FORECASTING OF WIND, PV GENERATION, AND MARKET PRICE FOR THE OPTIMAL OPERATIONS OF THE REGIONAL PEV CHARGING STATIONS

by

PIAMPOOM SARIKPRUECK

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

MAY 2015

Acknowledgements

I would like to express my deepest gratitude to my supervisor professor, Dr. Wei-Jen Lee, for his invaluable guidance, support, patience and consistent encouragement throughout my Ph.D. program. Also, I am very grateful for his faith for giving me an incredible opportunity to participate in multiple research works. I have learned not only research skills from him but also his following characters: self-discipline, hardworking, responsibility, and continual self-improving. These will be imprinted in my memory throughout my life.

I would like to extend my grateful appreciation to Dr. William E. Dillon, Dr. Rasool Kenarangui, Dr. Victoria C.P. Chen, Dr. Ali Davoudi and Dr. David A. Wetz for their valuable instruction and for serving as my committee member. Besides, I would like to thanks Dr. Jay M. Rosenberger and Dr. Asama Kulvanitchaiyanunt for their helpful advice and suggestion during working in the PHEV charging infrastructure project.

I wish to acknowledge to Pichit Lumyong for his initiation and guidance for my research skills throughout my M.S. study and academic career at King Mongkut's Institute of Technology Ladkrabang, Thailand.

I also want to deeply appreciate to Royal Thai Government for financial support and giving me the opportunity to pursue my Ph.D. Degree. I also thank all members of Energy Systems Research Center at the University of Texas at Arlington for their kind assistance, administrative work and in sharing memorable time during my study.

Finally, I dedicate my dissertation to my parents and family members for their unconditional love and support. Special thanks should be given to my girlfriend, Chodchanok Attaphong for her love, patience and sacrifice in every part for my Ph.D. study.

March 25, 2015

Abstract

FORECASTING OF WIND, PV GENERATION, AND MARKET PRICE FOR THE OPTIMAL OPERATIONS OF THE REGIONAL PEV CHARGING STATIONS

Piampoom Sarikprueck, PhD

The University of Texas at Arlington, 2015

Supervising Professor: Wei-Jen Lee

The transition from the conventional spark ignition engine vehicles to the electric vehicular technologies helps reduce greenhouse gas (GHG) emissions as well as improve the energy efficiency in the transportation sector. In the transformation of the electric vehicle, the hybrid electric vehicle (HEV) has evolved into the plug-in electric vehicle (PEV) due to the advancement in battery technologies that extend the electric driving distance of vehicles; however, this trend also creates concern among PEV users about how long or how far they might travel per battery charge.

A well-planned charging infrastructure with a fast (level 3) charging station is critical to overcome the range anxiety of PEV users, which can then promote the deployment and public acceptance of PEV. In addition, the PEV charging station must be considered from a regional point of view, especially in terms of operation optimization and support for the high penetration of PEVs in metro areas. Integrating renewable energy sources such as wind and solar PV power generation with electricity from the grid into PEV charging stations is critical for sustainable future development. A PEV charging station with a distributed energy storage system will be able to participate in the deregulated market to support the power system and optimize its operational cost. However, sufficient accuracy in the forecasting of energy sources and market prices are prerequisite to achieving the above mentioned benefits and goals.

Using the Dallas/Fort Worth (DFW) as an example, this dissertation develops novel approaches for the wind/PV generation and market price predictions. These predictions are calculated every 15 minutes (15-minute ahead prediction) for the following 15-minute settlement interval set by the Electric Reliability Council of Texas (ERCOT) market. Support Vector Classification (SVC) and Support Vector Regression (SVR) of Support Vector Machines (SVMs) are adopted for the prediction of categorical and continuous values, respectively.

SVR is used to predict the wind/PV generation because they are considered continuous functions. The validations of the estimation performance for these two predictions are illustrated using the wind power data from a wind farm in Oklahoma (a virtual wind farm for this study) and the PV generation from Dallas Redbird airport, respectively. The proposed method improves the forecasting performance of both predictions compared to the persistence model.

In order to achieve accurate market price predictions in the deregulated market, a hybrid market price forecasting method (HMPFM) including SVC and SVR with data clustering techniques is proposed. SVC is adopted to predict spike price occurrence, and SVR is used for market price magnitude prediction of both non-spike and spike prices. Additionally, three clustering techniques including Classification and Regression Trees, K-means, and Stratification methods are introduced to mitigate the higher error of spike magnitude estimation. The performance of the proposed hybrid method is validated with the ERCOT wholesale market price. The results from the proposed method show significant improvement over typical approaches.

To fulfill the comprehensive study, the characteristics of the forecast uncertainty have to be investigated to understand their stochastic nature for optimizing the benefits of operating PEV charging stations. In this dissertation, the Martingale Model Forecast

Evolution (MMFE) is used for the investigation, since it explores the multivariate random vector of the forecast change, which can apply to the multivariate case in this problem. Finally, the results show the effectiveness of the MMFE to generate the stochastic nature of the proposed predictions.

Table of Contents

List of Illustrations

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Energy is a prime mover of economic growth and is also one of the key factors in the development of a country. Over the years, the demand for energy has increased due to industrialization, modernization, and the growth of population. However, these issues have created environmental concerns about greenhouse gas (GHG) emissions, global warming, and climate changes. Considering the increased demand for energy and the concerns it creates, considerable attention has been paid to investigating solutions to energy and environmental issues.

In terms of energy, one study from Energy Information Administration (EIA) indicates that the transportation sector consumes about 28 % of the total energy consumption among various end-use sectors in the U.S., the second highest behind electric power, as illustrated in Figure 1-1 [1]. Most of the energy supply for the transportation sector is petroleum. Transportation's dependence on petroleum has led to a developing concern in this sector due to the limited availability of crude oil and the fluctuations in oil price. Hence, a sustained effort is needed in the transportation sector to overcome its dependence on petroleum.

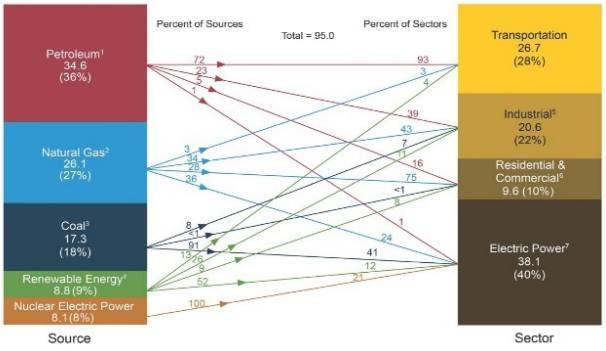


Figure 1-1 U.S. energy source and energy consumption by sector

In terms of GHG emissions, the U.S. government pledges to reduce its domestic GHG emissions by approximately 17 % by 2020 [2]. According to a report from the Environmental Protection Agency (EPA) [3], the transportation sector, as the second largest source of emissions, contributes up to 27 % of the U.S. GHG emissions, as depicted in Figure 1-2. Consequently, the transportation sector is highly motivated to reduce its GHG emissions.

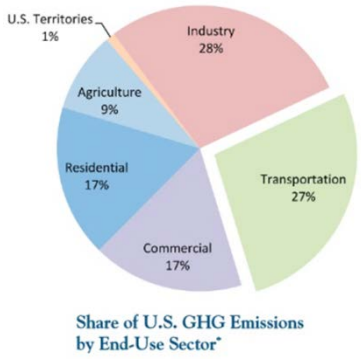


Figure 1-2 U.S. GHG emissions by end-use sector

As an evolving remedy for both the critical energy supply and environmental problems, a fundamental transformation from conventional oil-based vehicles to electrical powered ones has been proposed and is being implemented. The electrical powered vehicles can be classified into two categories: Hybrid Electric Vehicle (HEV) and Plug-in Electric Vehicle (PEV). Developed in the 1990s, the HEV uses a small electric battery to supplement a standard internal combustion engine and increase fuel efficiency by about 25 % over the conventional light-duty vehicles. Currently, the evolution of the electric vehicle has rapidly changed from HEV to PEV. The PEV can be categorized into two groups consisting of Plug-in Hybrid Electric Vehicle (PHEV) and Battery Electric Vehicle (BEV). The PHEV is one stage advanced from the HEV with two additional improvements, the increased battery size and the battery recharging capability from the power grid by adding a plug. Therefore, the PHEV can be driven using electricity farther

than the traditional HEV. In contrast to the PHEV, the BEV has no internal combustion engine, so this type of vehicle is fully driven by the electricity and must be plugged into the electric power grid for recharging.

Table 1-1 shows the comparison of the total energy efficiency of several types of vehicles [4]. As depicted in this table, the energy efficiency of electric vehicles including PHEV and BEV has been improved over the conventional oil-based vehicles. The energy efficiency of the BEV is more than double that of the diesel engine vehicle.

Table 1-1 Total Energy Efficiency of Several Types of Vehicles

Technology	Total Energy Efficiency(km/MJ)
Diesel engine	0.48
Gasoline engine	0.51
PHEV	0.64
BEV	1.14

According to a study from MIT, by 2035 all electric vehicles should be able to significantly reduce their life cycle GHG emissions compared to the traditional spark ignition engine (SIE) vehicles [5]. The researchers consider GHG emissions of the vehicles in three stages. In the first stage, the emissions are produced from the vehicle's material production process. In the second stage, which is well-to-tank, the emissions are released from the energy resource generation. For example, the high GHG emissions of the BEV in this stage comes from the electricity power plant that produces the electricity for recharging the BEV. In the last stage, which is tank-to-wheel, the emissions are generated from the driving engine. The BEV and Fuel Cell vehicle (FCV) have no GHG emissions in this stage. This study's results are illustrated in Figure 1-3.

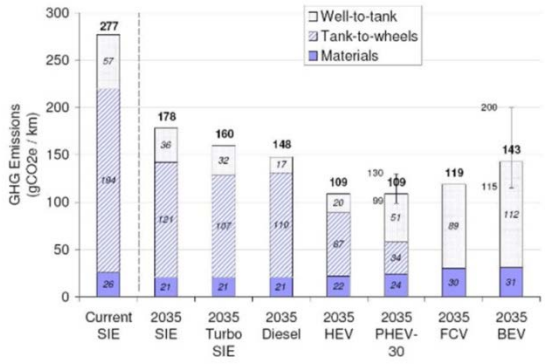
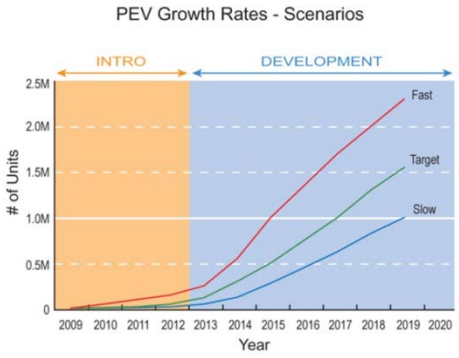


Figure 1-3 Life-cycle GHG emissions of vehicles in 2035

Currently, the PEV is promoted in the U.S. to improve the energy efficiency and reduce the GHG emissions of the transportation sector. The projection from the ISO/RTO Council (IRC) on the PEV growth rate under different scenarios and its penetration into metro areas are shown in Figure 1-4 [6]. At the target growth rate scenario, the amount of PEVs will be more than one million vehicles in the U.S. by 2017. Based on the same projection, there will be approximately 10,000 PEVs in the DFW metro area in this same time frame.



City	Consumer PEVs	Fleet PEVs	Total PEVs
New York	40,000	14,069	54,069
Los Angeles	105,000	14,069	119,069
Chicago	20,000	7,892	27,892
Washington, DC	31,000	6,520	37,520
San Francisco	85,000	6,005	91,005
Philadelphia	13,000	5,319	18,319
Boston	27,000	4,976	31,976
Detroit-Ann Arbor	6,000	4,718	10,718
Dallas-Fort Worth	6,500	4,461	10,961
Houston	8,000	4,032	12,032
Atlanta	4,500	3,517	8,017
Miami	8,000	3,346	11,346
Seattle-Tacoma	23,000	3,088	26,088
Phoenix	13,000	2,831	15,831
Minneapolis	8,000	2,574	10,574
Cleveland-Akron	6,000	2,574	8,574
San Diego	20,000	2,445	22,445
St. Louis	3,500	2,230	5,730
Denver-Boulder	9,000	2,230	11,230
Tampa-St. Pete	7,000	2,059	9,059

Note: Metro areas located within the ISO/RTO study are **bold**; other metro areas are in gray

Figure 1-4 PEV growth rate projections in metro areas

To encourage the acceptance of PEVs, PEV users must be able to drive their cars without suffering from range anxiety. A well-planned charging infrastructure plays a critical role in serving this purpose. For improving the development of the charging infrastructure, a study framework for the design of PEV charging stations is proposed in [7], as depicted in Figure 1-5. This framework is an integration design of three important tasks including system simulation, system design, and dynamic control. However, this dissertation only focuses on the system simulation task. The system simulation is a key prerequisite task to provide the important information for system design and dynamic control. The system design aims to minimize the charging station installation cost. The objective of dynamic control is to maximize the operational profit of the charging station. In addition, the PEV charging station system is considered from a regional point of view for global optimization and support for high PEV penetration. Finally, renewable energy resources are incorporated in the charging station system to promote the use of these renewable energies and the abatement of GHG emissions in order to create a sustainable future.

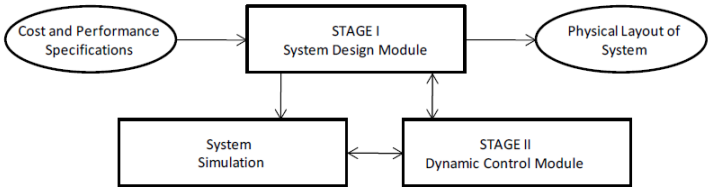


Figure 1-5 The PEV charging station design framework

Considering the above factors, it is important for the system simulation task to provide the configuration of the regional charging stations with the integration of renewable energy resources and the utility grid for the system design. Also, the forecasting of renewable energy resources and power market prices as well as the stochastic nature of these predictions are the critical outputs of the system simulation

task for the dynamic control task. The details of the configuration and critical outputs of the system simulation are described in the following sections.

1.2 The Regional PEV Charging Station Configuration

Current public charging stations rarely consider the installation of energy storage systems or the integration of renewable energy resources. This is because these charging stations are small and local due to the current low PEV demand. However, to support the increasing penetration of PEV users and to promote sustainable energy, the proposed PEV charging station is designed to be equipped with a distributed energy storage system charged by wind/solar PV generation and electricity from the power grid, which can simultaneously charge multiple PEVs.

The proposed distributed energy storage system is used as a buffer for the charging station to alleviate the load strain due to high numbers of PEVs charging, which can defer the need for distribution upgrade if the charging stations have an insufficient renewable energy supply. Also, the proposed system can be used to mitigate the mismatch between the renewable energy resources and the PEVs' demand by storing excessive wind/solar energy for future demand arriving at the station. In addition and very importantly, these proposed systems enable the charging station to participate in the deregulated market.

The participation of a PEV charging station in the deregulated market highlights the benefit of wind and solar energy as well as distributed energy storage systems with optimal operational strategies [8]. However, the way the charging station operates should be determined from a regional point of view to achieve global optimization of the above benefits. In addition, one charging station is insufficient to serve all of the PEV users throughout a metro area. Hence, the configuration of a regional PEV charging station system with n stations is proposed as shown in Figure 1-6.

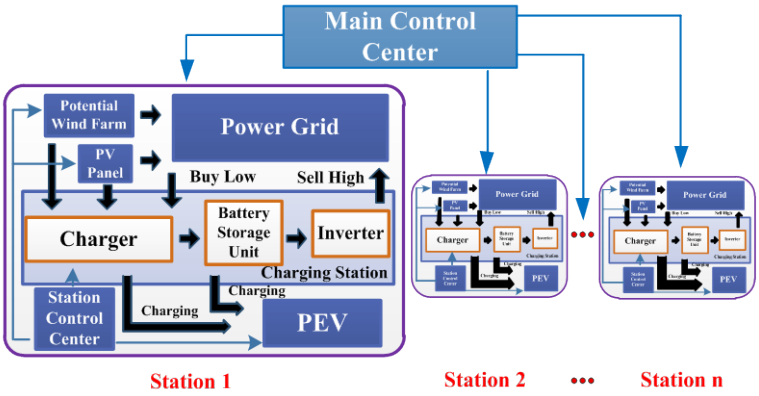


Figure 1-6 Regional PEV charging stations

According to Figure 1-6, all of the electricity from various sources is able to be directly used for charging PEVs, and the surplus can be either stored in the battery or sold back to the power grid. When a PEV arrives at the station, its demands can be served from both the direct charge and the battery storage. As proposed in this design, the global optimization can be achieved with the optimal operation strategies, which highly depend on the available wind/PV energy and the power market price at each charging station location. Details of PEV charging levels, wind/solar PV energy resources, real-time market prices, and battery technologies are described in the following sections.

1.2.1 PEV Charging Levels

According to SAE J1772 Std., the PEV charging infrastructure can be classified into three levels with two different voltage types, consisting of AC level 1, AC level 2, and DC fast charging as shown in Figure 1-7 [9]. PEV users may choose to charge their vehicles from any of these three charging levels based on their available time and driving needs.

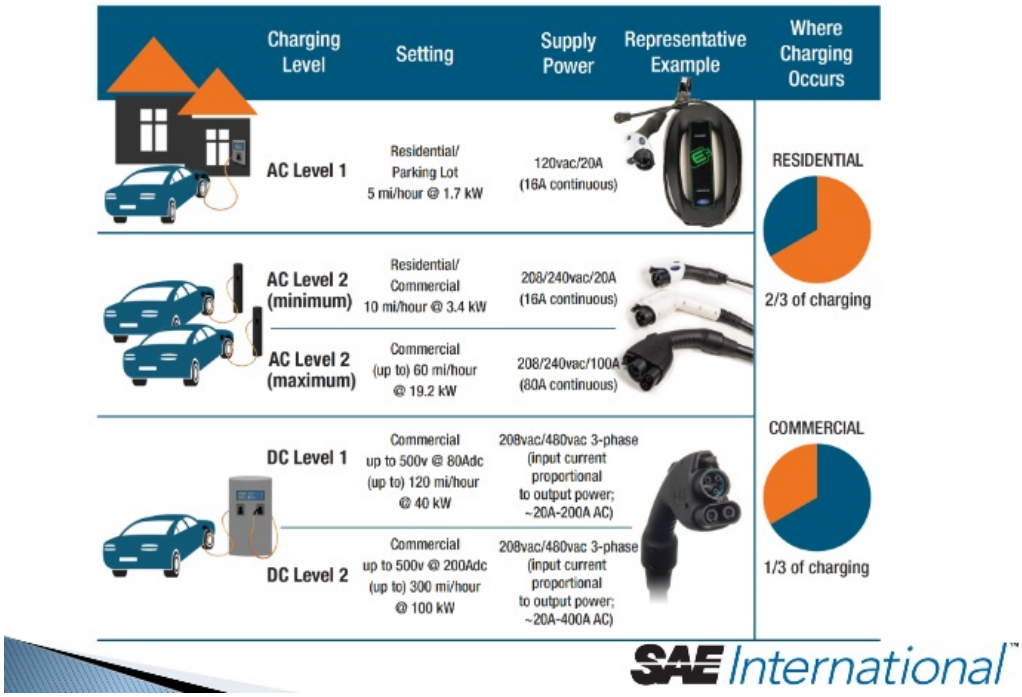


Figure 1-7 PEV Charging levels

1.2.1.1 AC level 1

The lowest common voltage charging level is AC level 1 using a standard 120 V, single phase power with 12 A of continuous maximum current for a 15 A branch circuit breaker or 16 A of continuous maximum current for a 20 A branch circuit breaker. This means that the PEV can be connected to National Electrical Manufacturers Association (NEMA) 5-15 and 5-20 outlets, which are the traditional home plugs in residential and commercial buildings in the U.S., as depicted in Figure 1-8 [10]. The maximum power of this charging level, which is only 1.7 kW, may take a long time to fully charge a PEV, approximately 8 to 15 hours depending on the size of the battery.

1.2.1.2 AC level 2

AC level 2 is a single-phase power with the continuous charging current of 16 A to 80 A from a 240 V outlet. The supplied voltage can be obtained from high power home

appliance voltage sources such as electric clothes driers, electric ovens, or pool pumps. Therefore, because of its fast charging time, AC level 2 is the preferred option for PEV users at home or at currently operating charging points in public areas. Figure 1-9 presents a PEV charging with AC level 2 connected to a high power home appliance plug [10]. An AC level 2 requires a higher level of safety equipment and an onboard charger for charging the PEV. In addition, due to the small onboard charging system in current PEVs, the maximum allowance of charging power is 19.2 kW. Based on this charging power limit, it takes approximately 2 to 4 hours to fully charge the vehicle depending on the size of the battery.

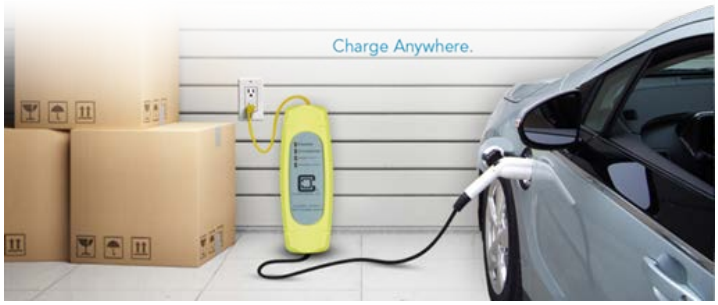


Figure 1-8 AC level 1 charging

Figure 1-9 AC level 2 charging

1.2.1.3 DC fast charging level

DC fast charging stations are designed to perform as commercial gasoline filling stations. Since DC fast charging uses high levels of voltage and current, a high speed of

charging can be achieved. Therefore, it is a suitable technique to be used to establish public charging stations. Figure 1-10 demonstrates the conceptual design of a public DC fast charging station with roof-topped PV panels [11].

DC fast charging includes DC level 1 and DC level 2, and can reach 500 VDC, 40-100 kW maximum power; therefore, a PEV is able to be fully charged in minutes. DC fast charging also requires an off board charger system as AC level 2, but it should be supplied by a DC current directly feeding to the PEV high voltage battery bus. The DC fast charging connecter follows the SAE J1772 standard, as shown in Figure 1-11 [12]. Considering the advantages from the quick charge feature of this charging level, this dissertation focuses on DC fast charging for regional PEV charging stations as a way to better serve the demand throughout the metro area. However, considering the high power consumption of DC fast charging stations, a critical situation in the distribution network may exist. Therefore, this dissertation proposes using energy storages devices in PEV charging stations to mitigate the high load strain in the distribution system, helping to defer the necessity of a distribution upgrade.

Figure 1-10 Public charging station

Figure 1-11 DC fast charging connector

1.2.2 Wind Energy Resource

Wind power has recently developed into the world's fastest growing renewable energy resource. The global wind power installed capacity exceeded 369 GW in 2014 [13]. The moderate scenario from the Global Wind Energy Council projects that the global cumulative wind power capacity could reach 1500 GW by 2030. Among that, the U.S. installed wind power capacity could be as high as 300 GW [14]. This indicates the significant potential for using wind power as one of the major energy sources in charging stations.

In this dissertation, the charging station acquires wind energy from a wind farm based on an energy procurement. Since the wind farm does not have to be on site, this dissertation uses the term "virtual wind farm" to describe the arrangement of purchasing the power from a remote wind farm. However, wind power has the characteristic of intermittency, as depicted in Figure 1-12, which may result in difficulty balancing the supply and demand while maximizing the charging station's financial benefit. To tackle this problem, accurate wind power predictions are necessary for the optimal operation of a regional PEV charging station system. Though it is located in a different market, a wind farm in Oklahoma with a 74.25 MW installation capacity is selected for the wind energy resource in this study. The algorithm proposed in this dissertation is testable because the necessary wind power and wind speed data from this wind farm are available.

