Title
Electric Vehicle - Smart Grid Integration: Load Modeling, Scheduling, and Cyber Security

Author
Chung, Yu-Wei

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Electric Vehicle - Smart Grid Integration: Load Modeling, Scheduling, and Cyber Security

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mechanical Engineering

by

Yu-Wei Chung

2020

ABSTRACT OF THE DISSERTATION

Electric Vehicle - Smart Grid Integration: Load Modeling, Scheduling, and Cyber Security

by

Yu-Wei Chung
Doctor of Philosophy in Mechanical Engineering
University of California, Los Angeles, 2020
Professor Rajit Gadh, Chair

The modern world has witnessed the surge of electric vehicles (EVs) driven by government policy worldwide to reduce transportation's dependence on fossil fuels. According to [SL19], the global EV market has grown sharply with the annual light-duty EV sales surpassing 2 million in 2018, which is about a 70% increase from 2017. The increase in EV population implies the rise in energy demand, and that introduces new challenges to the electricity sector. EV charging load demand in high penetration scenarios, which is foreseen, may lead to stability and quality issues in power grids. Generation capacity and the electricity infrastructure upgrade may be required to address those issues; however, it increases generation costs significantly. The most common EV chargers installed today deliver around 7 kW of power, which is over four times that of an averaged household power consumption in the US. EV charging load often shows two peaks in a day, one in the morning when people plug in the EV at the workplace and the other in the evening when people get home from work. Without proper energy management for EV charging, the vast power demand due to a large number of plugged-in EVs can stress the electric grid, degrade the electric power quality, and impact the wholesale electricity market. Although an EV battery may store energy up to 80 kWh, which requires more than 10 hours to charge at 7kW from empty, we found that most EVs need only 12 kWh per charge or 1.7 hours at 7 kW to meet daily commute requirement while they stay in the parking garage for a more extended period. This implies that EVs can have considerable time-flexibility for charging, and it is not necessary to start charging right after plugging in, which is likely to result in the charging power add-up. A proper EV

charging schedule can well allocate the charging load to prevent power peaks. Therefore, EV charging scheduling can play a significant role in mitigating the adverse effects of vast EV charging demand without upgrading the power grid capacity.

To optimize the EV charging schedule while satisfies EVs' charging demand, each EV's stay duration and energy need are essential parameters for the optimization. Those parameters are based on predictions to minimize human intervention. Nonetheless, the uncertainty of EV user behavior poses a challenge to the prediction accuracy. Therefore, this dissertation demonstrates an ensemble machine learning-based method to model and predict the EV loads accurately, thereby improving the performance of EV charging scheduling.

On the other hand, this smart EV-grid integration, which requires massive communication, including collecting, transmitting, and distributing real-time data within the network, makes it more susceptible to cyber-physical threats. Potential breaches could not only affect grid operation but also reduce consumers' willingness to adopting EVs over conventional fuel-powered vehicles. This dissertation also presents the vulnerability analysis and risk assessment for a smart EV charging system to develop the countermeasures to secure the network. Also, while it is inevitable that the security has flaws, this dissertation provides a novel anomaly detection approach based on the invariant correlations of different measurements within the EV charging network.

The dissertation of Yu-Wei Chung is approved.

Mani Srivastava

Xiaochun Li

Tetsuya Iwasaki

Rajit Gadh, Committee Chair

University of California, Los Angeles

2020

To my dearest parents, Jih-Sheng and Shiu-Yu,

my supportive sister, Yu-Yin,

my beloved wife, Yi-Chun,

and my lovely daughters, Olivia and Joanne.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to Prof. Gadh for his consistent support, guidance, and specifically for his effort in educating and preparing me for a career in smart grid research and development. It has been an enjoyable and fruitful learning experience at UCLA Smart Grid Energy Research Center (SMERC).

Further, I would like to thank my committee members Prof. Iwasaki, Prof. Li, and Prof. Srivastava, for their insightful comments and suggestions on my research and dissertation.

