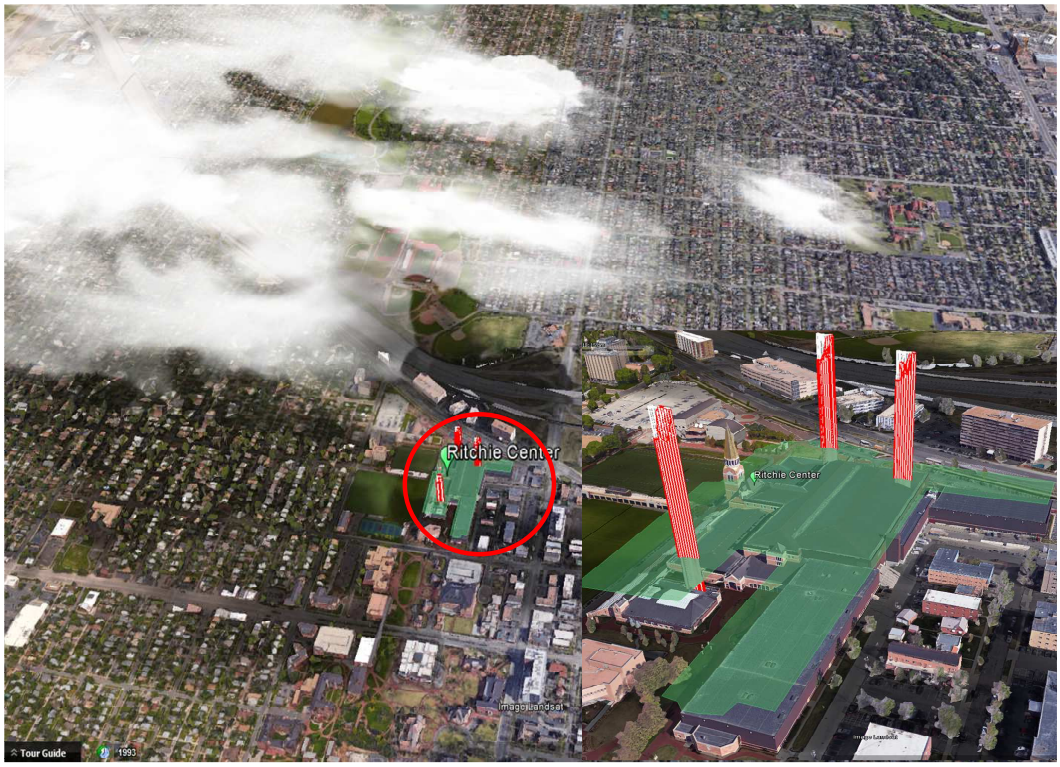



Figure 2.5: Real-time PV generation and total electrical power consumption on DU campus.

electrical power bought from utility, the white areas not covered by the red bars are the PV generations. As shown in Fig. 2.5, the PV generation is decreasing, and electrical power from utility is increasing in a cloudy day. With this real-time information, the customers, ISO, and power plant can schedule their behaviors and managements correspondingly.

### 2.5.3 Multivariate Linear Regression Analysis

The multivariate linear regression is used to analyze the correlationships between response variable $Y_t$ and explanatory variables $X_t^i$, where $Y_t$ and $X_t^i$ are time series variables, and $i \in \{1, 2, 3, \cdots\}$ indicates the different types of explanatory variables, $t$ is the time index [95]. The response variable $Y_t$ is the load profile of the Daniels College of Business building. The explanatory variables are collected from HVAC of the Daniels College of Business building including air pressure index 1 in HVAC ($X_t^1$), fan tuning index ($X_t^2$),

wind tunnel temperature 1 ($X_t^3$), wind tunnel temperature 2 ($X_t^4$), and air pressure index 2 ($X_t^5$). The regression output is a time series variable $\hat{Y}_t$. The collected data are normalized between 0 and 1.

Ordinary least squares (OLS) is a widely used to compute the coefficients of the explanatory variables, and optimize the squared differences between $Y_t$ and $\hat{Y}_t$. However, in the OLS, it is assumed that the error term is a constant variance, and ignores the heteroscedasticity, which concerns the expectations of the second moment of the errors [96]. This means the OLS is not a valid estimation approach. Considering the heteroscedasticity impact, the feasible generalized least squares (FGLS) can estimate the errors covariance matrix to improve the efficiency of the regression. Therefore, the FGLS is used to compute the multivariate linear regression in big data analysis [96].

Table 2.1: Regression Performance (Squared Error)

| | Original | Normalized |
|---|---|---|
| OLS | 635 | 0.457 |
| FGLS | 345 | 0.342 |

Table 2.2: The coefficients of explanatory variables with FGLS

| | $X_t^1$ | $X_t^2$ | $X_t^3$ | $X_t^4$ | $X_t^5$ |
|---|---|---|---|---|---|
| Coefficient | 0.0251 | 0.0826 | 0.0857 | 0.1719 | 0.0318 |

As shown in Table 2.1, compared to the original data without normalization, the regression error of the normalized data is much smaller. The regression errors of FGLS are less than the OLS in both original and normalized data regression. As shown in Table 2.2, the explanatory variable $X_t^4$, wind tunnel temperature 2, have the largest impact; and $X_t^1$, air pressure index 1, have the smallest impact to the response variable $Y_t$, the load profile of the Daniels College of Business building.

### 2.5.4 A consideration to solve the overflow problem in distribution power system

The test bench of University of Denver is used to describe how the operators monitor the real time system, and discuss the ways to duel with the unexpected happening. 6 main buildings, which always requires a big load power, are considered as the targets to overcome the problems. They are called Ritchie Center, Law, Sturm, Daniels, Newman and Olin Hall. In the following analysis, Sturm Hall and Ritchie Center are mainly considered. Bigger size of the green circle indicates a bigger power flow and small size indicates a smaller power flow. The flowing direction of the green circles represents the power flow direction in lines.

In power system, the power of transmission increases along with the increasing of load demand. When the load increases to the load equivalent impedance and the system equivalent impedance, the transmitted active power reaches the limit (when the impedance angle of the load equivalent impedance and the system equivalent impedance is opposite to each other), the power limit obtains the maximum value. on the other hand, when the load increases beyond the critical point, there will be a voltage collapse, which results the power flow does not converge. Because of that, the power flow equation has no solution or a infinite solution. Due to the heavy load on transmission lines of distribution system, the large current causes a line heating, which makes line strain.

In Fig. 2.6, the demand load of Sturm Hall increase sharply, which requires more energy from the substation (”G” in the Figure). Based on the limitation of each building and transmission line, Sturm issues a warning with the sharply increasing load demand. The operators can access the visualization platform to monitor if there is an issue and which building has the issue.

In Fig. 2.7, there is another way to remind the operators to pay attention on the demand load of the building. Hourly demand load of Ritchie Center increase and decrease randomly, which highly depends on the events hold in the building. A continuous 8 hour demand load is shown in the figure, green color of the bar means the real time load demand is under


Figure 2.6: System Visualization Platform in General Situation

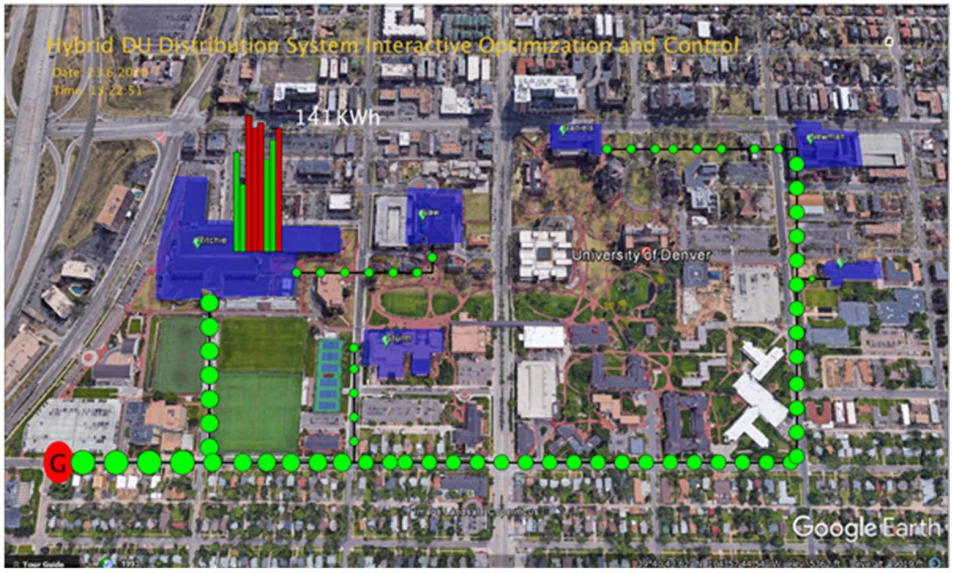



Figure 2.7: System Visualization Platform in Unexpected Situation

the maximum threshold, and the red color is warning the operators to take some actions if needed. It is convenient for operators to monitor the whole campus in directly. And they can decide how to due with it if unexpected problems happen.

There are several methods to avoid the sharply transmitted power reach the limits. The methods below are used to describe how to deal with the problems.

1. Add local renewable generators to satisfy the sharply increase load demand. Based on research, the sharply increasing demand load always happens in the daytime. Installing local PV panels is a potential way to reduce the power supply pressure of the school. Considering the prolonged sunshine of Denver and easy installation of solar energy, an efficient solar generation system could benefit both the power transmission and operation cost. Besides that, solar energy is clean with almost no pollution. In every year, the amount of solar radiation that reaches the earth's surface is about 130 trillion tons of coal, the largest energy available in the world today. It is a big advantage compare with the traditional energy, like coal. How to handle the uncertainty of solar energy is a challenge, which requires a high accuracy simulation model in future research work from the operators.

2. Compare with add extra power to the system, demand response supply a way for operators to work as a important role in electric power grid operation. The operators are allowed to operate the network by reducing or shift the electricity usage of peak time period base on some time series factors, such as the electricity price. For example at school, we can schedule the classes in off-peak period, or avoid all classes start at same periods. The problem is that the feasibility of plan from operators is doubted sometime. More constraints are necessary for objective model from the operators.

After a series operation on the model, Fig. 2.8 shows the system works in a general situation, the operators need to do nothing with the system right now. It is clear that the proposed visualization platform can help operators to monitor the system in real and make decision in a short time. The platform provides a convenient way for observers to get knowledge from the current power flow.

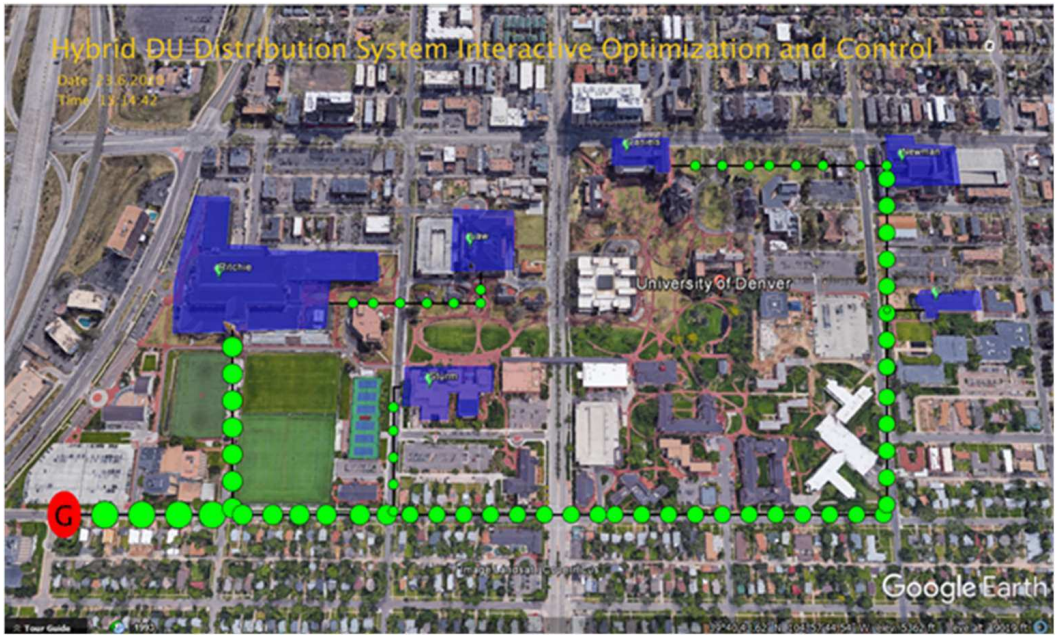



Figure 2.8: System Visualization in Monitoring Situation

## 2.6 Conclusion and Contributions

Data visualization platforms are constructed for operators to assess the load power and renewable energy in convenience. The results shows a novel way on how to use the platform to visualize the real time power flow and monitor the alert in system.

The **main contributions** of this section are:

1. Data visualization platforms are constructed for operators to assess the load power and renewable energy in convenience.

2. The visualization platforms provide the operators a clear understanding of the information in a short time. The platform can be used to visualize a large volumes of data, which helps the customers to make decision more efficiently.

# Chapter 3

# Load forecasting with Support Vector Regression (SVR)

## 3.1 Introduction

In modern power system, load forecasting is an indispensable role for power system operation and control. In traditional method, auto regressive integrated moving average is a widely used tool for time series analysis. Based on linear regression, support vector regression achieves higher performance with the kernel trick, which maps the input data into higher dimensions.

## 3.2 The Architecture of the Proposed Approach

In the left part of Fig. 3.1, the optimization of the hyper-parameters contains Grid traverse algorithm (GTA) and PSO to avoid be trapped into local minimum. And the right part is the distribution network reconfiguration part to validate the forecast results. In the first step, the global solution space is spitted by the GTA into the local solution spaces. In the Map phase, because of the independency among the local solution spaces, they are

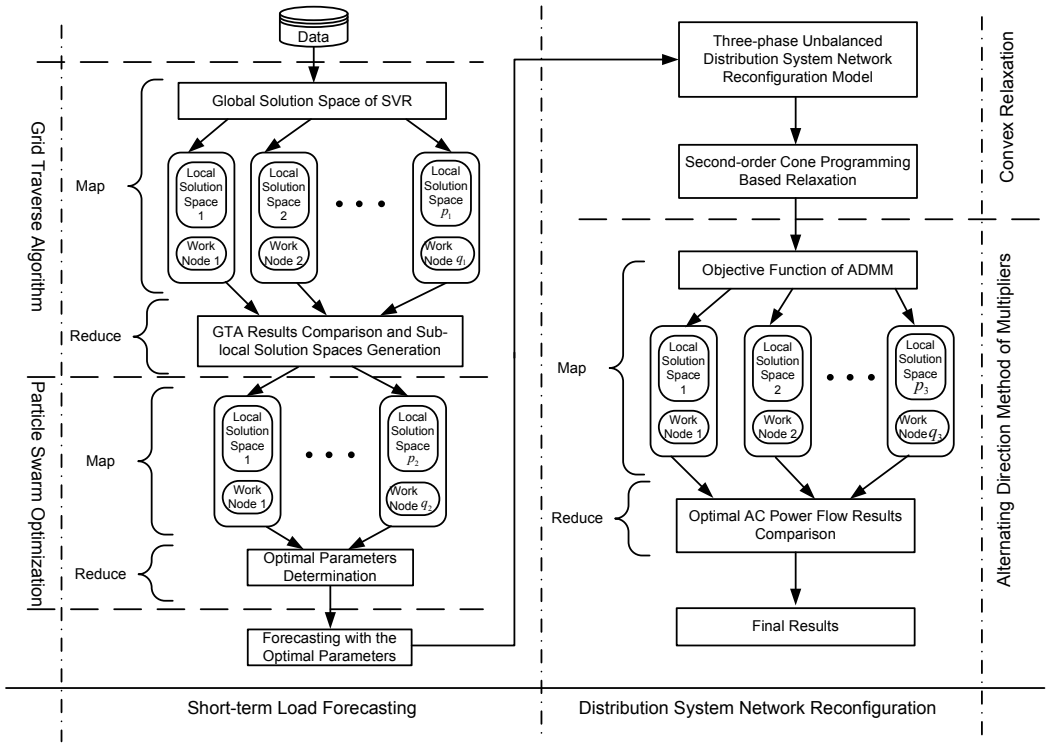



Figure 3.1: The flowchart of the proposed approach.

traversed by GTA in parallel. In the Reduce phase, one or several local solution spaces are selected with the minimum training errors. In the second step, the selected local solution spaces are optimized by the particle swam optimization (PSO) in the similar manner. The optimal parameters can be generated after comparison in the Reduce phase. In the right part of Fig. 3.1, in the part of distribution system network reconfiguration, a three-phase balanced distribution system model is built with the forecasted load profiles. The Second-order cone programming (SOCP) is used to relax the three-phase balanced optimal power flow problem into a convex problem. In the Map phase, the ADMM is used to compute the three-phase balanced optimal power flow in parallel.

## 3.3 SVR-Based Short-term Load Forecaster

Based on the architecture design, in this section, a SVR-based load forecaster is designed for short-term distribution system load forecast. Then, in the next step, the distribution network reconfiguration demo is used as an example to validate the forecast results.

### 3.3.1 SVR Formulation

In this part, the SVR-based short-term load forecaster is trained by the collected historical data to get the optimal hyper-parameters. The objective function of the Kernel based SVR can be built to minimize the forecast error with the soft margin as follows:

$$R_{risk} = \min_{\varepsilon,\omega,\xi_{i_1},\xi^*_{i_1},C,b,\gamma} \{\frac{1}{2}\omega^T\omega + C\sum_{i_1=1}^{n}(\xi_{i_1} + \xi^*_{i_1})\} \tag{3.3.1}$$

Subject to

$$\begin{cases} L'_{i_1} - f(x_{i_1}) \leqslant \varepsilon + \xi_{i_1}, \\ -L'_{i_1} + f(x_{i_1}) \leqslant \varepsilon + \xi^*_{i_1}, \\ \xi_{i_1}, \xi^*_{i_1} \geqslant 0. \end{cases} \tag{3.3.2}$$

where in (3.3.1),$f$ is a Kernel based regression function, $i_1$ is a time index, $\frac{1}{2}\omega^T\omega$ indicates the flatness of the regression coefficients, the second item is the tube violation, $C$ is a trade-off coefficient between the first two items, $\xi_{i_1}$ and $\xi^*_{i_1}$ indicates the two training errors. In [63], the risk function (3.3.1) with the constraint (3.3.2) can be derived to a dual problem with Karush-Kuhn-Tucker (KKT) condition. However, the parameters $\gamma$, $C$, and $\varepsilon$ are still need to be determined, which are the critical factors to the performance of the forecaster [63]. The detail derivative of the SVM or SVR and its dual forms can be found in [63, 97]. Then, as shown in Fig. 3.1, a two-step based parameter optimization approach is designed to compute the optimal parameters.

### 3.3.2 Two-step parameter optimization

Because the parameters $\gamma$, $C$, and $\varepsilon$ cannot be solved with the convex optimization, they are be defined as hyper-parameters in [98]. There are several approaches are proposed for the hyper-parameters such as random search and Gaussian process [98, 99]. Considering the complexity and feasibility, a grid traverse search based two-step hyper-parameter optimization is proposed for the SVR based short-term load forecasting [64].

**First Step: GTA Procedure**

As shown in Fig. 3.1, the GTA procedure is the first step for the hyper-parameters optimization.

1. In the second step, the local solution spaces can be searched with the PSO based approach. The proposed approach is based on the increasing computation capability and new computer cluster cooperation soughs. The three hyper-parameters are initialized with their upper bounds, lower bounds, and grid searching steps.

2. Then, a traversing vector $\mathbf{H}$ can be generated as a finite multi-Cartesian product, which is critical for the Mapreduce process. $H_{j_2}$ is an element in $\mathbf{H}$. For each $H_{j_2}$, the loss function of SVR $R_{risk}$ can be computed independently. As shown in Fig. 3.1, they can be computed in parallel to reduce the computation time.

3. In the last step, the minimum $R_{risk}$ is selected. In addition, if several $H_{j_2}$ are selected, all of them are transmitted to the second step for PSO optimization.

**Second Step: PSO Procedure**

The PSO procedure is designed as a ”fine” optimization for the hyper-parameters. It can be implemented as the scenarios with less time consumption requirements. The acceleration coefficients are defined as $\varphi_1$ and $\varphi_2$, $\theta_1$ and $\theta_2$ can be seemed as two independently

**Algorithm 1** GTA for Hyper-parameter Optimization

**Objective**: Shrink the global solution spaces into one or several local solution spaces.

**Initialization**: Hyper-parameters initialization and multi-Cartesian product generation for the GTA.

**Grid Traverse Searching**: For each core or each process, assign the $H_{j_2}$ to compute the $R_{risk}$, which can be computed in parallel with the Mapreduce model.

**Determine Local Solution Space**: Collected all the results, and select the local solution spaces with minimum $R_{risk}$.

weightiness coefficients. The best historical position and the best position are defined as $\boldsymbol{\eta}_{i_3}^{\Omega}$ and $\boldsymbol{\eta}_{g}^{\Omega}$, respectively. And hyper-parameter $\Omega = [\gamma\ C\ \varepsilon]$. $\boldsymbol{\alpha}_{i_3}^{\Omega}(t)$ and $\boldsymbol{\nu}_{i_3}^{\Omega}(t)$ are the position and velocity vectors, respectively.

$$\begin{aligned}\boldsymbol{\nu}_{i_3}^{\Omega}(t) =& \boldsymbol{\nu}_{i_3}^{\Omega}(t-1) + \varphi_1\theta_1(\boldsymbol{\eta}_{i_3}^{\Omega} - \boldsymbol{\alpha}_{i_3}^{\Omega}(t-1)) \\ &+ \varphi_2\theta_2(\boldsymbol{\eta}_{g}^{\Omega} - \boldsymbol{\alpha}_{i_3}^{\Omega}(t-1)),\end{aligned} \tag{3.3.3a}$$

$$\boldsymbol{\alpha}_{i_3}^{\Omega}(t) = \boldsymbol{\alpha}_{i_3}^{\Omega}(t-1) + \boldsymbol{\nu}_{i_3}^{\Omega}(t), \tag{3.3.3b}$$

## 3.4 Distribution System Network Reconfiguration

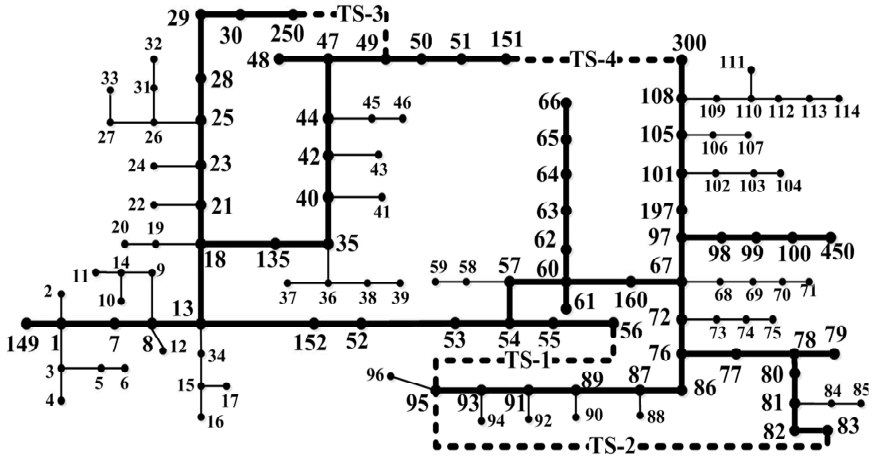
As shown in the Fig.3.1, the right part is distribution system network reconfiguration, which a demo to validate the forecasting results. The detailed information is described in this section with numerical results and analysis.

The topology of a distribution system can be represented in a graph with buses and branches: $\mathcal{G} = [\mathcal{V}, \mathcal{E}]$. Then the branch flow model can be built as follows [100, 101].

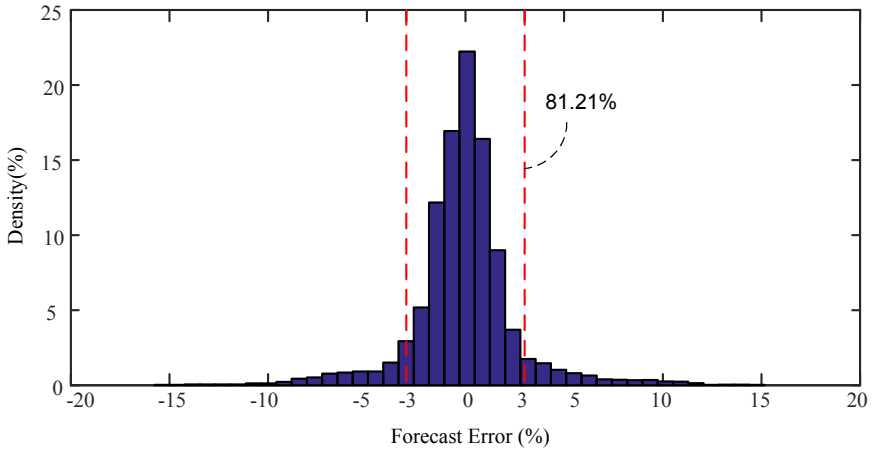
$$s_i = \sum_j S_{ij} - \sum_k (S_{ki} - l_{ki} z_{ki}), \tag{3.4.1a}$$

$$v_j = v_i - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2)l_{ij}, \tag{3.4.1b}$$

$$l_{ij} = (P_{ij}^2 + Q_{ij}^2)/v_j, \tag{3.4.1c}$$




(a)




(b)

Figure 3.2: (a) The IEEE 123-bus based distribution system, (b) the error distribution of short-term load forecasting.

where $l_{ij} := |I_{ij}|^2$, $v_i := |V_i|^2$, $S_{ij}$, $P_{ij}$, $Q_{ij}$ and $z_{ij}$ indicate the complex power flow, active power, reactive power, and impedance on branch $ij \in \mathcal{E}$, $S_{ij} = P_{ij} + \mathbf{i}Q_{ij}$, $P_{ij} = |I_{ij}|^2 r_{ij}$ and $z_{ij} = r_{ij} + \mathbf{i}x_{ij}$. In the formulations, it can be figured out that the branch flow model is used to compute in parallel. In the system, the target node only needs to exchange the information to its neighbors nodes instead of all other nodes. In traditional ways, the nodes need to get information from all nodes, which cannot be realized in parallel work.

Table 3.1: Results of Network Reconfiguration

| No. | Scenario | Bus No. | Opened Switches | Original $P_{Loss}$ | New $P_{Loss}$ | **Loss Reduction** |
|---|---|---|---|---|---|---|
| 1. | Load Increasing | 83 | TS-3, TS-4 | 54.2 kW | 36.7 kW | 32.28 % |
| 2. | Load Increasing | 300 | TS-2, TS-3 | 42.4 kW | 31.5 kW | 25.71 % |
| 3. | Load Increasing | 95 | TS-3, TS-4 | 78.5 kW | 67.0 kW | 14.65 % |
| 4. | Load Increasing | 49 | TS-1, TS-4 | 17.6 kW | 13.4 kW | 23.86 % |
| 5. | Load Decreasing | 47 | TS-2, TS-4 | 39.4 kW | 22.3 kW | 43.40 % |
| 6. | Load Decreasing | 108 | TS-1, TS-4 | 24.1 kW | 20.1 kW | 16.59 % |
| 7. | Load Decreasing | 250 | TS-3, TS-4 | 29.5 kW | 26.8 kW | 9.15 % |
| 8. | Load Decreasing | 56 | TS-1, TS-2 | 35.2 kW | 33.7 kW | 4.27 % |

During the operation of network reconfigurations, the topology of the distribution system is keeping radial and avoid any loops, which are characteristics of a model using branch flow model. Considering the characteristics of the three-phase balanced system, the SOCP relaxation inequalities are represented as follows [100, 101]:

$$\frac{|S_{ij}|^2}{v_i} \leq l_{ij}, \tag{3.4.2}$$

where (3.4.2) can be used to instead of (3.4.1c) as the inequalities constraints. The objective function is defined as total line loss as follows. The constraints contain (3.4.1a), (3.4.1b), (3.4.2) helps tp define the maximum voltage and power flow. The maximum amount of the voltage and power flow are important features for solving the problems here.

$$F = \sum_{\mathcal{E}} P_{ij}, \tag{3.4.3}$$

$$V_{i,min} \leq V_i \leq V_{i,max}, \tag{3.4.4a}$$

$$I_{ij} \leq I_{ij,max}. \tag{3.4.4b}$$

$$\mathbf{rank}(A) = N - d, \tag{3.4.5a}$$

$$\sum_{\mathcal{E}} a_{ij} = N - d, \tag{3.4.5b}$$

$$\sum_{\mathcal{E}_k} a_{ij} = M_k - 1. \tag{3.4.5c}$$

Considering the ADMM, the objective function (3.4.3) with the constraints (3.4.1a), (3.4.1b), (3.4.2), (3.4.4a), and (3.4.4b) can be decomposed into a dual problem. The detail derivatives of the ADMM can be found in [100, 101]. During the parallel traverse of all statuses of the switches, the topology of the distribution system is keeping as radial.

$A$ is the adjacency matrix of the graph $\mathcal{G}$. $d$ is the number of slack bus. $N$ is the number of buses $\mathcal{V}$. $M_k$ is the number of branches in path $\mathcal{E}_k$, and $a_{ij}$ is an element of $A$:

$$a_{ij} = \begin{cases} 1, \ if \ bus \ i \ and \ bus \ j \ are \ connected, \\ 0, \ else. \end{cases} \tag{3.4.6}$$

Considering the limited number of switches, the proposed approach is designed to traverse all the permutations and determine the optimal configuration of the distribution system. For example, the modified IEEE 123-bus system with 4 switches indicates 16 scenarios with all the permutations of the switches [54].

With the topology constraints discussed above, the number of scenarios can be reduced in different scenarios. Then, considering the independency of each configuration (permutation), all the permutation can be implemented into different cores or processes and computed independently, which dramatically reduces the computation time and keeps the convexity to get the guaranteed optimization results.

## 3.5 Numerical Results

As shown in Fig. 3.2(a), the test bench is based on the IEEE 123-bus distribution system. Four initially opened tie switches TS-1, TS-2, TS-3 and TS-4 are added to make the system topology changeable, and the detail information can be found in [54, 102].

### 3.5.1 Short-term Load Forecasting

The test load data contains four seasons of one year. The proposed approach is used for 1-hour-ahead sliding window forecasting with 1 second resolution. The training data is 5 times as the test data. The distribution of the forecaster errors is shown in Fig. 3.2(b). The MAPE is 2.23%, normalized root-mean-square error (NRMSE) is 4.03%, and more than 80% of the errors are accumulated between (-3.1%, 3.1%).

### 3.5.2 Network Reconfiguration

As shown in Table 3.1, considering the load increasing, for example, in scenario 1, with the forecasting results, there is a load increasing 20.31% in bus 83. The system loss reduces 32.28% with the proposed approach. For the load decreasing, for example, in scenario 5, with the forecasting results, there is a load decreasing 55.4% in bus 47. The system loss reduces 43.40% with the proposed approach. The average loss reduction for the load increasing scenarios is 24.13%, the average loss reduction for the load decreasing scenarios is 18.35%, and the total average for all scenarios is 21.24%.