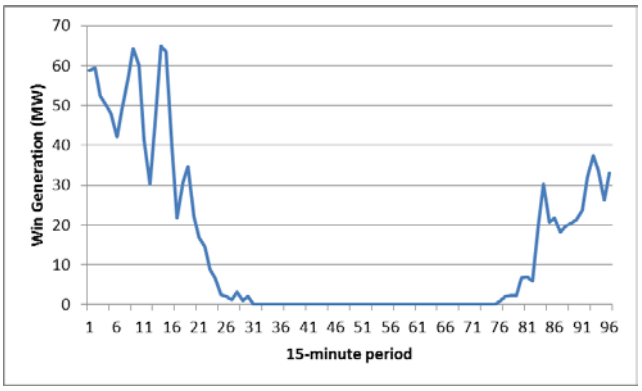


Figure 1-12 Intermittence of wind generation

1.2.3 Solar PV Energy Resource

In 2013, solar Photovoltaic (PV) energy ranked second among all of the renewable energies in the world with 139 GW installed capacity, and it achieved an average annual growth rate of 55 %, the highest of all renewable energies [15]. According to the IEA's highest projection [16], the global solar PV capacity could exceed 1700 GW installation capacity by 2030. The rapid development of solar energy has resulted in a large scale deployment of solar PV in residential, commercial, and utility-scaled sectors for serving the increased electric demand.

In this dissertation, it is assumed that roof-topped solar PV panels are installed on charging stations and supply part of the PEV energy demand. As the highest efficiency option among all of the available PV technologies, the single crystalline PV module is selected for evaluating the PV generation. The solar PV power output varies based on solar radiation and relevant weather. For instance, cloudy skies result in the fluctuation of the PV generation, as illustrated in Figure 1-13 . Also, the solar PV produces zero energy during the night. These factors highlight the need to accurately predict PV generation in order to balance supply and demand so as to optimize the operation of the charging station.

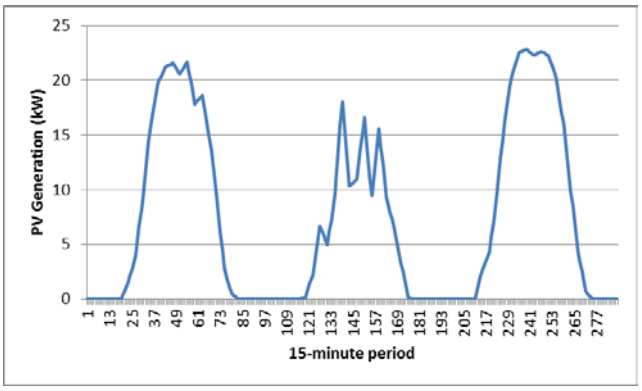


Figure 1-13 Intermittence and variation of PV generation

1.2.4 Wholesale Electricity Price in ERCOT Market

According to the above discussion, the regional PEV charging station system should be established in the metro area. The DFW (Dallas Fort-Worth) metroplex, under the jurisdiction of the Electric Reliability Council of Texas (ERCOT), is selected for the case study in this dissertation. The charging stations are designed to be built nearby the power nodes, which can serve as a Point of Interconnection (POI) of DC fast charging to the power grid. The power nodes in the DFW area are represented by red circles in Figure 1-14. There are 26 power nodes in 11 clusters. The nodal market prices are different at different clusters but similar inside each cluster.

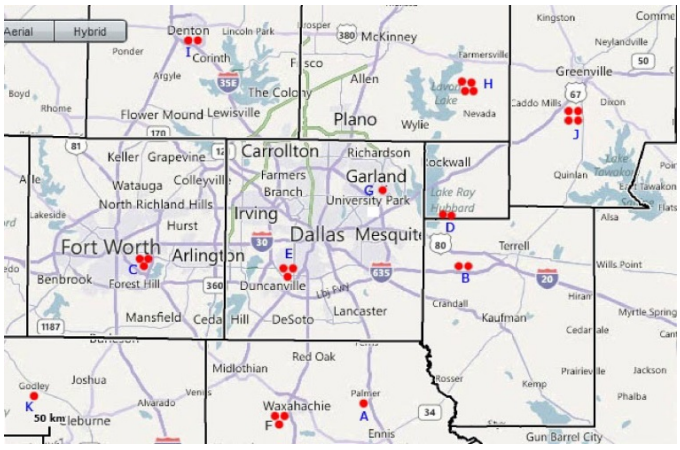


Figure 1-14 ERCOT Power nodes in DFW metroplex

To further investigate the price scenarios for these clusters, the wholesale market prices for each cluster in July 2011 are shown in Figure 1-15. For most of this period, the market prices are less than 50 $/MWh. However, their magnitudes can change suddenly from less than 50 $/MWh to 2000 $/MWh when the spike prices occur. Moreover, the spike prices can occur only once or repeatedly over many periods. Considering all of the volatile price scenarios in the ERCOT nodal deregulated market, it is important for the regional charging station system to forecast accurate prices in order to optimize its operation.

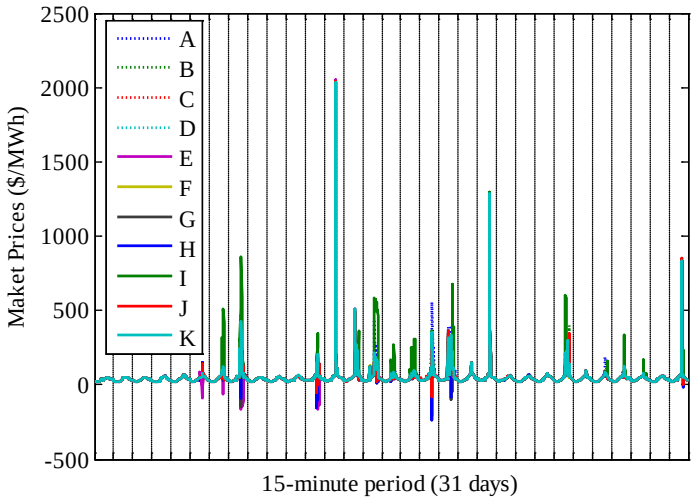


Figure 1-15 DFW Market price in July 2011

1.2.5 Battery Technologies

The design of a fast charging station requires its battery storage system to have a large capacity and a high charging/discharging rate. The technologies that meet these requirements consist of Sodium Sulphur (NaS), Lead Acid (Pbs), Lithium-ion (Li-ion), and Nickel-Metal Hydride (NiMH), and their characteristics are reported in Table 1-2 [17, 18]. According to Table 1-2, considering its long life cycle and high efficiency, NaS is the most

suitable battery technology at the present stage of technological development for storing the distributed energy in PEV charging stations.

Table 1-2 Potential Battery Technologies

Potential Technology	*Cycle life at 80% DOD*	*Efficiency*	*Advantages*	*Disadvantages*
NaS	4500 cycles	89 %	- Good for industrial and commercial sector - High efficiency	Creates high temperatures
Li-ion	3000 cycles	70-85 %	- High density - Low self-discharge rate - No memory effect	Expensive
NiMH	2000 cycles	50-80 %	- High density - Good abuse tolerance	Damage may occur with complete discharge
Lead Acid			Inexpensive	Limited cycling capability
Flooded	1500 cycles	70-80 %		
VRLA	500 cycles	70-80 %		

1.2.6 Conclusion

The proposed regional PEV charging station system equipped with distributed energy storage can participate in the deregulated market. The energy demand of this system can be served by electricity from wind/PV generation and the power grid. The charging station system configuration proposed in this dissertation is necessary for system design. Furthermore, wind/PV generation and electricity market prices need to be accurately forecast due to the intermittence of wind power, the variation of solar generation, and the volatility of electricity prices in the deregulated market. With more accurate predictions, the dynamic control can optimize the global operation of the proposed regional PEV charging stations. For this reason, it is critical to develop effective

forecasting techniques for wind/PV generation and electricity market prices to fully complete system simulation.

1.3 The Proposed Method

In this dissertation, Support Vector Machines (SVMs), effective artificial intelligent algorithms that consist of Support Vector Regression (SVR) and Support Vector Classification (SVC), are adopted to perform all requisite predictions. SVR is used to predict wind/PV generation, and the hybrid of SVR and SVC is used to predict electric market prices both in spike and non-spike price conditions. Moreover, this dissertation proposes using three well-known data clustering techniques, Classification and Regression Tree (CART), K-means, and Stratification methods, to classify the spike prices in several ranges for enhancing the market price prediction in the deregulated market.

Input parameters for SVMs are properly selected by correlation analysis evaluated with the Pearson correlation coefficient. The various impact parameters including historical wind/PV generation and market price as well as multiple weather parameters are taken into account. Due to the 15-minute settlement interval in the ERCOT deregulated market, all of the predictions are performed in the 15-minute ahead time period. The wind power data from a wind farm in Oklahoma, the PV production from Dallas Redbird airport, and the ERCOT wholesale electric prices are used to validate the prediction performance of the proposed approaches for the wind/PV generation and market price forecasting, respectively.

To achieve the complete study of the wind/PV power and market price prediction for the regional operation of the PEV charging station system, the forecast uncertainty also needs to be analyzed to help the charging station system operators understand how to optimize system benefits. The Martingale Model Forecast Evolution

(MMFE) is adopted to effectively examine the multivariate randomness of the predictions in this dissertation. The probability density function of the forecast changes is investigated by ARENA in order to confirm the multivariate random normal vector based on the MMFE.

1.4 Study Objectives

The first part of this dissertation focuses on the modeling of the regional PEV charging station system. The configuration of the charging station system is discussed. The second part focuses on developing the methods to predict the 15-minute ahead wind/PV generation and market price. For each method, different combinations of parameters are tested to determine the model of the best prediction performance. Finally, the third part investigates the forecast uncertainty by the MMFE to complete this comprehensive study of wind/PV generation and market price prediction for a regional PEV charging station system.

1.5 Synopsis of Chapter

The organizational structure of this dissertation is as follows:

Chapter 1 introduces the background and motivation for modeling the regional operation of the PEV charging station system. The configuration of the charging station is also proposed. In addition, Chapter 1 discusses the motivation for the wind/PV generation and market price forecasting. The proposed approaches and the objectives of this dissertation are presented.

Chapter 2 presents the wind power forecasting with the proposed SVR method. A case study from a selected wind farm in Oklahoma is used to validate the effective forecasting performance of the proposed approach.

Chapter 3 proposes the PV generation forecasting with the SVR, which is the same approach as the wind power forecasting. The energy conversion function from

solar radiation to electricity in a PV panel is derived. The PV production estimation from Dallas Redbird airport is used to verify the prediction accuracy of the proposed method.

Chapter 4 illustrates the framework of the novel hybrid market price forecasting method (HMPFM) with data clustering techniques. Three data clustering techniques, Classification and Regression Tree (CART), K-means, and Stratification method, are also discussed. Finally, Chapter 4 presents the results of efficacy tests of the proposed HMPFM with the ERCOT wholesale electricity price.

Chapter 5 shows how the Martingale Model Forecast Evolution (MMFE) is used to investigate the forecast uncertainty. The probability density function (pdf) of the forecast changes is examined to generate new forecast changes of wind/PV generation and market price predictions.

Chapter 6 presents the summary and contributions of this dissertation. The areas of possible further research are also discussed.

Chapter 2

Wind Generation Forecasting

2.1 Literature Review

Though there are numerous benefits, the integration of wind power into the electric power system introduces challenges to balance the supply and demand because of the intermittent nature of wind power. Wind power prediction plays an important role to overcome these challenges. With more accurate wind power forecasting results, the requirement of the system reserve margin can be reduced, thus lowering system operation costs. Considering the increasing installed wind power capacity, wind generation forecasting has gained more interest in the past few years.

Wind generation forecasting can be classified into two different techniques: the physical and statistical approaches [19]. The physical approach is characterized as a conversion model of wind speed and wind power output adopting physical characteristics such as the wind turbine power curve, wind condition, hub height, and others. For example, a short term wind power prediction with multiple observations implementing the wind turbine power curve for the forecasting in Australia has been proposed by [20]. For the statistical method, an analyzing time series model and a probability density function of the historical data for generating the correlation between the wind speed and the wind power output are discussed in [19, 21]. Artificial Intelligent (AI) algorithm, one of the statistical methods, is an effective tool to derive the implicit nonlinear complex relation between the historical data of the wind power and the relevant parameters. AI has been widely adopted to predict wind generation. For instance, Neural Network (NN) is used to predict the wind power generation from a wind farm to determine unit commitment [22]. Also, a study has proposed the hybrid model consisting of Adaptive Wavelet NN (AWNN)

for the wind speed prediction in the first stage, and Feed Forward NN (FFNN) for the wind power estimation in the second stage [23].

In addition to NN, many researchers have used Support Vector Machines (SVMs) combined with various techniques such as Bayesian clustering, Wavelet transform, and Grey relational analysis for wind power forecasting [24-26]. With decomposition based Wavelet transform, the prediction process adopts the combination of Particle Swarm Optimization (PSO) and Adaptive Network Fuzzy Inference System (ANFIS) to improve the forecast accuracy and to avoid the undesired overtraining process [27].

Because all of these studies present different time ahead wind power forecasting with different predicting models, choosing the correct model with the best prediction performance for 15-minute ahead wind power forecasting presents a challenge for this study. The discussions in [19, 25] indicate that SVMs have the best prediction accuracy compared to the other AI methods. Because of this and after thorough evaluation and comparison, this dissertation uses SVMs as the selected method to find the best prediction performance model for 15-minute ahead wind power forecasting.

2.2 Support Vector Machines (SVMs)

SVMs are machine learning algorithms that conduct the learning procedure by statistical theory. They can be separated into two groups, which are the classification and regression methods. A detailed overview of these two approaches [28] are presented in the following sections.

2.2.1 Support Vector Classification (SVC)

Figure 2-1 (a) illustrates how SVC determines the linear separability of the hyperplane $w \cdot x + b = 0$. The definition of $x = (x_1, x_2, ..., x_l)$ is the total number of events, w is the vector, and b is the scalar that defines the characteristics of the hyperplane.

Moreover, $y_i = +1$ and $y_i = -1$ represent two different classes, respectively. Thus, two constraints regarding this two-class separable hyperplane are shown in (2.1) and (2.2).

$$w^T \cdot x + b \geq +1 \tag{2.1}$$

$$w^T \cdot x + b \leq -1 \tag{2.2}$$

The target of the optimal separable hyperplane is to maximize the margin, so the objective function and constraint of this problem become (2.3) and (2.4)

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \tag{2.3}$$

Subject to

$$y_i(w^T \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1,...,l \text{ and } C > 0 \tag{2.4}$$

where C is a regularization parameter defined by the error penalty and ξ_i is a slack variable determined by the distance between the incorrectly classified x_i and the margin.

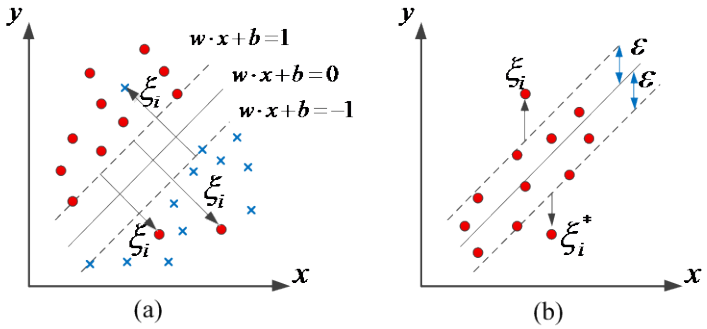


Figure 2-1 Support Vector Machine (a) Classification (b) Regression

The Lagrange multiplier is applied to solve (2.3) and (2.4). By solving the minimization problem, x_i becomes a dot product function. For a nonlinear separable hyperplane in high dimensional feature space, x_i can be mapped into $\phi(x_i)$, leading to a

linearly separable problem. This problem can be efficiently solved by a Kernel function. In fact, several Kernel functions can be adopted to solve this problem such as polynomial function (2.5), radial basic function (RBF) (2.6), hyperbolic tangent function (2.7), and others. In this dissertation, because RBF kernel has the best SVMs prediction performance [14, 15], it is used to perform all of the forecasting.

$$k(x, x_i) = \langle \phi(x) \cdot \phi(x_i) \rangle = (x_i + x_j)^d \tag{2.5}$$

$$k(x, x_i) = \langle \phi(x) \cdot \phi(x_i) \rangle = \exp(-\frac{\|x - x_i\|}{2\sigma^2})^2 \tag{2.6}$$

$$k(x, x_i) = \langle \phi(x) \cdot \phi(x_i) \rangle = \tanh(Kx \cdot x_i + c) \tag{2.7}$$

2.2.2 Support Vector Regression (SVR)

The concept of SVR is slightly different from SVC, as shown in Figure 2-1(b). The loss function insensitive band (ε) and slack variable (ξ_i) are introduced and defined as cost of errors. To maximize the margin, equations (2.8) and (2.9) describe the objective function and problem constraints regarding ε and ξ_i. The techniques to remedy this regression problem are similar to the classification solution: applying the Lagrange multiplier and Kernel function as explained in the previous section.

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i + \xi_i^* \tag{2.8}$$

Subject to

$$\begin{aligned} & y_i - (w^T \cdot x_i + b) \le \varepsilon + \xi_i, \\ & (w^T \cdot x_i + b) - y_i \le \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \ge 0, i = 1,...,l \text{ and } C > 0 \end{aligned} \tag{2.9}$$

Since the wind power is characterized as a continuous value, the SVR is suitable for predicting the wind power considering multiple impact parameters.

2.3 Case Study and Results

2.3.1 Parameters Selection

Prior to performing the wind generation forecasting, input parameters for training SVR should be properly selected. Several important parameters of the wind power generation, historical wind power generation, wind speed, gusty wind, wind direction, and temperature, are taken into consideration. The historical wind power is a significant parameter to the wind generation itself. In addition, the wind speed directly affects the wind power generation according to the wind turbine power curve. As other weather parameters, the efficiency of wind turbines can be enhanced by properly facing the wind turbine itself in the correct wind direction, and by appropriately considering the gusty wind at the wind turbine installation location. Also, different temperatures at the wind turbine area for different time periods can indirectly increase the wind speed at the site. Thus, the sample Pearson correlation coefficient, a correlation analysis, is used to examine the dependence of these parameters and the wind power generation (2.10).

$$r(x,y)=\frac{1}{n-1}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{s_x}\frac{(y_i-\bar{y})}{s_y}$$

$$s_x=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2.10}$$

$$s_y=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

where $\bar{x}$ and $\bar{y}$ are the sample mean of the series data of x and y, and s_x and s_y are the sample standard deviation of the series data of x and y, respectively.

The wind power generation and wind speed data are obtained from a wind farm in Oklahoma, while the other parameters consisting of gusty wind, wind direction, and temperature recorded at a station nearby the wind farm location are extracted from the National Climate Data Center (NCDC) website[29]. All of these data are from 2011.

Figure 2-2 presents the evaluation results of the autocorrelation between the wind generation and the wind generation at 15-minute prior time from February 2011 to December 2011 (not including January 2011 because this month's collective data are incomplete). It can be observed that the *t-15 minutes* of wind generation shows a strong correlation to the wind generation at time *t* with 0.9 coefficients throughout the entire time period.

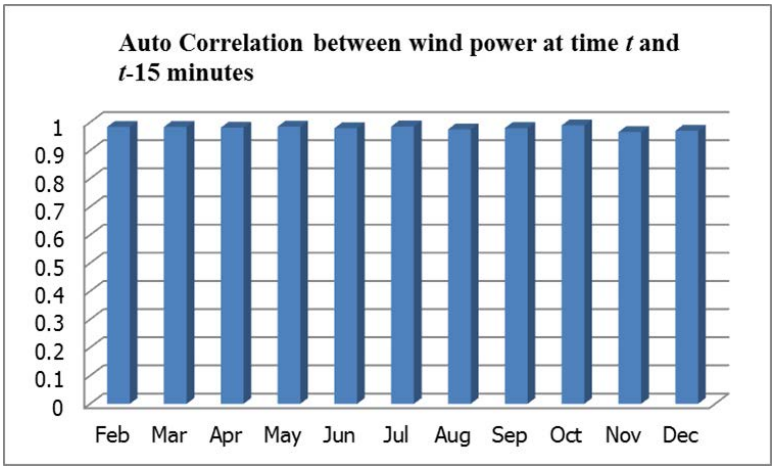


Figure 2-2 Autocorrelation analysis between wind power at time *t* and *t-15*

Figure 2-3 shows the correlation between wind generation and wind speed. Similar to the previous study, one can see that the wind generation also presents a strong correlation to the wind speed with correlation coefficients greater than 0.7 for the entire time period.

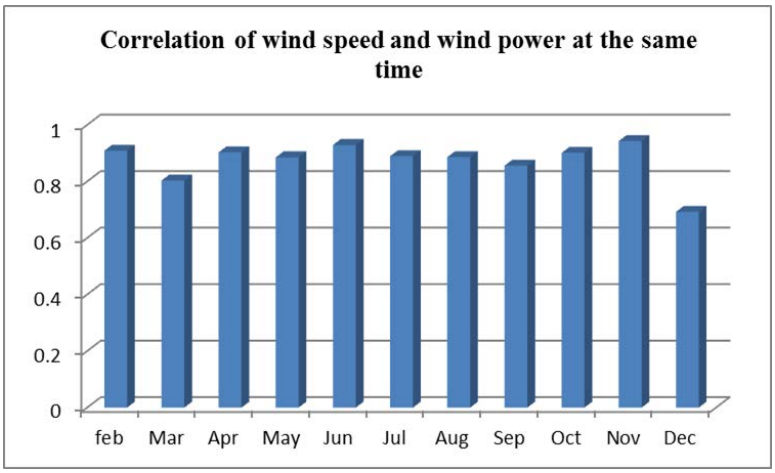


Figure 2-3 Correlation analysis between the wind generation and the wind speed

For the further evaluation over multiple time lags, Figure 2-4 illustrates the correlation between wind generation and historical wind power generation and wind speed from *t*-15 minute and *t*-24 hour. It can be observed that the correlation coefficients decrease steadily when the time lags increase for both the historical wind power generation and wind speed.

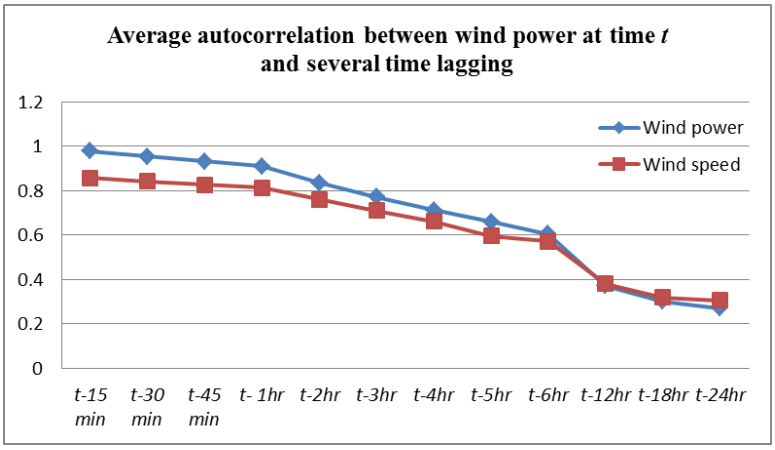


Figure 2-4 Average correlation coefficient for several time lags

Furthermore, three potential parameters consisting of wind direction, gusty wind, and temperature are used to evaluate the correlation coefficient of the wind power generation as reported in Table 2-1. There is a moderate correlation between wind generation and gusty wind with 0.494 coefficients, while the wind power generation illustrates weak correlation with the other two parameters.

Table 2-1 Correlation Analysis of the Relevant Weather Parameters and Wind Power

Parameter	*Average Correlation Coefficient*
Wind direction	0.103
Gusty wind	0.494
Temperature	0.069

Considering all of these correlation analyses, the historical wind generation and wind speed, which illustrate strong correlation with the wind power generation, are significant input parameters and therefore used first in the training of SVR. Even though the other parameters consisting of wind direction, gusty wind, and temperature show less significant correlation, they are still used as supplementary input parameters to examine the possible improvement of further prediction accuracy without loss of generality.

2.3.2 Results

Wind power data from a wind farm in Oklahoma with 74.25 MW installed capacity is used to validate the SVR performance for the 15-minute ahead wind generation forecasting in this dissertation. According to the aforementioned discussion, the wind power and wind speed data are acquired from the wind farm site, and the potential weather data are obtained from the NCDC website. The training data are 2010 data, while the data of February, May, July, and October in 2011, representing the four seasons, are used for testing.

Typically, the wind generation forecasting performance is evaluated by Mean Relative Error (MRE), as shown in (2.11). MRE defines the prediction error related to the maximum installed capacity of the wind farm.

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{W_{p,i} - W_{a,i}}{W_{\max}} \times 100\,\% \tag{2.11}$$

where W_p is predicted wind generation, W_a is actual wind generation, $W_{\max}$ is the maximum installed capacity of the wind farm, and N is the number of predictions.