Also, I am grateful for the friendly and collaborative research environment at SMERC. To Dr. Chi-Cheng Chu, Dr. Ching-Yen Chung, and Mr. Charlie Qiu, thanks for their kind support on my research; To Dr. Bin Wang, Dr. Yubo Wang, Dr. Hamidreza Nazaripouya, Dr. Yingqi Xiong, Dr. Tianyang Zhang, Dr. Behnem Khaki, Zhiyuan Cao, Shashank Gowda, Jesse Cha, and Amirhossein Ahmadian, thanks for their valuable experiences and suggestions on research ideas. Special thanks to Dr. Khaki for his efforts in various research collaborations, and Mervin Mathew, Cole Rodgers, and Zachary Lau for their assists in the cybersecurity project.

Moreover, I would like to thank my wife, Yi-Chun Chen, who serves my backbone of Ph.D. life. She's a real Ph.D., a.k.a Push Husband toward a Doctorate. Lastly, I want to thank my friends, Jinxin, Po-Ting, and Hung-Yun, who always by my side, providing strength, support, and excitement.

To conclude, thanks to my mentors, colleagues, family, and friends for all the unconditional support in these very intense academic years.

VITA

2013–2017 M.Sc. in Mechanical Engineering, University of California, Los Angeles.

2008–2010 M.Sc. in Applied Mechanics, National Taiwan University.

2004–2008 B.Eng. in Mechatronic Engineering, National Taiwan Normal University.

PUBLICATIONS

Journal Articles:

- **Chung, Yu-Wei**, Behnam Khaki, Tianyi Li, Chicheng Chu, and Rajit Gadh. "Ensemble machine learning-based algorithm for electric vehicle user behavior prediction." Applied energy 254 (2019):113732.
- Hamidreza Nazaripouya, **Yu-Wei Chung**, and Abbas Akhil. "Energy storage in microgrids: challenges, applications, and research need." International Journal of Energy and Smart Grid 3 (2019): 60-70.

Conference Papers:

- **Yu-Wei Chung**, Mervin Mathew, Cole Rodgers, Bin Wang, Behnam Khaki, ChiCheng Chu, and Rajit Gadh. "The Framework of Invariant Electric Vehicle Charging Network for Anomaly Detection." Accepted to the 2020 IEEE Transportation Electrification Conference & Expo (ITEC), Chicago, Illinois, 2020.,

- Khaki, Behnam, **Yu-Wei Chung**, Chicheng Chu, and Rajit Gadh. "Hierarchical distributed EV charging scheduling in distribution grids." ***Selected as one of the Best Papers***, 2019 IEEE Power & Energy Society General Meeting (PESGM). IEEE, 2019.

- Khaki, Behnam, **Yu-Wei Chung**, Chicheng Chu, and Rajit Gadh. "Probabilistic electric vehicle load management in distribution grids." 2019 IEEE Transportation Electrification Conference and Exp, Novi, Michigan, 2019.

- Reeh, Devin, Francisco Cruz Tapia, **Yu-Wei Chung**, Behnam Khaki, Chicheng Chu, and Rajit Gadh. "Vulnerability analysis and risk assessment of EV charging system under cyber-physical threats." 2019 IEEE Transportation Electrification Conference & Expo (ITEC), Novi, Michigan, 2019.

- **Chung, Yu-Wei**, Behnam Khaki, Chicheng Chu, and Rajit Gadh. "Electric vehicle user behavior prediction using hybrid kernel density estimator." 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, 2018.

- Khaki, Behnam, **Yu-Wei Chung**, Chicheng Chu, and Rajit Gadh. "Nonparametric user behavior prediction for distributed ev charging scheduling." 2018 IEEE Power & Energy Society General Meeting (PESGM). IEEE, 2018.

- Cao, Zhiyuan, **Yu-Wei Chung**, Yingqi Xiong, Chicheng Chu, and Rajit Gadh. "IoT based manufacturing system with a focus on energy efficiency." 2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia). IEEE, 2016.