### 3.5.3 Comparison

As shown in Table 3.2, compared with the traditional network reconfiguration approach with the genetic algorithm (GA), the proposed approach has less computation time and more loss reduction. Furthermore, the proposed approach is more intuitive, and convenient for implementation in different programming language such as python and Matlab.

Table 3.2: Performance comparison

| Methods | Loss Reduction | Computation Time (s) |
|---|---|---|
| GA based traditional | 17.87% | 107 |
| Proposed Approach | 21.24% | 30 |

## 3.6 Conclusion and Contributions

Consider these benefits, an advanced distribution load forecasting approach is proposed for load forecasting. In order to validate the forecast result, a distribution network reconfiguration demo is designed to demonstrate the feasible and efficient of the proposed approach.

The **main contributions** of this section are:

1. The novel method of grid traversal algorithm(GTA) with SVR is used to optimize the loss function with choosing the best hyperparameter combination. SVR is used to map the nonlinear data into high dimensions.

2. The two level optimized model is released with GTA and particle swam algorithm, which helps to figure out the best global minimal results. The grid search method efficiently figure out the best combination of hyper-parameters

# Chapter 4

# Multi-timescale distribution system optimization with renewable energy

## 4.1 Introduction

With a distribution system, a three phase unbalanced multi-timescale approach is proposed. The day-ahead dispatch of the substation level is used to minimize the operation cost with renewable energy hourly. The uncertainty of renewable energy is simulated with chance constraints and Gaussian Mixture Model is used to simulate the output of multiple renewable generation for a higher accuracy. Genetic Algorithm is used to further increase the accuracy of GMM results by automatically select the amount of components. At the feeder level, an OPF problem is formulated for the three-phase unbalanced system with considering the reality, the non-convex problem with three variables are relaxed by semidefinite programming. In the proposed multi-timescales model, the feeder scheduling is described with higher time resolution and update frequencies, which can be obtained near RT calculation.

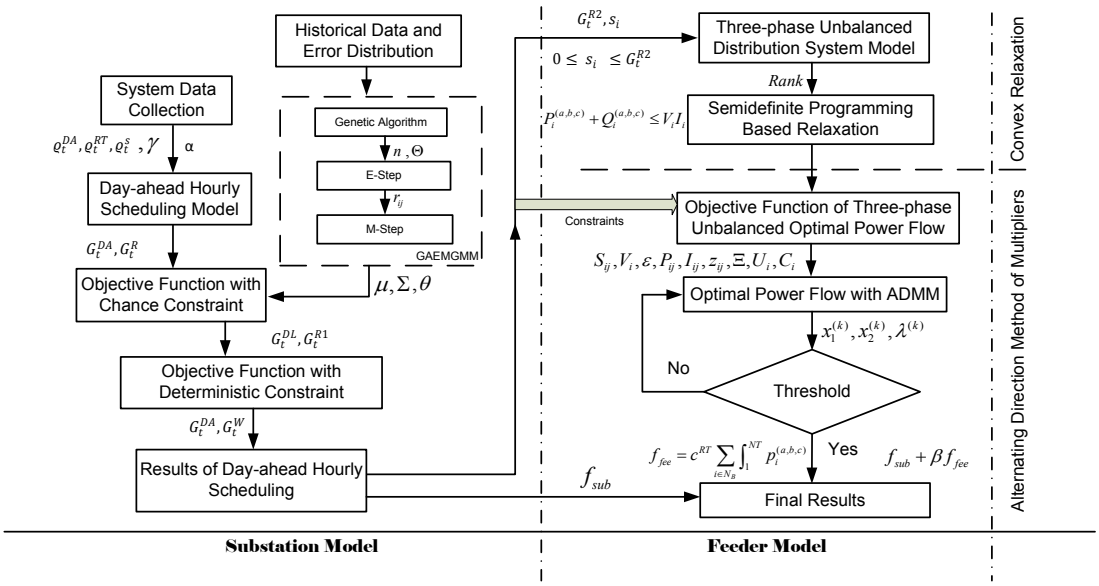



Figure 4.1: The flowchart of proposed approach.

## 4.2 The Flowchart of the Proposed Approach

In this method, multi-timescale stochastic algorithm [103] is applied to improve the operation cost of a distribution power system. We have proposed the configurable stochastic approach with the multi-timescale scheduling procedure, which consists of submodels including day-ahead power scheduling, real-time power trading constraint and OPF in the distribution system. The two parts stochastic optimization is implemented in this method.

As in Fig. 4.1, the proposed approach consists of two parts, the stochastic optimization for hourly scheduling at the substation, and the three-phase unbalanced OPF for minutes operation at feeders. By the proposed two-part framework, the optimal day-ahead scheduling power purchased from the utilities for the next 24 hours is determined in the first part. In the second part, based on the results of the first part, the OPF is computed in minutes to reduce the system loss and the total system cost within an hour. In the left part of Fig. 4.1, the error distribution of the forecast result is given by the historical data of the distribution renewable generation. Combined with the day-ahead hourly scheduling, an objective

function can be formulated with the chance constraints for the forecast errors. According to GAEMGMM, the error distribution model can be accurately modeled with several Gaussian components. And the chance constraints can be formulated into the deterministic forms for the stochastic optimization. Finally, the optimal hourly schedule can be determined and the optimal operation cost is computed at the substation level.

In the right part of Fig. 4.1, at feeder level, a three-phase unbalanced optimal power flow is used to model and compute the distribution system loss in minutes level. In this method, the renewable generation such as micro wind turbines and PV panels derated, $2.5\%$ power are reserved for the OPF regulation, which aims to minimize the system loss. Considering the nonconvexity of the three-phase unbalanced OPF, SDP is used to relax the problem. Then, the objective function of system loss can be solved with ADMM to further reduce computation time. Finally, the three-phase unbalanced OPF can be computed to minimize the system loss at feeder level.

The total cost is formulated in (4.2.1), which can be presented as the sum of the operation cost at the substation level and feeder level:

$$C_{total} = \min(f_{sub} + \beta f_{fee}) \tag{4.2.1}$$

where the cost of substation and feeder levels are ($f_{sub}$) and ($f_{fee}$), respectively, and $\beta$ is a weightness coefficient. Given three different ways to working with $\beta$, the operation cost at the substation level (feeder level) is attacking more attention if $\beta > 1$ ($\beta < 1$). $\beta$ is defined as 1, when the two levels are the same important.

## 4.3 Day-ahead hourly scheduling at the substation level

### 4.3.1 Problem formulation

$$f_{sub} = \sum_{t=1}^{T} \big( (\varrho_t^{DA} G_t^{DA} + \varrho_t^{R} G_t^{R}) + (\delta_t \cdot \varrho_t^{RT} G_t^{RT}) + (1 - \delta_t) \cdot \varrho_t^{s} G_t^{RT} \big) \tag{4.3.1}$$

$t = \{1, 2, \cdots, T\}$ represents the time intervals, $t$ indicates the $t$th time interval and $T$ is used to describe how many time intervals we have in this situation. $\varrho_t^{DA} G_t^{DA}$ is the generation cost of day-ahead scheduling and $\varrho_t^{R} G_t^{R}$ is the renewable generation cost. $\delta_t \cdot \varrho_t^{RT} G_t^{RT}$ presents the deviation power purchased from RT market to satisfy the power balance equation. And $(1 - \delta_t) \cdot \varrho_t^{s} G_t^{RT}$ is the corrective action (CA), which indicates the redundant energy will be reselled to the market in a lower price. It is noticed that $G^{R1}$ is used to supply the power consumption at the substation level for the first part, which is defined as $97.5\%$ of the total renewable power generation for each hour in this method.

A typical forecasting error model represented by a normal distribution is used to incorporate the uncertainty of renewable generation. Here, the load forecasting error model is also represented as normal distribution. The study of the data forecasting is not the major concentrate in this method, and the hourly forecasted renewable generation and the hourly load forecast are described as:

$$G^{R} = G_f^{R} + G_f^{R} * G_{err1} \tag{4.3.2}$$

$$G^{DL} = G_f^{DL} + G_f^{DL} * G_{err2} \tag{4.3.3}$$

$$G_t^{DA} + G_t^{RT} + G_t^W = G_t^{DL} \tag{4.3.4a}$$

$$\varrho_t^s < \varrho_t^R < \varrho_t^{DA} < \varrho_t^{RT} \tag{4.3.4b}$$

$$G^{RT,min} \leq G_t^{RT} \leq G^{RT,max} \tag{4.3.4c}$$

$$G^{DA,min} \leq G_t^{DA} \leq G^{DA,max} \tag{4.3.4d}$$

$$G^{R1,min} \leq G_t^{R1} \leq G^{R1,max} \tag{4.3.4e}$$

$$\delta_t = \begin{cases} 1 & G_t^{DA} + G_t^W \leqslant G_t^{DL} \\ 0 & G_t^{DA} + G_t^W > G_t^{DL} \end{cases} \tag{4.3.4f}$$

$$G_t^{RT} \begin{cases} \geqslant 0 & buy\ power\ from\ bulk\ system \\ < 0 & sell\ power\ to\ bulk\ system \end{cases} \tag{4.3.4g}$$

$$Pr(G_t^{DL} \leqslant G_t^{DA} + G_t^W) \geqslant \gamma \tag{4.3.4h}$$

$$Pr(\rho G_t^{R1} - G_t^W \leqslant 0) \geqslant \alpha \tag{4.3.4i}$$

The constraints include the modified power balanced equation, market price limitation, capacity limits for $G_t^{RT}, G_t^{DA}, G_t^{R1}$. $\delta_t$ is a binary variable to describe the relationship between $G_t^{DA} + G_t^W$ and $G_t^{DL}$. It is easy to understand that the $\varrho_t^s$ should be the lowest in case of purchasing the redundant power generation from the day-ahead market. (4.3.4h) defines that the demand load will not exceed the sum amount of the renewable generation and the day-ahead power scheduling with a prescribed probability $\gamma$. In (4.3.4i), for each hour, the amount of the used renewable generation for all buses should be larger than the renewable generation at chance $\alpha$, where $0 < \rho < 100\%$.

### 4.3.2 Genetic-based Expectation Maximization Algorithm for learning Gaussian Mixture Model

A brief description of the finite GAEMGMM is given below. Then, it is used to model the forecasting error distribution of the renewable generation.

GMM is an unique form of the finite mixture model. For the finite mixture model in (4.3.5), it is the sum of more than one components $(N > 1)$ with different weights $(\epsilon_n)$ in $\mathcal{R}^q$, which indicates:

$$p(x|\mathbf{\Theta}) = \sum_{n=1}^{N} \epsilon_n p(x|\theta_n) \tag{4.3.5}$$

The weights in (4.3.5) can be calculated in [104] and has the intuitive interpretation to be non-negative $(\epsilon_n \geqslant 0)$, and the sum equals to 1. Each component in the model obeys to a normal distribution, which is restricted by $\theta_n = (\mu_n, \mathbf{\Sigma}_n)$, the matrix of the means vector and the covariance.

The EM algorithm is used as the standard approach to calculate the parameters of the mixture model. It consists of an expectation-step (E-step) and a maximization-step (M-step).

For E-step, the complete data $\mathcal{Z} = \{(x_1, \xi_1), (x_2, \xi_2), \cdots, (x_m, \xi_m)\}$, $\{x_1, \cdots, x_m\}$ is known as the observed data and incomplete, $\xi_m$ is the component identity of $x_m$. The algorithm will be ended when the $\log$ likelihood function in (4.3.6) reaches the convergence.

$$\mathcal{L}(x|\Theta) = \sum_{m=1}^{M} \log \Big( \sum_{n=1}^{N} \xi_n(x_m|\theta_n) \Big) \tag{4.3.6}$$

The posterior probability $(\xi_n)$ at the $l$-th iteration is computed as equation below. It is used to calculate the hidden amount for next step.

$$\xi_n^{(l)} = \frac{\epsilon_n^{(l)} p(x_m|\theta_m)}{\sum_{n=1}^{N} \epsilon_n^{(l)} p(x_m|\theta_n^{(l)})} \tag{4.3.7}$$

For M-step, the parameters of the GMM is reestimated by $\xi_n^{(l)}$. The parameters can be updates as the equations below for Gaussian distributions.

$$\epsilon_n^{(l+1)} = \frac{\sum_{m=1}^{M} \xi_{m,n}}{m} \tag{4.3.8}$$

$$\mu_n^{l+1} = \frac{\sum_{m=1}^{M} \xi_{m,n} X_m}{\sum_{m=1}^{M} \xi_{m,n}} \tag{4.3.9}$$

$$\Sigma_n^{(l+1)} = \frac{\sum_{m=1}^{M} \xi_{m,n} (x_m - \mu_n^{(l+1)})(x_m - \mu_n^{(l+1)})}{\sum_{m=1}^{M} \xi_{m,n}} \tag{4.3.10}$$

For learning the GMM in [80], the GAEM shows its superiority on selecting the number of components based on the minimun description length (MDL) criterion. In [105], the MDL criterion is widely used on selection. Compared with the standard EM, the new algorithm is less sensitive to the initialization. GAEM is capable to explore the parameter more thoroughly, because of its population-based search skill. Meanwhile, the new algorithm still remain the property of monotonic convergence as before.

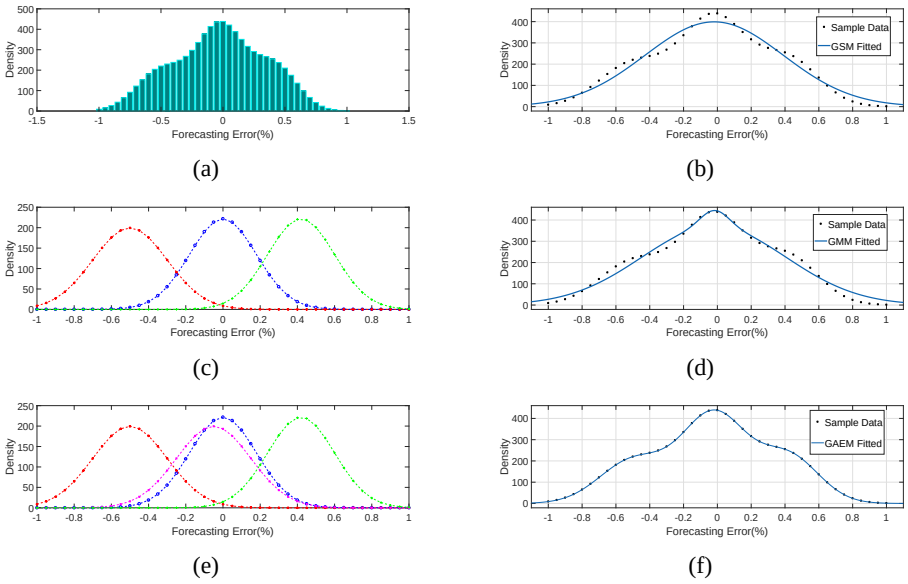



Figure 4.2: Comparisons between GSM, GMM and GAEMGMM: (a) the original aggregated error distribution, (b) the modeled result with GSM and comparing with the original original aggregated error distribution, (c) the original aggregated error distribution is modeled with a GMM consists of 3 GSMs, (d) the GMM fitted result compare with the original aggregated error distribution, (e) the original aggregated error distribution is modeled with a GMM consists of 4 GSMs, (f) the GAEMGMM fitted result compare with the original aggregated error distribution.

A comparison with the GSM, GMM, and the GAEMGMM is illustrated as shown in Fig. 4.2. Fig. 4.2(b) displays the modeled error distribution result by GSM, which cannot modeled the aggregated error distribution accurately. The results of standard GMM with 3 components are shown in Fig. 4.2(c) and Fig. 4.2(d). Because the standard GMM cannot determine the number of the components automatically, which also bring some errors in the error distribution modeling. The GAEMGMM are employed to model the original error distribution in Fig. 4.2(e) and Fig. 4.2(f), which can automatically determine the number of the components (like how many clusters in a Gaussian distribution) and reduce the modeling error in this process instead of the traditional ways.

$$\eta = \frac{(\nabla_d - \nabla_{org})^2}{\nabla_{org}^2} \cdot 100\% \tag{4.3.11}$$

As in (4.3.11), the ratio of the residual deviation $\eta$ is used to evaluate the performances of different approaches, where $\nabla_{org}$ is the envelope of the original forecasting error distribution, and $\nabla_d$ is the envelope of the deformed forecasting error model by GSM, general GMM and GAEMGMM. According to (4.3.11), the ratio of the residual deviation of GSM is 2.17%, GMM is 1.27%, the GAEMGMM is 0.13%.

### 4.3.3 Chance Constraint

According to discussion above, the proposed GAEMGMM is capable to accurately model the aggregated error distribution and determine the components number at each feeder. It can effectively convert the chance-constraint of the forecasting error model in (4.3.4h) (4.3.4i) into a deterministic problem.

Because $G_{err1}$ and $G_{err2}$ is normally distributed with mean $(\mu_1, \mu_2)$, covariance $(\Sigma_1)$ for multiple renewable generators and variance $(\sigma_2^2)$ for load forecast. The joint probability distribution of the renewable generators can be obtained by the proposed GAEMGMM model with a high accuracy, which demonstrates the efficiency of the proposed approach further more.

Based on the chance-constraint in (4.3.4h) and (4.3.4i), it is assumed that

$$y_1 = G^{DL} - G^{DA} - G^W \tag{4.3.12}$$

$$y_2 = \rho G^{R1} - G^W \tag{4.3.13}$$

The expectation and the variance can be calculated as:

$$E(y_1) = G_f(1 + \mu_2)^{DL} - G^{DA} - G^W \tag{4.3.14}$$

$$V(y_1) = G_f^{DL} * \sigma_2 \tag{4.3.15}$$

$$E(y_2) = \rho G_f^{R1}(1+\mu_1) - G^W \tag{4.3.16}$$

$$V(y_2) = G_f^{R^T} \Sigma_1 G_f^R \tag{4.3.17}$$

According to this, the chance-constraint in (4.3.4h), (4.3.4i) can be converted in a deterministic formulation as (4.3.18) and (4.3.19) with the proposed GMM model.

$$Pr(y_1 \leqslant 0) = \Phi\Big(\frac{0 - E(y_1)}{V(y_1)}\Big) \geqslant \gamma \tag{4.3.18}$$

$$Pr(y_2 \leqslant 0) = \Phi\Big(\frac{0 - E(y_2)}{V(y_2)}\Big) \geqslant \alpha \tag{4.3.19}$$

Where $\Phi(\cdot)$ indicates the cumulative distribution function of the standard normal distribution. Taking the inverse of $\Phi(\cdot)$ helps to rewrite the following analytical formulation.

$$G^W + G^{DA} \geqslant \Phi^{-1}(\gamma)\sigma_2 G_f^{DL} + G_f^{DL} + \mu_2 G_f^{DL} \tag{4.3.20}$$

$$G^W \geqslant \Phi^{-1}(\alpha)\rho\big((G_f^R)^T \Sigma_1 G_f^R\big)^{\frac{1}{2}} + \rho G_f^{R1} + \rho\mu_1 G_f^{R1} \tag{4.3.21}$$

## 4.4 Optimal power flow (OPF) and the ADMM based Semidefinite Programming (SDP) relaxation

The network of distribution power system is unbalanced typically because of its uneven distribution over the three phases. And the use of single phase generator in the power system will increase the unbalance additionally. The optimal power flow of three phase unbalanced power system is used to minimize the system loss in minute level. There are many factors can influence the performance of OPF with renewable generators. For example, the solar radiance, daily temperature, wind speed and etc. In order to overcome the problems, a minute level optimal power flow method is applied on the test bench of distribution network,

which involves the implementation of three-phase unbalanced with branch flow model, non-convex optimization method and semidefinite relaxation. The three-phase load flow method results in a non-convex problem, because of the optimization of three different variables.

### 4.4.1 Alternating Direction Method of Multipliers(ADMM)

Along with blending the decomposability of dual decomposition, ADMM shows the superior convergence properties of augmented Lagrangians [106]. For a general ADMM problem, the optimization problem is formulated as below.

$$\min_x f_1(x_1) + f_2(x_2) \tag{4.4.1}$$

$$\begin{aligned} A_1x_1 + A_2x_2 &= b \\ x_1 \in \mathcal{K}_1, x_2 &\in \mathcal{K}_2 \end{aligned} \tag{4.4.2}$$

$\mathcal{K}_1, \mathcal{K}_2$ are defined as convex sets. They are the necessary conditions for convex optimization. Then the objective is augmented as (the constraint is as (4.4.2)):

$$\min_x f_1(x_1) + f_2(x_2) + \frac{\rho}{2} \parallel A_1x_1 + A_2x_2 - b \parallel_2^2 \tag{4.4.3}$$

$\rho$ is a constant and never less than zero, which is used to decide if the objective is a augmented $(\rho > 0)$ or standard $(\rho = 0)$ Lagrangian. Then $\lambda$ is generally defined as the Lagrange multiplier for the equality constraint in (4.4.2).

After the formula deformation, ADMM repeats the iterations, for $k = 1, 2, 3 \cdots$

$$x_1^{(k)} = \arg\min_{x_1} L_\rho(x_1, x_2^{(k-1)}, \lambda^{(k-1)}), \tag{4.4.4a}$$

$$x_2^{(k)} = \arg\min_{x_2} L_\rho(x_1^{(k)}, x_2, \lambda^{k-1}), \tag{4.4.4b}$$

$$\lambda^{(k)} = \lambda^{(k-1)} + \rho(A_1x_1^{(k)} + A_2x_2^{(k)} - b). \tag{4.4.4c}$$

Refer to [106], ADMM is guaranteed to reach the converged optimal solution with less restricted conditions. It also helps to reduce the time consumption.

### 4.4.2 ADMM in Proposed Method

It is modeled as a tree radial topology $\mathcal{G}$, each bus $i$ in this system only has one parent bus $U_i$ and a few children buses $C_i$. $\mathcal{G} = (\mathcal{N}_B, \mathcal{N}_l)$, where $\mathcal{N}_B = \{0, 1, \cdots, n\}$ represents the set of buses. $\mathcal{N}_l$ is the set of the distributed lines, which are used to connect the buses in $\mathcal{N}_B$. Each $i \in \mathcal{N}_L = \{1, 2, \cdots, n\}$ indicates the line connected from bus $i$ to its parent bus $U_i$. Compared with bus injection model, branch power flow provides much more use for distribution network and stable computation results. For a three-phase power system, the branch power flow $\mathbf{S}_i^{(a,b,c)}$, complex branch current $\mathbf{I}_i^{(a,b,c)}$, voltage magnitude $\mathbf{V}_i^{(a,b,c)}$ and complex impedance $\mathbf{\Omega}_i^{(a,b,c)}$ are defined as follows, which expressed as instantaneous space vectors and indicates the power flow from bus $i$ to its parent bus $U_i$ :

$$\mathbf{S}_i = \begin{bmatrix} S_i^a \\ S_i^b \\ S_i^c \end{bmatrix} = \begin{bmatrix} P_i^a \\ P_i^b \\ P_i^c \end{bmatrix} + \mathbf{i} \begin{bmatrix} Q_i^a \\ Q_i^b \\ Q_i^c \end{bmatrix} \tag{4.4.5}$$

$$\mathbf{I}_i = \begin{bmatrix} |I_i^a|^2 \\ |I_i^b|^2 \\ |I_i^c|^2 \end{bmatrix} \tag{4.4.6}$$

$$\mathbf{V}_i = \begin{bmatrix} |V_i^a|^2 \\ |V_i^b|^2 \\ |V_i^c|^2 \end{bmatrix} \tag{4.4.7}$$

$$\mathbf{\Omega}_i = \begin{bmatrix} \Omega_i^a \\ \Omega_i^b \\ \Omega_i^c \end{bmatrix} = \begin{bmatrix} r_i^a \\ r_i^b \\ r_i^c \end{bmatrix} + \mathbf{i} \begin{bmatrix} x_i^a \\ x_i^b \\ x_i^c \end{bmatrix} \tag{4.4.8}$$

According to the definition of branch flow model, the network $\mathcal{G}$ is defined as:

$$\begin{aligned} V_{U_i}^{(a,b,c)} =& \mathbf{V}_i - 2(r_i^{(a,b,c)} P_i^{(a,b,c)} + x_i^{(a,b,c)} Q_i^{(a,b,c)}) \\ &+ \mathbf{I}_i((r_i^{(a,b,c)})^2 + (x_i^{(a,b,c)})^2) \end{aligned} \tag{4.4.9}$$

$$P_i^{(a,b,c)} + Q_i^{(a,b,c)} = \mathbf{V}_i \mathbf{I}_i \tag{4.4.10}$$

$$\sum_{j \in C_i} (P_j^{(a,b,c)}) - \mathbf{I}_i r_i^{(a,b,c)} + p_i^{(a,b,c)} = P_i^{(a,b,c)} \tag{4.4.11}$$

$$\sum_{j \in C_i} (Q_j^{(a,b,c)}) - \mathbf{I}_i x_i^{(a,b,c)} + q_i^{(a,b,c)} = Q_i^{(a,b,c)} \tag{4.4.12}$$

$$\begin{pmatrix} \mathbf{V}_i & \mathbf{S}_i \\ \mathbf{S}_i^H & \mathbf{I}_i \end{pmatrix} \in \mathbb{S}_+ \tag{4.4.13}$$

$$rank \begin{pmatrix} \mathbf{V}_i & \mathbf{S}_i \\ \mathbf{S}_i^H & \mathbf{I}_i \end{pmatrix} = 1 \tag{4.4.14}$$

Where $P$ and $Q$ ($p$ and $q$) indicates the bus $i \in \mathcal{N}_L$ ($i \in \mathcal{N}_B$). $S_0$ is the root of the topology with no parent bus. $\mathbb{S}$ and $\mathbb{S}_+$ are used to describe the hermitian and the positive semidefinite matrix, respectively. And $(.)^H$ denotes the hermitian transpose of the matrix. Because the proposed distributed system is radial with unique phase angle of the current and voltage at each bus, the branch flow model in (4.4.9)-(4.4.12) can be regarded as a complete AC power flow now [106]. The objective function is formulated as:

$$f_{fee} = \varrho^{RT} \sum_{i \in \mathcal{N}_B} \int_1^{NT} p_i^{(a,b,c)} \tag{4.4.15}$$

subject to: (4.4.9), (4.4.10), (4.4.11), (4.4.12), (4.4.13), (4.4.14), (4.4.16a), (4.4.16b) and (4.4.17).

The OPF problem in (4.4.15) cannot be regarded as a convex problem because of the rank constraint in (4.4.14). Due to the SDP relaxations in [107], the rank constraint in (4.4.14) can be removed and obtain a lower bound for the revised OPF problem in (4.4.18). It has been improved in [108], the semidefinite relaxed OPF is exact if the optimal solution of (4.4.18) still satisfy the rank constraint and the original OPF problem is also optimal.

$$\mathbf{V_i}_{min} \leqslant \mathbf{V_i} \leqslant \mathbf{V_i}_{max}, i \in \mathcal{N}_B \tag{4.4.16a}$$

$$s_i^{(a,b,c)} = p_i^{(a,b,c)} + \mathbf{i} q_i^{(a,b,c)}. \tag{4.4.16b}$$

$$0 \leqslant s_i^{(a,b,c)} \leqslant G^{R2} \tag{4.4.17}$$

In (4.4.16a), the magnitude of the voltage obeys to a reasonable range. The system controlling parameter $s_i^{(a,b,c)}$ is used to benefit on reducing the system loss, which is defined to be provided by the renewable generation and limited in (4.4.17). Based on the schedule results of first step, $G^{R2}$ is defined as $2.5\%$ of the total renewable generation and used at feeder level for step two.

The relaxation OPF formulation is summarized as follows. It subjects to: (4.4.9), (4.4.10), (4.4.11), (4.4.12), (4.4.13), (4.4.16a), (4.4.16b) and (4.4.17).

$$f_{fee} = \varrho^{RT} \sum_{i \in \mathcal{N}_B} \int_1^{NT} p_i^{(a,b,c)} \tag{4.4.18}$$

Table 4.1: Comparison on $\eta$ for different fitting model

| $\eta$ (%) | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| GAEMGMM | 0.09 | 0.13 | 0.16 | 0.15 |
| General GMM | 1.03 | 1.27 | 1.44 | 1.23 |
| GSM | 2.39 | 2.17 | 2.15 | 2.10 |

## 4.5 Numerical Simulation and Results

### 4.5.1 Numerical Results Analysis on IEEE 123-bus

#### Error Distribution Modeling Comparison

The predicted data, actual data, and error distribution are provided by National Renewable Energy Laboratory (NREL) [109], which can be modeled with different approaches as following. In table. 4.1, the proposed method can obtain a better ratio of the residual deviation $\eta$, as defined in (4.3.11), than others.

#### Day-Ahead Dispatching Cost at the Substation Model

The numerical results for evaluating the proposed method are tested based on the IEEE 123-bus system. Four wind turbines (100kW for each) are connected at bus 25, 35, 76 and 105, respectively. The PV panels (400kW in total) are installed at bus 28, 47, 49, 57, 64, 93 and 97, which are used to demonstrate that the hybrid power system can work efficiently and reliably with the proposed stochastic approach. The hourly renewable generation with the total operation cost is simulated in Fig. 4.3(a), where the blue (yellow) bars indicates the wind (solar) power generation hourly in a day. In this case, the lower limit of the forecasting error model in chance-constraint (4.3.4h) is set to $97.0\%$.

Fig. 4.3(a) describes the total operation cost with and without CA, which indicate with red and green curve, respectively. As shown in Fig. 4.3(a), in Golden, Colorado, a typical day with a windy night and a sunny daytime is selected with 24 hours data. The peak generation of the wind turbines (blue bars) and PVs (yellow bars) occur at midnight and

14:00, respectively. The red line represents the total operating cost with CA, which indicates the system can resell the redundant power at a lower price $\varrho_t^s$ to the current electricity market. Then, the total operation cost can be reduced with the benefit from the resell. The total operation costs at 13:00 and 14:00 for both scenarios (with and without CA) are the same, which indicates there are no redundant energy to resell and the CA doesnt occur. The similar scenarios also occur at 7:00 and 18:00. In the rest time, the redundant energy is resold to the market with a lower price to reduce the total operation cost.

$$a = \frac{G_t^{RT}}{G_t^{DL}} * 100\%, \tag{4.5.1a}$$

$$a = \frac{G_t^{DA} + G_t^{R}}{G_t^{DL}} * 100\%, \tag{4.5.1b}$$

The proposed approach is compared with the GSM based approach in Fig. 4.3(b) for 24 hours. For each hour, the left bar describes the GSM fitted model and its corresponding cost is displayed as the green dashed line.

The right bar describes the GA-EM fitted model and its corresponding cost is displayed as the red dashed line. It is clear that the GA-EM fitted model contains higher accuracy, the corresponding system operation cost is lower than the system operation cost of the GSM fitted model.

Specifically, in the left bar (with GSM fitted model), the yellow bar denotes the $G_t^{RT}$ (in percent, can be computed as (4.5.1) as following) and the rest is $G_t^{DA} + G_t^{R}$. Similarly, in the right bar, the orange and dark blue bars also indicate the $G_t^{RT}$ and $G_t^{DA} + G_t^{R}$ with the GA-EM fitted model. It is clearly that the orange bar is shorter than the yellow bar, which indicates the GA-EM fitted model contains less errors than the GSM fitted model, and the system requires less energy from the RT market. It is different from the signle Gaussian model and get better results.