To determine the best prediction performance model for wind generation forecasting, historical wind power generation and wind speed, the most impactful parameters, are the first used input parameters for SVR. After the best prediction performance model is determined, it is combined with the remaining three weather parameters to analyze possible improvement in forecasting accuracy. The comparison results of the wind generation forecasting from multiple models using SVR and the results from a persistence model defined as (2.12) are reported in Table 2-2.

$$WP(t) = WP(t-1) \tag{2.12}$$

According to Table 2-2, historical wind power generation and wind speed are independently used to predict wind power generation. It can be observed that the models using historical wind power give better prediction accuracy than the results from the wind speed models. Also, the MRE of the three time lags historical wind power model is higher than the two time lags historical wind power model. However, the MRE of the three time lags historical wind speed model is the same as the two time lags historical wind speed model. Because of this, the simulation is tested only up to the three time lags models. After testing the prediction accuracy of models with all possible combinations of wind power and wind speed, the model WP(*t*-1&*t*-2)&WS(*t*-1) gives the best prediction performance with 3.33 % MRE. Next, this best prediction performance model combined

with the other potential parameters is evaluated; however, the further combination models do not improve the wind generation forecasting accuracy any further. Also, this model with the best prediction performance gives better prediction accuracy compared to the persistence model.

Table 2-2 Wind Power Generation Forecasting Results (% MRE)

Model	*Feb*	*May*	*Jul*	*Oct*	*Average*
WP(*t*-1)	3.68	4.00	3.12	3.20	3.50
WP(*t*-1&*t*-2)	3.41	3.91	3.06	2.98	3.34
WP(*t*-1,…,*t*-3)	3.41	3.95	3.08	2.98	3.35
WS(*t*-1)	7.18	7.11	6.64	6.43	6.84
WS(*t*-1&*t*-2)	7.15	7.18	6.62	6.40	6.84
WS(*t*-1,…,*t*-3)	7.14	7.20	6.61	6.39	6.83
WP(*t*-1)&WS(*t*-1)	3.68	3.97	3.10	3.20	3.49
WP(*t*-1)&WS(*t*-1&*t*-2)	3.55	3.85	3.10	3.13	3.41
WP(*t*-1)&WS(*t*-1,…,*t*-3)	3.62	3.90	3.11	3.18	3.45
WP(*t*-1&*t*-2)&WS(*t*-1)	3.41	3.90	3.04	2.99	3.33
WP(*t*-1&*t*-2)&WS(*t*-1&*t*-2)	3.44	3.83	3.06	3.03	3.34
WP(*t*-1&*t*-2)&WS(*t*-1,…,*t*-3)	3.48	3.86	3.07	3.07	3.37
WP(*t*-1,…,*t*-3)&WS(*t*-1)	3.43	3.94	3.07	3.00	3.36
WP(*t*-1,…,*t*-3)&WS(*t*-1&*t*-2)	3.47	3.87	3.08	3.06	3.37
WP(*t*-1,…,*t*-3)&WS(*t*-1,…,*t*-3)	3.51	3.91	3.09	3.08	3.40
WP(*t*-1&*t*-2)&WS(*t*-1)&G(*t*-1)	3.41	3.90	3.04	2.99	3.33
WP(*t*-1&*t*-2)&WS(*t*-1)&WD(*t*-1)	3.41	3.90	3.04	3.01	3.34
WP(*t*-1&*t*-2)&WS(*t*-1)&T(*t*-1)	3.40	3.91	3.03	3.01	3.34
Persistence	3.62	3.95	3.09	3.13	3.45

The notations in Table 2-2 are defined as follows: WP is wind power; WS is wind speed; G is gusty wind; WD is wind direction; T is temperature; and (*t*-1, *t*-2, *t*-3) are 15, 30, and 45 minutes before the prediction time, respectively.

To demonstrate the effective wind generation forecasting approach by the proposed SVR, Figure 2-5 to Figure 2-8 illustrates the wind power forecasting for four months representing four seasons in a year.

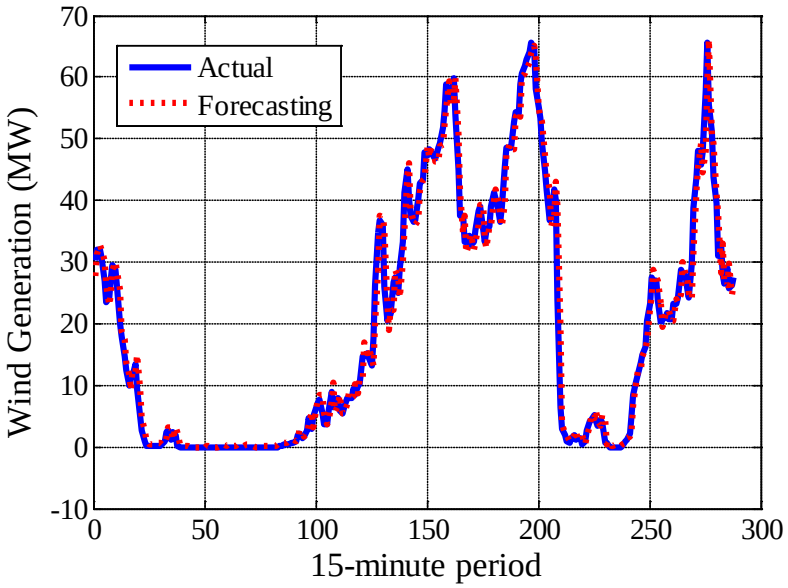


Figure 2-5 Wind generation forecasting in February

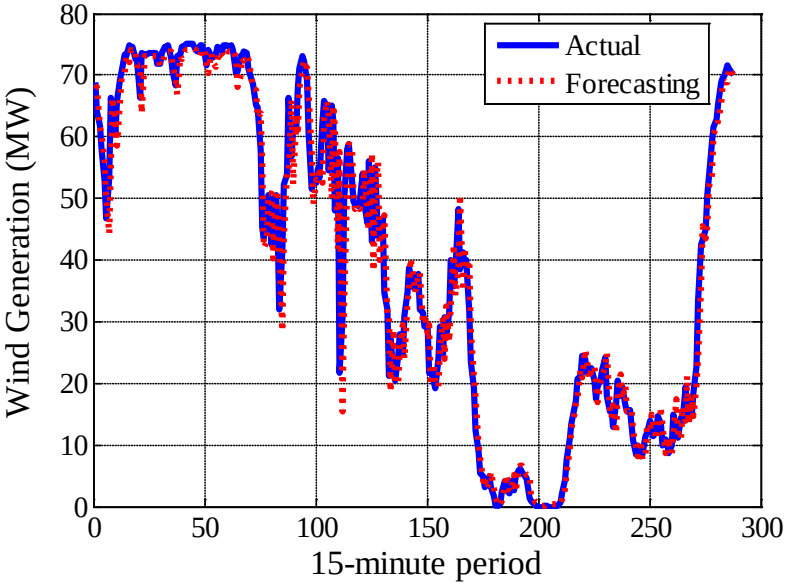


Figure 2-6 Wind generation forecasting in May

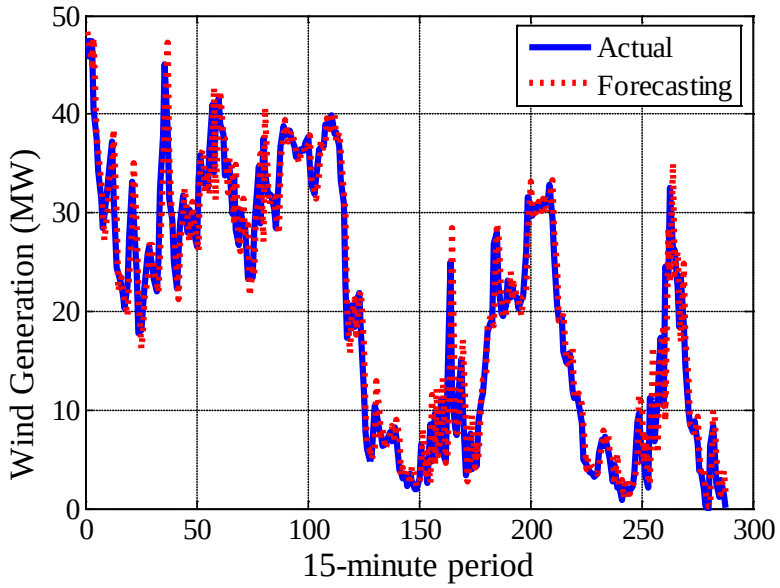


Figure 2-7 Wind generation forecasting in July

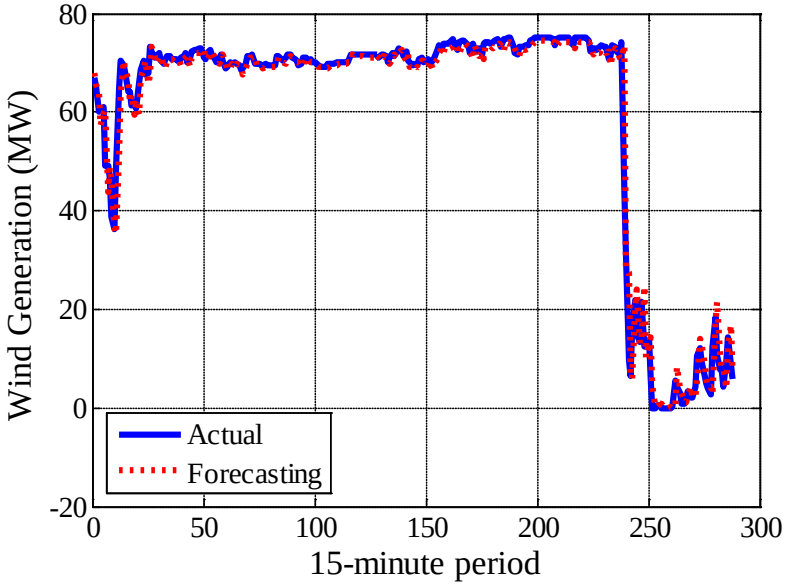


Figure 2-8 Wind generation forecasting in October

2.4 Summary

In order to accurately forecast wind generation, Support Vector Regression (SVR) is an effective method used to perform the predictions in this dissertation. Wind generation, wind speed and the relevant weather parameters including gusty wind, wind direction, and temperature are taken into consideration as the input parameters of the proposed SVR. The combination model of WP(*t*-1&*t*-2)&WS(*t*-1) yields the most accurate wind generation prediction with 3.33 % Mean Relative Error (MRE) evaluated in four months representing the four seasons for the entire year.

Chapter 3

PV Generation Forecasting

3.1 Literature Review

The PV industry has progressively developed from residential-scaled PV systems to utility-scaled PV power stations. However, due to its intermittent and variable nature, PV generation creates concern among system operators about how to equalize the supply and demand. Many studies address this concern by developing and testing more accurate PV generation prediction methods.

In the existing literature, the PV generation prediction method can be divided into two categories: the modeling and statistical approaches [30-46]. The first group, which is the modeling approach, develops the prediction model by deriving the PV generation from the amount of solar radiation, considering PV panel characteristics and weather conditions. Next, this prediction model is used to predict the next time ahead PV generation using the Numerical Weather Prediction (NWP) by the weather service provider. For example, one study investigates the impact of different weather conditions, including sunny, cloudy, rainy, and snowy, for the PV generation estimation [30]. Also, the PV generation prediction in [31] is evaluated using solar radiation and sky conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF). A study in [32] introduces Simulation model of Solar power forecast (SimSol) to predict PV power considering NWP combined with the installation details from the PV modules such as the tracking system, shadowing effect, and other factors.

On the other hand, the statistical and artificial intelligent methods with nonlinear machine learning have been frequently applied for predicting PV generation. For instance, Neural Network (NN) using the historical weather conditions and the solar irradiation for multiple time periods ahead is adopted to perform the PV generation

forecasting [33, 34]. The NN combined with the time delay function training by Levenberg-Marquardt algorithm is also used for predicting the PV generation [35]. A study in [36] proposes one type of NN, which is the Recurrent S-Cerebellar Model Articulation Controller (RSCMAC), for calculating the PV power forecasting.

Many PV generation predictions have been calculated using a statistical-based method with several regression models such as Auto Regressive Moving Average (ARMA) [37, 38] and H-Filtering [39]. In addition, two studies propose the regression model with additional complex algorithms including Kalman Filtering ARMA and Nonlinear Auto Regressive Model with Exogenous inputs (NARX) [40, 41]. Furthermore, several studies use data mining techniques consisting of K-means and Rough test methods to select the suitable characteristics for predicting PV generation [42, 43]. Support Vector Machines (SVMs) combined with several clustering techniques such as similar day approach [44], fuzzy inference method [45] and weather-based classification technique [46] have also been used to forecast PV generation.

Although these research works attempt to improve the accuracy of PV generation prediction, none of them have explicitly specified the forecasting model for very short term forecasting. Consequently, determining the most accurate prediction model for the 15-minute ahead PV generation forecasting is a challenge in this dissertation. Since the SVMs show the best prediction performance compared to other methods [45], this technique is adopted to calculate the 15-minute ahead PV generation estimation in this study. Therefore, because PV generation is considered as a continuous value, SVR can be effectively used to predict the PV generation.

3.2 Solar Radiation and PV Generation

The National Renewable Energy Laboratory (NREL) provides the hourly historical solar radiation and its associate weather data from 1991-2010 [47]. However,

prior to calculating the PV generation prediction, the energy conversion model from the solar radiation to the PV power output needs to be studied. Furthermore, the energy conversion model depends on the amount of solar radiation absorbed by the PV panel, the PV panel specification, and the dependent weather. The process of PV generation evaluation is described as follows.

There are three components of the solar radiation absorbed by the PV panel: direct radiation, diffuse radiation, and reflected radiation, as depicted in Figure 3-1 [48]. To calculate the PV power output, Plane of Array (POA) irradiance (E_{POA}) using these three solar radiation components is evaluated by (3.1).

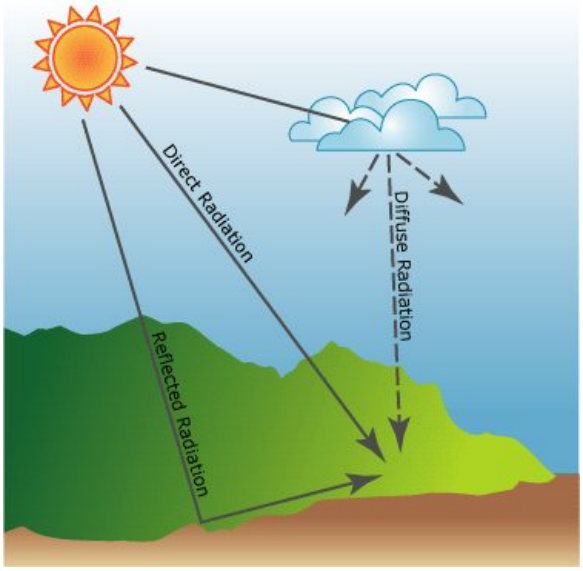


Figure 3-1 The components of solar radiation to PV panel

$$E_{POA} = E_b + E_g + E_d \tag{3.1}$$

where E_b is the POA beam component, E_g is the POA ground reflected component, and E_d is the POA sky diffuse component.

The POA beam component can be calculated by (3.2) using the incident angle modifier (IAM_B). The POA ground reflected component depends on $alberdo$, which varies based on PV panel installation surface, as formulated in (3.3). There are several methods to obtain the POA sky diffuse component such as Isotropic sky diffuse model, Perez sky diffuse model, and others. [49]. For simplicity without losing the accuracy of evaluation, the Simple Sandia sky diffuse model is adopted to calculate E_d in this dissertation, as shown in (3.4).

$$E_b = DNI \times IAM_B \times \cos(AOI) \tag{3.2}$$

$$E_g = GHI \times alberdo \times \frac{1-\cos(T_a)}{2} \tag{3.3}$$

$$E_d = DHI \times \frac{1+\cos(T_a)}{2} + GHI \times \frac{(0.12Z - 0.04) \times (1-\cos(T_a))}{2} \tag{3.4}$$

where DNI, GHI, and DHI denote the direct normal, global horizontal, and diffuse horizontal irradiances, respectively. The hourly data of these three parameters can be obtained from the actual measurements recorded by NREL [47]. $Alberdo$ is 0.35 for the galvanized steel that acts as the surface for installation of the PV panel. T_a and Z are the surface tilt angle (the longitude of the PV panel installation) and the sun zenith angle, respectively, depicted in Figure 3-2 [50]. The Azimuth angle (θ_A) is also shown in Figure 3-2. Finally, AOI is the angle of incident and is calculated using (3.5).

$$AOI = \cos^{-1}[\cos(Z)\cos(T_a) + \sin(Z)\sin(T_a)\cos(\theta_A - \theta_{A,array})] \tag{3.5}$$

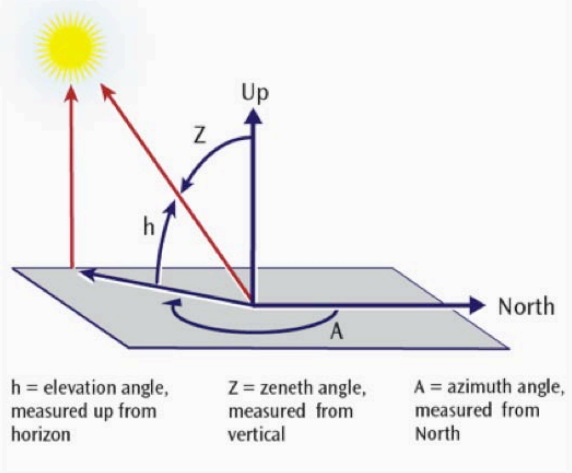


Figure 3-2 Zenith and Azimuth angle

Finally, the formula (3.6) is used to estimate the PV generation output considering the E_{POA}, the installation area (A), the adjusted PV panel efficiency (η) according to (3.7), and the 0.77 DC to AC conversion factor recommended by NREL.

$$PV = E_{POA} \times A \times \eta \times 0.77 \tag{3.6}$$

$$\eta = \eta_0 [1 - \gamma (T_C - T\gamma)] \tag{3.7}$$

where η_0 denotes the PV efficiency at reference temperature, γ is 0.005 for the temperature coefficient of the solar battery, and T_C and T_γ are cell and reference (25 C) temperatures, respectively.

Since the NREL provides hourly historical solar radiation and its associate weather data, the linear approximation method is used to convert the hourly data into 15-minute data corresponding to the PEV operational time period. NREL has recorded the solar radiation data for several locations in the DFW area; however, the PV generation calculated with this solar radiation data from these different locations is similar throughout the metroplex. Therefore, only the Dallas Redbird airport site is selected to calculate the PV generation predictions in this dissertation. Figure 3-3 shows an example of the power

output profile of a 33.82 kW single crystalline silicon panel at the Dallas Redbird airport. The panels have a 180 m^2 installation area (10 charging slots for one charging station) and 24.4% conversion efficiency [51].

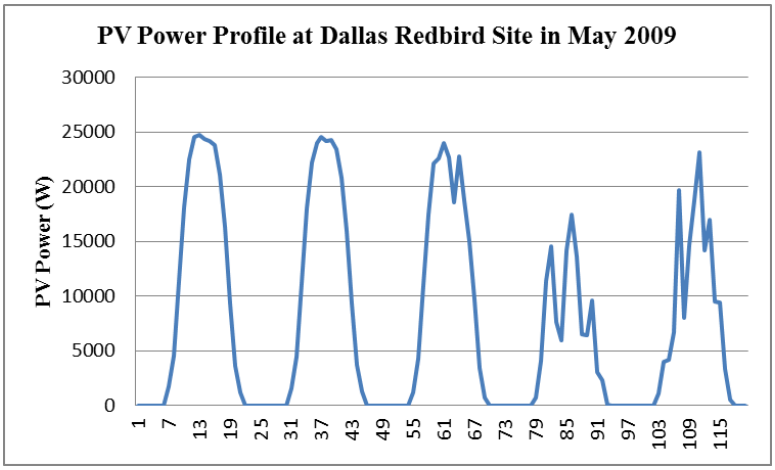


Figure 3-3 PV output power evaluated by solar radiation

3.3 Case Study and Results

3.3.1 Parameters Selection

The impact parameters to forecast the solar radiation have been identified using data mining techniques [52]. In addition, [52] indicates that the most significant parameter to predict the solar radiation is the historical solar radiation from both summer and winter. According to the aforementioned discussion, because the PV generation is a direct function of the solar radiation absorbed by the PV arrays, either the historical PV generation or solar radiation can be selected as an impact parameter for the prediction. In this dissertation, the historical PV generation is chosen due to its convenient implementation into the PEV charging station without losing the accuracy of the PV generation prediction.

Autocorrelation analysis according to (2.10) between the PV generation at time *t* and the PV generation at 15 minutes before is illustrated in Figure 3-4. The

autocorrelation demonstrates a strong dependence for the entire year with a 0.994 average correlation coefficient. As a result, the historical PV generation is considered as a significant input parameter of the forecasting algorithm.

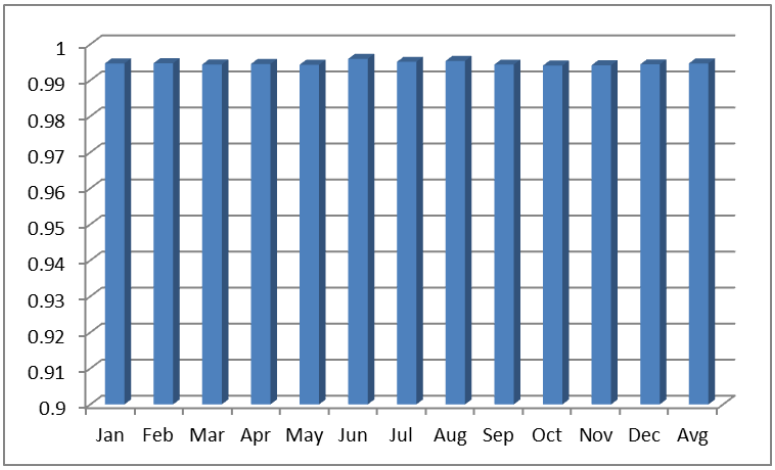


Figure 3-4 Auto correlation analysis between PV generation at time *t* and *t*-15 min

In addition, other parameters including sunshine duration, humidity, temperature, and altitude are considered as potential parameters for estimating the PV power output [52]. However, this dissertation does not consider altitude because charging stations have stationary PV panels. The cloud rating is taken into account for the PV power forecasting because PV panels generally produce less energy on cloudy days than on sunny days. Since the wind speed affects cloud conditions, it should be considered as one of the input parameters for PV generation forecasting.

Considering all these factors, the correlations between PV generation and the potential parameters of humidity, temperature, cloud rating, and wind speed are calculated and reported in Table 3-1. The humidity and cloud rating have inverse correlations with PV generation, while the temperature and wind speed show positive correlation with PV generation. Although all four of these parameters show lower correlation coefficients than the autocorrelation coefficient itself, they are still used as

input forecasting parameters to examine the further prediction accuracy improvement. However, as Pearson correlation analysis cannot be used to evaluate the correlation between sunshine duration from the previous day and PV generation, this parameter is not included in Table 3-1, but it is used in this dissertation to analyze the further prediction performance improvement of PV generation.

Table 3-1 Correlation Analysis of the Relevant Parameters and the PV Generation

Parameter	*Average Correlation Coefficient*
Humidity	-0.496
Temperature	0.298
Cloud rating	-0.122
Wind speed	0.188

3.3.2 Results

PV generation is calculated from the solar radiation based on the proposed evaluation method in section 3.2. Training data for SVR are 2008 data. To obtain the best prediction performance model, the data from the four months of January, April, July, and October 2009, representing the four seasons in a year, are used to test the model from the training stage.

The prediction performance evaluation of the PV generation forecasting typically adopts Mean Absolute Percentage Error (MAPE), as formulated in (3.8).

$$MAPE = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|PV_j^{true} - PV_j^{fst}\right|}{PV_j^{true,N}} \tag{3.8}$$

where PV_j^{true} is the actual PV generation at time j, PV_j^{fst} is the forecasted PV generation at time j, and $PV_j^{true,N}$ is the average of the recorded PV generation over N periods.