CHAPTER 1

Introduction

1.1 Background

The current electricity grid was designed to operate in a top-down structure. The single-direction power flow starts from power plants and is then followed by a transmission system and distribution system, and finally ends at the customers. However, increasing penetration of Renewable Energy Resources (RERs), particularly photovoltaic (PV), in the distribution grid has resulted in significant challenges for system operators who manage the grid. Fig. 1.1 shows the percentage of total capacity additions in the US over the past decade, and it also shows the rapid growth of solar's share of new capacity. 40% of the new capacity added to the grid came from solar in the year of 2019 [Sol20].

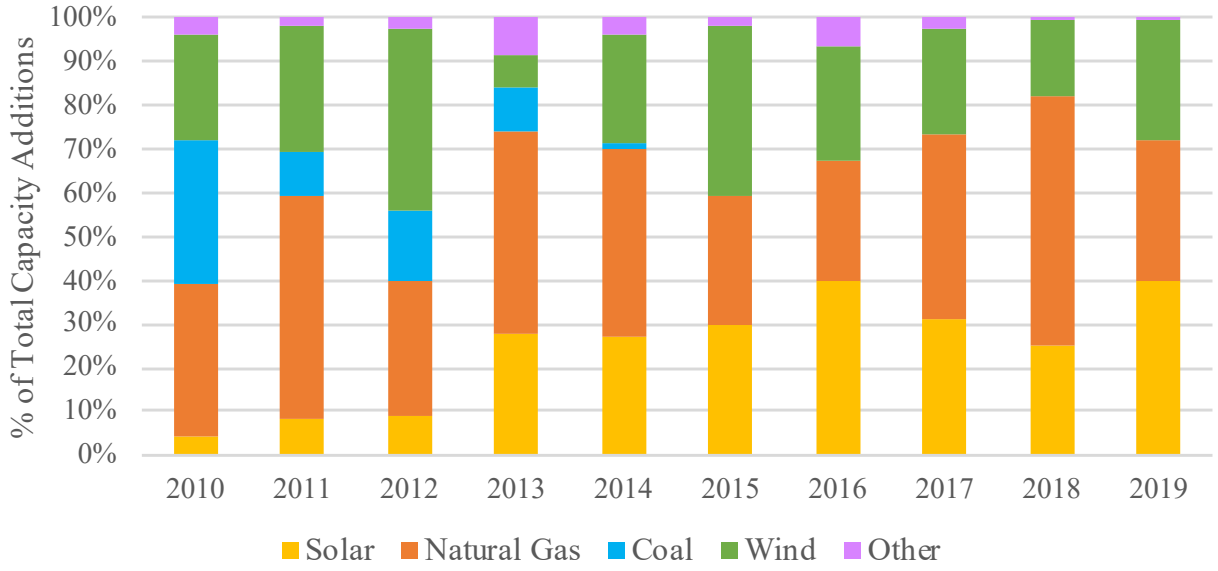

Figure 1.1: U.S. Annual Additions of New Electric Generating Capacity [Sol20]

California has a significant renewable energy adoption rate among the United States, as shown in Fig. 1.2 [Cal19, US19]. California has the largest solar market in the United States, and 19% of its electricity today comes from solar [Sol19]. However, high solar penetration also results in a new demand/supply challenge, known as the solar duck curve problem [Cal16] shown in Fig. 1.3. There is an over-generation risk during the midday when the sun is shining, and a steep ramp in the evening when the sun goes down along with increasing electricity demand. The valley goes deeper when PV penetration becomes higher, and this makes the problem more challenging. Flexible and controllable resources such as battery energy storage systems (BESS) are needed to ensure supply and demand matching all the time [NCP17].