**Case.1 Different Operation Costs for One Year Data at Substation Level.**

Table 4.2: Comparisons based on Fig. 4.2 at substation level

| Season | Test Days | Average Cost for Each Day($) | | |
|---|---|---|---|---|
| | | Proposed Method | GMM(General) | GSM |
| Spring | 30 | 3345.24 | 3662.58 | 4102.56 |
| Summer | 30 | 5170.86 | 5556.21 | 6321.99 |
| Autumn | 30 | 3123.97 | 3400.02 | 4112.66 |
| Winter | 30 | 5456.02 | 5822.21 | 6363.42 |

Table 4.3: Time Consumption Comparison

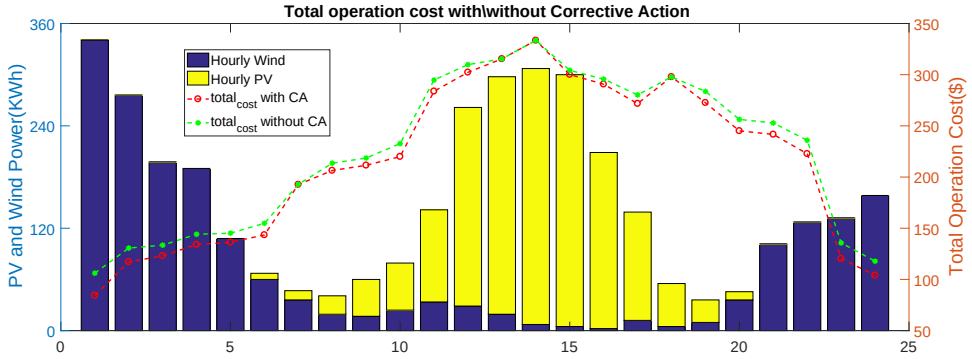
| Method | IEEE 13-bus | IEEE 34-bus | IEEE 123-bus |
|---|---|---|---|
| Genetic Algorithm | 210.77 s | 243.39 s | 507.53 s |
| Interior-Point | 7.72 s | 11.33 s | 27.67 s |
| Proposed Method | 1.59 s | 3.72 s | 10.27 s |

Based on the discussion above, one year data with four seasons are employed to validate the proposed approach. As shown in Table 4.2, the proposed approach contains the minimum cost for spring, summer, autumn, and winter. It is demonstrated that with the high accuracy fitted model, the corresponding operation cost can be reduced significantly, which also indicates the related fuel and carbon emission can be saved.

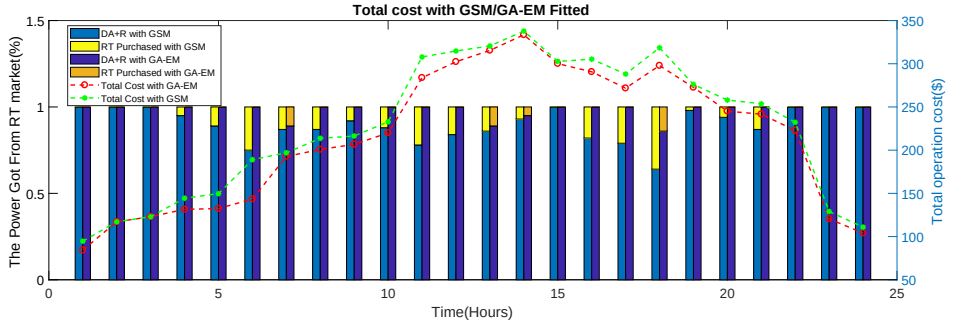
**Case.2 Total Operation Cost with Different Percentage Limit in Chance Constraint.**

The operating cost with the $\gamma$ of $95.0\%$ is lower than others. The highest cost is obtained with $\gamma$ = $99.0\%$. This illustrates that the higher reliability can result in higher total operation cost. This is due to that a bigger $\gamma$ places a stricter constraint that $G_t^{DA} + G_t^W$ is higher or equal to the $G_t^{DL}$ in (4.3.4h). On the contrary, the operating cost decrease if we define a smaller $\gamma$ for the system.

As shown in Fig. 4.4, the chance constraint probability $\gamma$ in (4.3.4h) is set as 95.0%, 97.0%, 99.0% with blue, red, and orange color. It is clearly that the operating cost with $\gamma$ = $95.0\%$ is the lowest among others, and the highest cost is obtained with $\gamma$ = $99.0\%$. These results illustrate that a high chance constraint probability $\gamma$ requires a high demand of $G_t^{DA}$ in equation (4.3.4h), which indicates the operation cost is also increasing.




(a)




(b)

Figure 4.3: (a) Total Operation Cost with and without the CA. (b) Total Cost with GSM based approach and proposed approach.

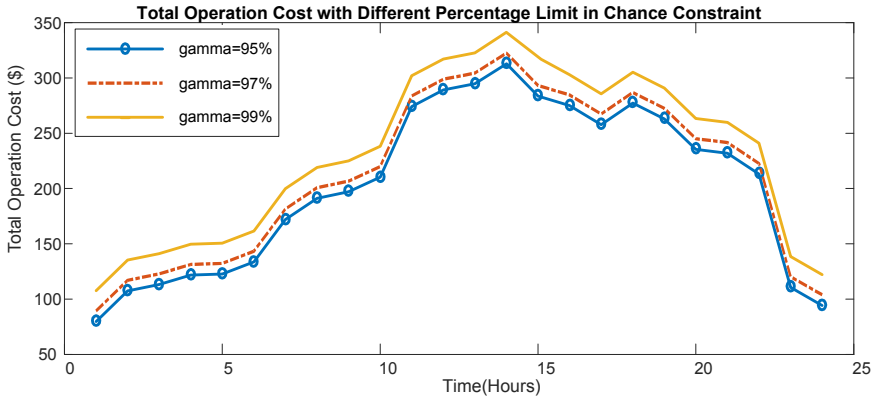



Figure 4.4: Total Operation Cost with Different $\gamma$ in Chance Constraint.

Table 4.4: Line Loss Comparison with or without OPF

| kWh | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| Line Loss with OPF | 227 | 324 | 226 | 521 |
| Line Loss without OPF | 316 | 432 | 307 | 657 |

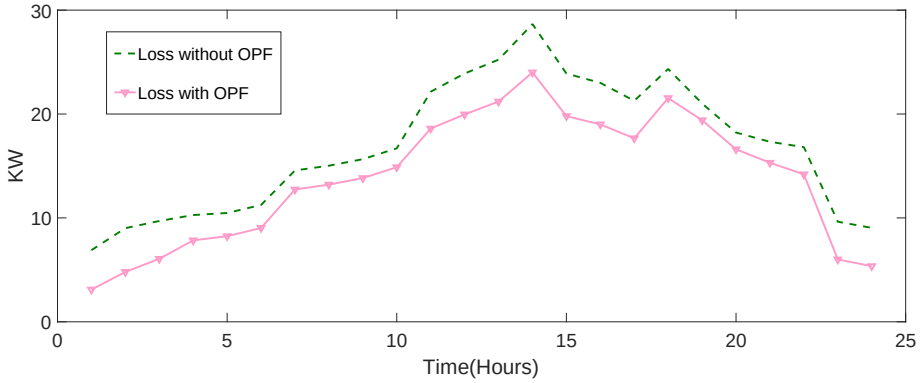



Figure 4.5: The Line Loss of the Distribution System.

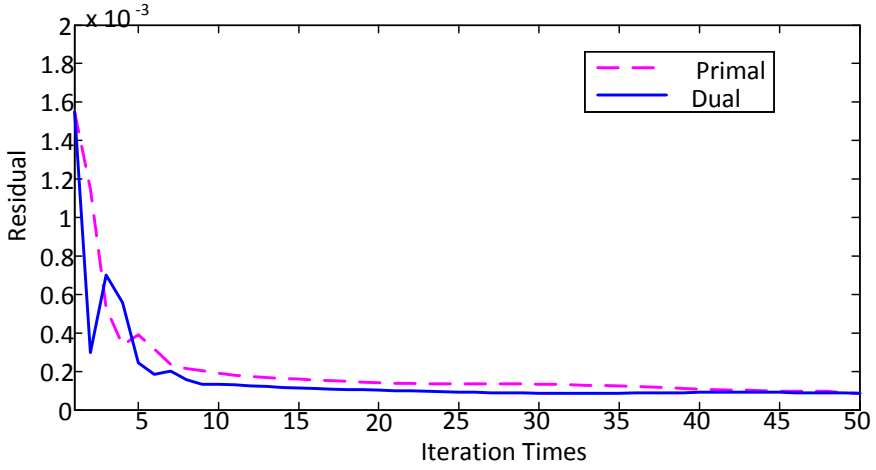



Figure 4.6: Convergence Analysis.

**Optimal Power Flow with ADMM**

The genetic algorithm (GA) is an artificial intelligence algorithm to simulate natural evolutionary processes. It retains a population of candidate solutions to search for the optimal one.

Some techniques are used to create candidate, which is inspired by crossover and selection. Genetic algorithm is usually implemented as a computer simulation method, which denotes that, for an optimization problem, the abstract representation of a number of candidate solutions use this algorithm to evolve toward better solutions. The evolution starts with completely random individual populations. In each generation, the fitness of the all creatures from the current generation is evaluated. Multiple individual(creatures) is randomly selected (based on their fitness) and generate new population through natural selection. The new generated population will repeat the same procedure to generate the next generation until satisfying the requirement of the system [110–112]. Considering the GA approach implemented in [113, 114], the population size is chosen as 600, which is enough to generate the new generation population with pinpoint accuracy. The probabilities of performing crossover and mutation are 0.8 and 0.08, respectively.

### 4.5.2 Numerical Results Analysis on Feeder J1 model

Feeder J1 is [91] selected for analysis because 1.7 MW of clients owned PV generators exists on the feeder. The system locates in the northeast of US, which serves 1300 commercial, small industrial and residential customers. There are 58 miles of primary lines. The topology of system is shown in Fig. 4.7. From the figure, the PV system is located at the south of the feeder. In Fig. 4.8, dark blue indicates the locations of customers, yellow dots indicate the existing PV on the feeder. PV plants are installed to support a small town with a population less than 1000. The surrounding area of the town like farms are served by Feeder J1. The PV panels(1.7MW in total) are installed as the yellow dots in Fig. 4.8, it is used to demonstrate that the proposed hybrid power system can work efficiently and reliable with the stochastic approach. The predicted data, actual data are provided from Electric Power Research Institute(EPRI), which is an American independent and nonprofit organization.

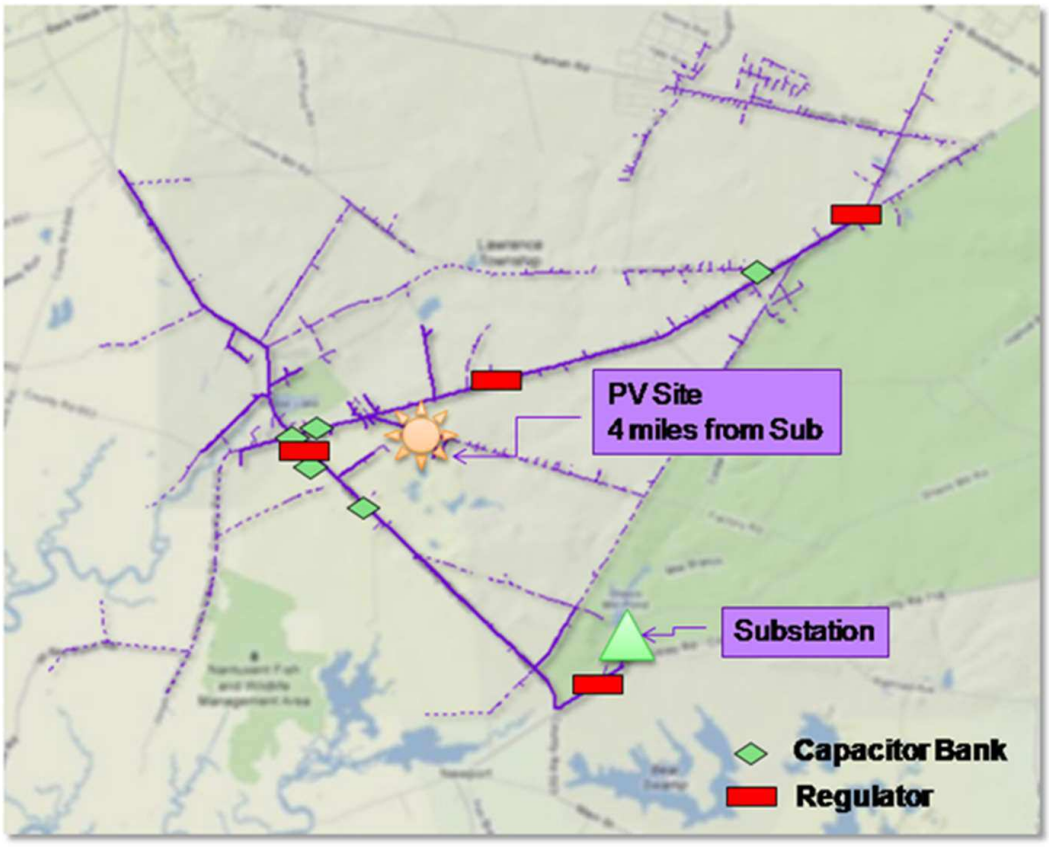



Figure 4.7: Topology of Feeder J1

**Error Distribution Modeling Comparison**

As defined in (4.3.11), the residual deviation $\eta$ is calculated to compare the difference between different simulation models. In table. 4.5, the GAEM model achieves a better ratio than other two ways, which indicates the simulation results of GAEM obtain a better fitted result than others. The reason is that genetic algorithm is used to find out the suitable amount of the components with Gaussian Mixture Model efficiently. In traditional ways, the amount of components are always selected by experience from human. It sometime lose the optimal number of the model, the automatic method (GA) increases the opportunity for obtaining a optimal number instead.

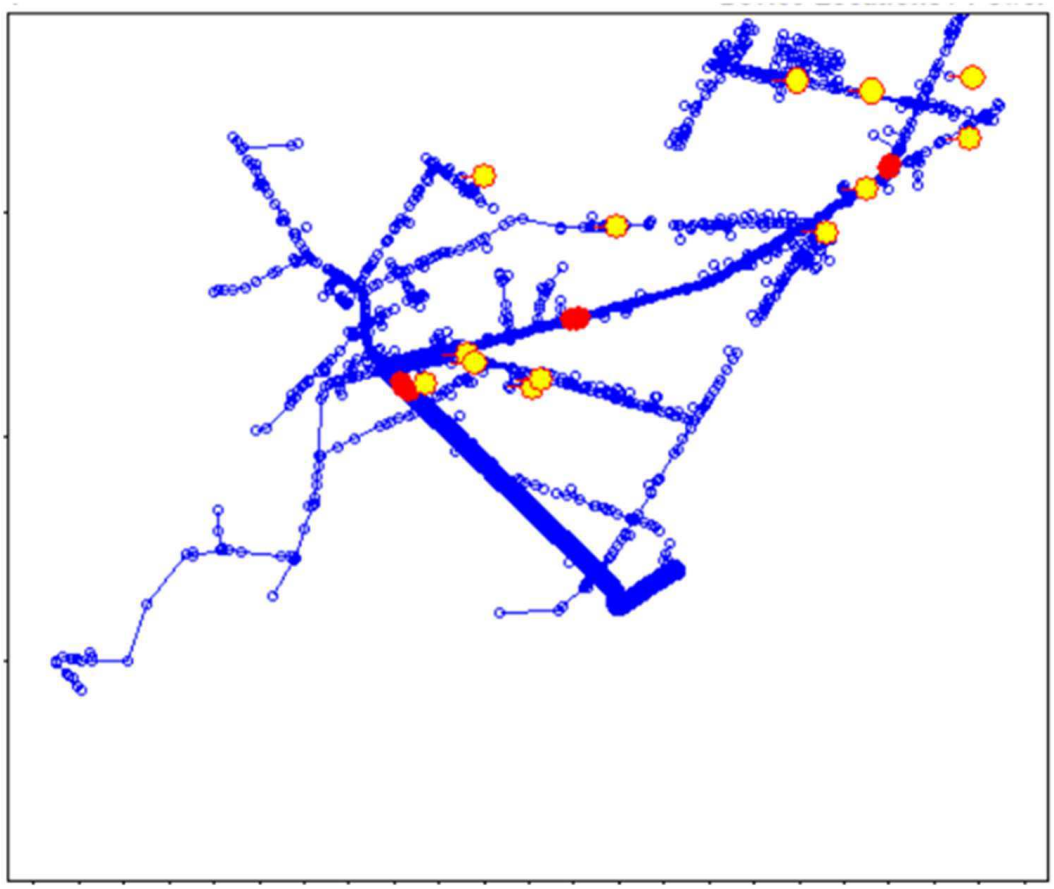


Figure 4.8: PV Installation of Feeder J1

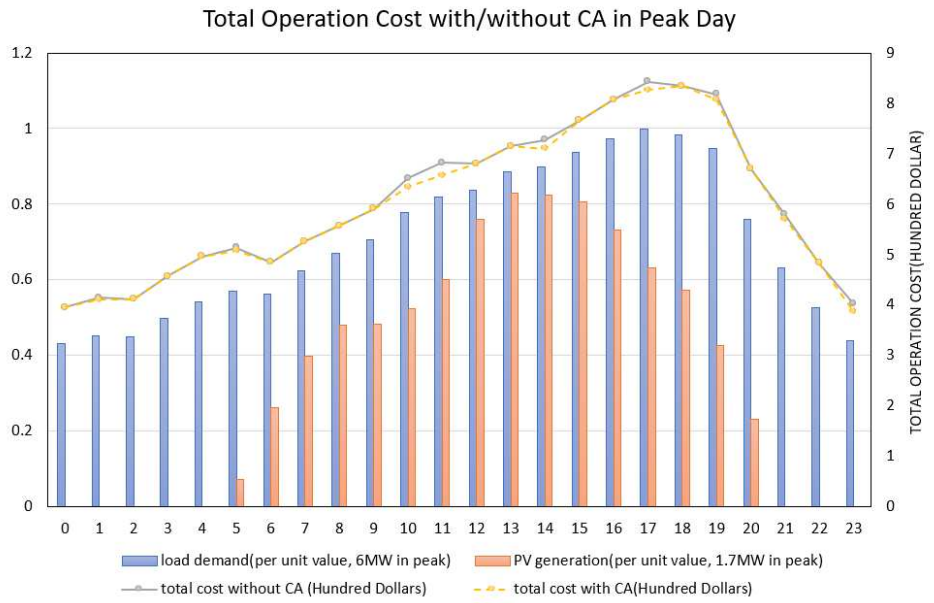
**Day-Ahead Dispatching with Feeder J1 Model** The numerical results for evaluate the proposed method are tested based on the Feeder J1 model. The total capacity of 1.7MW PV generators are installed and Fig. 4.8 shows the location of PV generators. The total operation cost with and without correct action are simulated in both peak day and off-peak day. In fig3, hourly renewable energy and load consumption are fitted and the total operation cost is compares between two different situations. As shown in Fig. 4.9(a), in the northeast of US, a peak load day is selected with 24 hours. The peak generation of PVs occur at 1pm and 2pm. Orange bar indicates the per-unit value of hourly renewable energy in 24 hours. Blue bars represents per-unit value of hourly demand load. It is clear that load consumption

Table 4.5: Comparison on $\eta$ for different fitting model

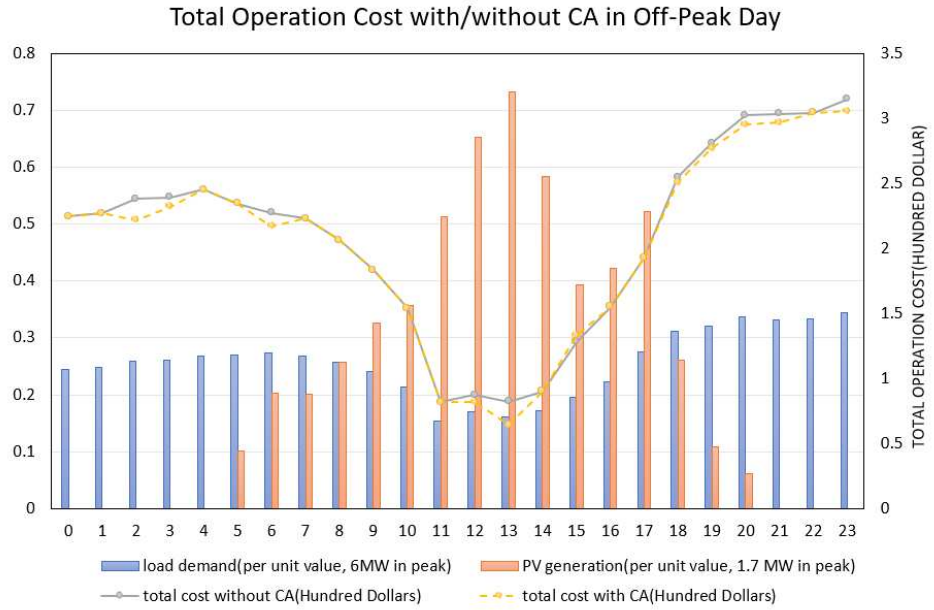
| $\eta$ (%) | Spring | Summer | Autumn | Winter |
|---|---|---|---|---|
| GAEMGMM | 0.08 | 0.11 | 0.17 | 0.11 |
| General GMM | 1.0 | 1.12 | 1.38 | 1.34 |
| GSM | 3.58 | 3.12 | 3.19 | 2.98 |

reaches the peak load (6MW) at 5pm in a day. Red line is used to describe how much the customers need to pay without correct action(CA) in (4.3.1). CA helps to sell the redundant power back to real time market at a lower price $\varrho_t^s$ hourly. It achieves saving extra money with the forecasting model in any accuracy. The yellow dash line represents the updated results the proposed method works, it is clear that total operation cost is reduced at 10am, 11am and etc.. No difference between red and yellow dots indicates no redundant energy is sold and correct action does not occur.

In Fig. 4.9(b), hourly renewable energy generation and load consumption are simulated within a off-peak load day of same location. The lowest load consumption occurs at 1pm, which is different from the records with only residents. It is because that light industrial customers and commercial customers are considered in this system. In off-peak day, they consume more energy at night. Different from the sunny day in peak load situation, the off-peak load day happens in a cloudy situation in the afternoon. It is obviously that the PV generation at 3pm and 4pm are not in steady descent along with the time because of the clouds. The load consumption continuously reduce from mid night to 1pm. After that, just a slight increasing of per-unit value of demand load from 0.18 to 0.34. The red line represents the total operation cost of off-peak day without correct action. The hourly operation cost decreases along with load consumption decreasing and renewable generation increasing. After 1pm, demand load starts increasing until the midnight. The results demonstrate that CA can still work well in a off-peak load day. In Fig. 4.9(b), more randomness happens in a cloudy day, which reduce the forecasting accuracy of predicting model. The big reduced cost at 1pm indicates the proposed methods helps to reduce the operation cost efficiently.




(a)




(b)

Figure 4.9: (a) Total Operation Cost with and without CA in Peak Day. (b) Total Operation Cost with and without CA in Off-Peak Day.

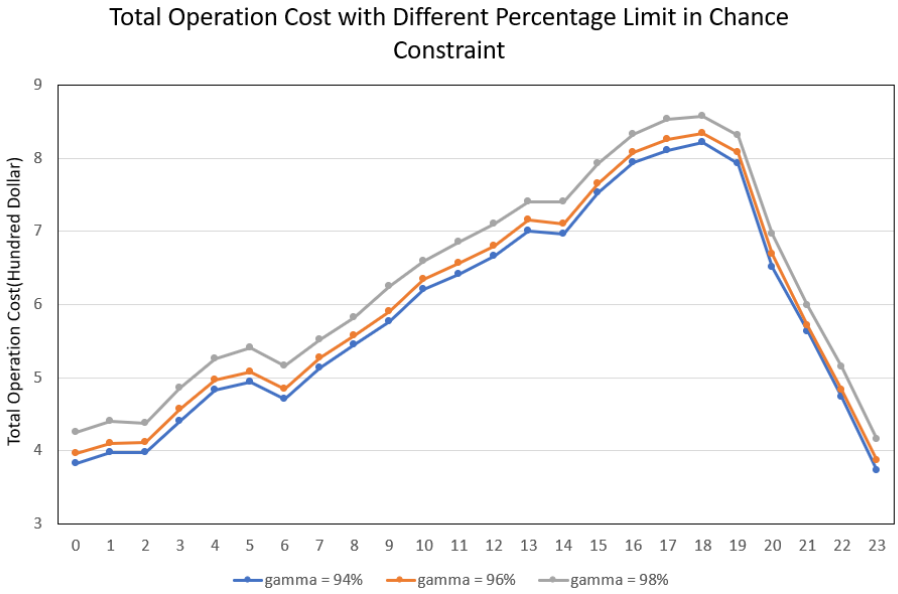



Figure 4.10: Total Operation Cost with Different Percentage Limit in Chance Constraint

**Total Operation Cost with Different Percentage Limit in Chance Constrains.** Three different $\gamma$ of $94\%$, $96\%$ and $98\%$ in (4.3.4h) are selected to analysis the results in Fig. 4.10. The operating cost with $\gamma$ of $94.0\%$ is lower than others. The highest cost is obtained with $\gamma$ = $98.0\%$. This indicates that higher reliability will lead in a higher operation cost. It is because that a bigger $\gamma$ requires a stricter constraint that $G_t^{DA} + G_t^W$ is higher or equal to the $G_t^{DL}$ in (4.3.4h). On the contrary, the operating cost decrease if we define a smaller $\gamma$ for the system. As shown in Fig. 4.10, chance constraint probability $\gamma$ in (4.3.4h) is set as 94.0%, 96.0%, 98.0% with blue, orange, and grey color, respectively. It is clearly that the operating cost with $\gamma$ = $94.0\%$ is the lowest among others, and the highest cost is obtained with $\gamma$ = $98.0\%$. These results demonstrates that a higher chance constraint probability $\gamma$ requires a high demand of $G_t^{DA}$ in equation (4.3.4h), which indicates the operation cost is also increasing. Because of that, a trade-off between the lower operation cost and a higher system reliability should be considered as a key problem in the future.

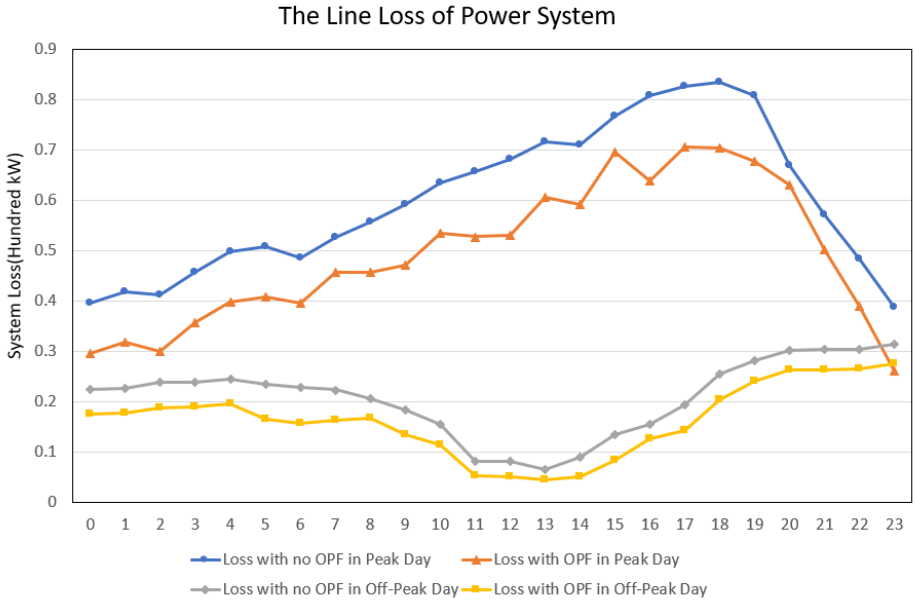



Figure 4.11: The Line Loss of Power System in Peak Day and Off-Peak Day

**Optimal Power Flow with System Loss** The system loss of feeder J1 model in peak load day and off-peak load day are shown in Fig. 4.11. The system loss without OPF in two situations are described in blue and grey color, respectively. Comparing to them, the system loss with OPF in peak day and off-peak day are represented in orange and yellow lines.

## 4.6 Conclusion and Contributions

A multi-timescale three-phase unbalanced approach is proposed. The day-ahead power dispatch of the substation level is used to minimize operation cost with renewable energy in each hour. The uncertainty of renewable energy is fitted with chance constraints. Gaussian Mixture Model is used to simulate the output of multiple renewable generation with a higher accuracy. At the feeder level, an optimal power flow problem is formulated for three-phase unbalanced system with considering the reality, non-convex problem with three variables are relaxed by semidefinite programming.

The **main contributions** of this section are:

1. The proposed method is used to optimize the system operation cost in two levels, corrective action in substation level is used to resell the redundant power back to real time market, which helps to save extra money. In feeder level, a part of renewable generation is used for optimal power flow on unbalanced distribution power system.

2. In substation level, Gaussian Mixture Model with Genetic Algorithm is used to simulate the output of renewable generation for each bus. Chance Constraints are used to fit the uncertainty of renewable energy, which define how much renewable generation need to be satisfied in a probability.

3. In feeder level, three phase unbalanced distribution system is simulated with ADMM. Semidefinite programming is used to relax the non convexity of the unbalanced model into convex.

4. Time consumption is one of the impotent indicators to evaluate the model. Compare with other methods, the proposed approach with ADMM efficiently decrease the calculation time.

# Chapter 5

# Improved distribution system optimization with image regression and ensemble learning

## 5.1 Solar Irradiance Capturing in Cloudy Sky Days –A Convolutional Neural Network Based Image Regression Approach

### 5.1.1 Introduction

With a low cost, the penetration of rooftop PV is increasing fast in recently years, which can be seemed as an indispensable component for modern power system integration [115–119]. Solar power, especially behind-meter solar power forecast is an important aspect for operating the grid with high level of PV penetration [21, 120]. The Global horizontal irradiance (GHI) is defined as the sum of the total received direct and diffuse shortwave solar irradiation on a horizontal plane. It is widely used to indicate the output power of PV [21, 120]. In traditional approaches, the local GHI can be collected with very expen-

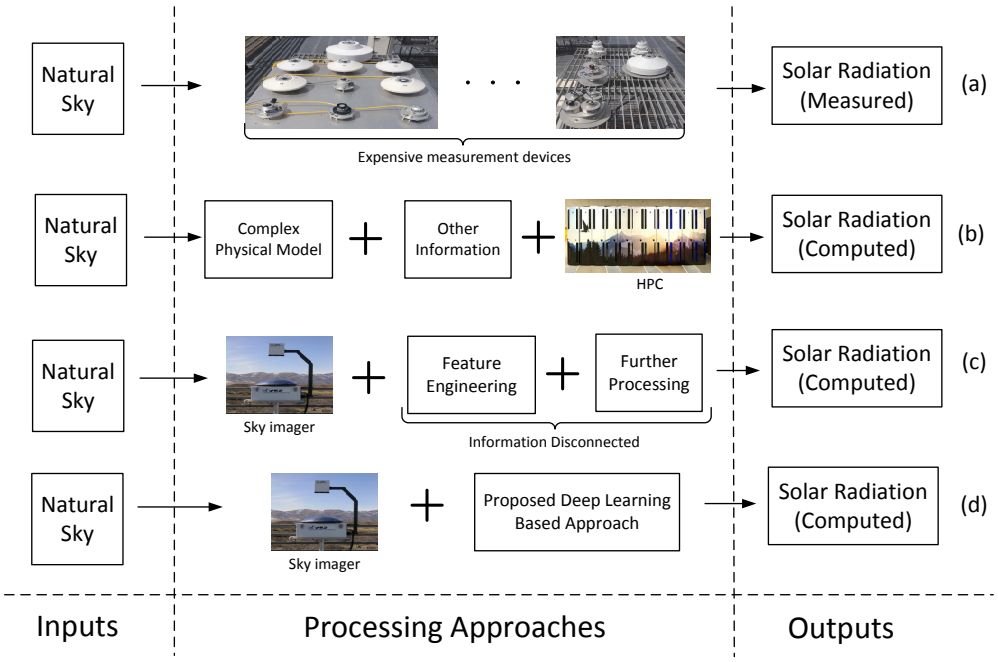



Figure 5.1: The proposed approach compared with the traditional approaches.

sive instruments, which are inconvenient to be moved and can hardly be used for large-area monitoring. For large-area monitoring, the radiative transfer models are widely used. However, these depend on the complex models and HPC. A deep learning-based approach, a CNN-based approach is proposed to capture the GHI conveniently and accurately, which can also be implemented into large-area monitoring.