The historical PV generation, the most significant impact parameter according to correlation analysis, is the first used input parameter for calculating the PV generation prediction by SVR. Then, the other potential parameters, humidity, temperature, cloud rating, wind speed, and the previous day of sunshine duration, are used as inputs combined with the model of the best prediction performance determined from the first step. The results of PV generation estimation from these models are compared with the results from the persistence model calculated by (3.9), as reported in Table 3-2.

$$PV(t) = PV(t-1) \tag{3.9}$$

Table 3-2 PV Generation Forecasting Results

Model	*Jan*	*Apr*	*Jul*	*Oct*	*Average*
PV(*t*-1)	8.24	7.93	7.16	8.77	8.03
PV(*t*-1&*t*-2)	2.61	2.76	2.37	3.29	2.76
PV(*t*-1,…,*t*-3)	2.66	2.76	2.29	3.34	2.76
PV(*t*-1&*t*-2)&H(*t*-1)	2.91	2.93	2.42	3.70	2.99
PV(*t*-1&*t*-2)&T(*t*-1)	2.68	2.84	2.48	3.39	2.85
PV(*t*-1&*t*-2)&C(*t*-1)	2.61	2.76	2.37	3.29	2.76
PV(*t*-1&*t*-2)&WS(*t*-1)	2.70	2.93	2.41	3.37	2.85
PV(*t*-1&*t*-2)&S(*d*-1)	2.46	2.69	2.36	3.10	2.65
Persistence	8.14	7.89	7.07	8.60	7.92

The notations in Table 3-2 are defined as follows: PV is PV generation; H is humidity; T is temperature; C is cloud rating; WS is wind speed; and S is previous day sunshine duration.

According to Table 3-2, the forecasting results of the PV(*t*-1&*t*-2) model is 2.76 % MAPE, which is significantly more accurate than the 8.03 % MAPE of the PV(*t*-1) model. However, the study shows that the prediction accuracy of the PV(*t*-1,…,*t*-3) model shows no improvement from 2.76 % MAPE. Therefore for simplicity, the PV(*t*-1&*t*-2) model is used to calculate the additional PV generation forecasting and is combined with humidity, temperature, cloud rating, wind speed, and the previous day of sunshine duration in the next step.

After evaluating the PV generation prediction accuracy with different combination models of the potential weather parameters, the PV(*t*-1&*t*-2)&S(*d*-1) model is shown to have the best estimation performance, and moderately improves the prediction accuracy to 2.65 % MAPE. This combination model also gives significantly better PV generation forecasting accuracy compared to the prediction accuracy of 7.92 % MAPE from the persistence model.

In order to illustrate the effective forecasting performance of the proposed approach with the best prediction model of PV(*t*-1&*t*-2)&S(*d*-1), Figure 3-5 to Figure 3-8 depict the comparison between the actual and forecasted PV generation for four months in 2009. It can be observed that prediction results by the proposed method are highly accurate for the PV generation in normal conditions. However, this model results in a few prediction errors when the PV generation experiences some fluctuations.

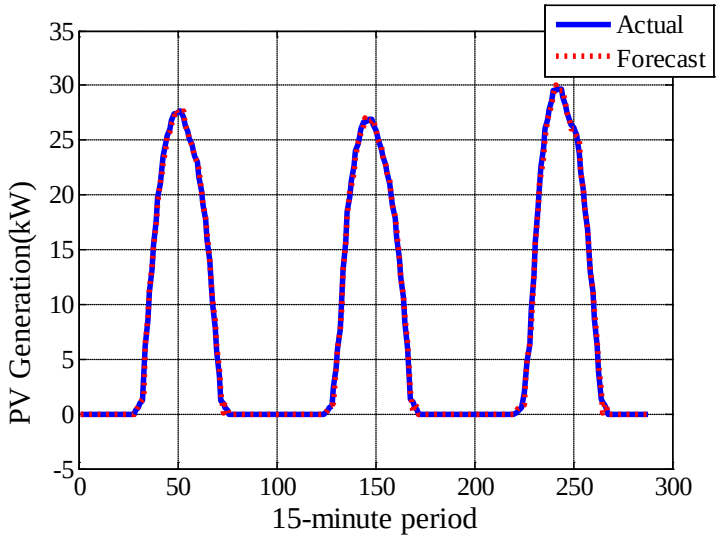
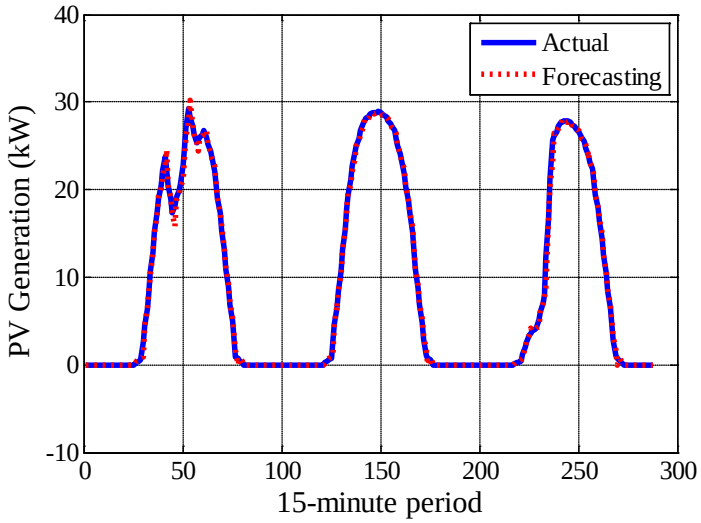


Figure 3-5 PV generation forecasting results in January



(b)

Figure 3-6 PV generation forecasting results in April

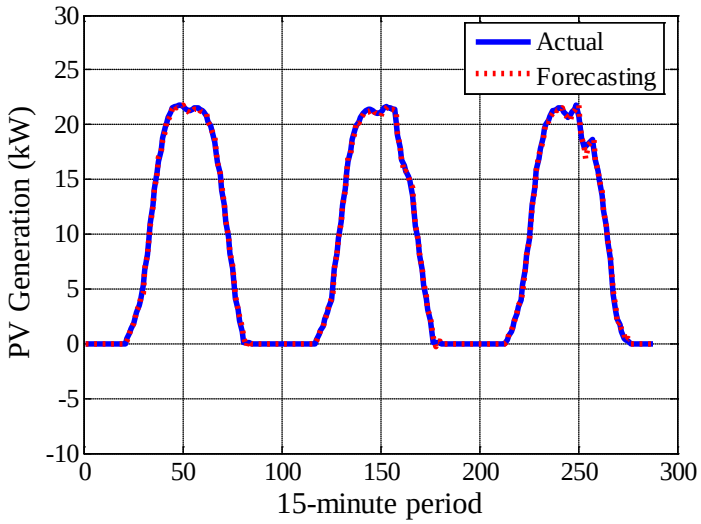


Figure 3-7 PV generation forecasting results in July

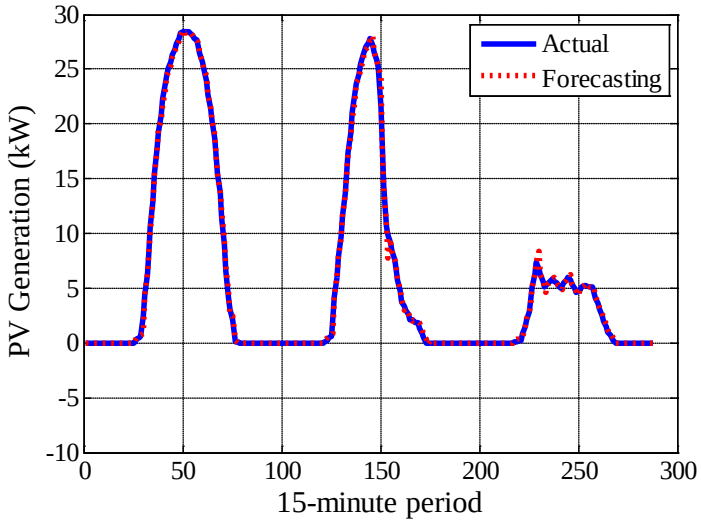


Figure 3-8 PV generation forecasting results in October

3.4 Summary

Either PV generation or solar radiation can be selected to forecast the future energy production from a PV system. However, because PEV charging stations directly use the amount of energy from PV generation and not from solar radiation, this dissertation uses historical PV generation to predict the future energy production from a PV system. The amount of energy from PV generation can be calculated using the energy conversion model of solar radiation, the PV panel specification, and the associated weather data. Historical PV generation, humidity, temperature, cloud rating, wind speed, and the previous day of sunshine are taken into consideration for the predictors. The model of the historical PV generation at 15 and 30 minutes before the prediction time and the previous day of sunshine, [PV(t-1&t-2)&S(d-1)], yields the best prediction performance, and demonstrates the effective prediction of the proposed SVR. This model significantly improves the forecasting accuracy compared to the persistence model from 7.92 to 2.65 % MAPE.

Chapter 4

Market Price Forecasting

4.1 Literature Review

In the deregulated electricity market, the interactions among supply, demand, constraints, and other market components are taken into account when settling the electricity price. The electricity price may experience volatility as these components change. This volatility in the deregulated market presents challenges since the occurrence and magnitude of the spike prices are difficult to estimate. Accurate market price prediction benefits power producers and consumers by optimizing their system costs. Consequently, the study of market price forecasting has gained increasing attention in recent years.

The electric price forecasting method in the deregulated market can be separated into simulation and statistical approaches[53]. The simulation method estimates market price considering generator dispatch patterns with various impact parameters. Although the simulation method can predict market price accurately, it needs a massive amount of data from actual electrical models for accurate calculation[54]. Furthermore, the statistical approach aims to map the connection between market price and its related parameters by either Artificial Intelligent (AI) algorithms or time series models. Due to the challenges present in the simulation method, the statistical method is more commonly used.

Neural Network (NN), one AI algorithm, combined with fuzzy logic has been proposed, considering demand, spinning reserve, and capacity shortfall as input parameters [54]. In addition, NN has been combined with Fuzzy c-mean for clustering different transaction periods into proper groups before calculating market price predictions[55]. Another study shows how Fuzzy inference with Least Square Estimator

(LSE) models the correlation between market price and historical predictors [56]. Taking demand load into consideration, the Euclidean norm is used to select the day with the most similar price characteristics, which is then used to predict market price by NN [57]. One study shows how the time series model is derived by Auto Regressive Integrated Moving Average (ARIMA) [58], and another shows how it is developed when combined with wavelet transform, which improve the forecasting accuracy [59]. All of these statistical approaches show sufficient forecasting accuracy for day ahead estimation, but they normally can only predict non-spike electric prices.

Consequently, a few hybrid models with classification algorithms have been developed to additionally predict the spike price occurrence. Radial Basic Function NN and Support Vector Machines (SVMs) hybrid models have been adopted to estimate the electric price both in spike and non-spike prices conditions in the deregulated market [60, 61]. However, the forecasting timeframes and training input parameters have not been clearly identified in these studies. In addition, the spike price forecasting in these hybrid models is typically calculated using only an AI method, which may lead to inaccuracies. These three important issues can significantly influence the accuracy of electric price prediction.

Therefore, this research proposes a hybrid market price forecasting method (HMPFM) with data clustering techniques. The goal of these clustering techniques is to dissect spike prices in several ranges before performing the spike price magnitude forecasting. This novel technique can improve the accuracy of spike price magnitude forecasting to enhance overall market price prediction. Since SVMs have been efficiently used for predicting both classification and regression in various applications such as the wind/PV generation prediction discussed in the previous chapters, Support Vector Classification (SVC) is adopted to predict spike price occurrence, and Support Vector

Regression (SVR) is used for market price magnitude prediction of both spike and non-spike prices. This study implements three clustering algorithms, Classification and Regression Trees (CART), K-means, and Stratification methods, because the CART and K-means approaches have been successfully used in clustering studies, and the Stratification method is the simplest clustering technique [62-65].

4.2 Hybrid Market Price Forecasting Method (HMPFM)

The framework of HMPFM with data clustering techniques is depicted in Figure 4-1. There are two main stages of the proposed method: spike price occurrence and price magnitude predictions. First, the spike price occurrence prediction is calculated. If the result of this prediction is yes, the spike price magnitude prediction will be calculated; otherwise, the non-spike price magnitude prediction is processed. The details of each prediction are described in the following discussion.

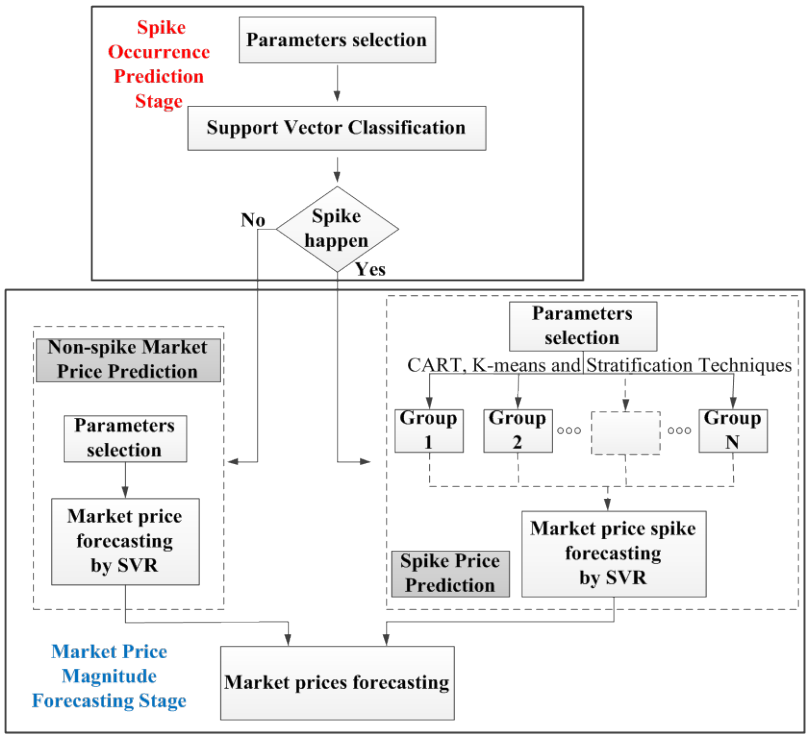


Figure 4-1 Hybrid market price prediction framework

4.2.1 Spike Market Price Occurrence Prediction

According to several previous research works [60, 61], there are three spike price definitions:

1) An abnormal high price is a price that is substantially higher than normal;

2) An abnormal jump price is the difference between two adjacent prices that is greater than a given threshold;

3) A negative price is when the price falls below zero.

An abnormal high price is the main focus of this dissertation. The levels of this type of spike price can be defined by statistical methods. The studies in [60, 61] show that it can be calculated by either one standard deviation threshold or two standard deviation thresholds. In order to escalate the spike event number for improving the forecasting accuracy, the spike price is defined by a one standard deviation threshold and is calculated by (4.1) in this dissertation.

$$spike = \mu \pm \sigma \tag{4.1}$$

where μ and σ are the mean and standard deviation of the market price, respectively (43.59 and 162.32 $/MWh for the DFW market price in 2011).

SVC is the selected algorithm in this dissertation to predict the spike price occurrence and considers several impact parameters such as historical market prices, load profiles, and others. The spike price occurrence forecasting is calculated by several models in this dissertation to identify the one with the best prediction performance.

4.2.2 Non-Spike Market Price Prediction

Due to the inconsequential change of the non-spike price in a 15-minute period, the typical AI forecasting method can be adequately adopted to predict the non-spike price condition. SVR is selected to estimate the magnitude of the non-spike price and considers the same impact parameters as the spike price occurrence prediction. In this

process, several models are tested with all spike prices removed prior to calculating the non-spike market price predictions to identify the model with the best prediction performance.

4.2.3 Spike Market Price Prediction

Spike prices in the DFW market fluctuated between less than -120 $/MWh and more than 3000 $/MWh in 2011[66]. Since this widespread distribution of spike prices can affect the ability to accurately estimate their magnitude by typical AI forecasting approaches, clustering methods are introduced in this dissertation to divide spike prices into appropriate clusters before SVR calculates their magnitude prediction. This dissertation implements three clustering algorithms: CART, K-means, and Stratification methods. The model with the best prediction performance considering impact parameters is obtained by performing the comprehensive HMPFM with these three proposed data clustering techniques.

4.3 Data Clustering Techniques

4.3.1 Classification and Regression Tree (CART)

CART is a binary recursive partitioning clustering technique[67, 68]. Target variables can be either categorical or continuous values in classification or regression scenarios, respectively. The clustering method in this dissertation focuses on the regression technique since the magnitude of the spike price is considered continuous. Regarding the regression algorithm itself, two main stages are used to determine the optimal clusters including growing and pruning processes. In the former stage, CART ultimately enforces maximum possible terminal nodes from their parents using the splitting rule as $x_i \leq d$. Thus, if a predictor value (x_i) is less than or equal to a setting value (d), this variable will be a member of the left children node. Conversely, this variable will be assigned to the right children node group. This rule is implemented with

the least square function and goodness of split as (4.2,4.3) for growing optimal terminal nodes. In the latter stage, a minimal cost tree by lowest mean square error is employed for pruning the generated tree from the first stage.

$$SS(t) = \sum (y_{i(t)} - \bar{y}_{(t)})^2 \tag{4.2}$$

$$\phi(t) = SS(t) - SS(t_R) - SS(t_L) \tag{4.3}$$

where $y_{i(t)}$ is the target of x_i in node t; $\bar{y}_{(t)}$ is the mean of target values in node t; $SS(t)$, $SS(t_R)$, and $SS(t_L)$ are the sum square errors of the parent node, right children node, and left children node, serially; and $\phi(t)$ is the goodness of split that shows the highest value for the best split.

4.3.2 K-means Clustering

This algorithm separates the d-dimensional vector space of data point (x_i), $D = \{x_i | i = 1,..., N\}$ into k partitions by minimizing the cost function as (4.4).

$$Cost = \sum_{i=1}^{n} (\arg\min_j \left\| x_i - c_j \right\|^2) \tag{4.4}$$

where c_j represents k -centroid clusters in set $C = \{c_j | j = 1,..., k\}$

To reach the aim of cost minimization, this algorithm iteratively performs two-step procedures. First, c_j is initialized randomly, and data points are assigned to the closest centroid by implementing a Euclidean distance function. Second, a new c_j is computed using the assigned data from the first step. This iteration is repeated until c_j is stabilized.

4.3.3 Stratification Method

Employing this clustering technique is a simple process based on statistical data. To have sufficient data in each group, this technique divides d -dimensional vector space

$D = \{x_i | i = 1,..., N\}$ equally into k clusters considering different target ranges that are different spike price ranges in this dissertation.

4.4 Case Study and Results

The regional PEV charging station system is designed to be built near the power nodes in the DFW area for DC fast charging. Since ERCOT's wholesale market prices in each cluster are similar, only one set of market prices is used for each cluster. Cluster E, which is near Dallas, is used to illustrate the proposed market prices prediction method. First, a correlation analysis is conducted to select the input parameters for the SVMs process. Then, the HMPFM with data clustering techniques is implemented following the framework in section 4.2. Finally, the comprehensive results are presented and analyzed to verify the prediction performance. The proposed approach is then applied to other power nodes to improve the forecasting accuracy for other PEV charging station locations in the DFW area.

4.4.1 Parameter Selection

It is possible to obtain historical market prices, temperatures, and load profiles before performing a 15-minute ahead market price forecasting, while several factors such as generator contingencies and transmission constraints remain unknown prior to predicting the market price. Other factors, such as fuel prices and day ahead load forecast, have less influence on very-short term market price forecasting. Therefore, correlation analyses of historical market prices, temperatures, and load profiles are studied. Historical market prices, temperature, and load profile are extracted from the National Climatic Data Center (NCDC) [29] and ERCOT websites. Figure 4-2 depicts the correlation results between market price and the three impact parameters at 15-minute intervals for a 12-hour time period.

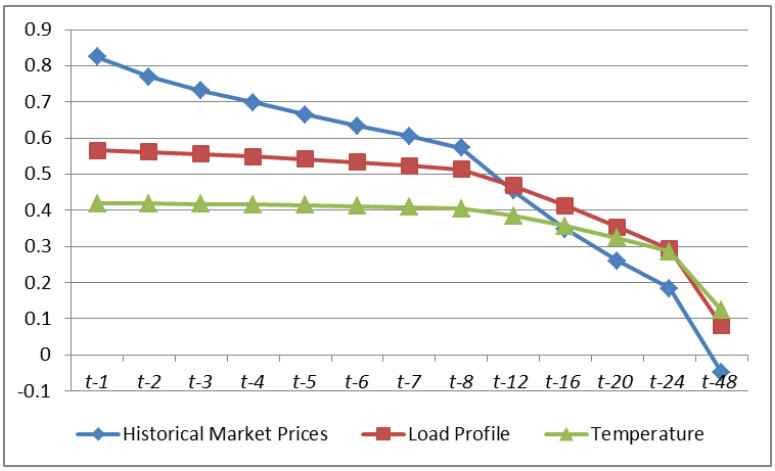


Figure 4-2 Correlation analyses between market price and three impact parameters

[(*t*-1),(*t*-2),…, (*t*-n) are 1,2,…,n prior times in a 15-minute period.]

According to Figure 4-2, all correlations decrease significantly when prior times increase. Historical market prices show a strong autocorrelation with coefficients of greater than 0.7 until *t*-4, so this parameter is identified as one important predictor. Moreover, both historical load profiles and temperatures give moderate correlations to market price with a coefficient exceeding 0.4. Although these two parameters present lower correlation coefficients than that of historical market prices, they are included as input parameters for further improving the forecasting accuracy.

4.4.2 Spike Market Price Occurrence Prediction

This study introduces $P(in)$ and $P(out)$ given by (4.5) and (4.6) in order to evaluate the spike occurrence prediction accuracy. These two indices provide classification precision of predicted spikes and incorrect classification of predicted non-spikes. The effective classification forecasting is determined by high $P(in)$ and low $P(out)$.

$$P(in) = P(correctly\ predicted\ spike \mid predicted\ spike) \tag{4.5}$$

$$P(out) = P(incorrectly\ predicted\ nonspike \mid predicted\ nonspike) \quad (4.6)$$

SVC is used to calculate the spike price occurrence prediction in several models following these steps. First, due to the most significant impact of historical market prices determined by a strong autocorrelation, they are selected to run spike price occurrence prediction for four time lag models. Second, the classification calculates the prediction separately for three time lags of temperature and load profile combined with the model of best prediction performance from the first step. Previous research shows that the dependence of temperature and load profile can be either strong or weak. In addition, it can have a positive or negative correlation [69]. Because of the different correlations of these parameters, the final evaluation step considers all combinations to examine the possible further prediction improvement. Two-thirds of the data from each month in 2011 is employed for training, while the remaining one third is used for testing. The spike price occurrence forecasting results are reported in Table 4-1.

Table 4-1 Spike Market Price Occurrence Prediction Results

Models	*P(in)*	*P(out)*
mp(*t*-1)	0.73	0.0046
mp(*t*-1&*t*-2)	0.78	0.0046
mp(*t*-1,…,*t*-3)	0.77	0.0049
mp(*t*-1,…,*t*-4)	0.75	0.0049
mp(*t*-1&*t*-2)&*T*(*t*-1)	0.78	0.0046
mp(*t*-1&*t*-2)&*L*(*t*-1)	0.78	0.0052
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)	0.78	0.0046
mp(*t*-1&*t*-2)&*L*(*t*-1&*t*-2)	0.80	0.0048
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)	0.78	0.0046
mp(*t*-1&*t*-2)&*L*(*t*-1,…,*t*-3)	0.78	0.0049
mp(*t*-1&*t*-2)&*T*(*t*-1)&*L*(*t*-1)	0.78	0.0052
mp(*t*-1&*t*-2)&*T*(*t*-1)&*L*(*t*-1&*t*-2)	0.80	0.0048
mp(*t*-1&*t*-2)&*T*(*t*-1)&*L*(*t*-1,…,*t*-3)	0.78	0.0049
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1)	0.78	0.0052
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1&*t*-2)	0.80	0.0048
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1,…,*t*-3)	0.78	0.0049
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1)	0.78	0.0052
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1&*t*-2)	0.80	0.0056
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1,…,*t*-3)	0.78	0.0049

A significantly low $P(out)$ in Table 4-1 is a result of high non-spike and low spike price number compared to the total amount of testing data. Following the aforementioned procedure, the model of historical market prices is simulated from *mp*(*t*-1) until *mp*(*t*-1,..,*t*-4). Spike price occurrence prediction by the model *mp*(*t*-1&*t*-2) yields the best results compared to other models and has the highest $P(in)$ and lowest $P(out)$ of 0.78 and 0.0046, respectively. Then, this model combined with *L*(*t*-1&*t*-2) enhances classification performance and provides the most accurate model compared to the other combination models. This model is selected for spike price occurrence prediction in the HMPFM.