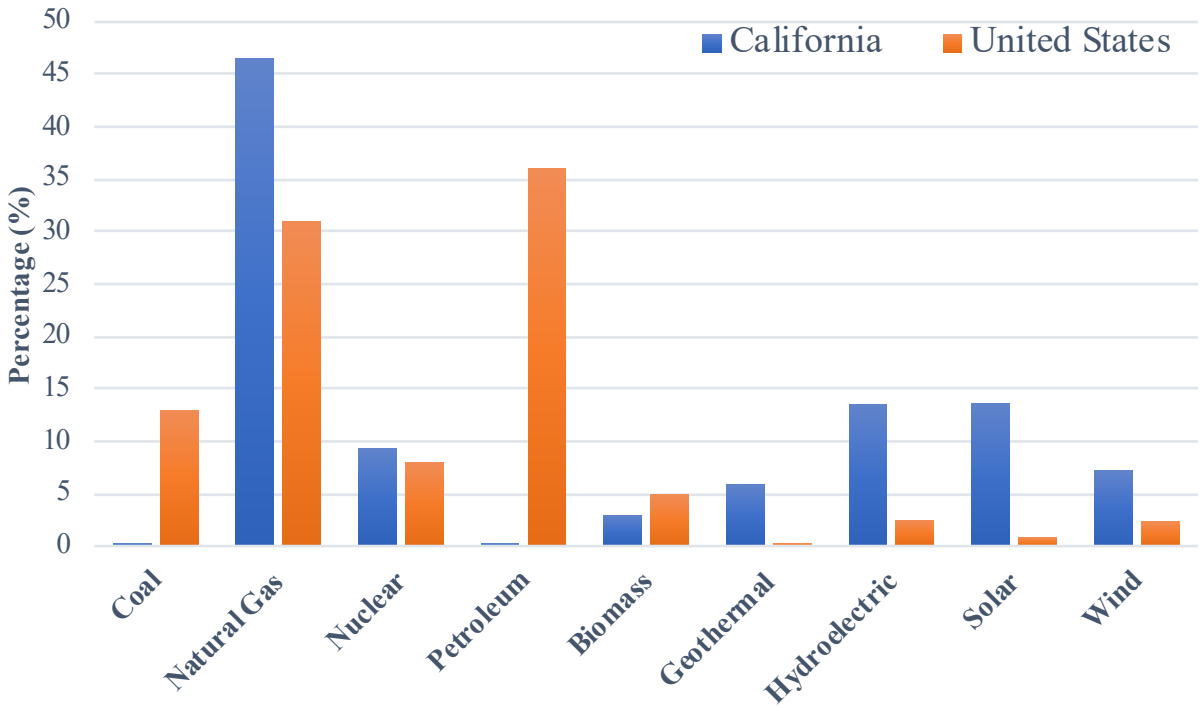

Figure 1.2: Comparison of energy generation sectors between California and the US as the year of 2018