1. In traditional approach (a), GHI is routinely measured by ground-based radiometers, e.g. pyranometers, use either thermoelectric or photoelectric detectors [121]. The reliability of this measurements is highly dependent on installation scheme, hardware maintenance, calibration technology and frequency [122]. The availability of measurements is often restricted by the high cost, which is always under consideration by operators.

2. In traditional approach (b), GHI can be numerically simulated by radiative transfer models, which account for the atmospheric absorption and scattering by air molecules, aerosols, and clouds [123]. However, accurate simulation of GHI replies on comprehensive measurements or retrievals of atmospheric constituents and land surface. Conventional radiative transfer models are often challenged by computation complexity by solving the radiative transfer equation and considering the interactions between the atmosphere and land surface.

3. In traditional approach (c), the processing approach includes two parts: feature engineering and further processing [19, 124–126]. Feature engineering is similar as feature extraction, which extracts the image features with different approaches such as filtering, Fourier transform, principle component analysis, wavelet analysis, and autoencoder related approaches [125, 127]. The extracted features are the inputs for next step, which is used for classification, pattern recognition, anomaly detection, and regression with different approaches such as hidden Markov model, support vector machine, neural network, logistical regression, random forest, and Bayesian network [125, 128–131]. The two parts framework in approach (c) is disconnected and has no communication between them, which results in a lower performance of [125, 132, 133].

4. In approach (d), it provides an End-to-End learning framework for the image regression task, which combines the feature extraction and regression part together to increase the efficiency and performance. As discussed above, the Pros and Cons of the 4 methods are summarized and presented in Table I.

In this section, three-month data (minutes level resolution data) are provided by the NREL. Considering the huge volume of data, a data cleaning procedure is used to eliminate the errors and inconsistencies for improving the input data quality, which contains data selection, calibration, missing data reconstruction, data standardization, and normal-

Table 5.1: The comparison of different approaches in Fig.1

| Approaches | Pros | Cons |
|---|---|---|
| (a) | The solar irradiation can be measured fast and accurately | Expensive device, inconvenient to move, calibration needed, weather fragile (broken by hails) and limited observation area |
| (b) | large-area solar irradiation can be computed | Complex physical model and atmosphere model, satellite image and HPC are needed |
| (c) | Can be extended to large-area monitoring | Low efficiency for parameter turning and low performance, large training data needed |
| (d) | Fast, accurate, cheap, and can be extended to large-area monitoring conveniently | large training data needed |


Figure 5.2: The flowchart of proposed approach.

ization. Gaussian mixture model is an useful method to detect the anomaly data, however, the number of the mixture components is difficult to determine. Based on the Dirichlet process, variational inference, and Bayesian theory, a Gaussian mixture model with Bayesian inference approach is employed to determine the number of the mixture components automatically [134–139]. The expectation propagation is very similar with the variational inference, which can be regarded as the same category [134, 136, 140]. Considering the real-application in power systems, the important factors are collected by devices to detect the anomaly data. In cloudy days, the profile of solar irradiation has a lot of stochastic deviations such as abrupt decreasing and increasing, which cause the deviations of output power of the PV. With the increasing penetration of PV, the large deviations of the PV out-

put power result in a series of problems such as voltage deviation, frequency oscillation, even unplanned islanding. Based on this [141–145], the proposed CNN based image regression model is focusing on building a relationship between the input cloudy sky images and the solar irradiations. Recently, the CNN based image processing, especially for image classification, is developing very fast. In 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC), the AlexNet (a type of CNN) [146] is proposed with 16.4 % error rate, which is considerably better than the previous shallow machine learning model (similar as approach (c) in Fig.5.1.1 with error 25.8 %). From then on, the CNN based image processing is attracting more attention in computer vision area. In 2014, the VGG with 19 layers and GoogleNet with 22 layer achieved 7.3 % and 6.7 % error rate, respectively, which further demonstrate the capability of CNN in computer vision processing [147, 148]. In 2015, a big improvement is provide by the 152 layers ResNet with 3.57 % error rate, which is better than human expert (5.1 % error rate) [149].

As shown in the Fig. 5.2, this method consists of three major components: data cleaning, CNN based image regression, and results validation, which are described as follows. In Section 5.1.2, data cleaning is introduced with 7 steps, and the Bayesian inference is used to detect the anomaly data in a nonparametric manner. In Section 5.1.3, the CNN based image regression approach is introduced for solar irradiation capturing, and the characteristic of the proposed CNN architecture is analyzed in detail. In Section 5.1.4, the numerical results are presented to validate the proposed approach.

### 5.1.2 Data cleaning

In machine learning, data cleaning is a significant component, which tremendously affect the performance of whole approach. The original data are directly collected from the devices in NREL's Solar Radiation Research Laboratory. As shown in Fig. 5.3, the proposed data cleaning approach contains 7 steps: data merge, data calibration, missing data recovery, data normalization, anomaly detection, data standard, and data verify and map.

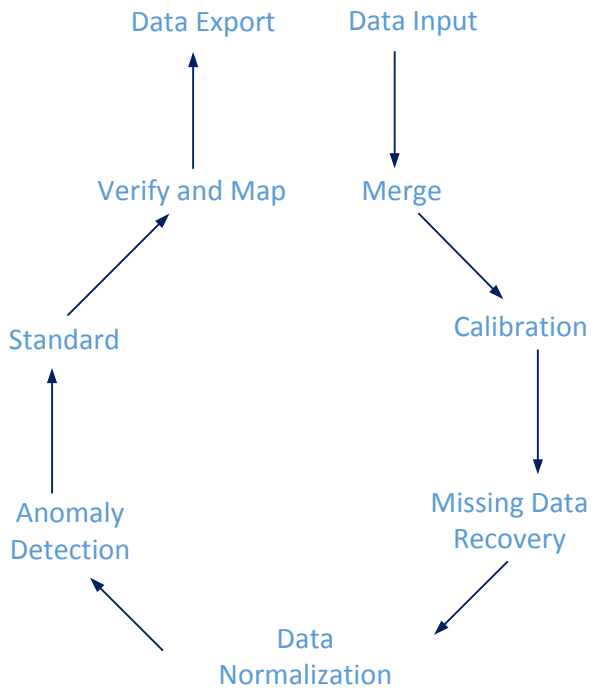



Figure 5.3: The flowchart of data cleaning in proposed approach.

**Data Merge, Calibration, and Missing Data Recovery**

The data merge, calibration, and missing data recovery are explained together as pre-data processing for data cleaning. The original data including the GHI data and sky image data are collected with sample rate 1 sample/min. Because the period of the sky image is 24 hours. We merge 24 hours data as a section for each day. Because the CNN is used to build a regression relationship between the sky images and GHI data. The original data need to be calibrated and recovered, and the problem are listed as follows:

1. The sky image camera only capturing the sky image from sunrise to sunset, the length is variant everyday (with some random offsets of the camera).

2. The GHI data are collected 24 hours, but the corresponding time index is Greenwich Mean Time (GMT).The way to get the data match is very important for training and calculation.

$$\beta_{miss}(t_2) = \frac{a_{j_2+1}(t_2 - t_{2,j_2})^3 + a_{j_2}(t_{2,j_2+1} - t_2)^3}{6h_{j_2}} \tag{5.1.1a}$$

$$+ \left(\frac{\beta_{j_2+1}}{h_{j_2}} - \frac{h_{j_2}}{6}a_{i_2+1}\right)(t_2 - t_{j_2}) + \left(\frac{\beta_{j_2}}{h_{j_2}} - \frac{h_{j_2}}{6}a_{i_2}\right)(t_{j_2+1} - t_2) \tag{5.1.1b}$$

$$h_{j_2} = t_{2,j_2+1} - t_{2,j_2},\ and\ a_0 = 0,\ and\ a_{n_2} = 0 \tag{5.1.1c}$$

$$h_{i_2-1}a_{j_2-1} + 2a_{j_2}(h_{i_2-1} + h_{i_2}) + a_{j_2+1}h_{j_2} = 6\left(\frac{\beta_{j_2+1} - \beta_{j_2}}{h_{j_2}} - \frac{\beta_{j_2} - \beta_{j_2-1}}{h_{j_2-1}}\right) \tag{5.1.1d}$$

3. There are random sky images and GHI data missing in the sky image series and GHI data everyday, which cause the mismatching between the sky images and GHI data.It will result incorrect results if we did not care about the mismatch problems. How to find out the mismatch image and solar data is very important. There will have incorrect results if the mismatching between image and solar are ignored.

The times of sunrise $t_{i_2}^{sr}$ and sunset $t_{i_2}^{ss}$ for day $i_2$ are collected from National Oceanic and Atmosphere Administration (NOAA). The accurate length between $t_{i_2}^{sr}$ and $t_{i_2}^{ss}$ can be calculated as: $\Delta t_{i_2}$ = $t_{i_2}^{ss}$ - $t_{i_2}^{sr}$, which can be used to eliminate the random offsets of the camera, and delete the useless sky image captured in the night. Then, the corresponding GHI data can be sectioned with $\Delta t_{i_2}$, and the time index is transfered from GMT to Mountain Time (MT).

In the collected original data, there are missing random sky images and GHI at some points. For the missing sky images, the corresponding GHI data are deleted to keep the one to one mapping relationship. Because the sky images is continuously changing and the sample rate 1 sample/min is relatively high, the missing GHI data can be recovered by the spline interpolation as given equation 5.1.1. $\beta_{miss}(t_2)$ is the missing GHI data, $t_{2,j_2}$ indicates the time index $j_2$ to discriminate with the missing GHI time $t_2$, $h_{j_2}$ and $a_{j_2}$ are coefficients. Then, the missing GHI data at time $t_2$ can be recovered.

**Anomaly Detection**

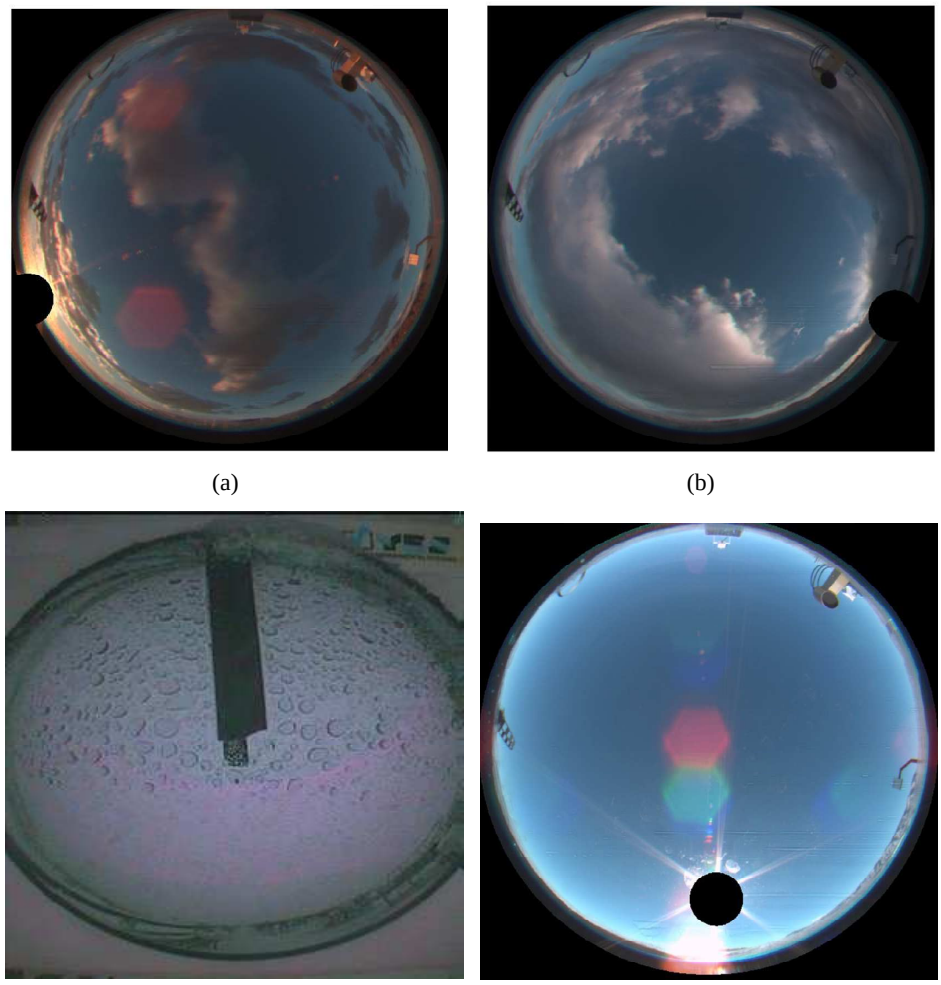
**Problem Description** We collected a large volume of original data from the sky image camera and GHI sensors, which contains a lot of anomaly data sets. These anomaly data sets can pollute the input data, and generate irrelevant information during the learning process [19, 150–153]. As in Fig.5.4, we are focusing on the solar irradiance capturing for the cloudy sky days (causing large deviations of PV output power), this means that the sky images of the sunrise, sunset, and clean days are seemed as the anomaly data and need to be cleaned. In oder to detect the anomaly conveniently, the data are normalized to range (0, 1) with the approach in [19], and the corresponding sky images are stored and tagged in a database. A nonparameter and fast anomaly detection method is employed to detect the anomaly data as shown in Fig. 5.4.

**Problem formulation** As shown in Fig. 5.5, the anomaly detection method is based on Gaussian mixture model with Bayesian inference [134, 136, 138]. The full joint distribution can be formulated with it.

$$p(X, Z, \pi, \mu, \Lambda) = p(X|Z, \mu, \Lambda)p(Z|\pi)p(\pi)p(\mu|\Lambda)p(\Lambda) \tag{5.1.2}$$

where

1. $X$ is the observation set, $X = \{x_1, \cdots, x_N\}$. $X$ is collected GHI.
2. $Z$ is the component index set $Z = \{z_1, \cdots, z_N\}$, total category is $K$.
3. $\pi$ is the mixing weights, in Dirichlet process $\pi = \{\pi_1, \pi_2, \cdots, \pi_\infty\}$.
4. $\mu$ is the mean of the normal distributions. It is with a normal distribution prior as shown in Fig. 5.5 with a mean $\mu_0$ and precision matrix $\lambda_0$.
5. $\Lambda$ is the precision matrix. It is a Wishart prior (it is equivalent to a one-dimensional Wishart distribution [154]) with a scale matrix $\Lambda_0$ and a degree of freedom $\nu_0$.



(a) (b)

(c) (d)

Figure 5.4: The selected anomaly sky images taken by different devices (a) the sky image of sunrise, (b) the sky image of sunset, (c) the sky image polluted by rain and snow, (d) the sky image of a clean day without any clouds.

In the Gaussian mixture model (GMM), the number of Gaussian components gives the prior distribution estimation of the whole distribution. If the amount of components is given, the whole distribution can be solved in a convenient manner. Therefore, the symbol $Z$ here is used as a latent variable to indicate the number of the Gaussian components, which is formulated as a categorical distribution and its conjugate distribution is the Dirichlet distribution.

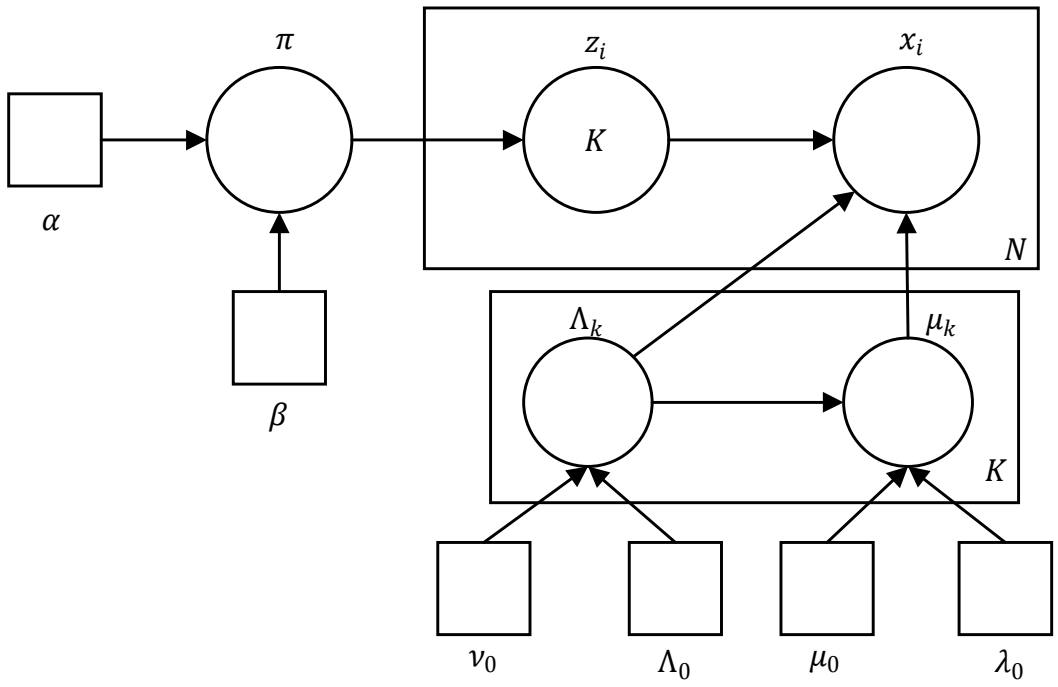



Figure 5.5: The graphic model of the Gaussian mixture model with Bayesian inference.

Based on (5.1.2), the relationship and probabilities can be formulated as follows. The conditional distribution of $Z$ given a mixing weight.

$$p(Z|\pi) = \prod_{n_3=1}^{N_3} \prod_{k_3=1}^{K_3} \pi_k^{z_{nk}} \tag{5.1.3}$$

Then, given the latent variables $Z, \mu, \Lambda$, the conditional distribution of the observed data can be formulated as:

$$p(X|Z, \mu, \Lambda) = \prod_{n_3=1}^{N_3} \prod_{k_3=1}^{K_3} \mathcal{N}(x_{n_3}|\mu_{k_3}, \Lambda_{k_3}^{-1})^{Z_{n_3,k_3}} \tag{5.1.4}$$

Here, the Bayesian inference is employed to estimate how many clusters are required for the observation data. It means that the hyperparameters can be generated with noninformative hyperprior distributions. Firstly, the nonparametic prior titled as Dirichlet distribution

is employed to build the finite Gaussians. Then, the Dirichlet process is used to generalize it into infinite Gaussians. It helps to decide the amount of components automatically.

$$p(\pi) = Dir(\pi|\alpha_0) = C(\alpha_0) \prod_{k_3=1}^{K_3} \pi_{k_3}^{\alpha_0 - 1} \tag{5.1.5}$$

$Dir$ is the Dirichlet distribution. $\alpha_0$ is the concentration parameter. And $C(\alpha_0)$ is a normalizing item which can be expressed with Gamma Function and also be named as multivariate Beta function [154, 155]. Then, according to (5.1.2), the Gaussian-Wishart prior can be introduced for the mean $\mu$ and precision matrix $\Lambda$. The equation below shows how it works and how it can get the results. $\mu_0$ is a mean, $\lambda_0$ is a precision matrix, $\Lambda_0$ is a scale matrix, and $\nu_0$ is degree of freedom.

$$p(\mu, \Lambda) = p(\mu|\Lambda)p(\Lambda) \tag{5.1.6a}$$

$$= \prod_{k_3=1}^{K_3} \mathcal{N}(\mu_{k_3}|\mu_0, (\lambda_0\Lambda_{k_3})^{-1})\mathcal{W}(\Lambda_{k_3}|\Lambda_0, \nu_0) \tag{5.1.6b}$$

As shown in (5.1.5), the Dirichlet distribution can be seen as the conjugate prior for the categorical distribution. To generalize it into infinite and nonparametric distribution, the Dirichlet process can be seen as its conjugate prior. For a sample space $\Theta$, $G_0$ is a distribution over $\Theta$ with a positive factor $\beta$, and the Dirichlet process can be generated with $G_0$ over $\Theta$.

$$(G(A_1), \cdots, G(A_{K_3})) \sim Dir(\beta G_0(A_1), \cdots, \beta G_0(A_{K_3})) \tag{5.1.7a}$$

$$G \sim DP(\beta, G_0) \tag{5.1.7b}$$

where $A_{K_3}$ is a finite measurable partition over $\Theta$, and the positive factor $\beta$ controls the density of $G$ [137, 156]. Here, the stick-breaking construction is used for the weights $\pi_{k_3}$. $\pi = \{\pi_1, \cdots, \pi_\infty\}$ indicates the original model to be generalized into infinite Gaussians

with Dirichlet process. It is finally realized automatically select the components number. The results demonstrates that automatic selection always give a better performance.

$$\pi'_{k_3} \sim Beta(1, \beta) \tag{5.1.8a}$$

$$\pi_{k_3} = \pi'_{k_3} \prod_{i_3=1}^{k_3-1} (1 - \pi'_{i_3}) \tag{5.1.8b}$$

**Variational Inference** The Bayesian theorem based variational inference algorithm is employed for the Dirichlet process based Gaussian mixture models. The posterior distribution of $Z$ can be computed as the equation below.

$$\log p(Z|X, \vartheta) = \log p(X, Z|\vartheta) - \log p(X|\vartheta) \tag{5.1.9}$$

where $\vartheta$ is the parameters, and this Bayesian theorem based variational inference provides a bridge to the likelihood function and prior function, which can also be seemed as a regularization item for the likelihood function.

Here, we introduce $q_\iota(W)$, $q_\iota(W)$ to be defined as a distribution family, for example, the exponential distribution family, and $\iota$ is the parameter. Then, according to Jensen's equation:

$$\log p(X|\vartheta) \geq E_q[\log p(Z, X|\vartheta)] - E_q[\log_{q_\iota}(Z)] \tag{5.1.10}$$

where (5.1.10) is the evidence of low bound (ELOB), and the gap is the Kullback-Leibler (KL) divergence between $q_\iota(Z)$ and $p(Z|X, \vartheta)$, which can be derived as

$$\begin{aligned} &D(q_\iota(Z)||p(Z|X, \vartheta)) \\ &= E_q[\log_{q_\iota}(Z)] - E_q[\log p(Z, X|\vartheta)] + \log p(X|\vartheta) \end{aligned} \tag{5.1.11}$$

In optimization, we can maximize the ELOB, which is an alternatively option for minimizing (5.1.11) [157, 158]. And the detailed information can be found in [134, 136, 137, 156].

Specifically, the temperature is an indispensable factor to impact the efficiency of the PV, PV battery, and related systems. We also collect corresponding temperature information to eliminate the anomaly data as shown in Fig. 5.4. Similarly, the Zenith angle is a critical factor for the GHI [159], which is also collected to eliminate the anomaly data.

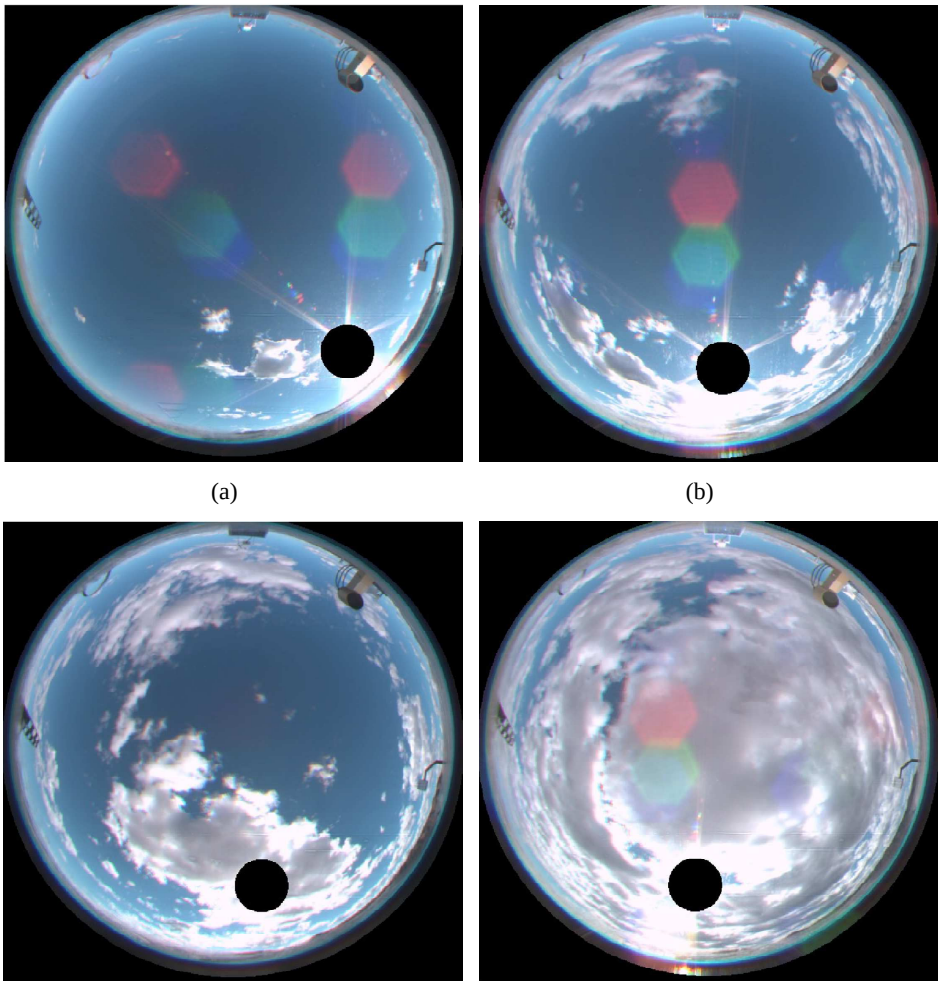
After this process, the anomaly data shown in Fig. 5.4 can be detected and only the sky images with clouds can be selected for next step. Then, all the selected sky images are merged together with the temporal and GHI information for CNN regression.

### 5.1.3 Image Regression Problem with Convolutional Neural Network

#### Problem Formulation

After the data cleaning as discussed above, the collected sky images and GHI values are formed as one-to-one correspondence. Then, an input space $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_{N_1}\}$ consists of the collected sky images. Because a sky image is a three dimensional matrix including red, green, and blue color, we use $\mathbf{x}_i$ to denote the $i$-th sky image. The corresponding output space $\mathcal{Y} = \{y_1, \cdots, y_{N_1}\}$ consists of the collected GHI values, where $y_i$ is the $i$-th GHI value. Given the training samples $D = \{\mathbf{x}_i, y_i\}_{i=1}^{N_1}$, the proposed CNN based image regression approach aims to find a mapping from images to GHI values $h(\cdot)$: $\mathcal{X} \mapsto \mathcal{Y}$ with a predefined cost function $C$: $\mathcal{X} \times \mathcal{Y} \mapsto R$.

From the traditional optimization perspective, in the training part, the cost function $C(h)$ needs to be minimized with several different approaches such as stochastic gradient descent (SGD), momentum, and Nesterov momentum, and in this chapter, ADAM is used [133, 160]. However, different from the traditional optimization, in machine learning, the distribution of the training space is usually different from the distribution in the testing space, which requires a good generalization characteristic for the selected $h(\cdot)$. If using the traditional optimization approaches without any revision during the training, several problems such as ill-condition, local minimum, cliffs, and etc. can dramatically degrade the testing performance.



(a) (b)

(c) (d)

Figure 5.6: The selected normal sky images with some clouds (a) few cloud with okta 1 to 2 in a sky image, (b) cloud with okta 3 to 4 in a sky image, (c) cloud with okta 5 to 6 in a sky image, (d) cloud with okta 7 to 8 in a sky image.

**The architecture of the proposed CNN**

After data cleaning, the sky images of normal cloudy days are selected as shown in Fig. 5.6. okta is a unit of measurement to describe the cloud cover in meteorology [161]. Compared with the images from ImageNet (a large-scale image database) [162], as shown in Fig. 5.6, the patterns of the cloud in the sky images are simpler.

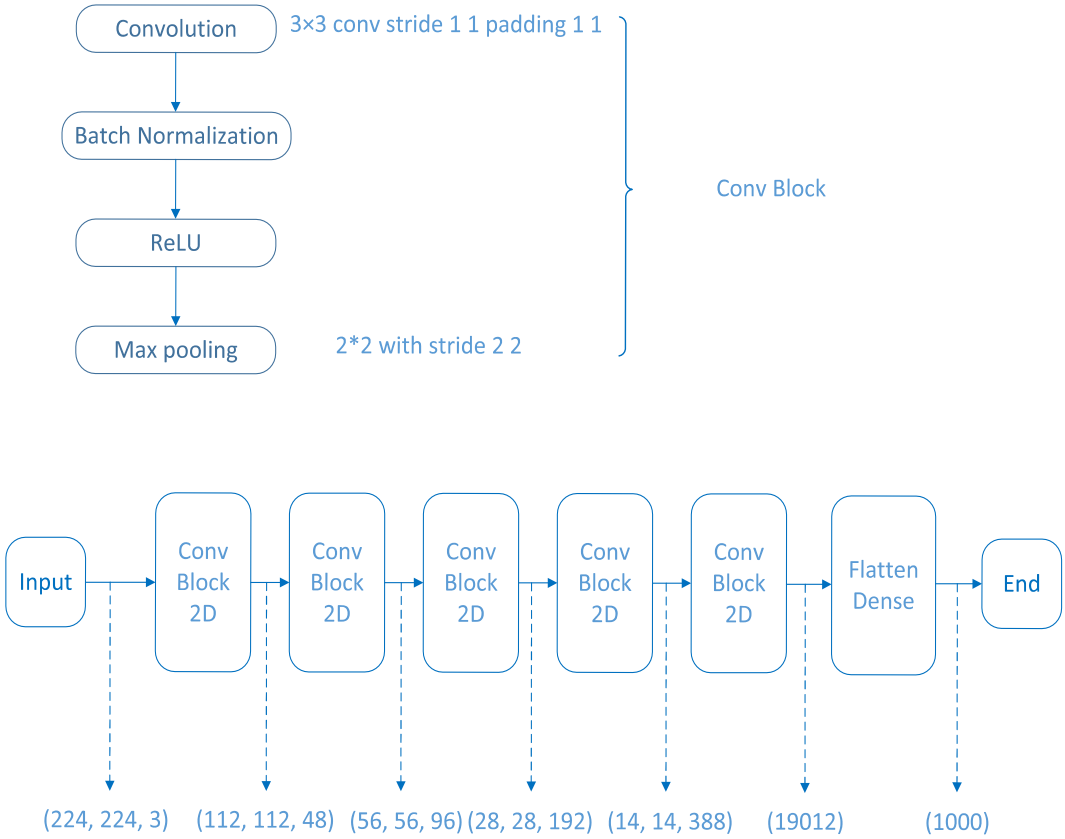



Figure 5.7: The architecture of proposed CNN.