In order to improve classification performance, two adjustable parameters in SVC including regularization (*C*) and bandwidth (*B*) are tuned. The model from the previous step is used to modify both parameters. The initial settings for *C* and *B* are 10 and 2, respectively. The modification results are shown in Table 4-2 and Table 4-3. The modification of regularization and bandwidth significantly improves the classification accuracy. The most efficient parameter settings (*C* = 5000 and *B* = 20) elevate $P(in)$ to 0.85 and stabilize $P(out)$ at 0.0046.

Table 4-2 Spike Occurrence Prediction Results by Parameter Modification (*C*=10)

B	0.4	2	10	20	30	100
P(in)	0.71	0.80	0.80	0.81	0.79	0.78
P(out)	0.0061	0.0048	0.0049	0.0048	0.0049	0.0052

Table 4-3 Spike Occurrence Prediction Results by Parameter Modification (*B*=20)

C	0.1	10	100	1000	5000	10000
P(in)	0.78	0.81	0.83	0.84	0.85	0.84
P(out)	0.0053	0.0048	0.0048	0.0049	0.0046	0.0045

4.4.3 Non-spike Market Price Prediction

SVR is used to estimate the magnitude of non-spike prices in the same procedure as the spike occurrence prediction. The prediction performance is evaluated by Mean Absolute Percentage Error (MAPE) calculated by (4.7). The forecasting results are shown in Table 4-4.

$$MAPE = \frac{1}{N}\sum_{j=1}^{N}\frac{\left|P_j^{true} - P_j^{fst}\right|}{P_j^{true,N}} \tag{4.7}$$

where P_j^{true} is an actual market price at time j, P_j^{fst} is a forecasting market price at time j, and $P_j^{true,N}$ is an average of recorded market prices over N period.

Table 4-4 Non-spike Market Price Prediction Results

Models	*MAPE (%)*
mp(*t*-1)	6.02
mp(*t*-1&*t*-2)	5.94
mp(*t*-1,…,*t*-3)	5.95
mp(*t*-1,…,*t*-4)	6.02
mp(*t*-1&*t*-2)&*T*(*t*-1)	5.94
mp(*t*-1&*t*-2)&*L*(*t*-1)	6.02
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)	5.93
mp(*t*-1&*t*-2)&*L*(*t*-1&*t*-2)	5.94
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)	5.93
mp(*t*-1&*t*-2)&*L*(*t*-1,…,*t*-3)	5.96
mp(*t*-1&*t*-2)&*T*(*t*-1)& *L*(*t*-1)	6.00
mp(*t*-1&*t*-2)&*T*(*t*-1)& *L*(*t*-1&*t*-2)	5.93
mp(*t*-1&*t*-2)&*T*(*t*-1)& *L*(*t*-1,…,*t*-3)	5.94
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1)	6.00
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1&*t*-2)	5.92
mp(*t*-1&*t*-2)&*T*(*t*-1&*t*-2)&*L*(*t*-1,…,*t*-3)	5.93
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1)	5.98
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1&*t*-2)	5.92
mp(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1,…,*t*-3)	5.93

The results in Table 4-4 show the prediction performance of SVR. Temperature and load profile can slightly enhance the forecasting precision. The model of *mp*(*t*-1&*t*-2) including *T*(*t*-1&*t*-2) and *L*(*t*-1&*t*-2) offers the best result with 5.92 % MAPE compared to

the results of the other models. This model is selected in the HMPFM for non-spike price estimation.

4.4.4 Spike Market Price Prediction

Three clustering techniques consisting of CART, K-means, and Stratification methods are utilized to enhance market price prediction in the deregulated market. This section presents the clustering selection results of three proposed approaches prior to performing the comprehensive HMPFM in the next stage.

4.4.4.1 Classification and Regression Tree (CART)

CART employs a ten-fold cross validation considering historical market prices, temperatures, and load profiles as predictors and market price as a target. The minimum number of target values in the parent nodes can vary from 10 to 70, and the number of values in the terminal nodes should be one-third the number of their parent node, as recommended by the software [70]. The optimal results specify the most efficient number of different terminal nodes for each model. CART provides regression tree rules for each terminal node to determine the best set of clusters prior to performing spike prediction. Example regression tree rules of the model including $mp(t\text{-}1)$ and $T(t\text{-}1,\ldots,t\text{-}3)$ are shown in Table 4-5. For instance, the rule for the 6th cluster is $mp(t\text{-}1)$, which falls between 2086.89 and 3000.6 $/MWh.

Table 4-5 Example Regression Tree Rules obtained by CART

Terminal Nodes	Rules
1	$mp(t\text{-}1)<=816.95$ and $T(t\text{-}3)<=3.3$
2	$mp(t\text{-}1)<=816.95$ and $T(t\text{-}3)>3.3$ and $T(t\text{-}3)<=28.05$
3	$mp(t\text{-}1)<=275.22$ and $T(t\text{-}3)>28.05$
4	$mp(t\text{-}1)> 275.22$ and $mp(t\text{-}1)<=816.95$ and $T(t\text{-}3)>28.05$
5	$mp(t\text{-}1)> 816.95$ and $mp(t\text{-}1)<=2086.89$
6	$mp(t\text{-}1)>2086.89$ and $mp(t\text{-}1)<=3000.66$
7	$mp(t\text{-}1)>3000.66$

4.4.4.2 K-means

K-means clustering is also used to determine the best set of clusters and yields different input parameters for each cluster. Then, the averages of input parameters are calculated and are used as the decision values in each cluster. The lowest Euclidean distance calculated by (4.8) is used to select the best set of clusters prior to predicting the magnitude of the spike price. An example result from K-means of the model including *mp*(*t*-1,…,*t*-3) is shown in Table 4-6.

$$d_n = \sqrt{\sum_{t=-1}^{-T} (X_{predicting(t)} - Y_{n(t)})^2} \tag{4.8}$$

where d_n is a Euclidean distance for n^{th} cluster, X is an input parameter value, Y is an average decision value, and T is a parameter at each several t prior times.

According to Table 4-6, it can be seen that K-means clustering is able to separate input parameters for each cluster effectively. All average decision values of the input parameters are less than 282.13 and more than 2762.85 $/MWh in clusters 1 and 3, respectively. In addition, the average decision values of the input parameters in cluster 2 show an increasing trend, while they show a decreasing trend in cluster 4. The suitable number of clusters is discussed in the comprehensive results.

Table 4-6 4 Clusters by K-means

Cluster	Cluster 1			Cluster 2		
Average decision values	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)
	169.90	199.05	282.13	520.67	1042.90	2144.89
Cluster	Cluster 3			Cluster 4		
Average decision values	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)
	2762.85	2929.57	2977.58	2655.70	2043.38	1059.57

4.4.4.3 Stratification

The Stratification method equally dissects the number of cluster members based on the total spike price number. According to different levels of spike prices specified by dissection, each input parameter is individually used to calculate the averages as the decision values at different times, such as *mp*(*t*-1), *T*(*t*-1), and others. The lowest Euclidean distance defined by (4.8) is employed to select the best set of clusters before calculating the prediction. Example results of four clusters of models including *mp*(*t*-1,…,*t*-3) are shown in Table 4-7. The proper number of clusters is discussed in the next section.

Table 4-7 4 clusters by Stratification Method

Cluster (no.of spike price)	Cluster 1 (66)			Cluster 2 (65)		
Range of spike price ($/MWh)	[-250,300)			[300-550)		
Average decision values	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)
	141.66	184.34	250.40	435.48	379.74	371.56
Group (no.of spike)	Cluster 3 (69)			Cluster 4 (72)		
Range of pike price ($/MWh)	[550,2000)			[2000,3500)		
Average decision values	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)	*mp*(*t*-3)	*mp*(*t*-2)	*mp*(*t*-1)
	634.93	661.46	739.53	2045.60	2270.62	2534.08

4.4.5 Comprehensive Results

The selected models from spike occurrence prediction and non-spike market price prediction are used to perform the HMPFM combined with the three proposed clustering techniques. CART is an algorithm that computationally assigns the optimal number of clusters, while the preliminary numbers of clusters for K-means and Stratification methods are set at four. Following the similar procedure for spike occurrence and non-spike price magnitude prediction, the results of spike price

magnitude prediction are obtained from the comprehensive HMPFM tested with several models, as reported in Table 4-8. The prediction performance is evaluated by MAPE.

Table 4-8 Comprehensive Market Prices Forecasting Results (MAPE %)

Models	CART	Models	K-means	Models	Stratification
mp(*t*-1)	15.65	*mp*(*t*-1)	15.86	*mp*(*t*-1)	16.00
mp(*t*-1&*t*-2)	15.76	*mp*(*t*-1&*t*-2)	16.69	*mp*(*t*-1&*t*-2)	15.68
mp(*t*-1,…,*t*-3)	15.87	*mp*(*t*-1,…,*t*-3)	15.75	*mp*(*t*-1,…,*t*-3)	16.30
mp(*t*-1,…,*t*-4)	15.87	*mp*(*t*-1,…,*t*-4)	15.83	*mp*(*t*-1,…,*t*-4)	16.17
mp(*t*-1) &*T*(*t*-1)	16.37	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1)	15.32	*mp*(*t*-1&*t*-2) &*T*(*t*-1)	16.55
mp(*t*-1) &*L*(*t*-1)	16.63	*mp*(*t*-1,…,*t*-3) &*L*(*t*-1)	16.28	*mp*(*t*-1&*t*-2) &*L*(*t*-1)	16.56
mp(*t*-1) &*T*(*t* -1&*t*-2)	16.09	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1&*t*-2)	15.40	*mp*(*t*-1&*t*-2) &*T*(*t*-1&*t*-2)	16.45
mp(*t*-1) &*L*(*t*-1&*t*-2)	16.50	*mp*(*t*-1,…,*t*-3) &*L*(*t*-1&*t*-2)	15.17	*mp*(*t*-1&*t*-2) &*L*(*t*-1&*t*-2)	16.48
mp(*t*-1) &*T*(*t*-1,…,*t*-3)	15.86	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1,…,*t*-3)	15.32	*mp*(*t*-1&*t*-2) &*T*(*t*-1,…,*t*-3)	16.41
mp(*t*-1) &*L*(*t*-1,…,*t*-3)	15.30	*mp*(*t*-1,…,*t*-3) &*L*(*t*-1,…,*t*-3)	15.19	*mp*(*t*-1&*t*-2) &*L*(*t*-1,…,*t*-3)	16.43
mp(*t*-1) &*T*(*t*-1)&*L*(*t*-1)	15.28	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1)&*L*(*t*-1)	15.21	*mp*(*t*-1&*t*-2) &*T*(*t*-1)&*L*(*t*-1)	16.41
mp(*t*-1) &*T*(*t*-1) &*L*(*t*-1&*t*-2)	15.25	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1) &*L*(*t*-1&*t*-2)	15.50	*mp*(*t*-1&*t*-2) &*T*(*t*-1) &*L*(*t*-1&*t*-2)	16.37
mp(*t*-1) &*T*(*t*-1) &*L*(*t*-1,…,*t*-3)	15.28	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1) &*L*(*t*-1,…,*t*-3)	15.19	*mp*(*t*-1&*t*-2) &*T*(*t*-1) &*L*(*t*-1,…,*t*-3)	16.36
mp(*t*-1) &*T*(*t*-1&*t*-2) &*L*(*t*-1)	15.31	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1&*t*-2) &*L*(*t*-1)	15.26	*mp*(*t*-1&*t*-2) &*T*(*t*-1&*t*-2) &*L*(*t*-1)	16.36
mp(*t*-1) &*T*(*t*-1&*t*-2) &*L*(*t*-1&*t*-2)	15.42	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1&*t*-2) &*L*(*t*-1&*t*-2)	15.24	*mp*(*t*-1&*t*-2) &*T*(*t*-1&*t*-2) &*L*(*t*-1&*t*-2)	16.34
mp(*t*-1) &*T*(*t*-1&*t*-2) &*L*(*t*-1,…,*t*-3)	15.35	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1&*t*-2) &*L*(*t*-1,…,*t*-3)	15.86	*mp*(*t*-1&*t*-2) &*T*(*t*-1&*t*-2) &*L*(*t*-1,…,*t*-3)	16.33
mp(*t*-1) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1)	15.32	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1)	15.51	*mp*(*t*-1&*t*-2) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1)	16.35
mp(*t*-1) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1&*t*-2)	15.34	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1&*t*-2)	15.88	*mp*(*t*-1&*t*-2) &*T*(*t*-1,…,*t*-3) &*L*(*t*-1&*t*-2)	16.34
mp(*t*-1) &*T*(*t*-1,..,*t*-3) &*L*(*t*-1,..,*t*-3)	15.29	*mp*(*t*-1,…,*t*-3) &*T*(*t*-1,..,*t*-3) &*L*(*t*-1,..,*t*-3)	15.32	*mp*(*t*-1&*t*-2) &*T*(*t*-1,..,*t*-3) &*L*(*t*-1,..,*t*-3)	16.33

According to Table 4-8, each model provides similar market price forecasting accuracy by the three proposed clustering approaches. The models with the best performance of spike price forecasting from the comprehensive HMPFM are *mp*(*t*-1)&*T*(*t*-1)&*L*(*t*-1&*t*-2), *mp*(*t*-1,…,*t*-3)&*L*(*t*-1&*t*-2), and *mp*(*t*-1&*t*-2) for the CART, K-means, and Stratification methods, respectively. These predictions have the lowest MAPE with 15.25 %, 15.17 %, and 15.68 %, respectively.

In addition, the number of clusters is adjusted from two to six with different temperature and load profile combinations to compare and select the optimal results from the K-means and Stratification methods. These results are illustrated in Figure 4-3 and Figure 4-4 . The maximum number of six clusters is chosen for ensuring sufficient data in each group. The model with the best prediction performance for K-means is the same as previously discussed, while two clusters of the model *mp*(*t*-1&*t*-2)&*T*(*t*-1,…,*t*-3)&*L*(*t*-1,…,*t*-3) for the Stratification method gives the best prediction performance and improves MAPE from 15.68 % to 15.19 %.

Figure 4-3 Comparison of market price prediction results with different clusters of K-means method

Figure 4-4 Comparison of market price prediction results with different clusters of Stratification method

To illustrate the significant improvement of the HMPFM combined with the three clustering techniques against other prediction methods, Figure 4-5 depicts the comparison results between the best cases of the three proposed approaches and the other general prediction methods including normal SVM (NSVM) and typical hybrid SVM (THSVM). Considering the same parameters and data as the proposed approaches, NSVM is used to forecast the market price with only the traditional SVM algorithm, and THSVM is used to forecast the market price with the hybrid of SVC and SVR but without the proposed data clustering techniques. These two methods are programmed with MATLAB. MAPEs are significantly reduced from 20.59 % and 16.95 % by NSVM and THSVM to 15.25 %, 15.17 %, and 15.19 % by the best prediction performance models of HMPFM with the CART, K-means, and Stratification methods, respectively

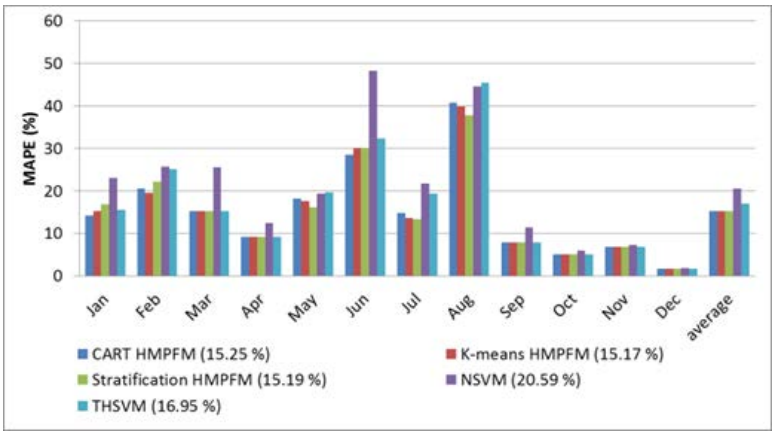
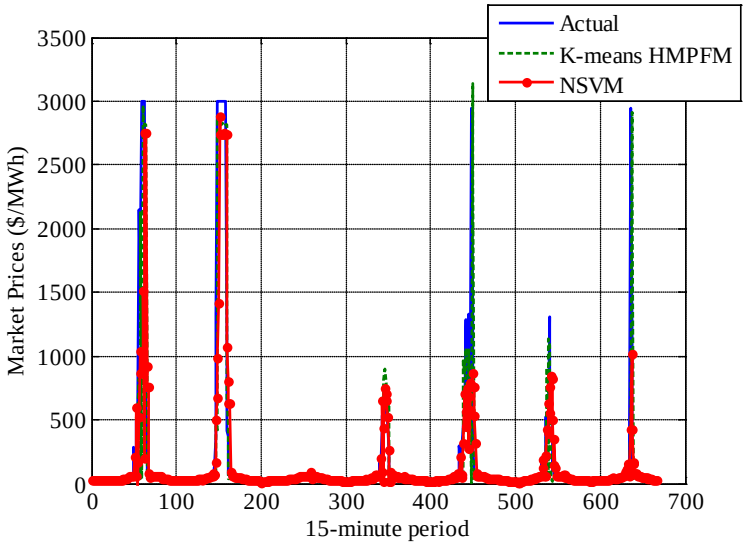


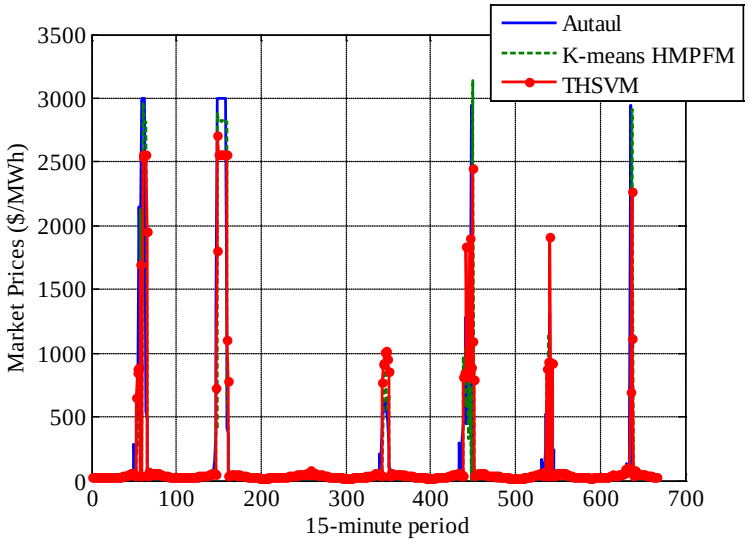
Figure 4-5 Market prices forecasting comparison results of various approaches.

As is also shown in Figure 4-6, since K-means HMPFM with four clusters of *mp*(*t*-1,…,*t*-3)&*L*(*t*-1&*t*-2) gives the most accurate result compared to the other two proposed data clustering techniques, it is selected as the proposed method in this dissertation. Figure 4-6 compares the results of this method with NSVM and THSVM. These three prediction methods yield comparable and satisfactory results of non-spike price estimation. However, Figure 4-6 (a) shows that while the NSVM is not able to predict the spike price accurately, the proposed approach can efficiently predict spike price occurrence and its magnitude. In addition, spike price magnitude prediction by THSVM has more error than the forecasting by K-means HMPFM, as depicted in Figure 4-6 (b).

Since ERCOT wholesale market prices of different clusters are different, it is necessary to verify the prediction performance of K-means HMPFM for all power nodes in the DFW area where PEV charging station systems will be built. The market prices for all power nodes are predicted by the proposed method and the prediction results are shown in Figure 4-7. The proposed method yields similar results for all power nodes with an average MAPE of 16.11 %.



(a)



(b)

Figure 4-6 Comparison of market price prediction results from

the proposed K-means hybrid SVM (a) with normal SVM (b) with typical hybrid SVM

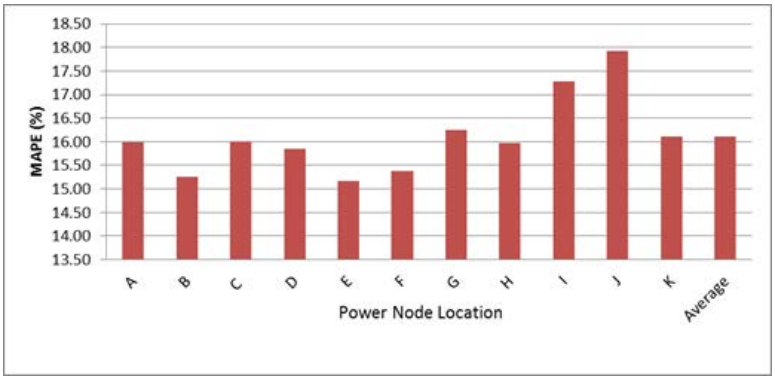


Figure 4-7 Market price prediction results from K-means HMPFM for all power nodes in the DFW area.

4.5 Summary

This study presents a novel HMPFM with three data clustering techniques, CART, K-means, and Stratification, to improve the accuracy of the wholesale electric price prediction in the deregulated market. The selected input models for SVMs in spike price occurrence and non-spike and spike price magnitude estimations consider three historical impact parameters consisting of market price, temperature, and load profile. The proposed K-means HMPFM shows the most effective prediction performance validated by the ERCOT wholesale market prices in the DFW area. This proposed approach significantly improves the prediction accuracy compared to general market price prediction approaches.

Chapter 5

Forecast Uncertainty Analysis

5.1 Literature Review

Although the proposed wind/PV generation and market price predictions in this dissertation yield satisfactory forecasting accuracy, the presence of prediction error is inevitable. The investigation of the prediction error can help the charging station system operators improve the optimal operation of charging station systems. In the recent years, many studies have investigated forecast uncertainty modeling for different applications.

In the literature, the forecast uncertainty analysis can be classified into two categories: the typical approach and the Martingale Model Forecast Evolution (MMFE)[71-76]. The typical approach uses the probability density function (pdf) analysis and then generates the randomness of forecast error by the Monte-Carlo simulation based on that analysis. For example, a study using this approach to evaluate the wind power forecast uncertainty is presented in [71]. This method is also implemented in [72] to characterize the stochastic nature of the wind power generation and electrical load predictions.

As introduced by [73], MMFE is applied to a safety stock analysis in production and distribution applications. Also, reservoir operations utilize MMFE to generate the steam flow forecast uncertainty [74, 75]. To solve the problem of stochastic dynamic programming for inventory, the demand forecast uncertainty can be characterized by MMFE [76].

The Monte-Carlo simulation based on the pdf analysis approach cannot investigate multiple related forecast uncertainties, while the advantage of the MMFE approach is the consideration of multivariate random vectors for the forecast changes with the exploration of their correlations. Due to the advantage of its multivariate

consideration, the MMFE is adopted for investigating the uncertainty of the proposed forecasting in this dissertation.

5.2 MMFE Framework

The MMFE framework is illustrated in Figure 5-1. There are five steps to analyze the forecast error and generate the randomness of the forecast uncertainty based on that analysis. The details of these five steps are described as follows.

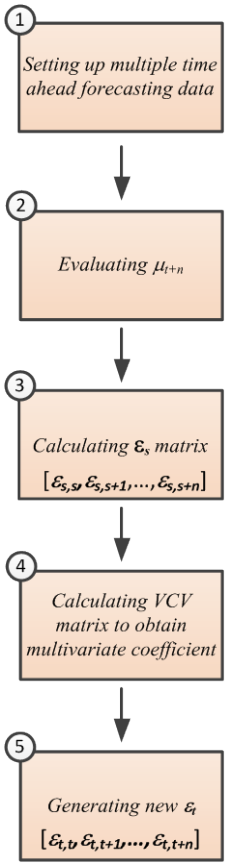


Figure 5-1 MMFE framework

5.2.1 Multiple Times Ahead Forecasting Selection

To select the suitable multiple time ahead predictions for investigating the correlation of the multiple time ahead forecast changes, this dissertation assumes the

initial number of multiple time ahead predictions is the same as the number of time lags of the best prediction performance model proposed in the previous chapters. However, although the number of these multiple time ahead predictions can be modified, the MMFE process still functions in the same framework.

More specifically, the best prediction model for the wind power forecasting is *WS*(*t*-1)&*WP*(*t*-1&*t*-2), which is two time lags of wind power, so one-period and two-period ahead estimations for the wind power generation are calculated. In addition, regarding the best prediction model identified in the previous chapters, one and two time ahead PV generation predictions are calculated, as well as one, two, and three time ahead market price predictions.