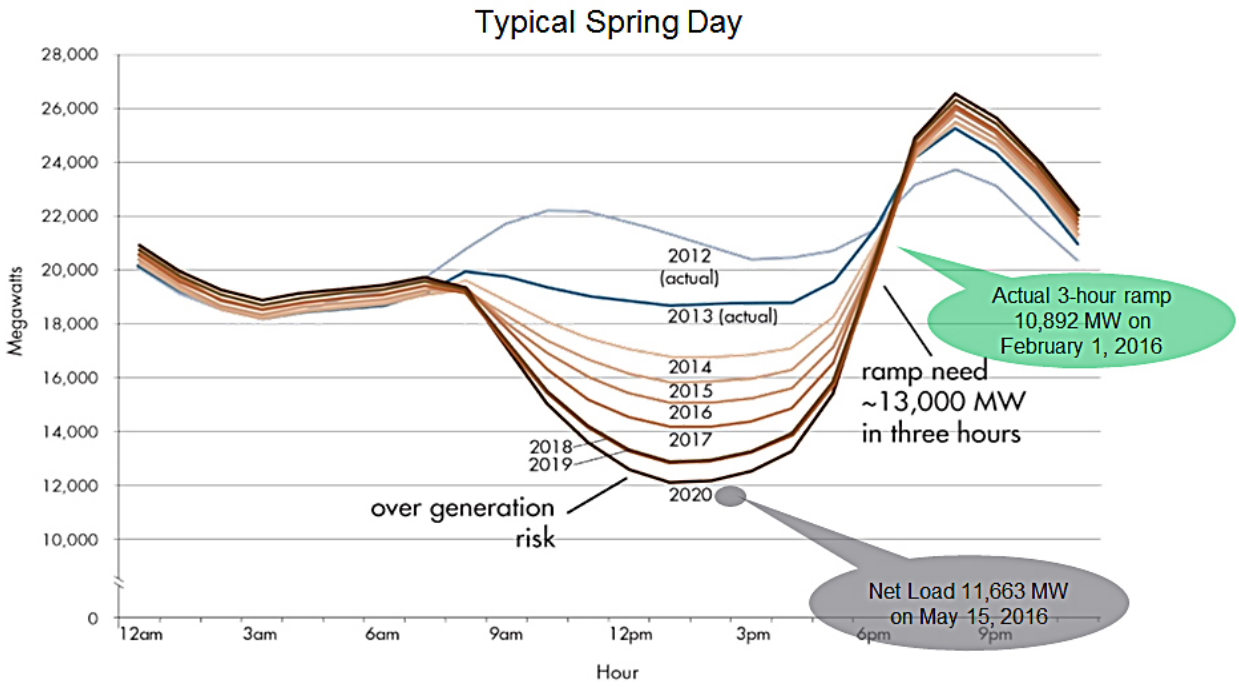

Figure 1.3: The duck curve shows steep ramping needs and over-generation risk [Cal16]

Distributed RERs have changed the electrical grid substantially. As Fig. 1.4 shown, with more RERs distributed across the power system, there is a transition from a centralized to decentralized topology. Because the RERs locally generate energy, single-direction power flow in the distribution system becomes bidirectional. The drastic changes in the power system stimulate the rapid development of smart grid technology, which becomes a game-changer for both markets and end-user players. Smart grid technology incorporated with the Internet of Things (IoT), such as smart meters, smart controllers, demand response (DR), results in the rapid digitalization of the power system. Therefore, there is a need for advanced communication and control schemes to enhance system reliability and resiliency[SXC14, SLC15]. Furthermore, the addition of communication, IoT, and Information Technology (IT) has resulted in a significant increase in potential vulnerabilities to cyber-attacks on the power sector, which is yet another challenge to the grid of the future.