As shown in Fig. 5.7, the architecture of the proposed CNN is designed with 5 convolution layers, batch normalization, rectified linear unit (ReLU) (activation function), and maxpooling, which is based on the VGG 16 architecture [147, 148]. For each convolutional block, the detailed design is shown in Fig. 5.7 and described in the figure.

1. The first step is the convolutional layer. Compared with $11 * 11$ or other $7 * 7$ perception fields of Alexnet, filters of $3 * 3$ in size are implemented to increase the nonlinear characteristics and reduce the computation load [147, 148]. At the same time, the stride step is set as 1 and padding is also set as 1, which keeps the size of the sky image for next block processing [133, 147, 148].

2. Next is batch normalization, which aims to improve interal covariance shift [163]. With the increasing layers of CNN, the distribution of inputs changes gradually. With batch normalization technique, the distribution shift can be reduced, training speed can be increased, and the data in same scales and more easy to get convergence [163].The inference part of the batch normalization can be found in [133, 163].

3. The third step is active function, it is used to bring nonlinear characteristics for the proposed method. Compared with other activation functions such as sigmoid, hyberbolic tangenet and Gaussian, Rectifier(ReLU) is more convenient to compute the derivative, fewer vanishing gradient, and fewer saturation parts [133]. Specifically, in this proposed network, the target is solar irraidance, which means all the output should be positive real numbers. ReLU can meet this requirement perfectly [133].

4. The parameter set for convolutional block, for example $(112, 112, 48)$, indicates that the output dimension of the convolutional block is $112 * 112$, and $48$ indicates that there are $48$ feature maps (or feature images). This means that the features can be selected in different feature maps, which provide a more convenient way for the feature extraction.

5. The last part is max pooling. It is design as $2 * 2$ with stride $2 * 2$, which select the maximum value over the $2 * 2$ part. This part can be regarded as a down-sampling job. It selects the feature of $2*2$ and non-overlapping area, reduce the dimensionality of input, and pass the selected features to next convolution block.

After 5 convolutional blocks, the dimension of the input image is changed from $224 * 224*3 = 150,528$ to $7*7*388 = 19,012$, which means that only $12.63\%$ data are needed to represent the features of the original image. Then, the output is flatten into a vector and treated as an input for the fully connected layer with $1000$ neurons for the final regression.

**Learning of the proposed CNN**

**Last Layer** In classification problem, the last layer are softmax activation function and cross entropy as objective function. Considering the regression objective, the mean-square-error is employed as the objective function, and linear function is used as activation function.

$$E = \frac{1}{2}(g_{final}(\chi_1^m) - y)^2 + reg_{l2} \tag{5.1.12}$$

where $g_{final}$ is the activation function, which is ReLU as shown in Fig. 5.7. $\chi_1^m$ is the only one output in last layer $m$, $y$ is a general form of the GHI value, and $reg_{l2}$ is a weighted L2 regularization item. The elastic (combined L1 and L2) can also be implemented here, and $E$ is the loss. Then, the error can be derived as:

$$\delta_1^m \equiv \frac{\partial E}{\partial \chi_1^m} = (g_{final}(\chi_1^m) - y) g'_{final}(\chi_1^m) \tag{5.1.13}$$

where the derivative of ReLU is

$$g'_{final}(\chi_1^m) = \begin{cases} 1, & if\ \chi_1^m > 0 \\ 0, & otherwise \end{cases} \tag{5.1.14}$$

**Hidden Layer** In the hidden layer, the error item can be derived as following:

$$\delta_j^k = g'(\chi_j^k) \sum_{l=1} w_{jl}^{k+1} \delta_l^{k+1} \tag{5.1.15}$$

where $k+1$ indicates the layer, $\delta^k$ is the error in last layer $k$, $w_{jl}^{k+1}$ is a weight for node $j$ for incoming node $l$ in layer $k+1$. And the weight can be derived as:

$$\frac{\partial E}{\partial w_{ij}^k} = o_i^{k-1} g'(\chi_j^k) \sum_{l=1}^{r^{k+1}} w_{jl}^{k+1} \delta_l^{k+1} \tag{5.1.16}$$

$o_i^{k-1}$ is the output for node $i$ in layer $k$. $r$ is the total number for node $l$.

**Extended to 2-dimensional CNN** The proposed approach is based on CNN and aims to build a mapping between sky images and GHI values. It requires an extension from 1 dimensional backward propagation to 2 dimensional backward propagation. The weight can be computed as

$$\frac{\partial E}{\partial w^k_{m',n'}} = \sum_{i_1=0}^{H_1-k_1} \sum_{j_1=0}^{W_1-k_2} \delta^k_{i_1,j_1} o^{k-1}_{i_1+m',j_1+n'} = \widehat{\delta^k_{i_1,j_1}} * o^{k-1}_{m',n'} \tag{5.1.17}$$

where $*$ indicates a 2-dimensional ($H_1$ and $W_1$ indicates the hight and width of the sky image) convolution in CNN [133, 160]. $k_1 \times k_2$ is the dimension of the filter, $\widehat{\delta^k_{i,j}}$ indicates the flipped kernel [164]. $\delta$ and $o$ are the same as defined above. Then, similar with equation (5.1.13) the error can be computed as:

$$\frac{\partial E}{\partial \chi^k_{i',j'}} = \sum_{m=0}^{k_1-1} \sum_{n=0}^{k_2-1} \delta^{k+1}_{i'-m,j'-n} w^{k+1}_{m,n} f'\left(\chi^k_{i',j'}\right) \tag{5.1.18}$$

where $\chi$ is defined as:

$$\chi^k_{i_1,j_1} = \sum_m \sum_n w^l_{m,n} o^{k-1}_{i_1+m,j_1+n} + b^k_{i_1,j_1} \tag{5.1.19}$$

where $b^k_{i_1,j_1}$ is the bias for node with position $i_1, j_1$ in layer $k$. In sum, the 2-dimensional CNN is formulated for the proposed sky image processing.

### 5.1.4 Results

The image data and GHI data are provided by NREL, which includes about thee-month data in the Autumn of 2016 and Summer of 2017. The sampling rate of the sky imager is 1 sample/min from sunrise to sunset per day. The sampling rate of the GHI measurement device is also 1 sample/min for 24 hours per day.

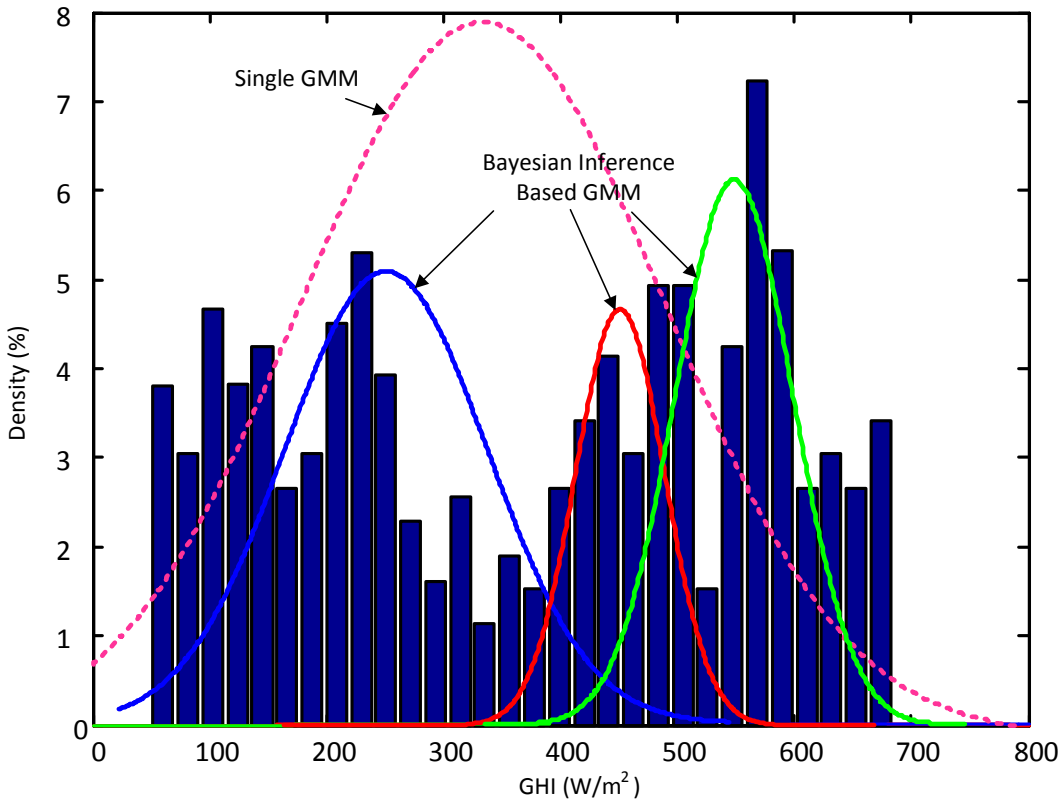



Figure 5.8: Anomaly detection for the GHI data.

**Anomaly Detection and Elimination**

First, the original GHI data are modeled with the Gaussian mixture model with Bayesian inference. Compared with the Gaussian mixture model without Bayesian inference (indicated as Single GMM with magenta), the GHI data can be more precisely modeled with 3 Gaussian models instead of 1 Gaussian model. According to our requirements, the GHI values located in the areas close to 0 and 800 can be considered as sunrise, sunset, and clean sky data. Second, as mentioned above, the Zenith angle and temperature are considered as useful factors, the Gaussian mixture model with Bayesian inference are employed to detect the anomaly data with them. To show the results clearly, the result of Zenith angle and temperate are shown separately as in Fig. 5.9 and Fig. 5.10. In the upper part of Fig. 5.9, the collected data can be modeled with a single Gaussian models, which means that a lot

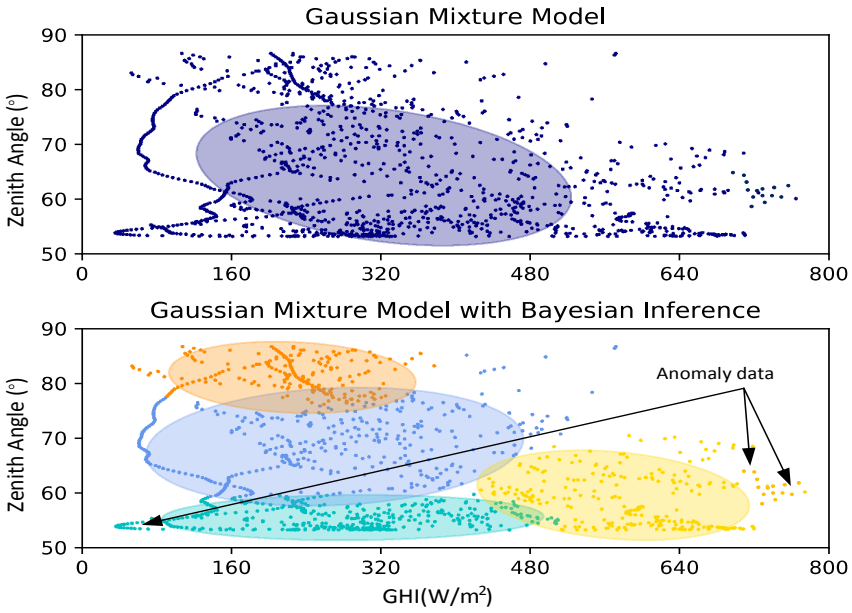



Figure 5.9: Anomaly detection for the GHI and Zenith Angle data.

of useful data are classified into anomaly data side. In the lower part, the collected data can be modeled with 4 Gaussian models, the anomaly data are indicated corresponding to the sky images of sunrise and sunset (with the green dots close to 0 GHI and $50^{\circ}$ Zenith angle). The yellow dots located close to 800 GHI are corresponding to the sky images of clean sky without clouds. Similarly, in the upper part of Fig. 5.10, the collected data can be modeled with 2 single Gaussian model, which means that a lot of useful data are classified into anomaly data side. In the lower part, the collected data can be modeled with 5 Gaussian models, the anomaly data are indicated corresponding to the sky images of sunrise and sunset (with the yellow dots close to 0 GHI and $10^{\circ}C$ temperature). The light blue dots located close to 800 GHI are corresponding to the sky images of clean sky without clouds. The results shows that GMM model works efficient on that.

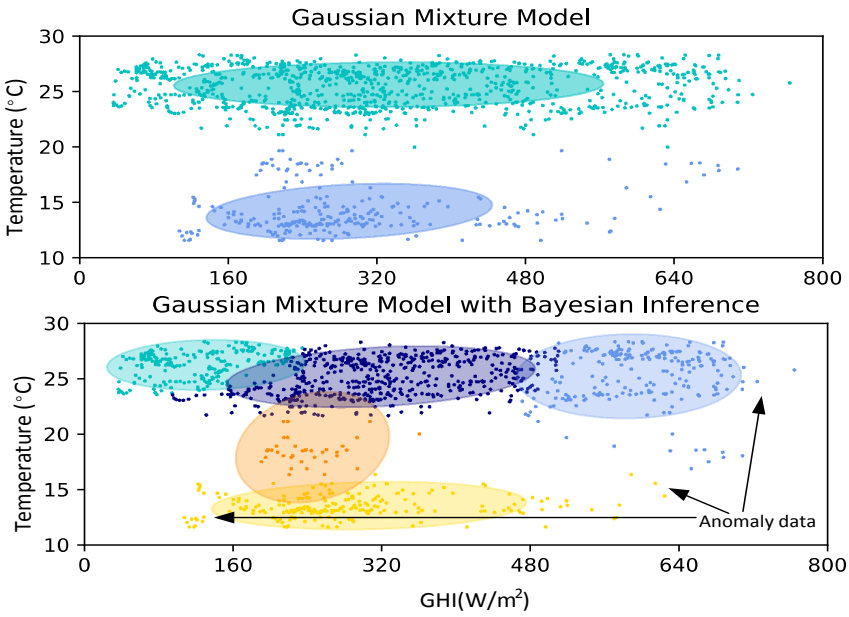



Figure 5.10: Anomaly detection for the GHI and Temperature data.

**CNN Based Regression**

Considering the limitation of the computer memory (64 GB), the data (including the sky images and corresponding GHI values) of 1, 2, 3, 4, and 5 days are used as training data, separately, and 70 randomly selected sky images and corresponding GHI values are used as testing data. As discussed above, the proposed CNN based image regression approach aims to build a mapping relationship from images to GHI values $h(\cdot)$: $\mathcal{X} \mapsto \mathcal{Y}$.

**Feature Analysis** To illustrate clearly the operation manner of CNN, its feature analysis or content reconstruction is shown in Fig. 5.11. In Fig. 11(a), the original figure is shown, which is extracted from the beginning side of the flowchart in Fig. 5.7. Fig. 11(b) is almost identical with Fig. 11(a), and the small and tiny crinkles indicates that the feature and edges of the clouds are extracted by the CNN. In Fig. 11(c), it is clearly that the edges of the clouds are selected by the crinkles, which means the features of the major parts are identified and located in the image. Furthermore, in Fig. 11(d) and Fig. 11(e), the features of the small

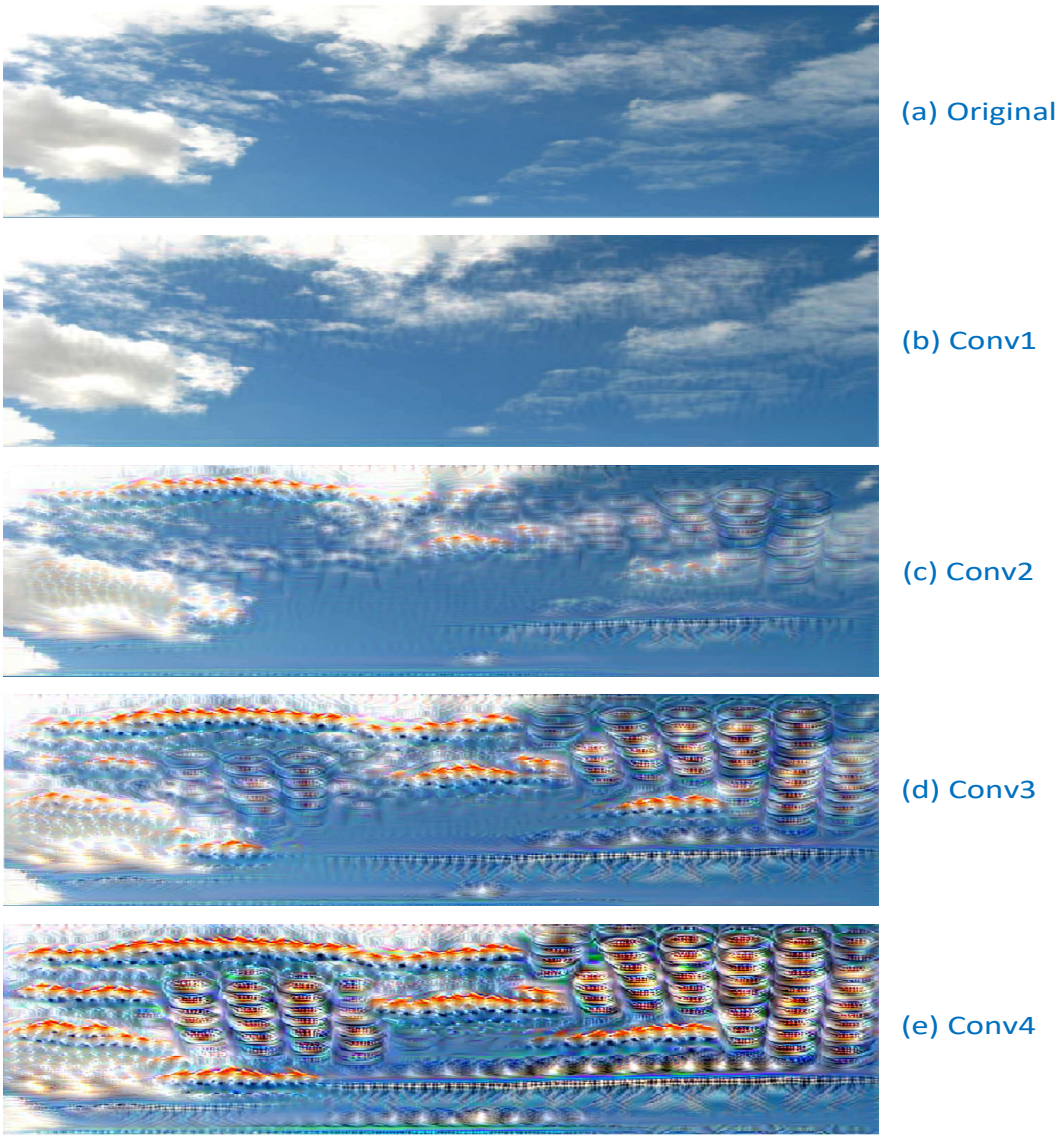



Figure 5.11: Feature expression of CNN (a) the original sky image demo, (b) the feature of conv1, (c) the feature of conv2, (d) the feature of conv3, (e) the feature of conv4.

parts are also identified and located. In brief, with deeper layers in the CNN, more and more features are extracted in different layers gradually. With the extracted features, the regression analysis is introduced in next step.

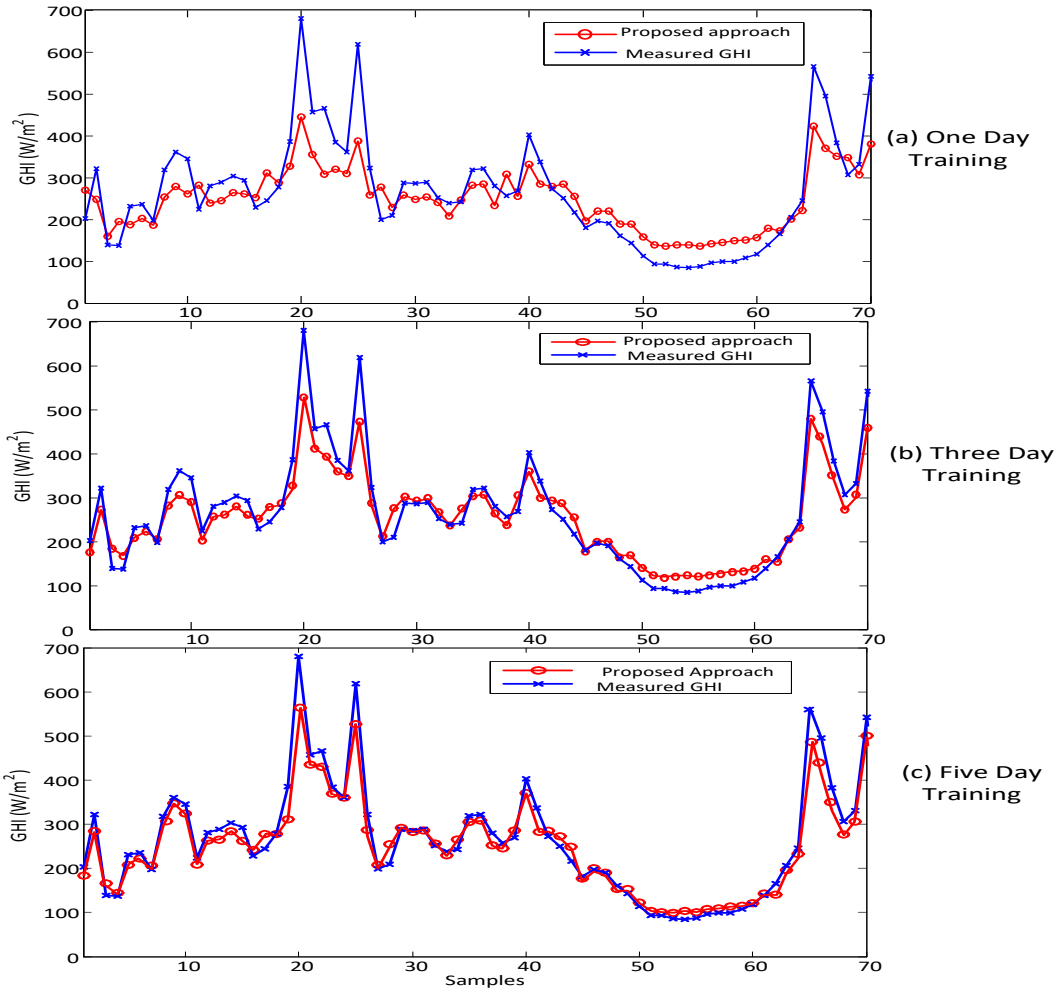



Figure 5.12: Image regression for GHI with different training data length: (a) 1 day data as training data, (b) 3 day data as training data, and (c) 5 day data as training data

**Regression Analysis** As shown in Fig.5.12, the top image regression results of training data with 1, 3, and 5 days are presented in red curves, and the testing data are presented with blue curves. The results of 1 day training data and testing data are presented in Fig. 5.12 (a), which contains relatively big errors. Specifically, as shown with the blue curve, there are 4 peaks located in samples 20, 25, 65, and 70, which indicate that high solar irradiance at these time slots. However, with 1 day training data, the results of the proposed approach are much lower, which indicates that a big mismatch between the proposed approach and

measured data. In addition, in the samples between 45 to 65, there is a big valley in the testing GHI values. However, as shown in Fig. 5.12 (a), the proposed approach also contains a big error from the measured GHI values. In Fig. 5.12 (a), the mean absolute percentage error (MAPE) is 21.8%, which indicates that the 1 day training data is not sufficient to achieve a acceptable result. Compared with Fig. 5.12 (a), the results presented in Fig. 5.12 (b) and Fig. 5.12 (c) with 3 and 5 days training data contain smaller errors. In Fig. 5.12 (b), it is obvious that the peak errors in samples 20, 25, 65, and 70 are much smaller than Fig. 5.12 (a), and the valley errors from samples 45 to 65 are also smaller than Fig. 5.12 (a). In Fig. 5.12 (c), the best regression results are presented with 5 days training data. The peak errors and valley errors are very small, and the blue curve and red curve are almost identical in the rest samples. It presents accurate image regression results for solar irradiance capturing. In Fig. 5.12 (b) and Fig. 5.12 (c), the MAPE are 12.6% and 8.1%, which indicates that the regression error is decreasing with the increasing training data length.

**Comparison and discussion**

To comprehensive evaluate the proposed approach, several high impacted algorithms are investigated, and the comparison and analysis are shown as follows.

**Benchmark Comparison** In [165], the persistence approach is introduced to evaluate the solar irradiance, which is a widely used approach to build the benchmark data [166]. We also implement the persistence approach to estimate the solar irradiance. The MAPE of the persistence approach is 27.8% and root-mean-square error (RMSE) is 29.2%, which are much larger than the proposed approach with MAPE 8.1% and RMSE 8.7%. In [166, 167]. It requires stationary for the time series. For the cloudy days data, the MAPE and the RMSE of the ARIMA approach are 31.2% and and 32.7%, which are also very high.

**Comparison With Other Approaches** In [168], a image regression approach is proposed to predict solar irradiance with sky image, which consists of feature extracting and regression model with promising results. The MAPE 8.1% of the proposed approach is lower than 21.91% with approach in [168]. In [169], a statistic based artificial neural network approach is used to provide solar irradiance forecasting with minimum root-mean-square error (RMSE) about 15%, which is higher than our proposed approach with 8.7%. In the review paper [166], a lot of recent solar irradiation capturing and forecasting approaches are collected and analyzed, and the best RMSE is about 10%, which is also higher than the proposed approach. Recently, in [170], a short-term solar irradiance forecasting approach is proposed with satellite and model coupling. In this approach, the sky images are captured by geostationary satellite and the typical errors ranges are from 8.5% to 17.2%, which is very close but still a little higher than the proposed approach. Furthermore, compared with the high-cost geostationary satellites, the proposed approach only requires the sky images, which can be easily captured by sky imagers, cameras, and even cellphones. In [171], several machine learning based approaches are used such as leapForward model and Spikeslab model to forecast the solar irradiance. Compared with the proposed approach, this proposed approach uses the measured solar irradiance as input, which indicates this method is heavily relies on the expensive solar irradiance measurement devices and not easy to extend on large-scale solar irradiance capturing. In [172], a solar irradiance forecasting approach is proposed with wavelet-based feature extraction, and this approach is focusing on the clear-sky days. Considering the large deviation profile of the PV output caused by clouds, the proposed approach is investigating in the cloudy days, which have larger impacts on power system operations. The randomness and stochastic characteristics of the cloud movements also increase the difficulties of the regression. The way on using CNN model helps to solve the problem. We only investigating in the sky images taken with visible light as a primer exploration. which indicates other images such as 10-12 $\mu m$ normalization radiance and R/B ratio mask are ignored. However, according to [166, 168, 169], these images contains

useful information and can be treated as a good assistance information for image regression. Beside the Zenith angle and temperature, the other parameters such as humidity, air condition, and solar periods can also be seemed as assistance inputs for the proposed regression approach. In this primer exploration, we only collected three month data in the Autumn of 2016 and Summer of 2017 to validate the proposed approach. In the next step, we will collect at least three years of data to further investigate the solar image regression.

As discussed above, the main advantages of the proposed approach are described one by one. Compare with the traditional methods, the proposed approach provide many different advantages.

1. Compared with the traditional solar irradiation capturing approaches, the proposed approach does not require the expensive high-resolution images captured by geostationary satellites. The data collection cost (sky images) of the proposed approach is lower, which indicates that the proposed approach is convenient to be widely used and extended for very large-scale solar irradiance monitoring.

2. Compared with the results of benchmark and traditional machine learning approaches, the proposed approach provides an end-to-end manner to capture the solar irradiances. This not only reduces the information loss between the feature engineering and regression processing, but also improves the learning efficiency of the proposed approach.

3. In power system operations, the clouds movements cause a lot of random deviations of PV outputs, which dramatically impact the health of the power systems. The proposed approach is focusing on the cloudy days solar irradiance capturing, which can be extended to PV output regression and power system stability forecasting in the next step research.

## 5.2 Ensemble model design with multiple machine learning and deep learning models

As discussed in Chapter 4, one critical issue which impacted the stochastic optimization performance dramatically is forecasting results. If the performance of the forecasting results can be improved, the total operation cost of the stochastic optimization can be further reduced. Considering traditional forecasting methods in power system forecasting such as linear regression, support vector regression, and neural network (also named as multi-layer perceptron), these machine learning based methods are very powerful but easy to overfit and usually have similar pattern in feature selection. It means all these types of methods only focus on limited features of the target, and other useful information are ignored. Ensemble learning is a method that uses a lot of weak learners together to reduce the forecast error and suppress the variance. To use the information comprehensively, a series of ensemble learning are employed such as random forest, extra tree, gradient boost decision tree, and Xgboost to improve the forecast performance. Furthermore, based on these ensemble learning methods, a novel stacking method with two step architecture is designed to further improve the forecast performance. Compare with the general forecasting model with machine learning, the novel stacking model is used to analysis the hidden relationship between different kinds of features. The proposed methods introduced a new way on how to efficiently construct a ensemble model with considering the different aspects of the data. After that, the output of first level models are used to further reduce the error between forecasting and actual results. The proposed model can be used as a pre-trained model for future forecasting work with a high accuracy without expensive cost on maintenance.

No matter energy generation, transmission or distribution, load forecasting is obviously very important for energy supplier and others. A power load forecasting model in high accuracy is essential to energy provider on their daily power planning and operation work. Load forecasting can be regarded as three different categories, long-term level [173], medium

level [174] and short-term level [50]. Based on the current research, short-term demand load forecasting is considered currently. It is usually form one hour to one week.

### 5.2.1 Load forecasting tools

**linear regression model** Linear regression [175] is a widely used statistical tool. It is very easy to be implemented and employed to model the relationship between different factors like temperature, zenith angle, kinds of renewable energy. The output is modeled as linear combination of inputs. The mathematical affine function between inputs and output. $\beta$ is linear parameter weights,$\epsilon$ is error. $X_i$ is the feature input, $y$ is the renewable generation.

$$y = \beta_0 + \sum \beta_i X_i + \epsilon_i \tag{5.2.1}$$

**Support Vector Machine** It is a machine learning technique based on the statistical learning method. It can be used for classification and regression problem, which analyzes the data and its category by support vectors computation. A SVM model always constructs one or several hyperplane in low, high or infinite dimensional space.

$$min \frac{w^2}{2} \tag{5.2.2}$$

$$s.t.\, y|w^T X + b| \geq 1 \tag{5.2.3}$$

where $w$ is the normal vector to hyperplane, $b$ is the bias, $X$ is the input data, $y$ indicates the classification of $X$. SVM only focus on the points which are classified correctly. For hard margin, two parallel hyperplane are figured out, which are used to separate the different classes data if the data are linearly separable. Otherwise, I have some other ways to deal with the problem if the input data is nonlinearly separable. Firstly, the tolerance parameter is added to the objective function, which also works as a regularization for the model. The tolerance parameter $C$ indicates how much the model will accept the wrong classification.