Actual historical data of wind/PV generation and market price are used to obtain the multiple time ahead predictions, defined as $D_s = [D_{s+1,s}, D_{s,s}, D_{s,s+1}, ..., D_{s,s+n}]$.

where $D_{s+1,s}$ is an actual value realized for time period *s*+1 (the end of time period *s*) from the beginning of time period *s,*

$D_{s,s}$ is a forecasting value for the end of time period *s* made at the beginning of time period *s* (One time ahead forecasting),

$D_{s,s+1}$ is a forecasting value for time period *s*+1 made at the beginning of time period *s* (Two time ahead forecasting), and

$D_{s,s+n}$ is a forecasting value for time period *s*+*n* made at the beginning of time period *s* (*n*+1 time ahead forecasting).

5.2.2 μ_{s+n} Evaluation

μ_{s+n} is the mean initial forecast at time *s*+*n* in the MMFE process. In this dissertation, the process for obtaining μ_{s+n} is by calculating the one time ahead

predictions repeatedly until μ_{s+n} is obtained. For example, the calculation of μ_{s+2} is shown in Figure 5-2. First, one time ahead prediction is calculated. Then, the next one time ahead prediction is calculated with the dynamical forecast update from the first prediction ($D_{s,s}$). This process is performed iteratively until obtaining μ_{s+2}.

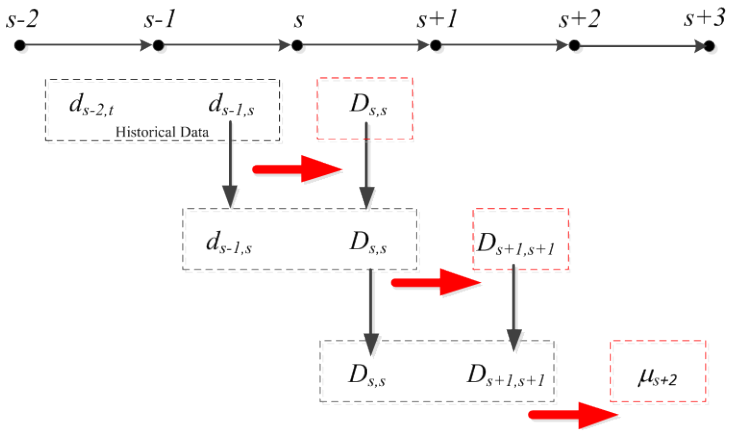


Figure 5-2 μ_{s+2} calculation

5.2.3 Forecast Change Matrix

Due to the associated forecast changes in the MMFE, the correlation between the variables for evaluating these forecast changes needs to be analyzed in order to construct an appropriate forecast change matrix. Considering the variables for the predictions of the regional PEV charging station application in this dissertation, the wind/PV generation and market price correlations are calculated by (2.10). Figure 5-3 depicts three scatter plots of these three parameters. It can be observed that these three parameters demonstrate weak correlations, as reported in Table 5-1. Consequently, the forecast change matrices for these three predictions are characterized separately.

Table 5-1 Correlation analysis between wind/PV generation and market price

Parameter	*Correlation Coefficients*
Wind generation - PV generation	-0.105
Wind generation - Market price	-0.054
PV generation - Market price	0.103

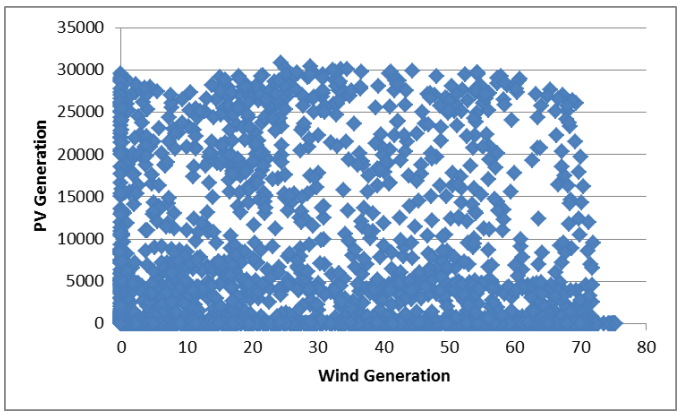


Figure 5-3 Scatter plot of Wind generation-PV generation

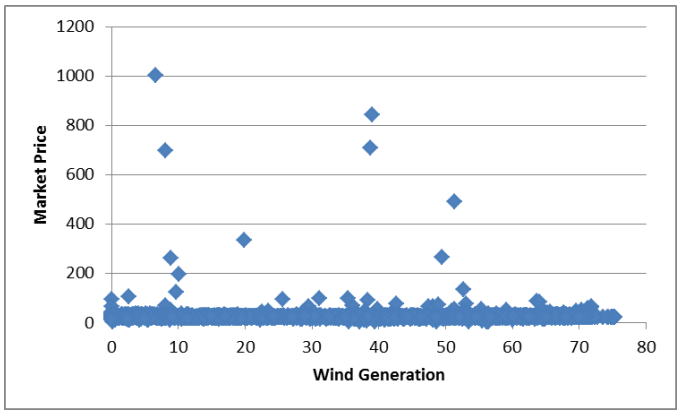


Figure 5-4 Scatter plot of Wind generation-Market price

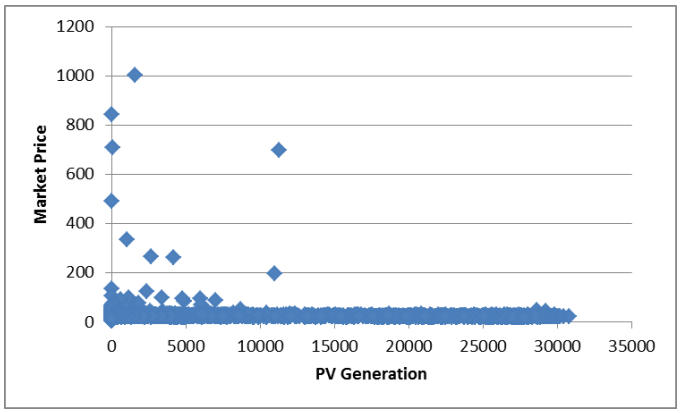


Figure 5-5 Scatter plot of PV generation-Market price

Because the forecast changes in each prediction have obvious strong correlations with each other, the construction of the forecast change matrix is $\varepsilon_s = [\varepsilon_{s,s}, \varepsilon_{s,s+1}, ..., \varepsilon_{s,s+n}]$. Each forecast change vector can be formulated by the log normal function and adjusting factors (5.1). The adjusting factor is a number that corrects the mean of forecast changes to 1, following MMFE requirements [77].

$$\varepsilon_{s,s} = \ln(\frac{D_{s+1,s}}{D_{s,s}}) / adjusting\ \ factor\ 1$$

$$\varepsilon_{s,s+1} = \ln(\frac{D_{s+1,s+1}}{D_{s,s+1}}) / adjusting\ \ factor\ 2 \qquad (5.1)$$

$$\varepsilon_{s,s+n} = \ln(\frac{D_{s+1,s+n}}{\mu_{s+n}}) / adjusting\ \ factor\ n+1$$

Based on the formulas (5.1), two time ahead predictions have three forecast changes $[\varepsilon_{s,s}, \varepsilon_{s,s+1}, \varepsilon_{s,s+2}]$, and three time ahead predictions have four forecast changes $[\varepsilon_{s,s}, \varepsilon_{s,s+1}, \varepsilon_{s,s+2}, \varepsilon_{s,s+3}]$. The wind/PV generation predictions are calculated at only one location in the DFW metroplex, and market price forecasting is executed for 11 power nodes at 11 different locations in this metro area. Therefore, the forecast

change matrices of wind, PV generation, and market price according to the multiple time ahead predictions in the first step are shown as follows:

$$\varepsilon_s^W = [\varepsilon_{s,s}^W, \varepsilon_{s,s+1}^W, \varepsilon_{s,s+2}^W], \varepsilon_s^{PV} = [\varepsilon_{s,s}^{PV}, \varepsilon_{s,s+1}^{PV}, \varepsilon_{s,s+2}^{PV}],$$

$$\varepsilon_s^{MP} = [\varepsilon_{s,s}^{MP1}, ..., \varepsilon_{s,s}^{MP11} \left| \varepsilon_{s,s+1}^{MP1}, ..., \varepsilon_{s,s+1}^{MP11} \right| \varepsilon_{s,s+2}^{MP1}, ..., \varepsilon_{s,s+2}^{MP11} \left| \varepsilon_{s,s+3}^{MP1}, ..., \varepsilon_{s,s+3}^{MP11} \right]$$

5.2.4 Variance Covariance Matrix (VCV)

After modeling the forecast change matrices, their VCV matrices can be obtained. Then, the eigenvalues ($U = [\lambda_1, \lambda_2, .., \lambda_n]$) and eigenvectors ($D = [u_1, u_2, ..., u_n]$) of these VCV matrices are calculated by the principle component analysis. The dimensions of the VCV matrices are 3 by 3 for wind/PV generation and 44 by 44 for market price, according to the forecast change matrices from the previous step. Finally, C , the multivariate coefficient, is calculated by (5.2) such that $CC' = VCV$.

$$C = UD^{1/2} = [\sqrt{\lambda_1} u_1, \sqrt{\lambda_2} u_2, ..., \sqrt{\lambda_n} u_n] \quad (5.2)$$

5.2.5 New Forecast Change Matrix Generation

The probability density function (pdf) of the forecast changes has to be analyzed to generate the randomness of the forecast changes based on MMFE. This dissertation uses ARENA [78] to investigate the pdf of the forecast changes because of its ease of use with comprehensive distribution models. The results of ε_s^W and ε_s^{PV} for the pdf evaluated by the square error are reported in Table 5-2. In addition, Table 5-3 to Table 5-8 report the results of the pdf analysis of ε_s^{MP} for 11 power nodes.

Table 5-2 Probability Density Function Analysis of ε_s^W and ε_s^{PV}

Distribution function	$\varepsilon_{s,s}^W$	$\varepsilon_{s,s+1}^W$	$\varepsilon_{s,s+2}^W$	$\varepsilon_{s,s}^{PV}$	$\varepsilon_{s,s+1}^{PV}$	$\varepsilon_{s,s+2}^{PV}$
Normal	0.1	0.23	0.05	0.393	0.234	0.068
Lognormal	0.154	0.29	0.058	0.453	0.234	0.087
Weibull	0.502	0.428	0.081	0.399	0.325	0.091
Beta	0.101	0.438	0.051	0.397	0.235	0.07
Erlang	0.206	0.486	0.11	0.532	0.338	0.161
Gamma	0.206	0.486	0.11	0.532	0.338	0.161
Triangular	0.273	0.533	0.169	0.604	0.455	0.209
Uniform	0.309	0.572	0.2	0.642	0.492	0.245
Exponential	0.32	0.585	0.212	0.653	0.498	0.258

Table 5-3 Probability Density Function Analysis of ε_s^{MP1} and ε_s^{MP2}

Distribution function	$\varepsilon_{s,s}^{MP1}$	$\varepsilon_{s,s+1}^{MP1}$	$\varepsilon_{s,s+2}^{MP1}$	$\varepsilon_{s,s+3}^{MP1}$	$\varepsilon_{s,s}^{MP2}$	$\varepsilon_{s,s+1}^{MP2}$	$\varepsilon_{s,s+2}^{MP2}$	$\varepsilon_{s,s+3}^{MP2}$
Normal	0.334	0.201	0.114	0.076	0.328	0.265	0.18	0.059
Lognormal	0.357	0.218	0.147	0.195	0.351	0.291	0.22	0.171
Weibull	1.06	0.925	0.157	0.22	1.09	1.04	0.875	0.056
Beta	0.661	0.457	0.372	0.306	0.648	0.534	0.459	0.453
Erlang	0.635	0.505	0.42	0.331	0.622	0.581	0.507	0.316
Gamma	0.635	0.505	0.42	0.331	0.622	0.581	0.507	0.316
Triangular	0.686	0.567	0.482	0.405	0.673	0.644	0.57	0.39
Uniform	0.725	0.606	0.521	0.444	0.712	0.683	0.609	0.428
Exponential	0.739	0.618	0.534	0.455	0.725	0.696	0.621	0.44

Table 5-4 Probability Density Function Analysis of ε_s^{MP3} and ε_s^{MP4}

Distribution function	$\varepsilon_{s,s}^{MP3}$	$\varepsilon_{s,s+1}^{MP3}$	$\varepsilon_{s,s+2}^{MP3}$	$\varepsilon_{s,s+3}^{MP3}$	$\varepsilon_{s,s}^{MP4}$	$\varepsilon_{s,s+1}^{MP4}$	$\varepsilon_{s,s+2}^{MP4}$	$\varepsilon_{s,s+3}^{MP4}$
Normal	0.334	0.159	0.142	0.096	0.337	0.273	0.165	0.072
Lognormal	0.358	0.183	0.179	0.218	0.356	0.307	0.201	0.134
Weibull	1.05	0.762	0.87	0.681	1.05	0.916	0.86	0.252
Beta	0.657	0.39	0.41	0.323	0.654	0.547	0.426	0.316
Erlang	0.631	0.438	0.459	0.346	0.627	0.593	0.474	0.366
Gamma	0.631	0.438	0.459	0.347	0.628	0.593	0.474	0.366
Triangular	0.682	0.5	0.521	0.421	0.679	0.656	0.536	0.426
Uniform	0.721	0.539	0.56	0.46	0.718	0.695	0.575	0.464
Exponential	0.734	0.551	0.573	0.471	0.731	0.708	0.588	0.477

Table 5-5 Probability Density Function Analysis of ε_s^{MP5} and ε_s^{MP6}

Distribution function	$\varepsilon_{s,s}^{MP5}$	$\varepsilon_{s,s+1}^{MP5}$	$\varepsilon_{s,s+2}^{MP5}$	$\varepsilon_{s,s+3}^{MP5}$	$\varepsilon_{s,s}^{MP6}$	$\varepsilon_{s,s+1}^{MP6}$	$\varepsilon_{s,s+2}^{MP6}$	$\varepsilon_{s,s+3}^{MP6}$
Normal	0.331	0.219	0.124	0.066	0.338	0.211	0.116	0.076
Lognormal	0.353	0.237	0.16	0.182	0.363	0.229	0.149	0.193
Weibull	1.02	0.25	0.875	0.045	1.05	0.843	0.85	0.788
Beta	0.657	0.471	0.379	0.303	0.663	0.465	0.373	0.307
Erlang	0.63	0.519	0.428	0.33	0.636	0.513	0.421	0.332
Gamma	0.63	0.519	0.428	0.33	0.636	0.513	0.422	0.332
Triangular	0.681	0.581	0.49	0.405	0.687	0.575	0.483	0.407
Uniform	0.72	0.621	0.529	0.443	0.727	0.614	0.523	0.445
Exponential	0.734	0.633	0.541	0.455	0.74	0.627	0.535	0.457

Table 5-6 Probability Density Function Analysis of ε_s^{MP7} and ε_s^{MP8}

Distribution function	$\varepsilon_{s,s}^{MP7}$	$\varepsilon_{s,s+1}^{MP7}$	$\varepsilon_{s,s+2}^{MP7}$	$\varepsilon_{s,s+3}^{MP7}$	$\varepsilon_{s,s}^{MP8}$	$\varepsilon_{s,s+1}^{MP8}$	$\varepsilon_{s,s+2}^{MP8}$	$\varepsilon_{s,s+3}^{MP8}$
Normal	0.344	0.22	0.128	0.063	0.336	0.21	0.103	0.054
Lognormal	0.366	0.243	0.166	0.17	0.355	0.23	0.141	0.169
Weibull	1.05	0.856	0.882	0.468	1.04	0.854	0.845	0.511
Beta	0.663	0.477	0.4	0.784	0.66	0.469	0.367	0.291
Erlang	0.638	0.525	0.449	0.33	0.633	0.517	0.416	0.32
Gamma	0.638	0.525	0.449	0.33	0.634	0.518	0.416	0.32
Triangular	0.689	0.588	0.511	0.405	0.685	0.58	0.478	0.394
Uniform	0.728	0.627	0.55	0.443	0.724	0.619	0.517	0.433
Exponential	0.741	0.639	0.562	0.455	0.737	0.631	0.53	0.444

Table 5-7 Probability Density Function Analysis of ε_s^{MP9} and ε_s^{MP10}

Distribution function	$\varepsilon_{s,s}^{MP9}$	$\varepsilon_{s,s+1}^{MP9}$	$\varepsilon_{s,s+2}^{MP9}$	$\varepsilon_{s,s+3}^{MP9}$	$\varepsilon_{s,s}^{MP10}$	$\varepsilon_{s,s+1}^{MP10}$	$\varepsilon_{s,s+2}^{MP10}$	$\varepsilon_{s,s+3}^{MP10}$
Normal	0.39	0.228	0.197	0.102	0.331	0.252	0.242	0.28
Lognormal	0.395	0.262	0.231	0.189	0.36	0.313	0.336	0.356
Weibull	0.904	0.744	0.85	0.232	1.13	0.441	0.86	0.928
Beta	0.678	0.436	0.442	0.307	0.6	0.492	0.494	0.589
Erlang	0.654	0.483	0.49	0.356	0.647	0.539	0.541	0.631
Gamma	0.655	0.484	0.49	0.356	0.647	0.54	0.542	0.631
Triangular	0.706	0.545	0.553	0.417	0.71	0.602	0.604	0.695
Uniform	0.745	0.584	0.592	0.456	0.749	0.641	0.644	0.735
Exponential	0.758	0.597	0.604	0.468	0.762	0.654	0.656	0.747

Table 5-8 Probability Density Function Analysis of ε_s^{MP11}

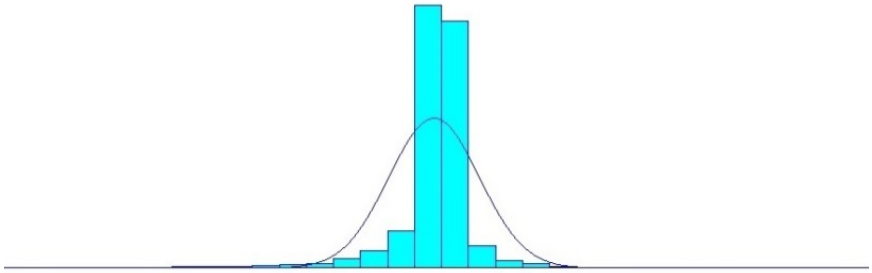
Distribution function	$\varepsilon_{s,s}^{MP11}$	$\varepsilon_{s,s+1}^{MP11}$	$\varepsilon_{s,s+2}^{MP11}$	$\varepsilon_{s,s+3}^{MP11}$
Normal	0.335	0.216	0.117	0.084
Lognormal	0.361	0.233	0.151	0.202
Weibull	1.05	0.858	0.881	0.816
Beta	0.66	0.469	0.375	0.314
Erlang	0.634	0.517	0.424	0.337
Gamma	0.634	0.517	0.424	0.337
Triangular	0.685	0.58	0.486	0.412
Uniform	0.724	0.619	0.525	0.451
Exponential	0.737	0.631	0.537	0.462

According to the results, most of the forecast changes are fit to the normal distribution function with the lowest square errors. Only $\varepsilon_{s,s+3}^{MP2}$ and $\varepsilon_{s,s+3}^{MP5}$ have the best fit with the Weibull distribution function; however, the square errors of the Weibull distribution function for these two forecast changes are close to those of the normal distribution function. Figure 5-6, Figure 5-7, and Figure 5-8 illustrate the fitting results for the normal distribution of ε_s^W , ε_s^{PV} and ε_s^{MP5} from ARENA.

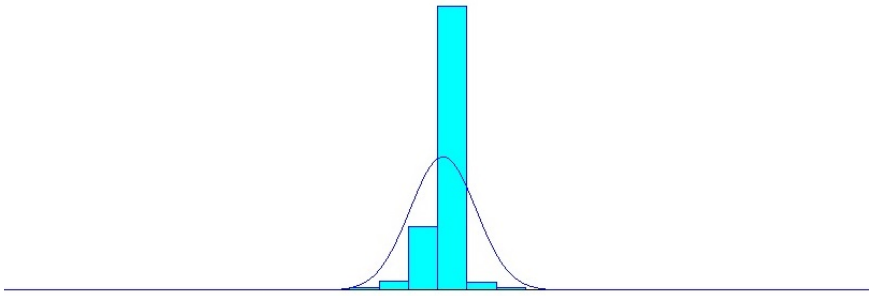
Though the study in [79] indicates that the forecast error has a better distribution fit with the Cauchy distribution function, the simulation in R confirms the null-hypothesis of the Cauchy distribution function for all of the forecast changes in this dissertation. As a result, it can be concluded that all of the forecast changes for the wind/PV generation and market price in this dissertation correspond to the normal distribution function; therefore, they follow the multivariate random normal vector based on MMFE. Hence, the new forecast change matrices (ε_t) can be generated by (5.3).

$$\varepsilon_t = C_1 z_1 + C_2 z_2 + ... + C_n z_n + \mu_i \tag{5.3}$$

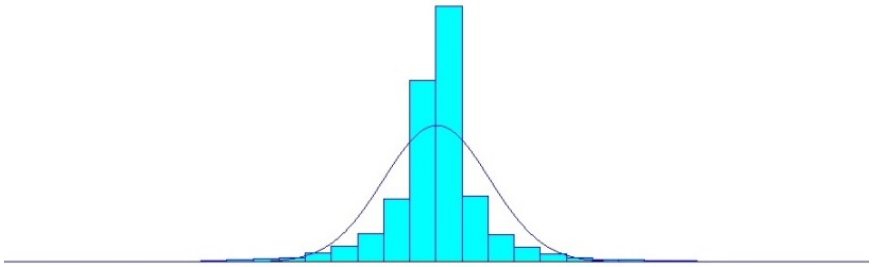
where C is a multivariate coefficient from the fourth step, z is an independent standard random normal variable, and μ_i is the mean value of historical ε_s.