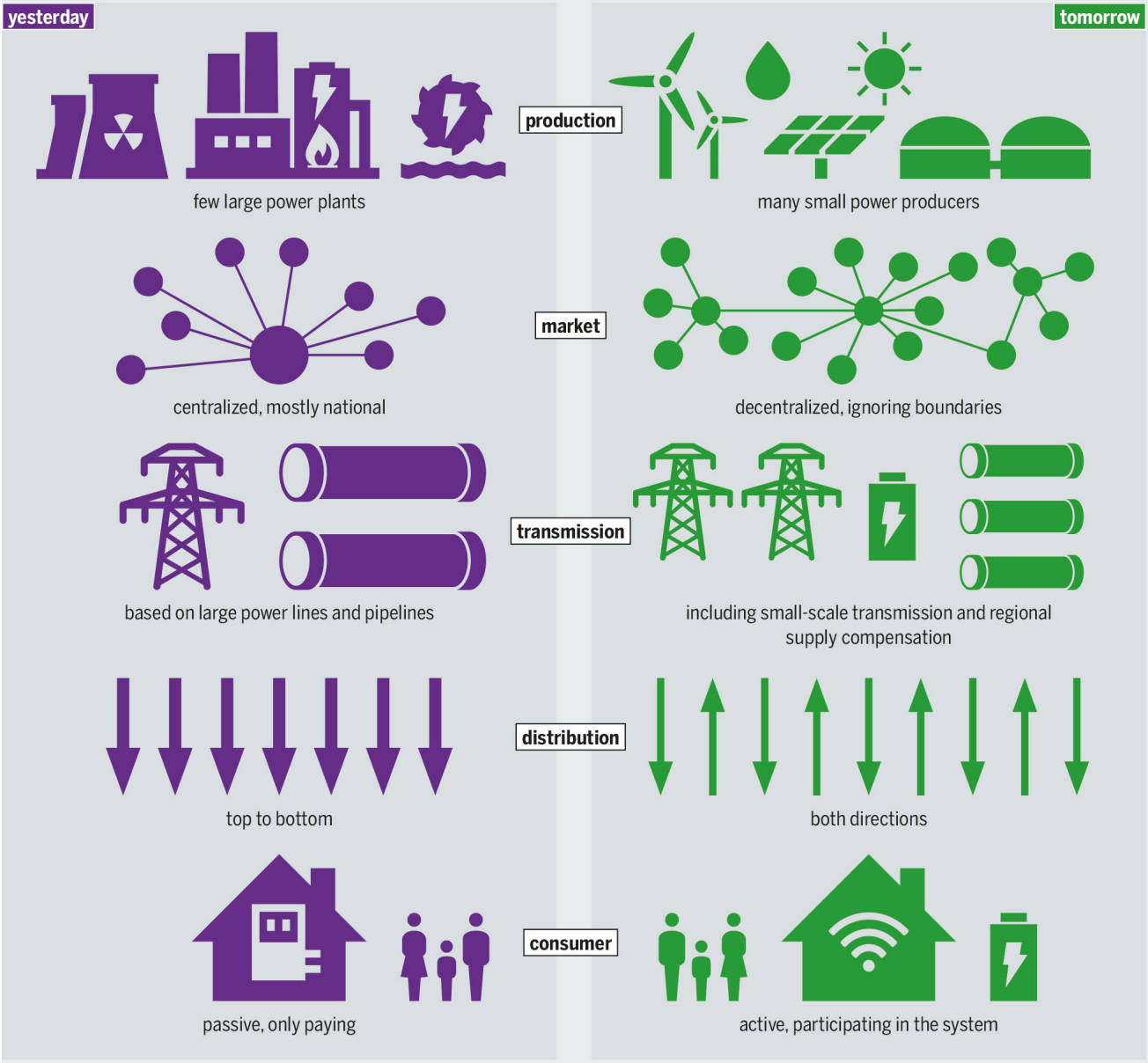

Figure 1.4: Characteristics of traditional grid versus smart grid [Sto18]

1.1.1 Smart Grid and EV Integration

The emergence of Electric vehicles (EV) comes with the evolution of smart grids, with the same goal to help the grid become greener. EV has great potential to significantly reduce the usage of petroleum since the transportation sector is the biggest portion of total energy

use [LL19]. There is a considerable expansion on the EV market in recent years, with annual light-duty EV sales surpassing two million in 2018 worldwide, a 70% increase from 2017 [SL19, Lov19]. The cumulative sales of EVs globally are presented in Fig. 1.5. EV sales by country vary worldwide. Widespread adoption of EVs is affected by consumer demand, government policy, and market prices. The United States is the third-largest EV market, and various government policies across the states have been made to support EV adoption. For example, California has adopted the Zero Emission Vehicle regulation, which requires increasing shares of electric vehicles through 2025. California continues to implement a wide array of policies and is home to most electric vehicle sales. These policy actions include consumer incentives, infrastructure deployment, information campaigns, and various local measures. All of this is to overcome EV adoption barriers related to higher upfront costs, functional electric range, range anxiety, and lack of awareness of the benefits [SL19]. The United States EV new sales market share and forecast are shown in Table1.1. The table shows the increasing trend of EV sales and the percentage of EVs in the US automotive. The table also shows that California shares a significant portion of the EV market in the US.