It is easy to see if $C$ is big enough, $\xi_n$ will be very small, which means that the model has a low tolerance on mistakes. It will lead the model to have low generalization and high accuracy.

$$min\,\frac{w^2}{2} + C\sum_{i=1}^{n}\xi_n \tag{5.2.4}$$

$$s.t.\,y|w^T X + b| \geq 1 - \xi_n \tag{5.2.5}$$

$$\xi_n \geq 0 \tag{5.2.6}$$

Secondly, SVM model can help to find out a linear separable hyperplane by using kernel conversion. It aims to solve the problems in a high dimension space if the hyperplane does not work in low dimension space.

**Artificial Neural Network** Artificial Neural Network(ANN) has been used to simulate the nonlinearly of human thoughts because of its strong computing ability. It has been found that it can be used to overcome the problem instead of a functional form model. The deeper layers construction of ANN model have been demonstrated with a high ability on nonlinear problem approximation. Multilayer perceptron network (MLP) in 5.13 is used here to simulate the results. In each model, there are several hidden layers in it. There are more than one neurons in each layer. Each input is multiplied by weights($\omega_i$) and the output of each neuron is decided by active function. The advantage of this model is that most of the forecasting model will not require a function model, but it needs more time on training and more data on pre-trained the model. The fully connection feed-forward(FCFF) model with an active function. MLP model describes a nonlinear relationship of weights which connect the input , hidden layers and the output data. In training of the model, for each iteration in each epoch, back propagation can be used to update the model parameter. In the forecasting work of the renewable energy, the historic data based supervised learning method is chosen to figure out the relationship between demand load and day temperature, human events, weather changing. The hourly electric load can be predicted with this model.

Besides that, several ANN models can be trained in parallel and the average results ba be used. It can help reduce the variance of the model. Transfer learning is a good way to solve the problem with not enough data, the pre-trained model parameter(weight) are applied and do fine tuning on the last few layers to modify the result in validation.

In 5.13, there are 4 feature units at the input layer, 5 at the first hidden layer, 4 at the second hidden layer, a01 indicates the first neuron at the first layer. a12 indicates the second neuron at the second layer. The first hidden layer computation is as follows. a01, a02, a03, 04 indicates the parameter of features, the output is the renewable generation in the forecasting model. $g$ is an active function.

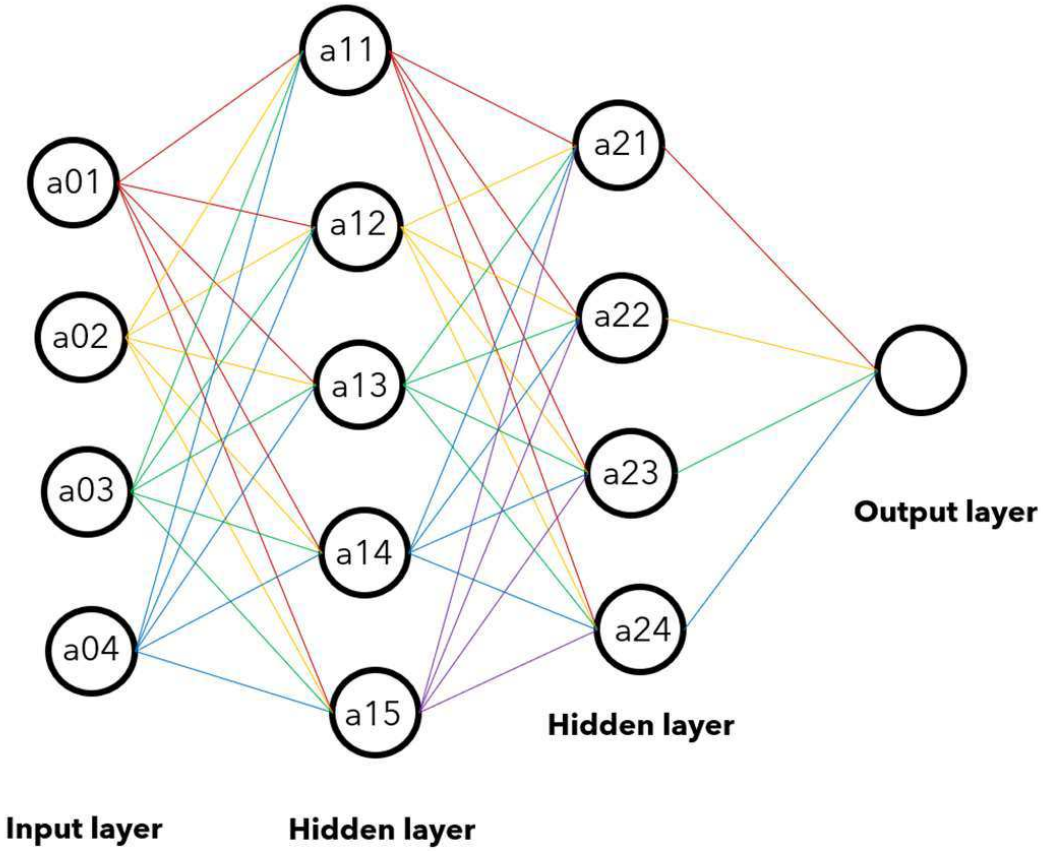



Figure 5.13: MLP network.

$$a_{11} = g(w_{01} * a_{01} + w_{02} * a_{02} + w_{03} * a_{03} + w_{04} * a_{04}) \tag{5.2.7}$$

$$a_{12} = g(w_{11} * a_{01} + w_{12} * a_{02} + w_{13} * a_{03} + w_{14} * a_{04}) \tag{5.2.8}$$

$$a_{13} = g(w_{21} * a_{01} + w_{22} * a_{02} + w_{23} * a_{03} + w_{24} * a_{04}) \tag{5.2.9}$$

$$a_{14} = g(w_{31} * a_{01} + w_{32} * a_{02} + w_{33} * a_{03} + w_{34} * a_{04}) \tag{5.2.10}$$

$$a_{15} = g(w_{41} * a_{01} + w_{42} * a_{02} + w_{43} * a_{03} + w_{44} * a_{04}) \tag{5.2.11}$$

**Random Forest** Random forest is an ensemble learning with several decision trees. The average results will be calculated for regression problems and vote results for classification. Random forest is an famous example for bagging, which means each time some of the samples are collected to train a decision tree. It is obviously that I can do parallel work on training if I need more than one decision tree. And the average results for regression can reduce the variance efficiently which help to avoid overfit. After many tests, it has been demonstrated that only 68.2% sample data has been used in training, the rest can be used for testing, which avoid the cross validation work. One more advantage for decision tree why we consider random forest as an important role in our forecasting work, the way to measure the best split point is not the same as other algorithm like SVM. The information gain and gini are used to find out the best split points which do not care the imbalanced data.

**ID3-maximum information gain** For dataset $D$, assume there are K categories, the empirical entropy of D $H(D)$. It can be calculated with the equation below.

$$H(D) = -\sum_{k=1}^{K} \frac{|C_k|}{|D|} log_2 \frac{|C_k|}{|D|} \tag{5.2.12}$$

$$H(D|A) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} H(D_i) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} \big( - \sum_{k=1}^{K} \frac{|D_{ik}|}{|D_i|} \log \frac{|D_{ik}|}{|D_i|} \big) \tag{5.2.13}$$

$$g(D, A) = H(D) - H(D|A) \tag{5.2.14}$$

The empirical conditional entropy of feature A to dataset D $H(D|A)$ is computed here. $D_i$ indicates the dataset from $D$ if feature A equals its ith value. $D_{ik}$ is the $k$th dataset in $D_i$. The information gain can be calculated as the equation below.

**C4.5- maximum information gain ratio** The definition of information gain ratio of feature A to dataset D is defined as below. The cross entropy is used to describe how much different between the distributions.

$$g_R(D, A) = \frac{g(D, A)}{H_A(D)} \tag{5.2.15}$$

$$H_A(D) = - \sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|} \tag{5.2.16}$$

**CART - maximum gini** Gini is used to describe the purity of data.

$$Gini(D) = 1 - \sum_{k=1}^{K} \big( \frac{C_k}{D} \big)^2 \tag{5.2.17}$$

**Adaboost** After testing in some independent models, a new way of ensemble learning model is constructed based on some different individual models. It is an efficient way to avoid overfitting and take more attention on errors. In every iteration, the model will reweight the error and each base learner. The weights will pay more attention on error and award the base learner who has a high accuracy. It is much different from bagging, which aims to reduce the bias and cannot be computed in parallel.

**Gradient Boosting Descent Tree(GBDT)** The main idea of using GBDT is that a CART is used to simulate the residual between forecasting value and the real data. In each iteration, the model aims to reduce the residual and finally eliminate it. The model is selected to set up with direction of negative gradient. As the industrial application of GBDT, XGBoost shows the more obvious advantages and better performance in our results.

$$L(\phi) = \sum_i l(\hat{y_i}, y_i) + \sum_k \Omega(f_k) \tag{5.2.18}$$

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda ||w||^2 \tag{5.2.19}$$

It is easy to see that the loss function of XGBoost has been added regularization, which includes the amount and value of the leaf nodes. The results shows it efficiently increase the generalize ability of the model on testing data.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{5.2.20}$$

The loss function at time $t$ tells us the loss is calculated as the difference between the real tag and the sum of the forecast tag of time $t-1$ plus new fitted model of time $t$. The second order taylor expandion: g is first-order derivative, h is the second order derivative.

$$L^{(t)} \simeq \sum_{i=1}^{n} \left[ l(y_i, \hat{y_i}^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{5.2.21}$$

$$g_i = \delta_{\hat{y_i}^{(t-1)}} l(y_i, \hat{y_i}^{(t-1)}) \tag{5.2.22}$$

$$h_i = \delta^2_{\hat{y_i}^{(t-1)}} l(y_i, \hat{y_i}^{(t-1)}) \tag{5.2.23}$$

Different from GBDT, XGBoost has its approximate algorithm on feature and split points selection, which is much faster than grid search of GBDT. XGBoost can realize parallel work in some step of the feature engineering, which is key points on industrial use.

## 5.3 Methods Analysis

### 5.3.1 Correlation Analysis

In order to analysis the relationship between different kinds of features and target GHI, the pearson correlation figure is shown. The Figure presents the relationship between different features.

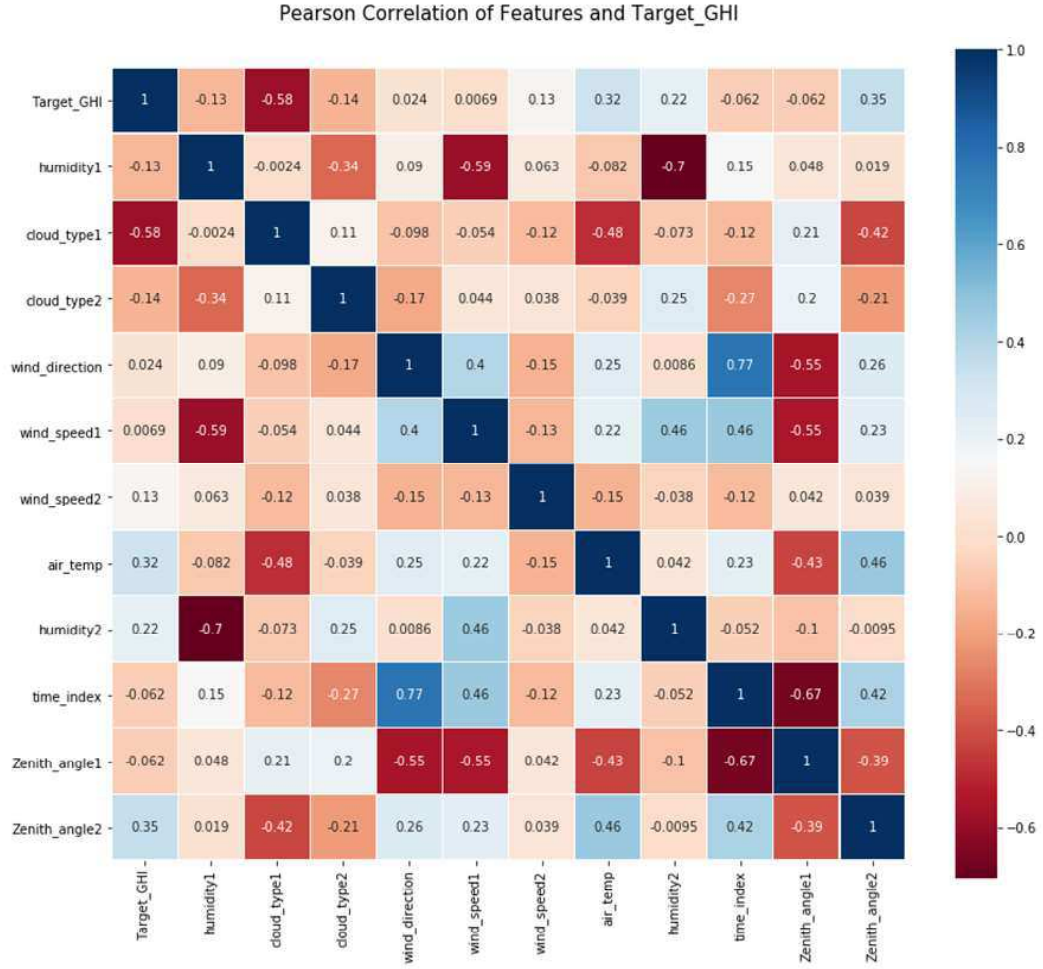



Figure 5.14: Pearson Correlation Analysis.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}. \tag{5.3.1}$$

- $cov(X, Y)$ is the covariance
- $\sigma_X$ is the standard deviation of $X$.
- $\sigma_Y$ is the standard deviation of $Y$.
- $X$, $Y$ are the feature data of the input.

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]. \tag{5.3.2}$$

- $\mu_X$ is the mean of $X$.
- $\mu_Y$ is the mean of $Y$.

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}. \tag{5.3.3}$$

The Pearson correlation can be computed here5.3.3. Compared with linear analysis, the Pearson correlation coefficients analysis is more accurate on describing the relationships between random variables and stochastic process, which is also widely used in many statistic analysis for random distributions. The Pearson correlations analysis contains two critical quantities for random variables: mean and variance, which indicate the first-order and second-order variances of the variables.

As shown in Fig. 5.14, 11 features such as 'humidity1','cloud type1','cloud type2','wind direction','wind speed1','wind speed2','air temperature','humidity2','time index','Zenith angle1','Zenith angle2' are selected as inputs to generate the Pearson correlation coefficients with the target GHI. In detail, there are several critical points: It is clear that the diagonal values are equals to 1, which means that they are highly related to themselves with equation ( 5.3.3. On the right side, the colored bar indicates that the numerical value from 1.00 to -1.00 with the color from deep blue to deep red. which indicate the positive relationships and negative relationships separately. The higher absolute value indicates a higher

correlation between these two features. On positive relationships, it is clear that the Zenith angle2 is 0.35. It means the target GHI will increase if the Zenith angle 2 increases. On negative relationship, it is clear that the cloud type1 is -0.58. It means the target GHI will decreases if the cloud type2 increases. In large-scale, Fig. 5.14 shows various relationships from positive to negative. It means our data collection contains many different aspects of GHI related data features and this will benefit the final results. Besides the target analysis, there are many other interesting relationships, for example, the wind direction is positive related to time index, which means the wind direction has strong temporal pattern.

### 5.3.2 Linear Regression Analysis

Linear Regression is very easy to be implemented and employed to model the relationship between different factors like temperature, wind speed, kinds of renewable energy. The output is modeled as linear combination of inputs, the mathematical model can be wrote as affine function between inputs and output. As shown in the Fig. 5.15, the linear regression performance is highly correlated to the Pearson correlation analysis, because linear regression only compute the weighted sum of the input features. The high important feature of linear regression are cloud type 1, air temp, and zenith angle 2. The right bar is used to measure how the features contribute to the results. The sum of all the weights is 1. The higher number obey to the bar means the feature plays an more important role in result.

### 5.3.3 Support Vector Regression Analysis

SVM can be used for classification and regression problem, which analyzes the data and its category by support vectors computation. A SVM model always constructs one or several hyperplane in low, high or infinite dimensional space. As shown in the Fig. 5.16, the SVR performs better than simple linear regression, which has high importance features on cloud type 1, air temp, and zenith angle 2. SVR has a kernel trick, which maps the input features into high-dimensional space.

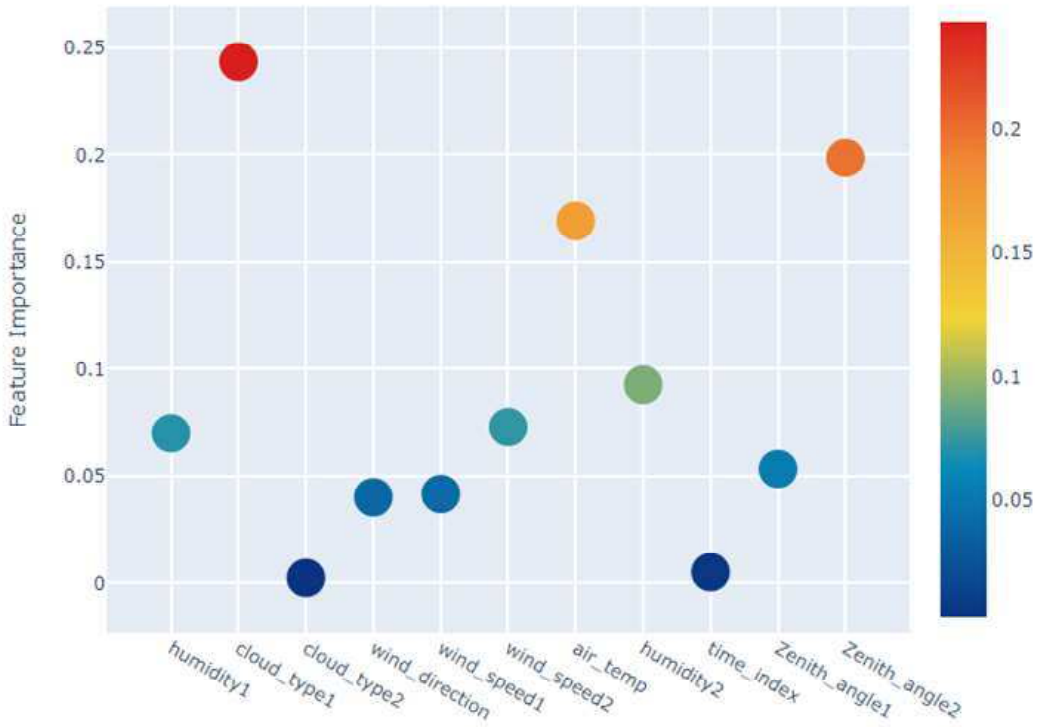



Figure 5.15: Feature importance analysis of Linear Regression.

The nonlinear mapping dramatically improves its regression and tracking capability for nonlinear analysis. It is obvious that SVR works much better than linear regression on this forecasting work.

### 5.3.4 Multi-layer Perceptron Analysis

The MLP model is a construction of fully connection layers. As shown in the Fig. 5.17, similar as SVR, the MLP has high importance features on cloud type 1, air temp, and zenith angle 2. In MLP, the nonlinear capability is introduced by the active function, which is a nonlinear and differentiable function. To sum, the linear regression, SVR, and MLP, these traditional methods have very similar pattern in feature selection. Because only the some group of feature data are studied. Therefore, several better methods are introduced, and then all the methods will serve as inputs for the final ensemble learning.

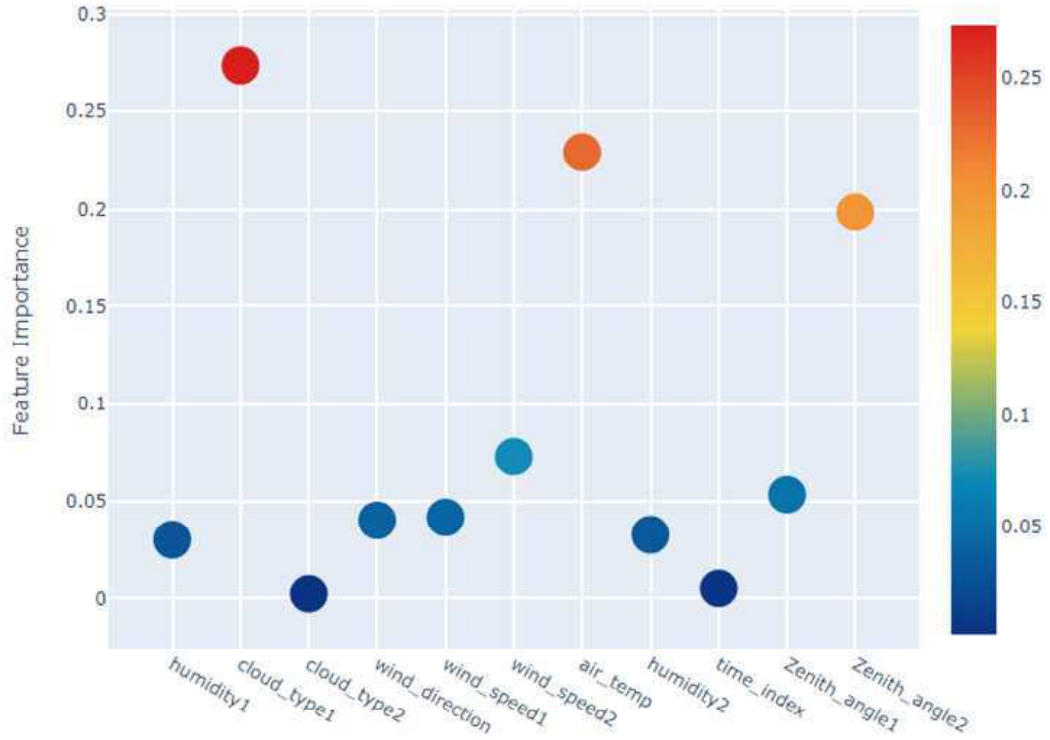



Figure 5.16: Feature importance analysis of Support Vector Regression.

### 5.3.5 Random Forest Analysis

Random forest is an ensemble learning with decision trees. The average results will be calculated for regression problems. And the results can reduce the variance efficiently which help to avoid overfitting. As shown in the Fig. 5.18, different from the traditional method, which is only sensitive on one or two input features, the random forest methods has several high scores on feature importance analysis. This means that the random forest method collects more features from the input data and evaluates the target GHI comprehensively. The importance feature for random forest are Zenith Angle2, Cloud typ1, and humidity1.

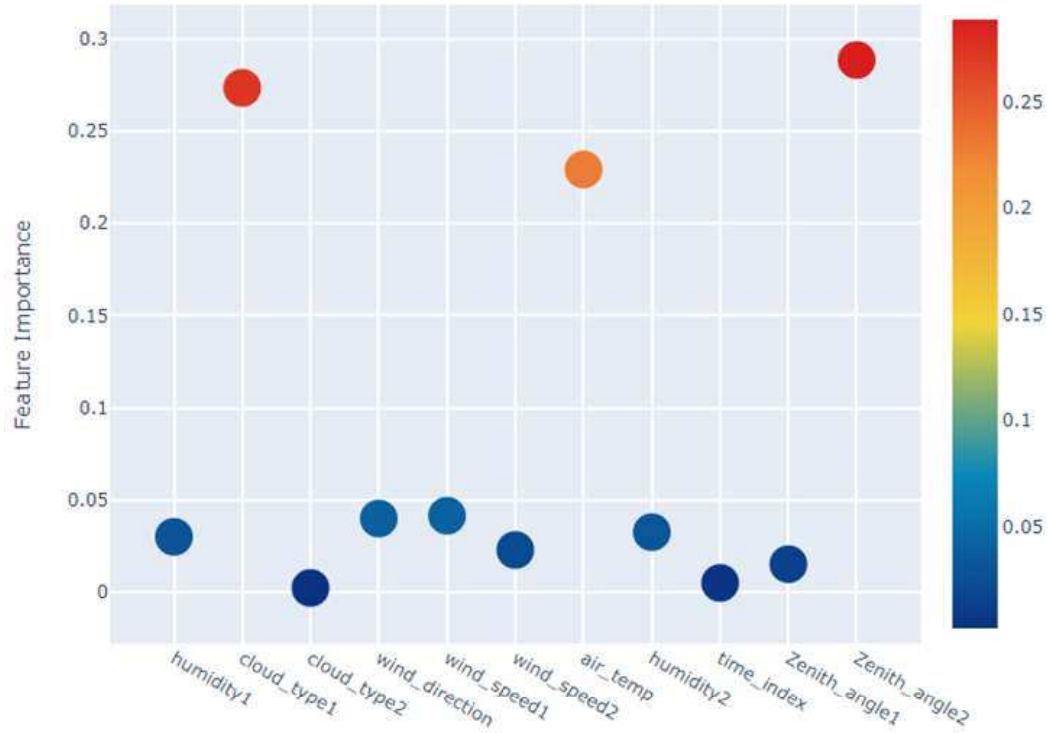



Figure 5.17: Feature importance analysis of Multi-layer Perceptron.

### 5.3.6 Extra Tree Analysis

A little different from the random forest, all the dataset is used instead of bootstrap of RF to train the decision tree. As shown in the Fig. 5.19, the extra tree methods has several high scores on feature importance analysis such as humidity1, cloud type1, air temperature , zenith angle 1 and zenith angle 2. Compared with random forest method, the extra tree collects different aspects of input features. This indicates that our proposed approach collect the bread information of the input data. With different features, the ensemble learning performances better than one or two feature collections.

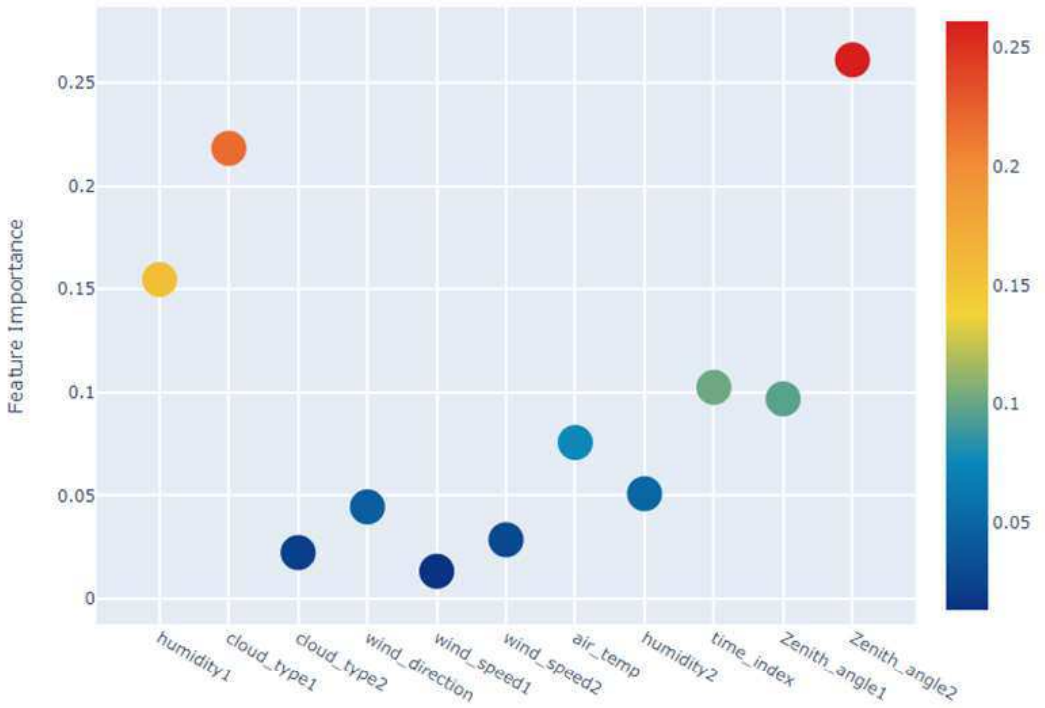



Figure 5.18: Feature importance analysis of Random Forest

### 5.3.7 Adaboost Analysis

It is an efficient way to avoid overfitting and take more attention on errors. In every iteration, the model will reweight the error and each base learner. The weights will pay more attention on error and award the base learner who has a high accuracy. As shown in the Fig. 5.20, the Adaboost is a type of boost method, which can regress any kind of curves with limited error. The key feature of boost technology is to estimate the residual for next step approximate. On one hand, this technology generate high-accuracy result; on another hand, this type of method is easy to overfit. As shown in the Fig. 5.20, the importance feature of Adaboost is only focusing on the zenith angle 2.

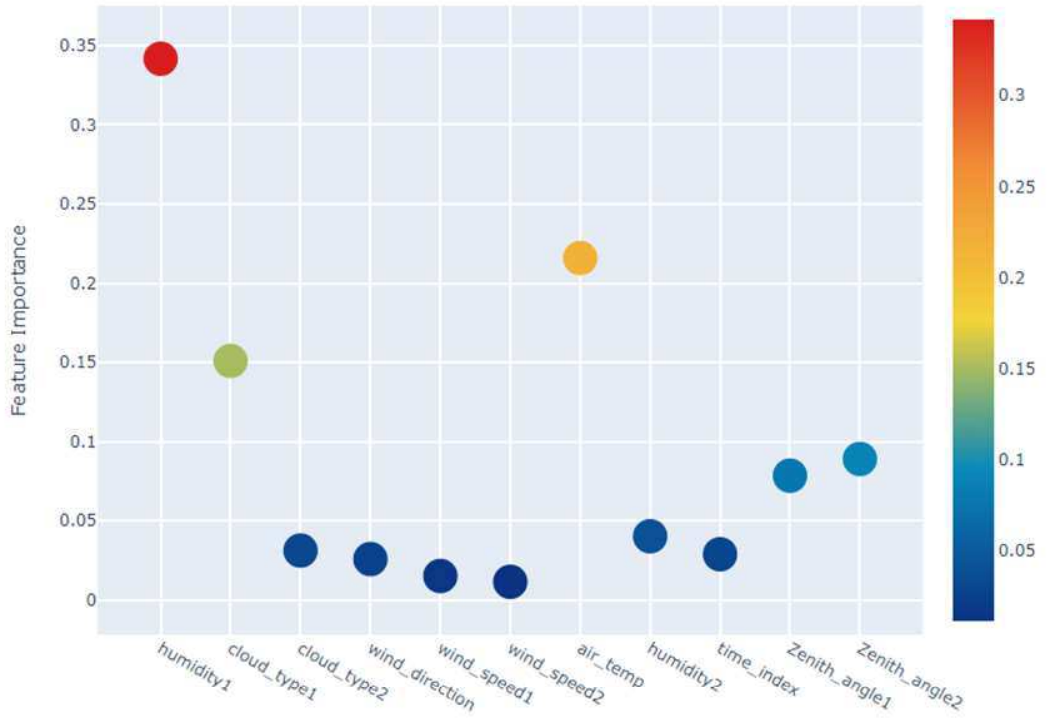



Figure 5.19: Feature importance analysis of Extra Tree

### 5.3.8 GBDT Analysis

It is an efficient way to avoid overfitting and take more attention on errors. In every iteration, the model will reweight the error and each base learner. The weights will pay more attention on error and award the base learner with a high accuracy. As shown in the Fig. 5.21, the GBDT method has highest import feature score is about 0.5. Compared with the Adaboost method, which focuses on only one feature, the GBDT split the pressure into several different features such as wind speed1, cloud type 2, zenith angle 1. With comprehensive feature collections, GBDT performs better than Adaboost methods. GBDT is widely used in current machine learning research. It is different from the boosting method, like adaboost. GBDT pays more attention on the incorrect results.