(a) $\varepsilon_{s,s}^{W}$

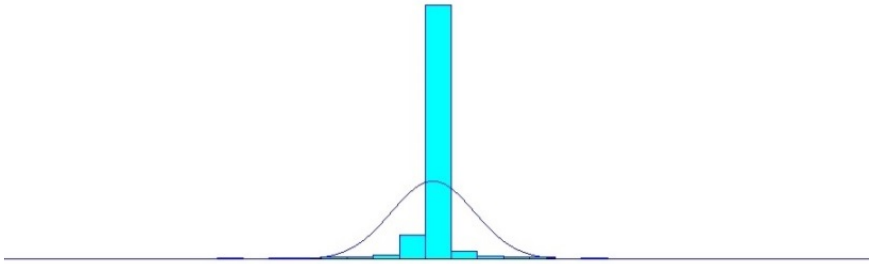


(b) $\varepsilon_{s,s+1}^{W}$

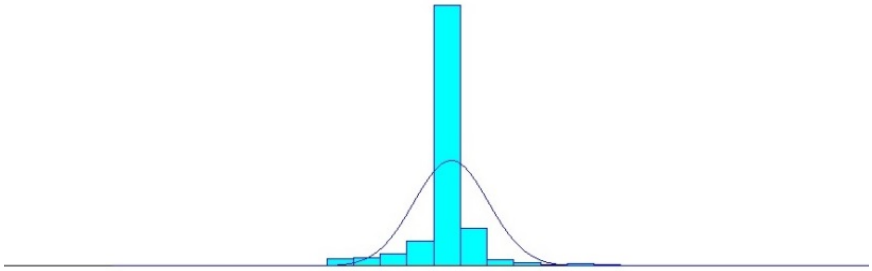


(c) $\varepsilon_{s,s+2}^{W}$

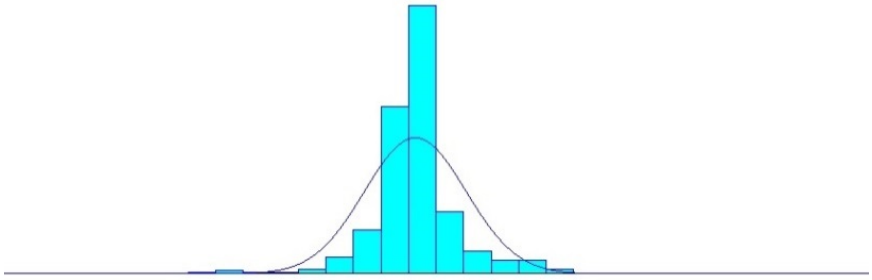
Figure 5-6 Fitting results of wind generation forecast change ε_s^W



(a) $\varepsilon_{s,s}^{PV}$

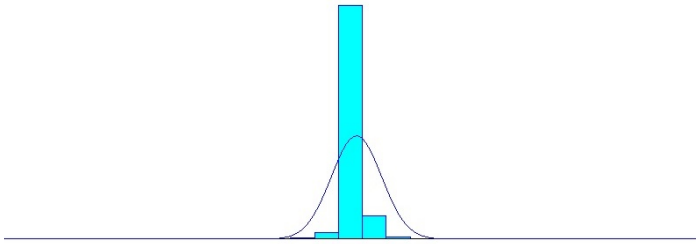


(b) $\varepsilon_{s,s+1}^{PV}$

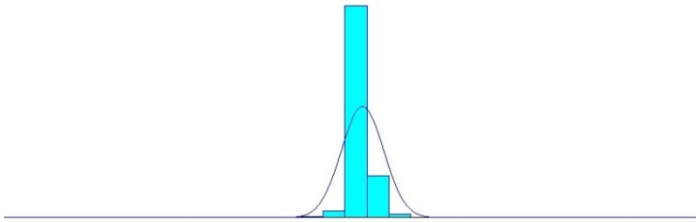


(c) $\varepsilon_{s,s+2}^{PV}$

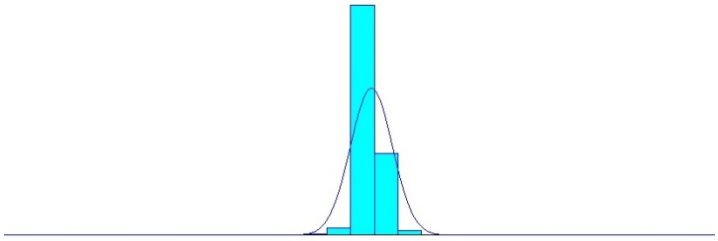
Figure 5-7 Fitting results of PV generation forecast change ε_{s}^{PV}



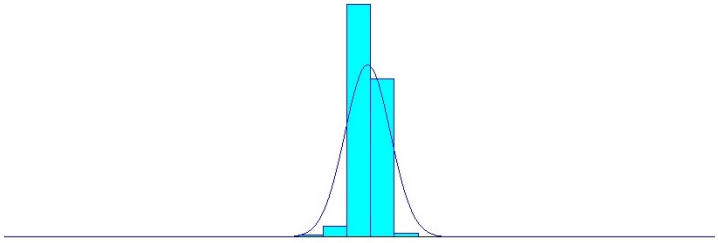
(a) $\varepsilon_{s,s}^{MP}$



(b) $\varepsilon_{s,s+1}^{MP}$



(c) $\varepsilon_{s,s+2}^{MP}$



(d) $\varepsilon_{s,s+3}^{MP}$

Figure 5-8 Fitting results of market price forecast change ε_s^{MP5}

To show the effectiveness of the MMFE approach, Table 5-9 reports the mean and standard deviations of the normal distribution function comparisons between ε_s^W, ε_s^{PV} and ε_t^W, ε_t^{PV}. Also, the comparisons of the mean and standard deviations between the actual forecast changes and the new forecast changes using this proposed method for the 11 locations of market price predictions are reported in Table 5-10 to Table 5-15. It can be observed that the mean and standard deviations of the normal distribution function of the actual forecast changes and the new forecast changes using the MMFE approach are similar for all of the predictions. These results demonstrate the effectiveness of the proposed method for investigating forecast uncertainty by analyzing the correlations between wind/PV generation and market price, and for generating the randomness of forecast change based on this investigation.

Table 5-9 Characteristics Comparison between ε_s^W, ε_s^{PV} and ε_t^W, ε_t^{PV}

Actual Forecast Changes	$\varepsilon_{s,s}^W$	$\varepsilon_{s,s+1}^W$	$\varepsilon_{s,s+2}^W$	$\varepsilon_{s,s}^{PV}$	$\varepsilon_{s,s+1}^{PV}$	$\varepsilon_{s,s+2}^{PV}$
Mean	-0.128	-0.017	-0.292	-0.112	-0.136	-0.567
Standard deviation	0.461	0.190	0.609	0.358	0.501	0.646
New Forecast Changes by MMFE	$\varepsilon_{t,t}^W$	$\varepsilon_{t,t+1}^W$	$\varepsilon_{t,t+2}^W$	$\varepsilon_{t,t}^{PV}$	$\varepsilon_{t,t+1}^{PV}$	$\varepsilon_{t,t+2}^{PV}$
Mean	-0.128	-0.018	-0.298	-0.098	-0.134	-0.566
Standard deviation	0.455	0.191	0.612	0.328	0.495	0.642

Table 5-10 Characteristics Comparison between ε_s^{MP1}, ε_s^{MP2} and ε_t^{MP1}, ε_t^{MP2}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP1}$	$\varepsilon_{s,s+1}^{MP1}$	$\varepsilon_{s,s+2}^{MP1}$	$\varepsilon_{s,s+3}^{MP1}$	$\varepsilon_{s,s}^{MP2}$	$\varepsilon_{s,s+1}^{MP2}$	$\varepsilon_{s,s+2}^{MP2}$	$\varepsilon_{s,s+3}^{MP2}$
Mean	-0.055	-0.059	-0.051	-0.042	-0.057	-0.068	-0.056	-0.041
Standard deviation	0.233	0.231	0.218	0.221	0.237	0.242	0.219	0.213
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP1}$	$\varepsilon_{t,t+1}^{MP1}$	$\varepsilon_{t,t+2}^{MP1}$	$\varepsilon_{t,t+3}^{MP1}$	$\varepsilon_{t,t}^{MP2}$	$\varepsilon_{t,t+1}^{MP2}$	$\varepsilon_{t,t+2}^{MP2}$	$\varepsilon_{t,t+3}^{MP2}$
Mean	-0.055	-0.059	-0.050	-0.041	-0.057	-0.068	-0.054	-0.038
Standard deviation	0.235	0.232	0.218	0.221	0.239	0.243	0.220	0.212

Table 5-11 Characteristics Comparison between ε_s^{MP3}, ε_s^{MP4} and ε_t^{MP3}, ε_t^{MP4}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP3}$	$\varepsilon_{s,s+1}^{MP3}$	$\varepsilon_{s,s+2}^{MP3}$	$\varepsilon_{s,s+3}^{MP3}$	$\varepsilon_{s,s}^{MP4}$	$\varepsilon_{s,s+1}^{MP4}$	$\varepsilon_{s,s+2}^{MP4}$	$\varepsilon_{s,s+3}^{MP4}$
Mean	-0.055	-0.064	-0.051	-0.044	-0.059	-0.065	-0.056	-0.042
Standard deviation	0.235	0.243	0.217	0.226	0.240	0.241	0.224	0.218
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP3}$	$\varepsilon_{t,t+1}^{MP3}$	$\varepsilon_{t,t+2}^{MP3}$	$\varepsilon_{t,t+3}^{MP3}$	$\varepsilon_{t,t}^{MP4}$	$\varepsilon_{t,t+1}^{MP4}$	$\varepsilon_{t,t+2}^{MP4}$	$\varepsilon_{t,t+3}^{MP4}$
Mean	-0.055	-0.063	-0.050	-0.042	-0.059	-0.065	-0.055	-0.039
Standard deviation	0.237	0.246	0.218	0.227	0.243	0.243	0.224	0.217

Table 5-12 Characteristics Comparison between ε_s^{MP5}, ε_s^{MP6} and ε_t^{MP5}, ε_t^{MP6}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP5}$	$\varepsilon_{s,s+1}^{MP5}$	$\varepsilon_{s,s+2}^{MP5}$	$\varepsilon_{s,s+3}^{MP5}$	$\varepsilon_{s,s}^{MP6}$	$\varepsilon_{s,s+1}^{MP6}$	$\varepsilon_{s,s+2}^{MP6}$	$\varepsilon_{s,s+3}^{MP6}$
Mean	-0.053	-0.063	-0.053	-0.037	-0.055	-0.061	-0.052	-0.043
Standard deviation	0.233	0.237	0.221	0.211	0.235	0.234	0.218	0.220
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP5}$	$\varepsilon_{t,t+1}^{MP5}$	$\varepsilon_{t,t+2}^{MP5}$	$\varepsilon_{t,t+3}^{MP5}$	$\varepsilon_{t,t}^{MP6}$	$\varepsilon_{t,t+1}^{MP6}$	$\varepsilon_{t,t+2}^{MP6}$	$\varepsilon_{t,t+3}^{MP6}$
Mean	-0.053	-0.058	-0.051	-0.037	-0.056	-0.061	-0.051	-0.042
Standard deviation	0.234	0.234	0.217	0.212	0.237	0.237	0.219	0.219

Table 5-13 Characteristics Comparison between ε_s^{MP7}, ε_s^{MP8} and ε_t^{MP7}, ε_t^{MP8}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP7}$	$\varepsilon_{s,s+1}^{MP7}$	$\varepsilon_{s,s+2}^{MP7}$	$\varepsilon_{s,s+3}^{MP7}$	$\varepsilon_{s,s}^{MP8}$	$\varepsilon_{s,s+1}^{MP8}$	$\varepsilon_{s,s+2}^{MP8}$	$\varepsilon_{s,s+3}^{MP8}$
Mean	-0.059	-0.067	-0.050	-0.039	-0.056	-0.059	-0.047	-0.039
Standard deviation	0.239	0.236	0.212	0.204	0.236	0.231	0.210	0.207
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP7}$	$\varepsilon_{t,t+1}^{MP7}$	$\varepsilon_{t,t+2}^{MP7}$	$\varepsilon_{t,t+3}^{MP7}$	$\varepsilon_{t,t}^{MP8}$	$\varepsilon_{t,t+1}^{MP8}$	$\varepsilon_{t,t+2}^{MP8}$	$\varepsilon_{t,t+3}^{MP8}$
Mean	-0.058	-0.066	-0.050	-0.038	-0.055	-0.059	-0.047	-0.037
Standard deviation	0.241	0.237	0.211	0.203	0.237	0.232	0.211	0.207

Table 5-14 Characteristics Comparison between ε_s^{MP9}, ε_s^{MP10} and ε_t^{MP9}, ε_t^{MP10}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP9}$	$\varepsilon_{s,s+1}^{MP9}$	$\varepsilon_{s,s+2}^{MP9}$	$\varepsilon_{s,s+3}^{MP9}$	$\varepsilon_{s,s}^{MP10}$	$\varepsilon_{s,s+1}^{MP10}$	$\varepsilon_{s,s+2}^{MP10}$	$\varepsilon_{s,s+3}^{MP10}$
Mean	-0.089	-0.079	-0.071	-0.054	-0.068	-0.069	-0.075	-0.065
Standard deviation	0.266	0.272	0.243	0.264	0.244	0.256	0.250	0.228
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP9}$	$\varepsilon_{t,t+1}^{MP9}$	$\varepsilon_{t,t+2}^{MP9}$	$\varepsilon_{t,t+3}^{MP9}$	$\varepsilon_{t,t}^{MP10}$	$\varepsilon_{t,t+1}^{MP10}$	$\varepsilon_{t,t+2}^{MP10}$	$\varepsilon_{t,t+3}^{MP10}$
Mean	-0.087	-0.078	-0.069	-0.052	-0.067	-0.070	-0.076	-0.063
Standard deviation	0.268	0.275	0.243	0.262	0.245	0.256	0.251	0.230

Table 5-15 Characteristics Comparison between ε_s^{MP11} and ε_t^{MP11}

Actual Forecast Changes	$\varepsilon_{s,s}^{MP11}$	$\varepsilon_{s,s+1}^{MP11}$	$\varepsilon_{s,s+2}^{MP11}$	$\varepsilon_{s,s+3}^{MP11}$
Mean	-0.054	-0.062	-0.052	-0.043
Standard deviation	0.234	0.235	0.218	0.222
New Forecast Changes by MMFE	$\varepsilon_{t,t}^{MP11}$	$\varepsilon_{t,t+1}^{MP11}$	$\varepsilon_{t,t+2}^{MP11}$	$\varepsilon_{t,t+3}^{MP11}$
Mean	-0.055	-0.061	-0.050	-0.042
Standard deviation	0.236	0.237	0.219	0.221

5.4 Summary

This chapter presents the forecast uncertainty analysis using Martingale Model Forecast Evolution (MMFE). MMFE considers the multivariate random normal vector of the forecast changes evaluated from the Variance Covariance matrix, which offers the benefit of analyzing the correlations of the forecast changes. Based on this correlation analysis, the forecast change matrices are constructed independently for wind/PV generation and market price with the same number of multiple time period estimations as the number of time lags models from the best prediction performance model of the three proposed predictions. The actual forecast uncertainty of wind/PV generation and market price fit well with the normal distribution function, which satisfies MMFE requirements. The similar means and standard deviations of the actual forecast changes and the new forecast changes generated based on MMFE illustrate the effectiveness of the proposed approach.

Chapter 6

Conclusions and Future Research

6.1 Conclusions

To reduce/mitigate greenhouse gas (GHG) emissions and improve the energy efficiency of the transportation sector, the transition from conventional oil-based fleets to Plug-in Electric Vehicles (PEV) has attracted more attention and shows high acceptance in metro areas. A well-planned charging infrastructure plays a critical role for supporting this developing transition. The charging infrastructure should be considered from a regional point of view to optimize operations and support PEV users driving their cars without range concerns.

Virtual wind farms, roof-topped solar panels, and utility grids are three main energy resources in the proposed fast charging station equipped with a NaS battery that acts as the distributed energy storage system. This proposed charging station can also participate in the deregulated market.

Due to the intermittence of wind energy, the variations in PV generation, and the volatility of electricity prices in the deregulated market, the wind/PV generation and electricity market price forecasting need to be calculated. The purpose of these predictions is to address the variances of energy resources and price so that optimal operation of the proposed regional charging station system can be achieved.

This dissertation proposes a regional charging station to be built in the DFW metro area. Due to the 15-minute settlement interval of the ERCOT deregulated market in this metroplex, the 15-minute ahead predictions are calculated for all forecasting. Support Vector Machines (SVMs), effective machine learning algorithms, are used to calculate the predictions considering various relevant parameters. The best prediction

performance models for wind/PV generation and market price are derived in this dissertation.

In wind power forecasting, the best prediction performance model from the proposed Support Vector Regression (SVR) considerably improves the forecasting accuracy compared to the results from the persistence model, which is verified by data from a sample wind farm in Oklahoma. Furthermore, the PV generation at Dallas Redbird airport is used to examine the PV generation prediction accuracy. The corresponding forecasting via the proposed SVR significantly enhances the prediction performance compared to the estimation from the persistence model.

A novel Hybrid Market Price Forecasting Method (HMPFM) with three data clustering techniques, Classification and Regression Tree (CART), K-means, and Stratification methods, is proposed to improve the accuracy of the wholesale electric price prediction in the deregulated market. After calculating the predictions considering the significant dependent parameters, the proposed K-means HMPFM gives the best prediction performance among these three data clustering techniques. The proposed K-means HMPFM also shows significant improvement in prediction accuracy compared to typical forecasting approaches.

Finally, the forecast uncertainty is investigated to help charging station operators better understand its stochastic nature for optimal system benefit. The Martingale Model Forecast Evolution (MMFE) is used to characterize prediction uncertainties because of its ability to analyze multivariate problems. The probability density function of the forecast change is investigated by ARENA. Based on the best fit of the forecast changes to the normal distribution function, the forecast uncertainty can be modeled by the multivariate random normal vector for all of the wind/PV generation and market price predictions following the MMFE technique. New forecast changes generated by MMFE show similar

means and standard deviations with the actual forecast, which illustrates the effectiveness of the proposed approach.

6.2 Dissertation Contributions

This comprehensive study describes the critical components for configuring a regional PEV charging station, develops novel approaches for wind/PV generation and market price predictions, and analyzes the forecast uncertainty of these predictions. These novel forecasting methods demonstrate efficient prediction performance, improving the forecasting accuracy as compared to typical estimation approaches. The major contributions of this dissertation are as follows:

- Establish an efficient configuration of the regional PEV charging station system based on the studies of a virtual wind farm concept, solar PV energy, wholesale deregulated market price, and battery storage technology;
- Identify the most significant parameters for predicting wind/PV generation and electricity market price through analysis of the correlations of these parameters;
- Determine the best prediction performance models by calculating the 15-minute ahead wind/PV generation forecasting using SVR;
- Propose a novel hybrid market price forecasting method (HMPFM) with data clustering techniques to accurately predict the electricity prices in both non-spike and spike conditions in the deregulated market;
- Derive the models of best prediction performance for spike price occurrence, spike price magnitude, and non-spike price magnitude of the proposed HMPFM;
- Use the MMFE process to determine the best fit for the probability density function (pdf) of the forecast uncertainty;
- Characterize the MMFE with the multivariate random normal vector based on the pdf analysis.

6.3 Possible Future Research

As mentioned in the review of literature, numerous methods for wind/PV generation and electricity market price forecasting have been proposed. This dissertation aims to improve upon these proposed methods and obtain the models of the best prediction accuracy by SVMs. Although the forecasting approaches presented in this dissertation demonstrate effective prediction accuracy, there remain possibilities for the further improvement. For instance, the accuracy of the spike price occurrence prediction may be further improved by using a hybrid model instead of the sole SVC algorithm.

In addition, the market price prediction in this dissertation uses ERCOT wholesale market prices in 2011 which the high system-wide offer cap (HCAP) is 3000 $/MWh. However, ERCOT will raise the HCAP to 9000 $/MWh in June 2015 and will continuously increase its value. This presents a challenge for a further study to accurately predict the market price for higher HCAP levels.

Moreover, the probability density functions of forecast changes derived in this dissertation may not be applicable in different forecast scenarios, for example, with different prediction time frames or locations of PV generation. Therefore, the forecast uncertainty should be investigated in each scenario to find that scenario's specific best fit function. Also, the various types of multivariate random vectors based on the results of these investigations into probability density functions may present a challenging and interesting topic for further examination.

Finally, the operational optimization of the regional PEV charging station system implementing the forecast uncertainty by MMFE is an important topic that should be studied in future research. The results of this research may increase the ability to develop and build regional PEV charging station systems, which could lead to a significant increase in the number of PEV users in the future.

Appendix A

Matlab Training SVMs Code

```
%%%%% This code for training Wind/PV generation and market price
prediction %%%%%
%%%%% Coding by Piampoom Sarikprueck updated on 1/30/15 %%%%%

   tic;
    [m_x1,n_x1]=size(n1);
    [m_x2,n_x2]=size(x1);
    %[m_y1,n_y1]=size(n2);
    %[m_y2,n_y2]=size(x2);
    nMax=max(max(n1));
    nMin=min(min(n1));
    n1=(n1-nMin)/(nMax-nMin);
    %n2=(n2-nMin)/(nMax-nMin);
     xMax=max(max(x1));
    xMin=min(min(x1));
    x1=(x1-xMin)/(xMax-xMin);
    %x2=(x2-xMin)/(xMax-xMin);

 trnX = n1;          %
 trnY = x1;          %
 %tstX = n2;           %
 %tstY = x2;           %

%-------------------------------------------------------
% Coefficient Setting

e = 0.001;             %     - insensitivity
p1=2; % 20 for market price occurence predition
C=10; % 500 for market price occurence prediction
%-------------------------------------------------------
type='function approximation';
[alpha,b]=trainlssvm({trnX,trnY,type,C,p1,'RBF_kernel'});


toc;
```

Appendix B

Matlab Testing SVMs Code

```
%%%%% This code for testing Wind/PV generation and market price
prediction %%%%%
%%%%% Coding by Piampoom Sarikprueck updated on 1/30/15 %%%%%



   clearvars -except alpha b n1 x1

   [m_y1,n_y1]=size(n2);
    nMax=max(max(n1));
    nMin=min(min(n1));
    n1_nor=(n1-nMin)/(nMax-nMin);
    n2_nor=(n2-nMin)/(nMax-nMin);
     xMax=max(max(x1));
    xMin=min(min(x1));
    x1_nor=(x1-xMin)/(xMax-xMin);
    x2=(x2-xMin)/(xMax-xMin);

 trnX = n1_nor;          %
 trnY = x1_nor;          %
 tstX = n2_nor;           %
 tstY = x2;           %

%-------------------------------------------------------
% Coefficient Setting

e = 0.001;             %     - insensitivity
p1=2;
C=10;
%-------------------------------------------------------
type='function approximation';
tstY1 =
simlssvm({trnX,trnY,type,C,p1,'RBF_kernel','preprocess'},{alpha,b
},tstX);  % ²âÊÔ
finalY1=tstY1*(xMax-xMin)+xMin;
finalY=tstY*(xMax-xMin)+xMin;

%% MRE for Wind power forecasting performance evaluation
err=(finalY1-finalY)*100/(74.25);
MRE=abs(err);
err2=sum(MRE)/m_y1;

%% MAPE for PV power forecasting performance evaluation
MAPE_err=(finalY-finalY1);
avgfinalY=sum(finalY)/m_y1;
MAPE_err1= abs(MAPE_err)*100/avgfinalY;
MAPE= sum(MAPE_err1)/m_y1;
```

```
%%%% Start of Market price forecasting %%%%
%%%% Market price forecasting by hybrid SVM with predicted spike
occurrance
%%%% the spike level prediction has been dissected into 4 ranges
by
%%%% stratification and k-mean methods.

%data=xlsread('D:\Ph.D Project\PHEV\Forecasting\wind
forecasting\BlueCanyon\reforecasting\results\4monthstrain\4months
','wst-3sc-3');

  %%% XXX_c is for classification
  %%% XXX_n is for normal SVM forecasting
  %%% XXX_s is for spike SVM forecasting


  for i=0:11
   clearvars -except alpha_c b_c n1_c x1_c alpha_n b_n n1_n x1_n
data_c data_n data_s i E m1 m2 m3 m4 m5 m6 m7...
       alpha_s_1 b_s_1 n1_s_1 x1_s_1 alpha_s_2 b_s_2 n1_s_2
x1_s_2 ...
       alpha_s_3 b_s_3 n1_s_3 x1_s_3 alpha_s_4 b_s_4 n1_s_4
x1_s_4 alpha_s_5 b_s_5 n1_s_5 x1_s_5 alpha_s_6 b_s_6 n1_s_6
x1_s_6 ...
       %alpha_s_7 b_s_7 n1_s_7 x1_s_7
   if i==0 ||i==2 || i==4 ||i==6 || i==7 ||i==9 || i==11
       n2_c=data_c(1:1051,8*i+1:8*i+4);
       x2=data_c(1:1051,8*i+6);
       n2_n=data_n(1:1051,10*i+1:10*i+6);
       n2_s=data_s(1:1051,9*i+1:9*i+5);
   end
   if i==3 ||i==5 || i==8 || i==10
       n2_c=data_c(1:955,8*i+1:8*i+4);
       x2=data_c(1:955,8*i+6);
       n2_n=data_n(1:955,10*i+1:10*i+6);
       n2_s=data_s(1:955,9*i+1:9*i+5);
   end
    if i==1
       n2_c=data_c(1:763,8*i+1:8*i+4);
       x2=data_c(1:763,8*i+6);
       n2_n=data_n(1:763,10*i+1:10*i+6);
       n2_s=data_s(1:763,9*i+1:9*i+5);
    end


        m=42.5896; % mean of market price in 2011
        s=162.3179; % standard deviation of market price in 2011
        t1=m+(s); % threshold prices = m+s
        t2=m-(s);
```

```
%%% set up output to be 1 if nonspike and -1 if spike for result
comparison
    x2_c=x2;
      [o,p]=size(x2_c);

for k=1:o
    if x2_c(k)>t1 || x2_c(k)<t2
        x2_c(k)=-1;
    else x2_c(k)= 1;
    end
end

x2_c=reshape(x2_c,o*p,1);

if i==0

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EB
2');

end
if i==1

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EC
2');

end
if i==2

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','ED
2');

end
if i==3

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EE
2');

end
if i==4

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EF
2');

end
if i==5

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EG
2');
```

```
end
if i==6

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EH
2');

end
if i==7

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EI
2');

end
if i==8

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EJ
2');

end
if i==9

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EK
2');

end
if i==10

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EL
2');

end
if i==11

xlswrite('C:\PHEV\classification_PM_64',x2_c,'classification','EM
2');

end


%%% starting 2 stages method

   [m_y1_c,n_y1_c]=size(n2_c);
    nMax_c=max(max(n1_c));
    nMin_c=min(min(n1_c));
    n1_nor_c=(n1_c-nMin_c)/(nMax_c-nMin_c);
    n2_nor_c=(n2_c-nMin_c)/(nMax_c-nMin_c);
    xMax_c=max(max(x1_c));
    xMin_c=min(min(x1_c));
    x1_nor_c=(x1_c-xMin_c)/(xMax_c-xMin_c);
```

```
    x2_c=(x2_c-xMin_c)/(xMax_c-xMin_c);

     trnX_c = n1_nor_c;
     trnY_c = x1_nor_c;
     tstY_c = x2_c;

      %%%% SVM Classification for predicted spike occurance
   for k=1:m_y1_c
       tstX_c=n2_nor_c(k,:);
       e = 0.001;             %     - insensitivity
       p1=20;
       C=5000;

       type='function approximation';
       tstY1_c(k) =
simlssvm({trnX_c,trnY_c,type,C,p1,'RBF_kernel','preprocess'},{alp
ha_c,b_c},tstX_c);  % ²âÊÔ
       finalY1_c(k)=tstY1_c(k)*(xMax_c-xMin_c)+xMin_c;
       if finalY1_c(k)==1 %%% normal marketprice forecasting

           n2_nn=n2_n(k,:);
    [m_y1_n,n_y1_n]=size(n2_nn);
    nMax_n=max(max(n1_n));
    nMin_n=min(min(n1_n));
    n1_nor_n=(n1_n-nMin_n)/(nMax_n-nMin_n);
    n2_nor_n=(n2_nn-nMin_n)/(nMax_n-nMin_n);
    xMax_n=max(max(x1_n));
    xMin_n=min(min(x1_n));
    x1_nor_n=(x1_n-xMin_n)/(xMax_n-xMin_n);

     trnX_n = n1_nor_n;
     trnY_n = x1_nor_n;
     tstX_n = n2_nor_n;

     p1_n=2;
     C_n=10;
     type='function approximation';

       tstY1_n(k) =
simlssvm({trnX_n,trnY_n,type,C_n,p1_n,'RBF_kernel','preprocess'},
{alpha_n,b_n},tstX_n);  % ²âÊÔ
       finalY1(k)=tstY1_n(k)*(xMax_n-xMin_n)+xMin_n;
       else              %%% market prices spike forecasting
      % starting disect with finding the nearest neighbor by
Euclidean norm
        E(k,1)=sqrt((m1(1)-n2_s(k,1))^2+(m1(2)-
n2_s(k,2))^2+(m1(3)-n2_s(k,3))^2+(m1(4)-n2_s(k,4))^2+(m1(5)-
n2_s(k,5))^2);%+(m1(6)-n2_s(k,6))^2+(m1(7)-n2_s(k,7))^2+(m1(8)-
n2_s(k,8))^2);
```

```
        E(k,2)=sqrt((m2(1)-n2_s(k,1))^2+(m2(2)-
n2_s(k,2))^2+(m2(3)-n2_s(k,3))^2+(m2(4)-n2_s(k,4))^2+(m2(5)-
n2_s(k,5))^2);%+(m2(6)-n2_s(k,6))^2+(m2(7)-n2_s(k,7))^2+(m2(8)-
n2_s(k,8))^2);
        E(k,3)=sqrt((m3(1)-n2_s(k,1))^2+(m3(2)-
n2_s(k,2))^2+(m3(3)-n2_s(k,3))^2+(m3(4)-n2_s(k,4))^2+(m3(5)-
n2_s(k,5))^2);%+(m3(6)-n2_s(k,6))^2+(m3(7)-n2_s(k,7))^2+(m3(8)-
n2_s(k,8))^2);
        E(k,4)=sqrt((m4(1)-n2_s(k,1))^2+(m4(2)-
n2_s(k,2))^2+(m4(3)-n2_s(k,3))^2+(m4(4)-n2_s(k,4))^2+(m4(5)-
n2_s(k,5))^2);%+(m4(6)-n2_s(k,6))^2+(m4(7)-n2_s(k,7))^2+(m4(8)-
n2_s(k,8))^2);
        %E(k,5)=sqrt((m5(1)-n2_s(k,1))^2+(m5(2)-
n2_s(k,2))^2);%+(m5(3)-n2_s(k,3))^2+(m5(4)-n2_s(k,4))^2+(m5(5)-
n2_s(k,5))^2
        %E(k,6)=sqrt((m6(1)-n2_s(k,1))^2+(m6(2)-
n2_s(k,2))^2);%+(m6(3)-n2_s(k,3))^2+(m6(4)-n2_s(k,4))^2+(m6(5)-
n2_s(k,5))^2
        %E(k,7)=sqrt((m7(1)-n2_s(k,1))^2+(m7(2)-
n2_s(k,2))^2+(m7(3)-n2_s(k,3))^2);

          [minimum,I]=min(E(k,:));
   % select cluster
   if I==1;  n1_s=n1_s_1; x1_s=x1_s_1; alpha_s=alpha_s_1;
b_s=b_s_1; end
   if I==2;  n1_s=n1_s_2; x1_s=x1_s_2; alpha_s=alpha_s_2;
b_s=b_s_2; end
   if I==3;  n1_s=n1_s_3; x1_s=x1_s_3; alpha_s=alpha_s_3;
b_s=b_s_3; end
   if I==4;  n1_s=n1_s_4; x1_s=x1_s_4; alpha_s=alpha_s_4;
b_s=b_s_4; end
   %if I==5;  n1_s=n1_s_5; x1_s=x1_s_5; alpha_s=alpha_s_5;
b_s=b_s_5; end
   %if I==6;  n1_s=n1_s_6; x1_s=x1_s_6; alpha_s=alpha_s_6;
b_s=b_s_6; end
   %if I==7;  n1_s=n1_s_7; x1_s=x1_s_7; alpha_s=alpha_s_7;
b_s=b_s_7; end


   % start spike price forecasting
          n2_sn=n2_s(k,:);
          [m_y1_s,n_y1_s]=size(n2_sn);
    nMax_s=max(max(n1_s));
    nMin_s=min(min(n1_s));
    n1_nor_s=(n1_s-nMin_s)/(nMax_s-nMin_s);
    n2_nor_s=(n2_sn-nMin_s)/(nMax_s-nMin_s);
    xMax_s=max(max(x1_s));
    xMin_s=min(min(x1_s));
    x1_nor_s=(x1_s-xMin_s)/(xMax_s-xMin_s);
```

```
     trnX_s = n1_nor_s;
     trnY_s = x1_nor_s;
     tstX_s = n2_nor_s;
       p1_s=2;
       C_s=10;

     type='function approximation';
       tstY1_s(k) =
simlssvm({trnX_s,trnY_s,type,C_s,p1_s,'RBF_kernel','preprocess'},
{alpha_s,b_s},tstX_s);  % ²âÊÔ
       finalY1(k)=tstY1_s(k)*(xMax_s-xMin_s)+xMin_s;
       end
   end
finalY1_c=reshape(finalY1_c,m_y1_c,1);
finalY=x2;
finalY1=reshape(finalY1,m_y1_c,1);

MAPE_err=(finalY-finalY1);
avgfinalY=sum(finalY)/m_y1_c;
MAPE_err1= abs(MAPE_err)*100/avgfinalY;
MAPE= sum(MAPE_err1)/m_y1_c;

%% record data for 12 month
if i==0

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IB2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','B2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BB2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AB2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','B3');
end
if i==1

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IC2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','C2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BC2');
```

```
xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AC2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','C3');
end
if i==2

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','ID2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','D2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BD2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AD2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','D3');
end
if i==3

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IE2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','E2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BE2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AE2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','E3');
end
if i==4

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IF2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','F2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BF2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AF2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','F3');
end
```

```
if i==5

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IG2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','G2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BG2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AG2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','G3');
end
if i==6

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IH2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','H2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BH2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AH2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','H3');
end
if i==7

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','II2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','I2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BI2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AI2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','I3');
end
if i==8

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IJ2');
```

```
xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','J2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BJ2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AJ2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','J3');
end

if i==9

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IK2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','K2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BK2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AK2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','K3');
end
if i==10

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IL2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','L2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BL2');
```

```
xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AL2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','L3');
end
if i==11

xlswrite('C:\PHEV\classification_PM_64',finalY1_c,'classification
','IM2');

xlswrite('C:\PHEV\classification_PM_64',finalY1,'hybrid_kmean_Inp
_GL','M2');

xlswrite('C:\PHEV\classification_PM_64',MAPE_err1,'hybrid_kmean_I
np_GL','BM2');

xlswrite('C:\PHEV\classification_PM_64',finalY,'hybrid_kmean_Inp_
GL','AM2');
   xlswrite('C:\PHEV\classification_PM_64',MAPE,'summary','M3');
end
  end
```

References

[1] E. I. Administration, *Annual Energy Review 2011*: Government Printing Office, 2012.

[2] T. Stern. (2010). *U.S. Special Envoy for Climate Change*. Available: http://unfccc.int/files/meetings/application/pdf/unitedstatescphaccord_app.1.pdf

[3] O. o. t. a. a. quality, "Fast Fact: US transportation sector greenhouse gas emission 1991:2011," EPA-42D-F-13-033a, 2013.

[4] T. M. Inc. *The 21st Century Electric Car*. Available: http://www.evworld.com/library/Tesla_21centuryEV.pdf

[5] A. Bandivadekar, *On the road in 2035: reducing transportation's petroleum consumption and GHG emissions*: Massachusetts Institute of Technology, 2008.

[6] K. Fell, K. Huber, B. Zink, R. Kalisch, D. Forfia, D. Hazelwood, N. Dang, D. Gionet, M. Musto, and W. Johnson, "Assessment of plug-in electric vehicle integration with ISO/RTO systems," *KEMA, Inc. and ISO/RTO Council,* 2010.

[7] A. Kulvanitchaiyanunt, V. C. Chen, J. Rosenberger, P. Sarikprueck, and W.-J. Lee, "Control for a System of PHEV Charging Stations," ed. IIE Annual conference and expo, The Industrial and Systems Engineering Research Conference (ISERC) 2013, May 2013.

[8] F. Huang, P. Sarikprueck, Y. Cheng, and W.-J. Lee, "Design optimization of PHEV charging station," in *Industrial & Commercial Power Systems Technical Conference (I&CPS), 2012 IEEE/IAS 48th*, 2012, pp. 1-7.

[9] T. Borne., "PEV Charging Standards Status Including AC, DC, and Wireless Technologies." Available: http://www.sae.org/events/gim/presentations/2013/pev_charging_standards_status.pdf

[10] I. ClipperCreek. Available: http://www.clippercreek.com/

[11] EVRUS. Available: http://www.evrus.net/

[12] E. E. Time. Available: http://m.eet.com/media/1200054/sae-combo.jpg

[13] GWEC. Global Wind Stats 2014 Report [Online]. Available: http://www.gwec.net/wp-content/uploads/2015/02/GWEC_GlobalWindStats2014_FINAL_10.2.2015.pdf

[14] G. W. E. Council, "Global wind energy outlook 2014," *GWEC, November,* 2014.

[15] P. S. REN21, *Renewables 2014: Global Status Report*: REN21 Secretariat Paris, France, 2014.

[16] P. Frankl and S. Nowak, *Technology roadmap: solar photovoltaic energy*: OECD/IEA, 2013.

[17] D. H. Doughty, P. C. Butler, A. A. Akhil, N. H. Clark, and J. D. Boyes, "Batteries for large-scale stationary electrical energy storage," *The Electrochemical Society Interface,* pp. 49-53, 2010.

[18] S. Vazquez, S. M. Lukic, E. Galvan, L. G. Franquelo, and J. M. Carrasco, "Energy storage systems for transport and grid applications," *Industrial Electronics, IEEE Transactions on,* vol. 57, pp. 3881-3895, 2010.

[19] B. Ernst, B. Oakleaf, M. L. Ahlstrom, M. Lange, C. Moehrlen, B. Lange, U. Focken, and K. Rohrig, "Predicting the Wind," *Power and Energy Magazine, IEEE,* vol. 5, pp. 78-89, 2007.

[20] M. Khalid and A. V. Savkin, "A Method for Short-Term Wind Power Prediction With Multiple Observation Points," *Power Systems, IEEE Transactions on,* vol. 27, pp. 579-586, 2012.

[21] S. Tewari, C. J. Geyer, and N. Mohan, "A Statistical Model for Wind Power Forecast Error and its Application to the Estimation of Penalties in Liberalized Markets," *Power Systems, IEEE Transactions on,* vol. 26, pp. 2031-2039, 2011.

[22] K. Methaprayoon, C. Yingvivatanapong, W.-J. Lee, and J. R. Liao, "An Integration of ANN Wind Power Estimation Into Unit Commitment Considering the Forecasting Uncertainty," *Industry Applications, IEEE Transactions on,* vol. 43, pp. 1441-1448, 2007.

[23] K. Bhaskar and S. N. Singh, "AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network," *Sustainable Energy, IEEE Transactions on,* vol. 3, pp. 306-315, 2012.

[24] S. Fan, J. R. Liao, R. Yokoyama, L. Chen, and W.-J. Lee, "Forecasting the Wind Generation Using a Two-Stage Network Based on Meteorological Information," *Energy Conversion, IEEE Transactions on,* vol. 24, pp. 474-482, 2009.

[25] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-Term Wind-Power Prediction Based on Wavelet Transform-Support Vector Machine and Statistic-Characteristics Analysis," *Industry Applications, IEEE Transactions on,* vol. 48, pp. 1136-1141, 2012.

[26] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, and M. Zhang, "Hybrid Forecasting Model for Very-Short Term Wind Power Forecasting Based on Grey Relational Analysis and Wind Speed Distribution Features," *Smart Grid, IEEE Transactions on,* vol. 5, pp. 521-526, 2014.

[27] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Wind Power Forecasting in Portugal," *Sustainable Energy, IEEE Transactions on,* vol. 2, pp. 50-59, 2011.

[28] X. Wu and V. Kumar, *The top ten algorithms in data mining*: CRC Press, 2010.

[29] N. C. D. Center. *Historical Weather Data* Available: http://www.ncdc.noaa.gov/

[30] M. H. Rahman, J. Nakayama, K. Nakamura, and S. Yamashiro, "A viable grid-connected PV-ECS system with load leveling function using a day-ahead weather

forecast," in *Power System Technology, 2004. PowerCon 2004. 2004 International Conference on*, 2004, pp. 311-316 Vol.1.

[31] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of,* vol. 2, pp. 2-10, 2009.

[32] B. Ernst, F. Reyer, and J. Vanzetta, "Wind power and photovoltaic prediction tools for balancing and grid operation," in *Integration of Wide-Scale Renewable Resources Into the Power Delivery System, 2009 CIGRE/IEEE PES Joint Symposium*, 2009, pp. 1-9.

[33] A. Yona, T. Senjyu, and T. Funabashi, "Application of Recurrent Neural Network to Short-Term-Ahead Generating Power Forecasting for Photovoltaic System," in *Power Engineering Society General Meeting, 2007. IEEE*, 2007, pp. 1-6.

[34] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and K. Chul-Hwan, "Application of Neural Network to One-Day-Ahead 24 hours Generating Power Forecasting for Photovoltaic System," in *Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on*, 2007, pp. 1-6.

[35] N. Al-Messabi, Y. Li, I. El-Amin, and C. Goh, "Forecasting of photovoltaic power yield using dynamic neural networks," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-5.

[36] C.-T. Chiang, Y.-S. Lee, X. R. Li, and C.-C. Liao, "A RSCMAC based forecasting for Solar Irradiance from local weather information," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-7.

[37] A. Moreno-Munoz, J. J. G. De la Rosa, R. Posadillo, and F. Bellido, "Very short term forecasting of solar radiation," in *Photovoltaic Specialists Conference, 2008. PVSC '08. 33rd IEEE*, 2008, pp. 1-5.

[38] H. Cheng, W.-s. Cao, and P.-j. Ge, "Forecasting Research of Long-Term Solar Irradiance and Output Power for Photovoltaic Generation System," in *Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on*, 2012, pp. 1224-1227.

[39] Y. Hosoda and T. Namerikawa, "Short-term photovoltaic prediction by using H filtering and clustering," in *SICE Annual Conference (SICE), 2012 Proceedings of*, 2012, pp. 119-124.

[40] M. Hassanzadeh, M. Etezadi-Amoli, and M. S. Fadali, "Practical approach for sub-hourly and hourly prediction of PV power output," in *North American Power Symposium (NAPS), 2010*, 2010, pp. 1-5.

[41] T. Cai, S. Duan, and C. Chen, "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement," in *Power Electronics for Distributed Generation Systems (PEDG), 2010 2nd IEEE International Symposium on*, 2010, pp. 773-777.

[42] K. Min-Cheol, S. Jin-Man, P. Jong-young, L. Song-Keun, and Y. Yong-Tae, "Development of algorithm for day ahead PV generation forecasting using data mining method," in *Circuits and Systems (MWSCAS), 2011 IEEE 54th International Midwest Symposium on*, 2011, pp. 1-4.

[43] Y.-z. Li, R.-q. Nie, and J.-c. Niu, "Forecast of Power Generation for Grid-Connected Photovoltaic System Based on Knowledge Representation of Rough Sets," in *Power and Energy Engineering Conference (APPEEC), 2012 Asia-Pacific*, 2012, pp. 1-4.

[44] R. Xu, H. Chen, and X. Sun, "Short-term photovoltaic power forecasting with weighted support vector machine," in *Automation and Logistics (ICAL), 2012 IEEE International Conference on*, 2012, pp. 248-253.

[45] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Huang, "A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output," 2014.

[46] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and W. Peng, "Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines," *Industry Applications, IEEE Transactions on,* vol. 48, pp. 1064-1069, 2012.

[47] N. R. E. Laboratory. *Solar radiation data base*. Available: http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2010/

[48] C. Chalkias, A. Faka, and K. Kalogeropoulos, "Assessment of the Direct Sun-Light on Rural Road Network through Solar Radiation Analysis Using GIS," *Open Journal of Applied Sciences,* vol. 3, pp. 224-231, 2013.

[49] S. N. L. a. U. S. D. o. Energy. *An Industry and National Laboratory collaborative to Improve Photovoltaic Performance Modeling*. Available: https://pvpmc.sandia.gov/

[50] N. E. Research. *Basic Airmass Models*. Available: http://bxhorn.com/air-mass-models/

[51] J. Zhao, A. Wang, M. A. Green, and F. Ferrazza, "19.8% efficient “honeycomb” textured multicrystalline and 24.4% monocrystalline silicon solar cells," *Applied Physics Letters,* vol. 73, pp. 1991-1993, 1998.

[52] H. Mori and A. Takahashi, "A data mining method for selecting input variables for forecasting model of global solar radiation," in *Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES*, 2012, pp. 1-6.

[53] J. Bastian, Z. Jinxiang, V. Banunarayanan, and R. Mukerji, "Forecasting energy prices in a competitive market," *Computer Applications in Power, IEEE,* vol. 12, pp. 40-45, 1999.

[54] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the Ontario competitive power system market," *Power Systems, IEEE Transactions on,* vol. 19, pp. 366-374, 2004.

[55] Y. Y. Hong and C. Y. Hsiao, "Locational marginal price forecasting in deregulated electricity markets using artificial intelligence," *Generation, Transmission and Distribution, IEE Proceedings-,* vol. 149, pp. 621-626, 2002.

[56] G. Li, C.-C. Liu, C. Mattson, and J. Lawarree, "Day-Ahead Electricity Price Forecasting in a Grid Environment," *Power Systems, IEEE Transactions on,* vol. 22, pp. 266-274, 2007.

[57] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi, and A. K. Srivastava, "A Novel Approach to Forecast Electricity Price for PJM Using Neural Network and Similar Days Method," *Power Systems, IEEE Transactions on,* vol. 22, pp. 2058-2065, 2007.

[58] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *Power Systems, IEEE Transactions on,* vol. 18, pp. 1014-1020, 2003.

[59] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *Power Systems, IEEE Transactions on,* vol. 20, pp. 1035-1042, 2005.

[60] Y. Baez-Rivera, B. Rodriguez-Medina, and A. K. Srivastava, "An Attempt to Forecast Price Spikes in Electric Power Markets," in *Power Symposium, 2006. NAPS 2006. 38th North American*, 2006, pp. 143-148.

[61] J. H. Zhao, Z.-Y. Dong, X. Li, and K. P. Wong, "A Framework for Electricity Price Spike Analysis With Advanced Data Mining Methods," *Power Systems, IEEE Transactions on,* vol. 22, pp. 376-385, 2007.

[62] K. R. Skinner, D. C. Montgomery, G. C. Runger, J. W. Fowler, D. R. McCarville, T. R. Rhoads, and J. D. Stanley, "Multivariate statistical methods for modeling and analysis of wafer probe test data," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 15, pp. 523-530, 2002.

[63] M. Seera, L. Chee Peng, D. Ishak, and H. Singh, "Fault Detection and Diagnosis of Induction Motors Using Motor Current Signature Analysis and a Hybrid FMM&-CART Model," *Neural Networks and Learning Systems, IEEE Transactions on,* vol. 23, pp. 97-108, 2012.

[64] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers," *Power Systems, IEEE Transactions on,* vol. 22, pp. 1120-1128, 2007.

[65] P. K. Dash, S. R. Samantaray, and G. Panda, "Fault Classification and Section Identification of an Advanced Series-Compensated Transmission Line Using Support Vector Machine," *Power Delivery, IEEE Transactions on,* vol. 22, pp. 67-73, 2007.

[66] ERCOT. *ERCOT Market Information*. Available: http://www.ercot.com/mktinfo

[67] D. Steinberg and P. Colla, "CART: classification and regression trees," *The Top Ten Algorithms in Data Mining,* vol. 9, p. 179, 2009.

[68] Y. Yohannes and P. Webb, *Classification and regression trees, CART: a user manual for identifying indicators of vulnerability to famine and chronic food insecurity* vol. 3: Intl Food Policy Res Inst, 1999.

[69] S. Fan, L. Chen, and W.-J. Lee, "Machine learning based switching model for electricity load forecasting," *Energy Conversion and Management,* vol. 49, pp. 1331-1344, 2008.

[70] D. Steinberg and M. Golovnya, "CART 6.0 user's guide," *Salford Systems, San Diego, CA,* 2007.

[71] H. Keko, M. A. da Rosa, J. Sumaili, and V. Miranda, "Wind power forecast uncertainty in daily operation of wind park combined with storage," in *Energy Market (EEM), 2011 8th International Conference on the European*, 2011, pp. 773-778.

[72] M. E. Khodayar, W. Lei, and M. Shahidehpour, "Hourly Coordination of Electric Vehicle Operation and Volatile Wind Power Generation in SCUC," *Smart Grid, IEEE Transactions on,* vol. 3, pp. 1271-1279, 2012.

[73] D. C. Heath and P. L. Jackson, "Modeling the evolution of demand forecasts ITH application to safety stock analysis in production/distribution systems," *IIE transactions,* vol. 26, pp. 17-30, 1994.

[74] T. Zhao, X. Cai, and D. Yang, "Effect of streamflow forecast uncertainty on real-time reservoir operation," *Advances in Water Resources,* vol. 34, pp. 495-504, 2011.

[75] T. Zhao, J. Zhao, D. Yang, and H. Wang, "Generalized martingale model of the uncertainty evolution of streamflow forecasts," *Advances in Water Resources,* vol. 57, pp. 41-51, 2013.

[76] V. C. Chen, "Application of orthogonal arrays and MARS to inventory forecasting stochastic dynamic programs," *Computational statistics & data analysis,* vol. 30, pp. 317-341, 1999.

[77] P. Jackson, D. Heath, C. Loew, D. Marschke, L. Rieth, and J. Shunk, "Modelling the Impact of Assembly Demand Forecast Accuracy on Customer Service, Inventory, and Expediting," Cornell University Operations Research and Industrial Engineering1992.

[78] W. D. Kelton, R. P. Sadowski, and D. A. Sadowski, *Simulation with ARENA* vol. 3: McGraw-Hill New York, 2002.

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

Biographical Information

Piampoom Sarikprueck received the B.Eng and M.Eng. degrees in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand in 2002 and 2005, respectively. From 2002 to 2005, he was a researcher in the Engineering Service and Developing Center at KMITL, where he had served as a lecturer in the Electrical Engineering Department during 2005-2010.

Piampoom Sarikprueck was a recipient of the Royal Thai Government Scholarship between 2010 and 2015. In 2010, he came to the Energy Systems Research Center (ESRC) at the University of Texas at Arlington to pursue his Ph.D. study in Electrical Engineering. His research areas of interest include statistical forecasting, renewable energy application, power systems, rotating electric machines, and energy conversion technology.