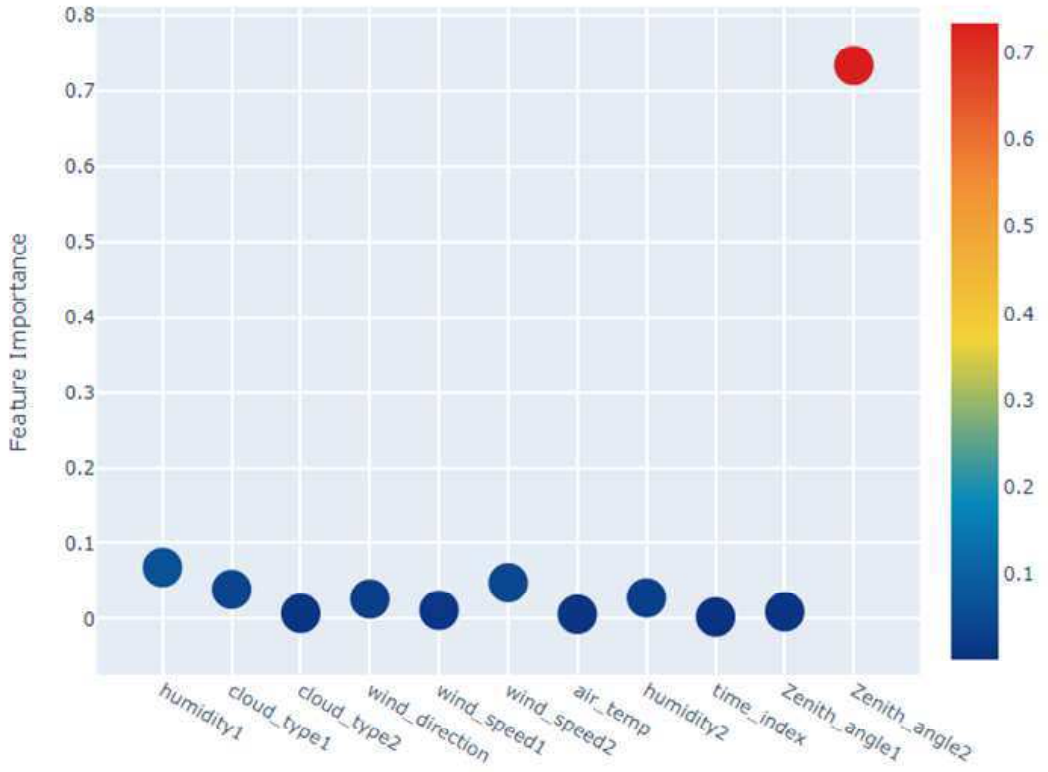



Figure 5.20: Feature importance analysis of Extra Tree

### 5.3.9 XGBoost Analysis

As the industrial application of GBDT, XGBoost shows the more obvious advantages and better performance in our results. As shown in the Fig. 5.22, the Xgboost methods has several high scores on feature importance analysis such as cloud type1, air temperature , and zenith angle 2. Compared with Adaboost and GBDT, the Xgboost has better suppression of single tree grows and leaf splits, which avoid to focusing on one or two features as shown above. At the same time, Xgboost is also the method which performs best in the first level method. Consider its capability, a lot of ensemble method contains it as a key method to improve their final performance.

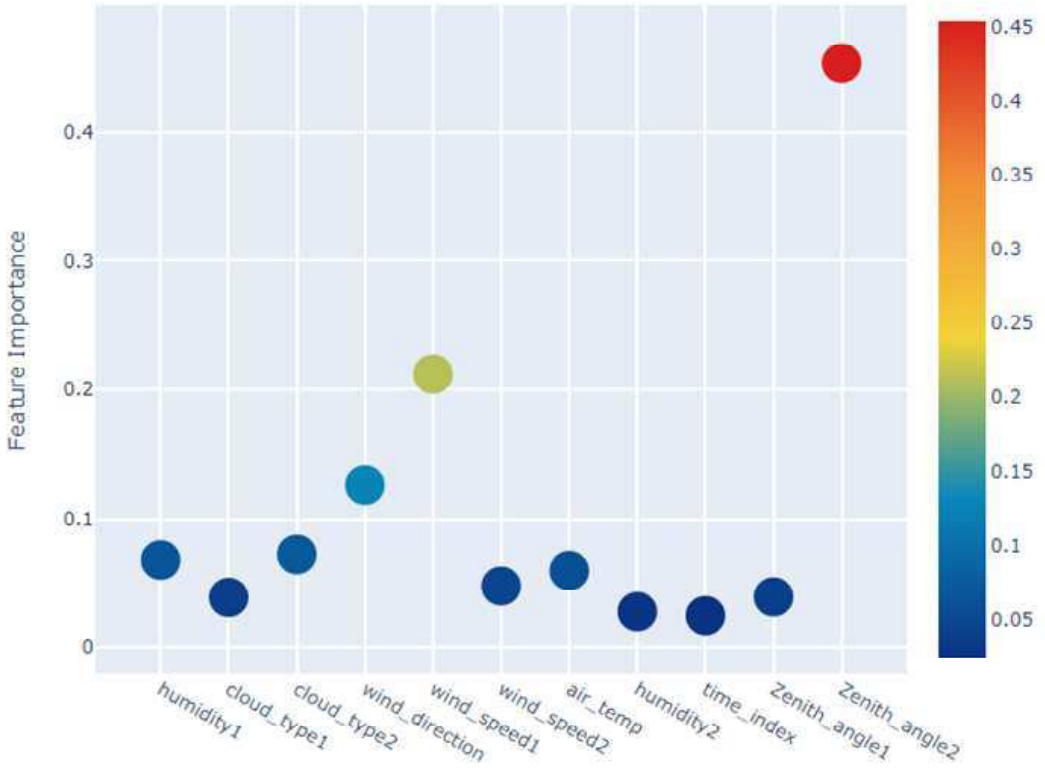



Figure 5.21: Feature importance analysis of GBDT.

## 5.4 Ensemble Learning - Stacking Regression

As discussed above, the basic models such as linear regression, SVR, and random forest contain different characteristics, which perform the importance of different feature. The performances of the basic models in different feature importances also indicate that none of them can comprehensively extract all the features of the input data. As shown in Fig. 5.24, a stacking regression approach is proposed to learn widely from the strong points of the basic models. Specifically, in the first step, the basic models are used to extract the feature from the input data. In the second step, a stacking regression method is used to find a nonlinear relationship between the extracted feature and given labels. The detailed process will be introduced step by step. The architecture of stacking regression is shown in Fig. 5.25 in detail.

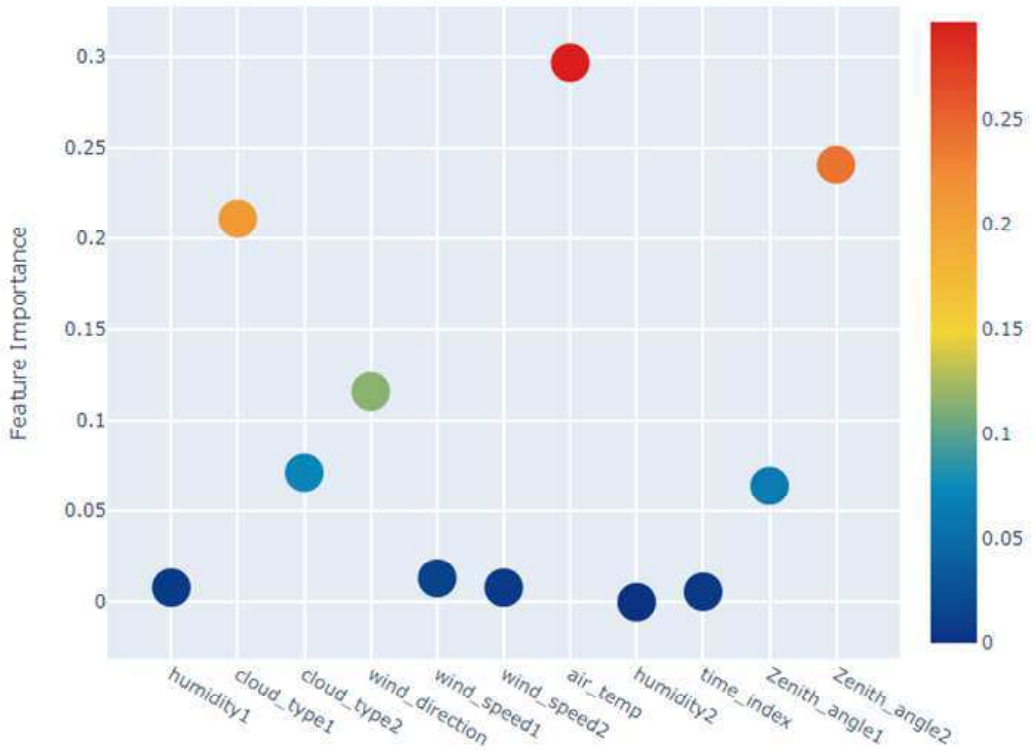



Figure 5.22: Feature importance analysis of Xgboost.

The upper part indicates the training part, and the lower part indicate the testing part. In the training part, the original input data is divided as K-fold format for training. Each model uses different training set and prediction set (also be named as validation set) sequentially. Then, all the prediction sets are collected as new features for next step process. In this step, the basic models are employed for feature extraction. In the second step, the extracted features are treated as input data to train the advanced model, which can be any regression models such as MLP, linear regression, and Xgboost. In the testing part, the large gray arrows indicate the testing data are predicted by the trained basic models, which can be seen as feature extraction for the testing data. The average step is used to keep the dimension of the testing data. Last, the trained advanced model is employed to generate the final prediction results, which evaluates all the input data comprehensively.


Figure 5.23: Feature extraction correlations.

## 5.5 Bootstrapping Based Prediction Intervals Computation

### 5.5.1 The Theory of Bootstrapping

To quantify the forecast uncertainty of the proposed approach, the bootstrapping method is used to compute the prediction intervals. The resampling data can provide the empirical distribution from the forecasting results, which also brings high computational cost.

For $n_1$ forecasting results, the noise variable $\varepsilon_{j_1}$ can be approximated as the equation below. $j_1 \in \{1, 2, 3, \cdots, n_1\}$. $\hat{\Psi}$ is the approximated function of $\Psi$.

$$\hat{\varepsilon}_{j_1} = \gamma(j_1) - \hat{\Psi}(j_1). \tag{5.5.1}$$

Considering the systematic error, $\hat{\varepsilon}_{j_1}$ can be centered as 5.5.2. The equation shows a process of iteration of the whole project. In general, each model has its own advantages, it

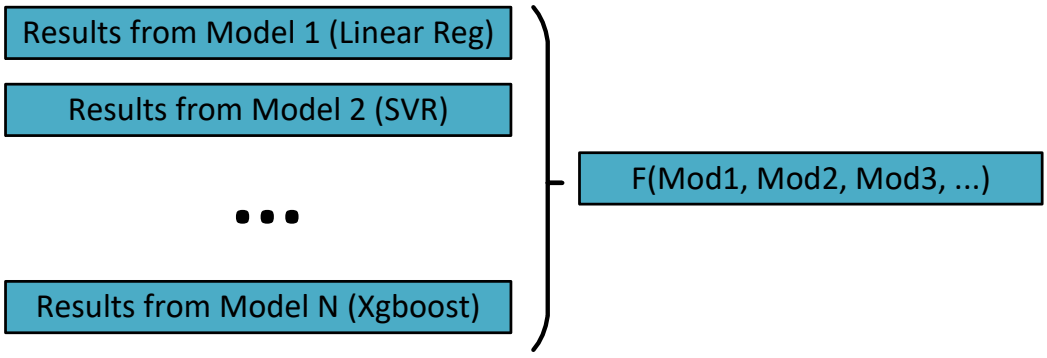



Figure 5.24: Major Idea of Stacking Regression.

always prefer some special aspects of the data. And how to construct a reasonable ensemble model based on these advantages from different models. The proposed model describe a efficient way on constructing a stacking model to achieve a comprehensive model.

$$\tilde{\varepsilon}_{j_1} = \hat{\varepsilon}_{j_1} - \frac{1}{n_1}\sum_{j_2=1}^{n_1}\hat{\varepsilon}_{j_2}. \tag{5.5.2}$$

where $j_2 \in \{1, 2, 3, \cdots, n_1\}$. As shown in Fig. 3.1, in Block 8, the Bootstrap resampling can be generated as following. The residual $\delta^*$ can be drawn randomly with replacement from the set $\{\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \cdots, \tilde{\varepsilon}_{n_1}\}$ with the probability $1/n_1$ as following

$$\delta^*_{i_1} = \tilde{\varepsilon}_{i_1}\ with\ probability\ \frac{1}{n_1}. \tag{5.5.3}$$

Then the output can be computed as

$$\gamma^*_{i_1} = \hat{\Psi}(i_1) + \delta^*_{i_1}. \tag{5.5.4}$$

$n_2$ is a given number of bootstrapping replicates generated by formula (5.5.3) and (5.5.4). $n_2$ is a large number, such as 5,000 or 10,000. According to the Law of Large Numbers (LLN) and Central Limit Theorem (CLT), if the number of bootstrap replications

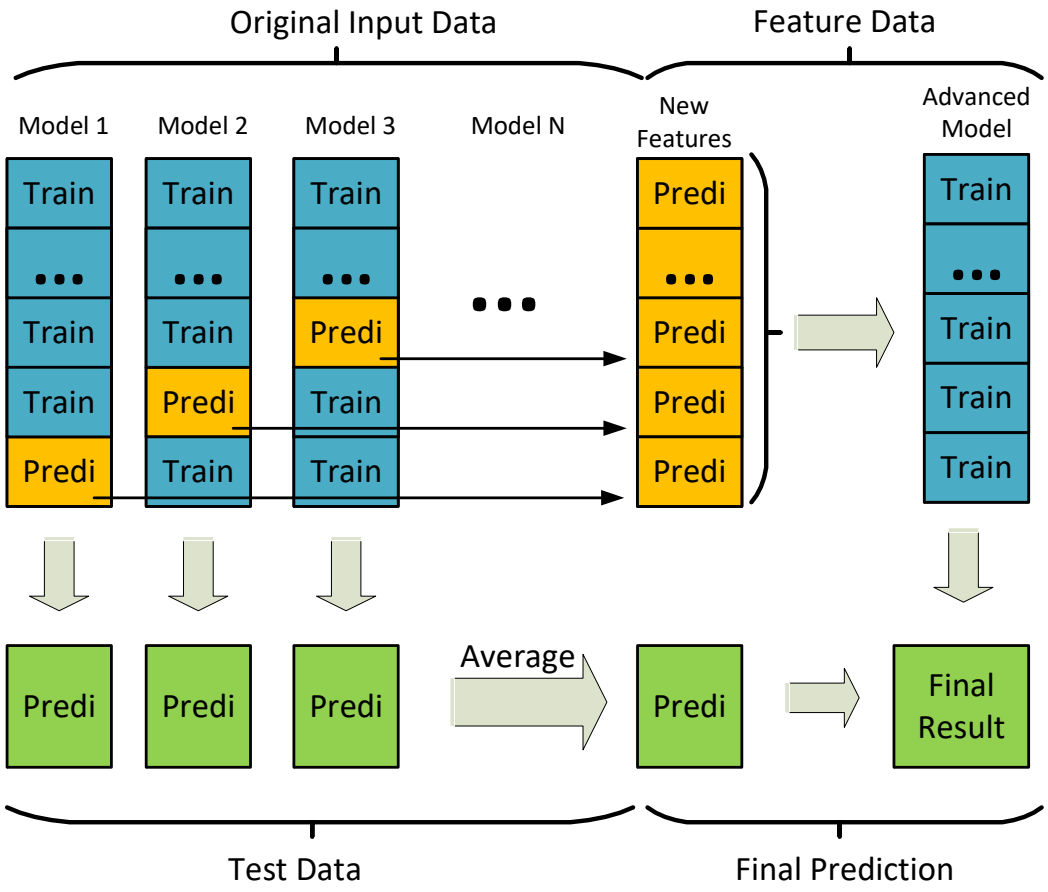



Figure 5.25: Detail Architecture of Stacking Regression.

is large, the forecast errors are normally distributed. The $100(1-2\alpha)$ prediction intervals can be computed as equations below. And the parameters in the proposed project are defined as $\alpha$.

$$P\Big\{\underline{\mathbf{Q}}(\Phi_\alpha) < \gamma_{i_1} < \overline{\mathbf{Q}}(\Phi_{1-\alpha})\Big\} = 1-2\alpha. \tag{5.5.5}$$

where $\underline{\mathbf{Q}}$ is a function of $\Phi_\alpha$, which indicates the lower prediction interval. $\overline{\mathbf{Q}}$ is a function of $\Phi_{1-\alpha}$, which indicates the upper prediction interval. The algorithm of the Bootstrapping Based Prediction Intervals Computation is shown in Algorithm 1.

**Algorithm 2** Bootstrapping Based Prediction Intervals Computation

**Objective**: Compute the prediction intervals from the prediction results.

**Data Collection**: Collect the prediction results.

**Residuals Processing**: Compute residuals with formula (5.5.1) and re-center residuals with formula (5.5.2).

**Resampling**: Resample the residuals with formula (5.5.3) and compute $\hat{\Psi}$ with formula (5.5.4). Repeat this step, and generate $n_2$ set of bootstrapping replicates.

**Computation of Prediction Intervals**: Given a probability $\alpha$, compute the prediction intervals $\underline{\mathbf{Q}}$ and $\overline{\mathbf{Q}}$ with formula (5.5.5).

Table 5.2: Performance of Proposed Approach

| Methods | MAPE (%) | Variance |
|---|---|---|
| Linear Regression | 21.1 | 27.5 |
| SVR | 17.7 | 16.9 |
| MLP | 17.9 | 16.4 |
| Random Forest | 15.2 | 14.2 |
| Extra Tree | 14.3 | 15.1 |
| Adaboost | 14.1 | 15.2 |
| GBDT | 12.21 | 13.4 |
| Xgboost | 11.8 | 12.5 |
| Stacking Approach | 7.8 | 7.2 |

### 5.5.2 Forecast Results and Statistic Analysis

As shown in Table5.2, all the methods are compared together with Mean absolute percentage error (MAPE) and variance, which indicate the 1st order and 2 order error in statistic learning. Specifically, the traditional approach linear regression, SVR, and MLP has largest MAPE and variance. With ensemble learning introduced, the MAPE and variance also reduced dramatically. Furthermore, based on all these approaches, the proposed method performs best with smallest MAPE and variance.

As shown in Fig.5.26, a cloudy day normalized GHI is shown as blue curve. The red curve is the forecasted GHI with stacking approach. It is clear that the red curve and blue

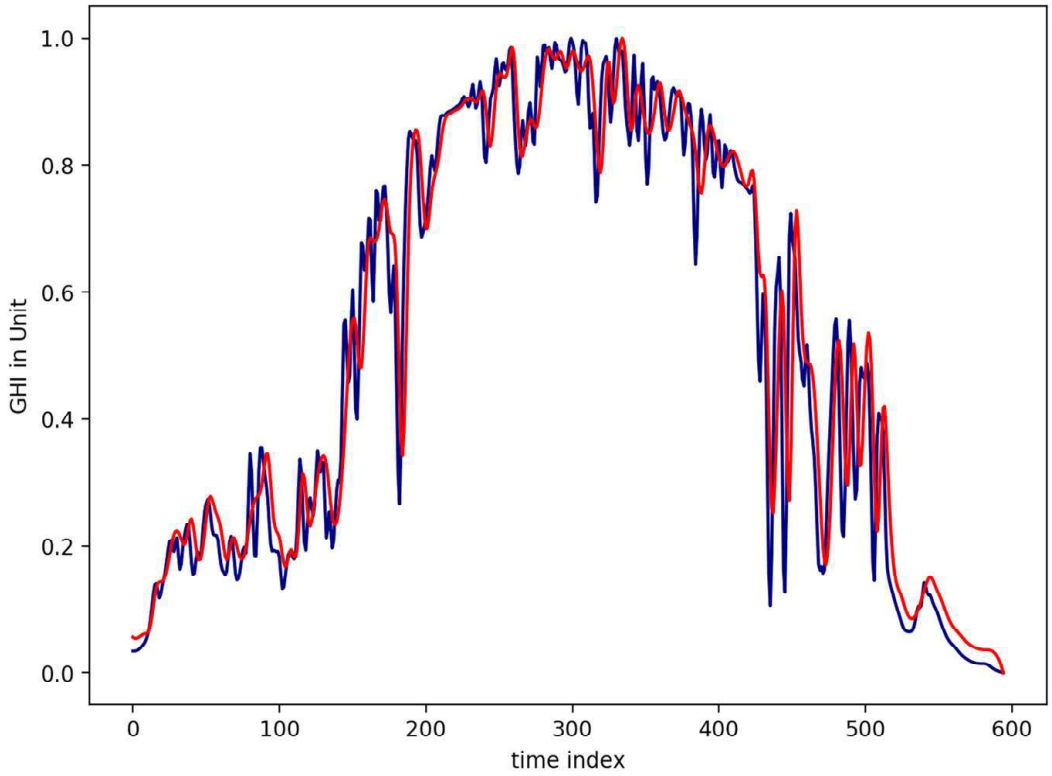



Figure 5.26: One day forecast example of proposed method.

curve are almost identical, except at several large deviation points. With this high accuracy forecast results as input, the numerical results of stochastic optimization will be presented in next step. As shown in Fig.5.27 and Table5.2, the proposed method and Xgboost rank in Top 1 and 2. Compared with Xgboost, the proposed approach has smaller variance, and higher kurtosis, which indicates the proposed approach provides better results with narrow confidential intervals.

## 5.6 Results

The numerical results for evaluating the proposed method in chapter 4 with the proposed ensemble learning forecasting model are tested based on the IEEE 123-bus system. The PV panels(55 kw for each) are installed at bus 28, 47, 49, 64, 93 and 97.

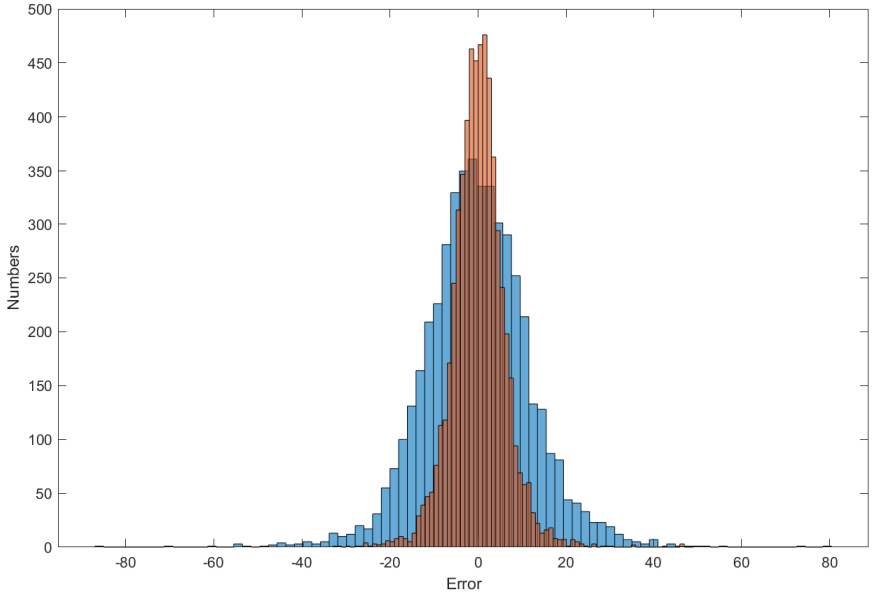



Figure 5.27: Error Distribution between Proposed Method and Xgboost.

### 5.6.1 Total cost with ensemble learning

The hourly renewable energy generation is predicted with the proposed forecasting model. And the total operation cost are compared in Fig.5.28 for 24 hours. For each hour, the left bar describe the ratio of $G_t^{DA}$ + $G_t^{R1}$ (light blue) and $G_t^{RT}$(yellow), which is fitted by multi-timescale model with traditional forecasting model. The corresponding total cost is displayed as a green dashed line. Similarly, the right bar describes the ratio of $G_t^{DA}$ + $G_t^{R1}$ (dark blue) and $G_t^{RT}$(orange), which is fitted by multi-timescale model with proposed forecasting model. The corresponding total cost is displayed as a red dashed line. It is clear that the orange part is shorter than the yellow one, which indicates that the proposed forecasting model achieves a higher accuracy on hourly predicting of the renewable energy, and the system requires less energy from the RT market. The difference between the total operation cost demonstrates the proposed forecasting method reduces the operation.

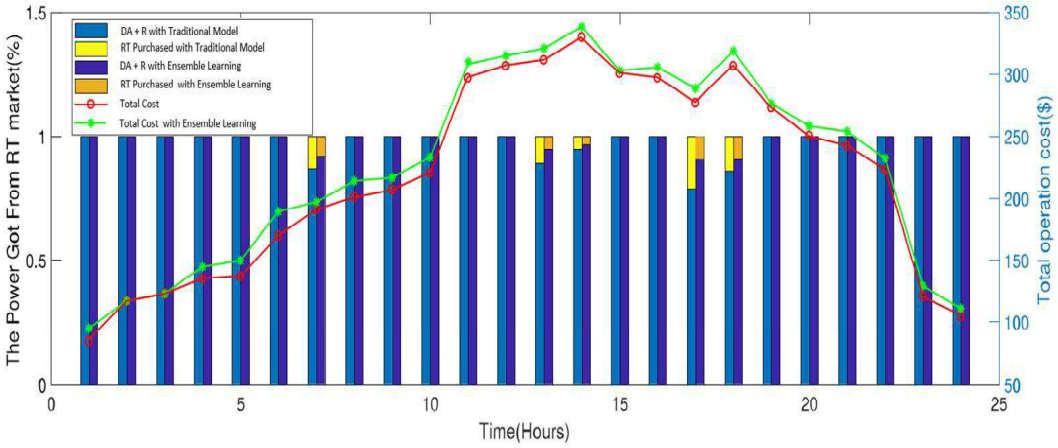



Figure 5.28: Total cost with Ensemble learning.

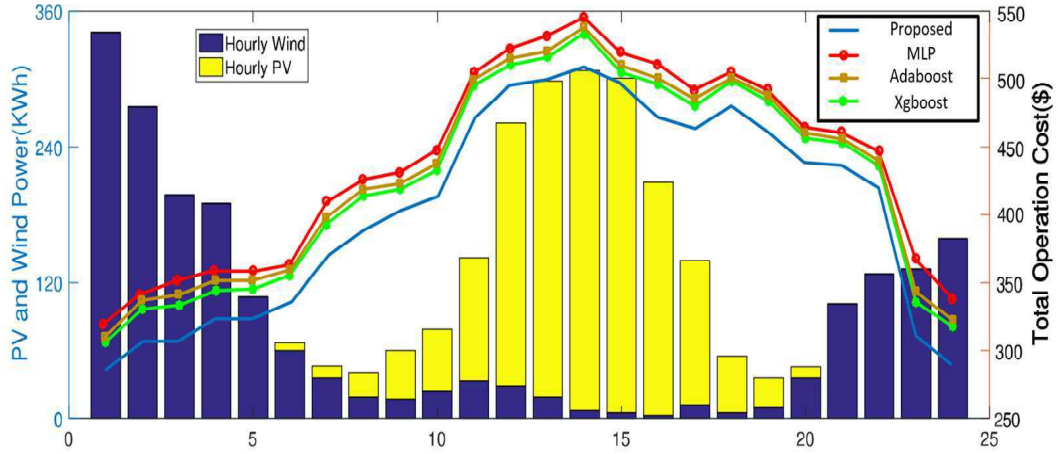



Figure 5.29: Total cost with different machine learning models and proposed model.

### 5.6.2 Total costs comparison with different machine learning models

Fig.5.29 describes the total operation cost with different forecasting model with machine learning technology. Multilayer perceptron model, adaboost, XGBoost and the proposed method are used to compare the results. As shown in 5.29, in Golden, Colorado, a typical day with a windy night and sunny daytime is selected with 24 hours. The peak generation of the wind turbines (blue bars) and PVs (yellow bars) occurs at midnight and

14:00, respectively. The total operation cost decreases from MLP, adaboost, XGBoost and the proposed method, which represents MLP obtains a lowest forecasting accuracy in this test, and the stacking forecasting model achieves much better results on load forecasting. The reason is that the stacking model is an ensemble learning of different kinds of machine learning model, it brings the different advantages from the several models. And how to find out the combination of models is a big problem, which helps to get a high accuracy with avoid overfitting.

## 5.7 Conclusion and Contributions

The proposed image regression problem is similar with the age estimation problem in computer vision area, which aims to build a map between the input human images and the corresponding ages. In [176], the age estimation is studied with a shallow machine learning model, support vector machine. In [177], a CNN based regression approach is proposed to estimate human age with the images of human face as the input data. However, the proposed multi-output CNN is also a classification problem, and people with different ages are classified into different small groups. Therefore, in this chapter, a CNN based image regression approach is proposed to provide a fast and accurate solar irradiation capturing.

The **main contributions** of this method are:

1. A big data processing approach is used in renewable energy area for solar irradiation capturing, which contains big data cleaning and deep learning based image regression. Compared with the traditional solar irradiation capturing approach, the CNN based approach is cheap, fast, accurate, and convenient to be extended for large-area monitoring.

2. Based on the Dirichlet process, variational inference, and Bayesian theory, a Gaussian mixture model with Bayesian inference approach is employed to determine the mixture components automatically. It decides the components amount automatically.

3. Based on the CNN architecture for classification, a new regression CNN architecture is designed for image regression problem. The input of the proposed approach is the image set and the output is the continuous variable set, which can be extended for multiple regression problems. According to Deep dream related algorithms, the CNN working manner is illustrated layer by layer with Figures, which help researchers deeply understand the working principle of deep learning.

# Chapter 6

# Conclusion and Future Work

A big data visualization platform is designed for engineers to monitor the system in an convenient way, which helps to discover the hidden useful knowledge for smart grid operation, control, and situation awareness. A short-term load forecasting based network reconfiguration is proposed to reduce the distribution system loss dynamically. Instead of the static load measurements at the scheduled time spots, the short-term load forecasting approach can provide accurate future load profiles, which contains more information for the network reconfiguration. The whole proposed approach is designed as a distributed computation approach. Considering the reality of three-phase unbalanced power system, a multi-timescale approach is proposed aiming to obtain a higher accuracy with SDP relaxation. The numerical results are calculated in a short time with parallel work by ADMM. In order to overcome the missing data in power system, a novel solar irradiance capturing approach is proposed with a CNN framework. Based on the new trend of artificial intelligence, a proposed convolution neural network is used to capture the GHI with the sky images. Compared with the traditional approaches, the proposed approach is accurate, flexible and convenient to be widely deployed for large-area solar irradiance capturing. Furthermore, the proposed approach can also be used as an automatic and multi-functional platform for other image regression projects such as tidal and geothermal power estimation.

After that, a new stacking model is built for renewable energy forecasting. The models aims to set up an efficient way to do the forecasting work of different kinds of renewable energy. It has been demonstrated that a reasonable stacking model can achieve a higher accuracy in forecasting. Using more artificial intelligence in the research of power system will be a new trend in the next decades.

Based on the research, application of artificial intelligence has been an active techniques for power system. After significant successes were achieved, it must be a continuous trend in next decades. Considering the industrial use of power system, the problems of improving system reliability and dynamic state estimation are essential to be solved with AI. AI is developed with complex computer tools and used to resolve all the above problems. In power system, many projects consist of different non-feasible requirements. AI is a useful way to solve this kind of problem, such as controlling of frequency and power flow, forecasting work for renewable energy, energy trading; scheduling for reliability of power system, transmission expansion, and reactive power. AI-based technology can be used to closely monitor the consumption of power system equipments in higher accuracy. It is indispensable to achieve a reliable and efficient power supply. A big amount of research has demonstrated that deep learning is an emerging technology with powerful ability in many aspects; however, it also contains many problems such as initial point setting, optimizer selection, and architecture design. In the next step, we will collect a bigger dataset, and focus on improving the training efficiency, reducing the network complexity, and increasing its capability to provide more useful information for power system operation and control.

# References


[1] N. J. Nilsson, *Principles of artificial intelligence*. Morgan Kaufmann, 2014.

[2] R. Ploetz, R. Rusdianasari, and E. Eviliana, "Renewable energy: Advantages and disadvantages," in *Proceeding Forum in Research, Science, and Technology (FIRST) 2016*. Politeknik Negeri Sriwijaya, 2016.

[3] E. Y. Gorodov and V. V. Gubarev, "Analytical review of data visualization methods in application to big data," *Journal of Electrical and Computer Engineering*, vol. 2013, 2013.

[4] A. Almalaq, J. Hao, J. J. Zhang, and F.-Y. Wang, "Parallel building: a complex system approach for smart building energy management," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1452–1461, 2019.

[5] J. Hao, X. Dai, A. Stroder, J. J. Zhang, B. Davidson, M. Mahoor, and N. McClure, "Prediction of a bed-exit motion: Multi-modal sensing approach and incorporation of biomechanical knowledge," in *Signals, Systems and Computers, 2014 48th Asilomar Conference on*. IEEE, 2014, pp. 1747–1751.

[6] J. Hao, X. Dai, Y. Zhang, J. Zhang, and W. Gao, "Distribution locational real-time pricing based smart building control and management," in *2016 North American Power Symposium (NAPS)*, Sept 2016, pp. 1–6.

[7] J. Hao, Y. Gu, Y. Zhang, J. J. Zhang, and D. W. Gao, "Locational marginal pricing in the campus power system at the power distribution level," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, July 2016, pp. 1–5.

[8] Y. Zhang, J. Zhang, W. Gao, X. Zheng, L. Yang, J. Hao, and X. Dai, "Distributed electrical energy systems: Needs, concepts, approaches and vision," *Acta Automatica Sinica*, vol. 43, no. 9, 9 2017.

[9] F. He, J. Hao, X. Dai, J. J. Zhang, J. Wei, and Y. Zhang, "Composite socio-technical systems: A method for social energy systems," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2017, pp. 426–431.

[10] J. J. Zhang, D. W. Gao, Y. Zhang, X. Wang, X. Zhao, D. Duan, X. Dai, J. Hao, and F. Wang, "Social energy: mining energy from the society," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 466–482, 2017.

[11] K. Brodlie, R. A. Osorio, and A. Lopes, "A review of uncertainty in data visualization," in *Expanding the frontiers of visual analytics and visualization*. Springer, 2012, pp. 81–109.

[12] T. J. Overbye and J. D. Weber, "Visualization of power system data," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. IEEE, 2000, pp. 7–pp.

[13] ——, "New methods for the visualization of electric power system information," in *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*. IEEE, 2000, pp. 131–16c.

[14] D. V. Nga, O. H. See, C. Y. Xuen, L. L. Chee *et al.*, "Visualization techniques in smart grid," *Smart Grid and Renewable Energy*, vol. 3, no. 03, p. 175, 2012.

[15] H. Jiang, J. J. Zhang, and D. W. Gao, "Fault localization in smart grid using wavelet analysis and unsupervised learning," in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2012, pp. 386–390.

[16] H. Jiang, J. J. Zhang, A. Hebb, and M. H. Mahoor, "Time-frequency analysis of brain electrical signals for behvior recognition in patients with parkinson's disease," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 1843–1847.

[17] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2014.

[18] H. Jiang, L. Huang, J. J. Zhang, Y. Zhang, and D. W. Gao, "Spatial-temporal characterization of synchrophasor measurement systemsa big data approach for smart grid system situational awareness," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 750–754.

[19] H. Jiang, Y. Zhang, E. Muljadi, J. Zhang, and W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Transactions on Smart Grid*, 2016.

[20] H. Jiang and Y. Zhang, "Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine," in *2016 Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[21] Y. Gu, H. Jiang, Y. Zhang, J. J. Zhang, T. Gao, and E. Muljadi, "Knowledge discovery for smart grid operation, control, and situation awarenessa big data visualization platform," in *2016 North American Power Symposium (NAPS)*. IEEE, 2016, pp. 1–6.

[22] F. Ding, H. Jiang, and J. Tan, "Automatic distribution network reconfiguration: An event-driven approach," in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[23] R. Yang, H. Jiang, and Y. Zhang, "Short-term state forecasting-based optimal voltage regulation in distribution systems: Preprint," NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States)), Tech. Rep., 2017.

[24] H. Jiang, Y. Li, Y. Zhang, J. J. Zhang, D. W. Gao, E. Muljadi, and Y. Gu, "Big data-based approach to detect, locate, and enhance the stability of an unplanned microgrid islanding," *Journal of Energy Engineering*, vol. 143, no. 5, p. 04017045, 2017.

[25] H. Jiang, F. Ding, and Y. Zhang, "Short-term load forecasting based automatic distribution network reconfiguration: Preprint," National Renewable Energy Laboratory (NREL), Golden, CO (United States), Tech. Rep., 2017.

[26] Y. Zhang, R. Yang, K. Zhang, H. Jiang, and J. J. Zhang, "Consumption behavior analytics-aided energy forecasting and dispatch," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 59–63, 2017.

[27] Y. Gu, H. Jiang, J. J. Zhang, Y. Zhang, E. Muljadi, and F. J. Solis, "Load forecasting based distribution system network reconfiguration-a distributed data-driven approach," *arXiv preprint arXiv:1711.10690*, 2017.

[28] ——, "Chance-constrained day-ahead hourly scheduling in distribution system operation," *arXiv preprint arXiv:1711.10687*, 2017.

[29] H. Jiang, Y. Zhang, Y. Chen, C. Zhao, and J. Tan, "Power-traffic coordinated operation for bi-peak shaving and bi-ramp smoothing-a hierarchical data-driven approach," *arXiv preprint arXiv:1712.00211*, 2017.

[30] Y. Gu, H. Jiang, Y. Zhang, and D. W. Gao, "Statistical scheduling of economic dispatch and energy reserves of hybrid power systems with high renewable energy penetration," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 530–534.

[31] H. Jiang, J. J. Zhang, D. W. Gao, Y. Zhang, and E. Muljadi, "Synchrophasor based auxiliary controller to enhance power system transient voltage stability in a high penetration renewable energy scenario," in , *2014 IEEE Symposium Power Electronics and Machines for Wind and Water Applications (PEMWA)*. IEEE, 2014, pp. 1–7.

[32] H. Jiang, Y. Zhang, J. J. Zhang, D. W. Gao, and E. Muljadi, "Synchrophasor-based auxiliary controller to enhance the voltage stability of a distribution system with high renewable energy penetration," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 2107–2115, 2015.

[33] M. K. Banavar, J. J. Zhang, B. Chakraborty, H. Kwon, Y. Li, H. Jiang, A. Spanias, C. Tepedelenlioglu, C. Chakrabarti, and A. Papandreou-Suppappola, "An overview of recent advances on distributed and agile sensing algorithms and implementation," *Digital Signal Processing*, vol. 39, pp. 1–14, 2015.

[34] H. Jiang, Y. Zhang, J. J. Zhang, and E. Muljadi, "PMU-aided voltage security assessment for a wind power plant," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.

[35] A. Bose, "Smart transmission grid applications and their supporting infrastructure," *IEEE Transcations on Smart Grid*, vol. 1, pp. 11–19, 2010.

[36] J. Hao, X. Dai, A. Stroder, J. J. Zhang, B. Davidson, M. Mahoor, and N. McClure, "Prediction of a bed-exit motion: Multi-modal sensing approach and incorporation of biomechanical knowledge," in *2014 48th Asilomar Conference on Signals, Systems and Computers*,. IEEE, 2014, pp. 1747–1751.

[37] X. Dai, Z. Zhou, J. J. Zhang, and B. Davidson, "Ultra-wideband radar based human body landmark detection and tracking with biomedical constraints for human motion measuring," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 1752–1756.

[38] H. Jiang, “Synchrophasor sensing and processing based smart grid security assessment for renewable energy integration,” 2015.

[39] F. He, Y. Gu, J. Hao, J. J. Zhang, J. Wei, and Y. Zhang, “Joint real-time energy and demand-response management using a hybrid coalitional-noncooperative game,” in *2015 49th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2015, pp. 895–899.

[40] Y. Gu, H. Jiang, Y. Zhang, and D. W. Gao, “Statistical scheduling of economic dispatch and energy reserves of hybrid power systems with high renewable energy penetration,” in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 530–534.

[41] Y. Song, G. Zhou, and Y. Zhu, “Present status and challenges of big data processing in smart grid,” *Power System Technology*, vol. 37, no. 4, pp. 927–935, 2013.

[42] D. Vesset, B. Woo, H. D. Morris, R. L. Villars, G. Little, J. S. Bozman, L. Borovick, C. W. Olofson, S. Feldman, S. Conway *et al.*, “Worldwide big data technology and services 2012–2015 forecast,” *IDC Report*, vol. 233485, 2012.

[43] M. Kezunovic, L. Xie, and S. Grijalva, “The role of big data in improving power system operation and protection,” in *Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*. IEEE, 2013, pp. 1–9.

[44] H. Jiang, Y. Zhang, J. J. Zhang, D. W. Gao, and E. Muljadi, “Synchrophasor-based auxiliary controller to enhance the voltage stability of a distribution system with high renewable energy penetration,” *IEEE Transactions on Smart Grid*, vol. 6, pp. 2107–2115, 2015.

[45] M. K. Banavar, J. J. Zhang, B. Chakraborty, H. Kwon, Y. Li, H. Jiang, A. Spanias, C. Tepedelenlioglu, C. Chakrabarti, and A. Papandreou-Suppappola, “An overview

of recent advances on distributed and agile sensing algorithms and implementation," *Digital Signal Processing*, vol. 39, pp. 1–14, 2015.

[46] H. Jiang, L. Huang, J. J. Zhang, Y. Zhang, and D. W. Gao, "Spatial-temporal characterization of synchrophasor measurement systems-a big data approach for smart grid system situational awareness," *2014 Conference Record of the 48th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 750 – 754, 2014.

[47] H. Karau, A. Konwinski, P. Wendell, and M. Zaharia, *Learning Spark: Lightning-Fast Big Data Analysis*. " O'Reilly Media, Inc.", 2015.

[48] C. Lam, *Hadoop in action*. Manning Publications Co., 2010.

[49] M. N. K. Boulos, "Web gis in practice iii: creating a simple interactive map of england's strategic health authorities using google maps api, google earth kml, and msn virtual earth map control," *International Journal of Health Geographics*, vol. 4, no. 1, p. 1, 2005.

[50] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on power systems*, vol. 16, no. 1, pp. 44–55, 2001.

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

[52] N.-H. Liao, Z.-H. Hu, Y.-Y. Ma, and W.-Y. Lu, "Review of the short-term load forecasting methods of electric power system," *Power system protection and control*, vol. 39, no. 1, pp. 147–152, 2011.

[53] A. R. Khan, A. Mahmood, A. Safdar, Z. A. Khan, and N. A. Khan, "Load forecasting, dynamic pricing and dsm in smart grid: A review," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 1311–1322, 2016.

[54] F. Ding, H. Jiang, and J. Tan, "Automatic distribution network reconfiguration: An event-driven approach," in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[55] Y. Luo, M. Panwar, M. Mohanpurkar, and R. Hovsapian, "Real time optimal control of supercapacitor operation for frequency response," in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[56] L. Bai, F. Li, H. Cui, T. Jiang, H. Sun, and J. Zhu, "Interval optimization based operating strategy for gas-electricity integrated energy systems considering demand response and wind uncertainty," *Applied Energy*, vol. 167, pp. 270–279, 2016.

[57] X. Fang, Q. Hu, F. Li, B. Wang, and Y. Li, "Coupon-based demand response considering wind power uncertainty: A strategic bidding model for load serving entities," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1025–1037, 2016.

[58] Y. Zhang, R. Yang, K. Zhang, H. Jiang, and J. J. Zhang, "Consumption behavior analytics-aided energy forecasting and dispatch," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 59–63, 2017.

[59] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 422–433, 2015.

[60] M. Cui, J. Zhang, Q. Wang, V. Krishnan, and B.-M. Hodge, "A data-driven methodology for probabilistic wind power ramp forecasting," *IEEE Transactions on Smart Grid*, 2017.

[61] X. Zhu, J. Yan, and N. Lu, "A probabilistic-based pv and energy storage sizing tool for residential loads," in *Transmission and Distribution Conference and Exposition (T&D), 2016 IEEE/PES*. IEEE, 2016, pp. 1–5.

[62] Y. Chen and J. Borken-Kleefeld, "No x emissions from diesel passenger cars worsen with age," *Environmental science & technology*, vol. 50, no. 7, pp. 3327–3332, 2016.

[63] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[64] H. Jiang, Y. Zhang, E. Muljadi, J. Zhang, and W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Transactions on Smart Grid*, 2016.

[65] Y. Chen and J. Borken-Kleefeld, "Real-driving emissions from cars and light commercial vehicles–results from 13 years remote sensing at zurich/ch," *Atmospheric Environment*, vol. 88, pp. 157–164, 2014.

[66] Y. Jiang, J. Xu, Y. Sun, C. Wei, J. Wang, D. Ke, X. Li, J. Yang, X. Peng, and B. Tang, "Day-ahead stochastic economic dispatch of wind integrated power system considering demand response of residential hybrid energy system," *Applied Energy*, vol. 190, pp. 1126–1137, 2017.

[67] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets." *HotCloud*, vol. 10, no. 10-10, p. 95, 2010.

[68] Y. Gu, H. Jiang, Y. Zhang, J. J. Zhang, T. Gao, and E. Muljadi, "Knowledge discovery for smart grid operation, control, and situation awarenessa big data visualization platform," in *North American Power Symposium (NAPS), 2016*. IEEE, 2016, pp. 1–6.

[69] W. Yan, W. Gao, T. Gao, D. W. Gao, S. Yan, and J. Wang, "Distributed cooperative control of virtual synchronous generator based microgrid," 2017.

[70] H. Jiang, L. Huang, Y. Zhang, J. J. Zhang, and D. W. Gao, "Spatial-temporal characterization of synchrophasor measurement systems - A big data approach for smart

grid system situational awareness," in *2014 Conference Record of the Forty Eighth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, 2014, pp. 750–754.

[71] M. N. Faqiry, A. K. Zarabie, F. Nassery, H. Wu, and S. Das, "A day ahead market energy auction for distribution system operation," in *Proceedings of the IEEE Conference on Electro-Information Technology, Lincoln, NE, USA*, 2017, pp. 14–17.

[72] P. Siano and D. Sarno, "Assessing the benefits of residential demand response in a real time distribution energy market," *Applied Energy*, vol. 161, pp. 533–551, 2016.

[73] A. Shahsavari, A. Sadeghi-Mobarakeh, E. M. Stewart, E. Cortez, L. Alvarez, F. Megala, and H. Mohsenian-Rad, "Distribution grid reliability versus regulation market efficiency: An analysis based on micro-pmu data," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2916–2925, 2017.

[74] H. Wu, I. Krad, A. Florita, B.-M. Hodge, E. Ibanez, J. Zhang, and E. Ela, "Stochastic multi-timescale power system operations with variable wind generation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3325–3337, 2017.

[75] S. Choi, "Practical coordination between day-ahead and real-time optimization for economic and stable operation of distribution systems," *IEEE Transactions on Power Systems*, 2017.

[76] H. Farzin, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "A stochastic multi-objective framework for optimal scheduling of energy storage systems in microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 117–127, 2017.

[77] F. S. Gazijahani, H. Hosseinzadeh, N. Tagizadeghan, and J. Salehi, "A new point estimate method for stochastic optimal operation of smart distribution systems considering demand response programs," in *Electrical Power Distribution Networks Conference (EPDC), 2017 Conference on*. IEEE, 2017, pp. 39–44.

[78] K. Baker, G. Hug, and X. Li, "Energy storage sizing taking into account forecast uncertainties and receding horizon operation," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 331–340, 2017.

[79] D. Ke, C. Chung, and Y. Sun, "A novel probabilistic optimal power flow model with uncertain wind power generation described by customized gaussian mixture model," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 200–212, 2016.

[80] F. Pernkopf and D. Bouchaffra, "Genetic-based em algorithm for learning gaussian mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1344–1348, 2005.

[81] H. Wu, M. Shahidehpour, and A. Al-Abdulwahab, "Hourly demand response in day-ahead scheduling for managing the variability of renewable energy," *IET Generation, Transmission & Distribution*, vol. 7, no. 3, pp. 226–234, 2013.

[82] Z. Zhu, D. Liu, Q. Liao, F. Tang, J. J. Zhang, and H. Jiang, "Optimal power scheduling for a medium voltage ac/dc hybrid distribution network," *Sustainability*, vol. 10, no. 2, p. 318, 2018.

[83] P. Karlsson and J. Svensson, "Dc bus voltage control for a distributed power system," *IEEE Transactions on Power Electronics*, vol. 18, no. 6, pp. 1405–1412, 2003.

[84] J.-Y. Kim, K.-J. Mun, H.-S. Kim, and J. H. Park, "Optimal power system operation using parallel processing system and pso algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 33, no. 8, pp. 1457–1461, 2011.

[85] Z. Tang and K. K. Bagchi, "Globally convergent particle swarm optimization via branch-and-bound," *Computer and Information Science*, vol. 3, no. 4, p. 60, 2010.

[86] M. T. Hagan and S. M. Behr, "The time series approach to short term load forecasting," *IEEE transactions on power systems*, vol. 2, no. 3, pp. 785–791, 1987.

[87] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Transactions on power systems*, vol. 16, no. 3, pp. 498–505, 2001.

[88] J.-F. Chen, W.-M. Wang, and C.-M. Huang, "Analysis of an adaptive time-series autoregressive moving-average (arma) model for short-term load forecasting," *Electric Power Systems Research*, vol. 34, no. 3, pp. 187–196, 1995.

[89] M. Hamlich, N. eddine Belbounaguia *et al.*, "Short-term load forecasting using machine learning and periodicity decomposition," 2019.

[90] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "A deep learning framework for short-term power load forecasting," *arXiv preprint arXiv:1711.11519*, 2017.

[91] E. Feeder, "J1, distributed pv (dpv) monitoring and feeder analysis," 2013.

[92] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets." *HotCloud*, vol. 10, pp. 10–10, 2010.

[93] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[94] N. Chadil, A. Russameesawang, and P. Keeratiwintakorn, "Real-time tracking management system using gps, gprs and google earth," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, vol. 1. IEEE, 2008, pp. 393–396.

[95] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact." *MIS quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.

[96] W. H. Greene, *Econometric analysis*. Pearson Education India, 2003.

[97] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[98] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, no. Feb, pp. 281–305, 2012.

[99] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems*, 2011, pp. 2546–2554.

[100] Q. Peng and S. H. Low, "Distributed algorithm for optimal power flow on a radial network," in *2014 IEEE 53rd Annual Conference on Decision and Control (CDC)*. IEEE, 2014, pp. 167–172.

[101] S. H. Low, "Convex relaxation of optimal power flowpart i: Formulations and equivalence," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[102] F. Ding and K. A. Loparo, "Hierarchical decentralized network reconfiguration for smart distribution systemspart i: Problem formulation and algorithm development," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 734–743, 2015.

[103] K. G. Boroojeni, M. H. Amini, S. Bahrami, S. Iyengar, A. I. Sarwat, and O. Karabasoglu, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon," *Electric Power Systems Research*, vol. 142, pp. 58–73, 2017.

[104] W. Peng, "Model selection for gaussian mixture model based on desirability level criterion," *Optik-International Journal for Light and Electron Optics*, vol. 130, pp. 797–805, 2017.

[105] G. McLachlan and D. Peel, *Finite mixture models*. John Wiley & Sons, 2004.

[106] Q. Peng and S. H. Low, "Distributed algorithm for optimal power flow on a radial network," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 167–172.

[107] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *Power Systems Computation Conference (PSCC), 2014*. IEEE, 2014, pp. 1–9.

[108] Q. Peng and S. H. Low, "Distributed algorithm for optimal power flow on an unbalanced radial network," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 6915–6920.

[109] B.-M. Hodge and M. Milligan, "Wind power forecasting error distributions over multiple timescales," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–8.

[110] H. Heinz, "Parallel genetic algorithms, population genetics and combinatorial optimization," *Parallelism, Learning, Evolution*, pp. 398–406, 1991.

[111] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[112] D. Lawrence, *Handbook of genetic algorithms*, 1991.

[113] A. G. Bakirtzis, P. N. Biskas, C. E. Zoumas, and V. Petridis, "Optimal power flow by enhanced genetic algorithm," *IEEE Transactions on power Systems*, vol. 17, no. 2, pp. 229–236, 2002.

[114] M. Osman, M. A. Abo-Sinna, and A. Mousa, "A solution to the optimal power flow using genetic algorithm," *Applied mathematics and computation*, vol. 155, no. 2, pp. 391–405, 2004.

[115] M. Cui, J. Zhang, A. Florita, B.-M. Hodge, D. Ke, and Y. Sun, "Solar power ramp events detection using an optimized swinging door algorithm," in *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers Digital Collection, 2015.

[116] H. Jiang, Y. Li, Y. Zhang, J. J. Zhang, D. W. Gao, E. Muljadi, and Y. Gu, "Big data-based approach to detect, locate, and enhance the stability of an unplanned microgrid islanding," *Journal of Energy Engineering*, vol. 143, no. 5, p. 04017045, 2017.

[117] X. Fang, Q. Hu, F. Li, B. Wang, and Y. Li, "Coupon-based demand response considering wind power uncertainty: A strategic bidding model for load serving entities," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1025–1037, 2015.

[118] Y. Luo, M. Andrus, J. Lian, S. K. Srivastava, and D. Cartes, "Power impact analysis of pulse power loads in the future integrated shipboard power system," in *ASNE Electric Electric Machines Technology Symposium*, 2012.

[119] X. Geng and L. Xie, "A data-driven approach to identifying system pattern regions in market operations," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.

[120] W. Yan, W. Gao, T. Gao, D. W. Gao, S. Yan, and J. Wang, "Distributed cooperative control of virtual synchronous generator based microgrid," in *2017 IEEE International Conference on Electro Information Technology (EIT)*. IEEE, 2017, pp. 506–511.

[121] G. M. Stokes and S. E. Schwartz, "The atmospheric radiation measurement (arm) program: Programmatic background and design of the cloud and radiation test bed," *Bulletin of the American Meteorological Society*, vol. 75, no. 7, pp. 1201–1222, 1994.

[122] T. Stoffel, I. Reda, D. Myers, D. Renne, S. Wilcox, and J. Treadwell, "Current issues in terrestrial solar radiation instrumentation for energy, climate, and space applications," *Metrologia*, vol. 37, no. 5, p. 399, 2000.

[123] Y. Xie, M. Sengupta, and J. Dudhia, "A fast all-sky radiation model for solar applications (farms): Algorithm and performance evaluation," *Solar Energy*, vol. 135, pp. 435–445, 2016.

[124] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.

[125] A. Mark Nixon, "Feature extraction & image processing for computer vision," 2012.

[126] Y. Gu, H. Jiang, J. J. Zhang, Y. Zhang, H. Wu, and E. Muljadi, "Multi-timescale three-phase unbalanced distribution system operation with variable renewable generations," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4497–4507, 2018.

[127] M. Pipattanasomporn, H. Feroze, and S. Rahman, "Multi-agent systems in a distributed smart grid: Design and implementation," in *2009 IEEE/PES Power Systems Conference and Exposition*. IEEE, 2009, pp. 1–8.

[128] H. Jiang, J. J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transactions on Smart Grid*, vol. 5, pp. 2947 – 2956, 2014.

[129] X. Geng and L. Xie, "Learning the lmp-load coupling from data: A support vector machine based approach," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1127–1138, 2016.

[130] Q. Lu and K.-D. Kim, "A mixed integer programming approach for autonomous and connected intersection crossing traffic control," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–6.

[131] H. Jiang, X. Dai, D. W. Gao, J. J. Zhang, Y. Zhang, and E. Muljadi, "Spatial-temporal synchrophasor data characterization and analytics in smart grid fault detection, identification, and impact causal analysis," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2525–2536, 2016.

[132] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[133] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[134] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[135] T. D. Nielsen and F. V. Jensen, *Bayesian networks and decision graphs*. Springer Science & Business Media, 2009.

[136] D. M. Blei, M. I. Jordan *et al.*, "Variational inference for dirichlet process mixtures," *Bayesian analysis*, vol. 1, no. 1, pp. 121–143, 2006.

[137] K. Yoshii and M. Goto, "A nonparametric bayesian multipitch analyzer based on infinite latent harmonic allocation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 717–730, 2011.

[138] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing clusters among related groups: Hierarchical dirichlet processes," in *Advances in neural information processing systems*, 2005, pp. 1385–1392.

[139] Y. Gu, H. Jiang, Y. Zhang, and D. W. Gao, "Statistical scheduling of economic dispatch and energy reserves of hybrid power systems with high renewable energy penetration," in *2014 48th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2014, pp. 530–534.

[140] Q. Lu and K.-D. Kim, "Intelligent intersection management of autonomous traffic using discrete-time occupancies trajectory," *Journal of Traffic and Logistics Engineering Vol*, vol. 4, no. 1, pp. 1–6, 2016.

[141] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Applied Energy*, vol. 190, pp. 1245–1257, 2017.

[142] M. Cui, J. Zhang, A. R. Florita, B.-M. Hodge, D. Ke, and Y. Sun, "An optimized swinging door algorithm for identifying wind ramping events," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 150–162, 2015.

[143] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar, "Fully decentralized multi-agent reinforcement learning with networked agents," *arXiv preprint arXiv:1802.08757*, 2018.

[144] K. Zhang, A. Koppel, H. Zhu, and T. Başar, "Global convergence of policy gradient methods to (almost) locally optimal policies," *arXiv preprint arXiv:1906.08383*, 2019.

[145] Q. Lu and K.-D. Kim, "Autonomous and connected intersection crossing traffic management using discrete-time occupancies trajectory," *Applied Intelligence*, vol. 49, no. 5, pp. 1621–1635, 2019.

[146] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[147] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[148] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[149] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[150] Y. Zhang, R. Yang, K. Zhang, H. Jiang, and J. J. Zhang, "Consumption behavior analytics-aided energy forecasting and dispatch," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 59–63, 2017.

[151] J. Zhao, G. Zhang, K. Das, G. N. Korres, N. M. Manousakis, A. K. Sinha, and Z. He, "Power system real-time monitoring by using pmu-based robust state estimation method," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 300–309, 2015.

[152] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended kalman filter for power system dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205–3216, 2016.

[153] F. Ding, H. Jiang, and J. Tan, "Automatic distribution network reconfiguration: An event-driven approach," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[154] D. Görür and C. E. Rasmussen, "Dirichlet process gaussian mixture models: Choice of the base distribution," *Journal of Computer Science and Technology*, vol. 25, no. 4, pp. 653–664, 2010.

[155] B. A. Frigyik, A. Kapila, and M. R. Gupta, "Introduction to the dirichlet distribution and related processes," *Department of Electrical Engineering, University of Washignton, UWEETR-2010-0006*, no. 0006, pp. 1–27, 2010.

[156] C. Bishop, "Pattern recognition and machine learning (information science and statistics)(springer-verlag new york, inc., secaucus, nj, usa, 2006)."

[157] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[158] X. Zhou, M. Farivar, and L. Chen, "Pseudo-gradient based local voltage control in distribution networks," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 173–180.

[159] R. Marquez and C. F. Coimbra, "Intra-hour dni forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, 2013.

[160] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[161] A. Dai, T. R. Karl, B. Sun, and K. E. Trenberth, "Recent trends in cloudiness over the united states: A tale of monitoring inadequacies," *Bulletin of the American Meteorological Society*, vol. 87, no. 5, pp. 597–606, 2006.

[162] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[163] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[164] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[165] J. Kleissl, *Solar energy forecasting and resource assessment*. Academic Press, 2013.

[166] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013.

[167] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, vol. 168, pp. 60–101, 2018.

[168] C.-L. Fu and H.-Y. Cheng, "Predicting solar irradiance with all-sky image features via regression," *Solar Energy*, vol. 97, pp. 537–550, 2013.

[169] R. Marquez and C. F. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database," *Solar Energy*, vol. 85, no. 5, pp. 746–756, 2011.

[170] S. D. Miller, M. A. Rogers, J. M. Haynes, M. Sengupta, and A. K. Heidinger, "Short-term solar irradiance forecasting via satellite/model coupling," *Solar Energy*, vol. 168, pp. 102–117, 2018.

[171] A. Sharma and A. Kakkar, "Forecasting daily global solar irradiance generation using machine learning," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2254–2269, 2018.

[172] T. Zhu, L. Xie, H. Wei, H. Wang, X. Zhao, and K. Zhang, "Clear-sky direct normal irradiance estimation based on adjustable inputs and error correction," *Journal of Renewable and Sustainable Energy*, vol. 11, no. 5, p. 056101, 2019.

[173] K. Lindberg, P. Seljom, H. Madsen, D. Fischer, and M. Korpås, "Long-term electricity load forecasting: Current and future trends," *Utilities Policy*, vol. 58, pp. 102–119, 2019.

[174] W.-J. Lee and J. Hong, “A hybrid dynamic and fuzzy time series model for mid-term power load forecasting,” *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 1057–1062, 2015.

[175] K.-B. Song, Y.-S. Baek, D. H. Hong, and G. Jang, “Short-term load forecasting for the holidays using fuzzy linear regression method,” *IEEE transactions on power systems*, vol. 20, no. 1, pp. 96–101, 2005.

[176] D. Cao, Z. Lei, Z. Zhang, J. Feng, and S. Z. Li, “Human age estimation using ranking svm,” in *Chinese Conference on Biometric Recognition*. Springer, 2012, pp. 324–331.

[177] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, “Ordinal regression with multiple output cnn for age estimation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4920–4928.