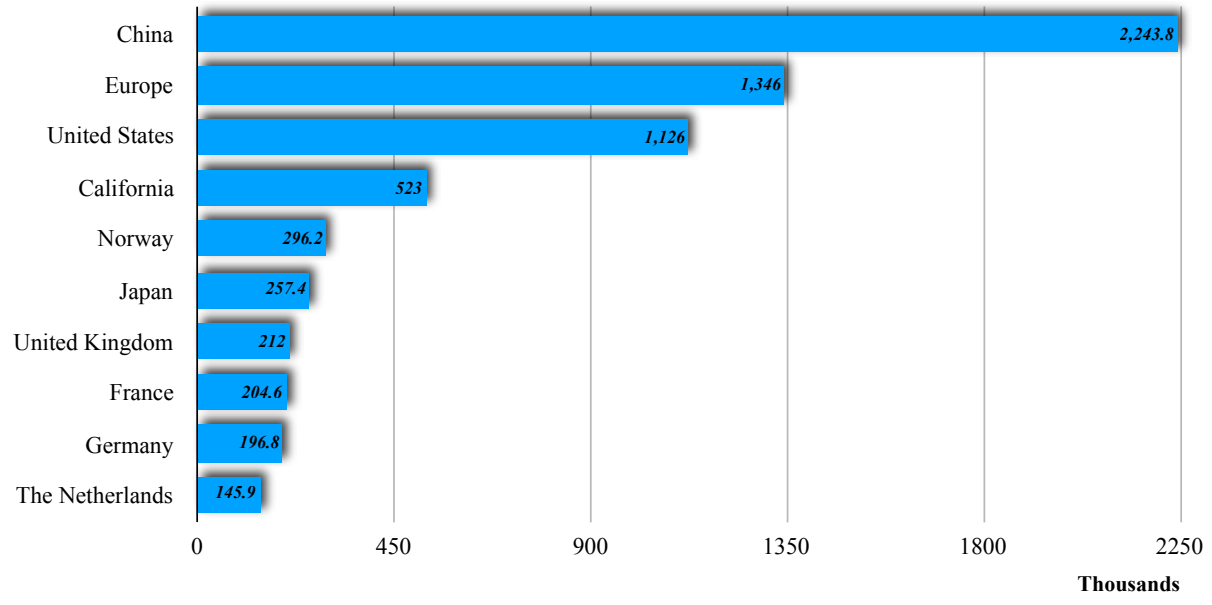

Figure 1.5: Top-selling light-duty plug-in EV global markets (cumulative sales through December 2018 by country/region)[Ort18]

Table 1.1: US and California EV Sales & Market Share Projections [EVA18b]

	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
US EV Sales	158,000	199,826	325,000	450,000	600,000	875,000	1,250,000	1,800,000	2,500,000	3,500,000
Total US Auto Sales	17,550,000	17,208,748	16,800,000	16,500,000	16,000,000	16,000,000	16,000,000	16,000,000	16,000,000	16,000,000
US EV % of Sales	0.90 %	1.16 %	1.93 %	2.73 %	3.75 %	5.47 %	7.81 %	11.25 %	15.63 %	21.88 %
CA EV Sales	75,165	110,000	200,000	275,000	350,000	450,000	600,000	800,000	1,050,000	1,350,000
Total CA Auto Sales	2,086,966	2070,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000	2,000,000
CA EV % of Sales	3.60 %	5.31 %	10.00 %	13.75 %	17.50 %	22.50 %	30.00 %	40.00 %	52.50 %	67.50 %
CA % of US EV Sales	47.57 %	55.05 %	61.54 %	61.11 %	58.33 %	51.43 %	48.00 %	44.44 %	42.00 %	38.57 %
CA EV % of US Autos	0.43 %	0.64 %	1.19 %	1.67 %	2.19 %	2.81 %	3.75 %	5.00 %	6.56 %	8.44 %

There is a concern with the sharply increasing number of plug-in EVs (PEVs) in the distribution grid. The power supply may be insufficient to meet the additional EV charging demand. According to [SBC04], the additional EV charging demand in the US will increase the existing load by 18% by 2040. This increased load will eventually cause the degrade of power quality, requiring a distribution infrastructure upgrades for mitigation. Fig. 1.6 illustrates the typical EV loads for three scenarios [fle19]. The study was based on the real-world charging data from 650 battery electric vehicles over four weeks, representing 13,000 charging events. These account for 120 MWh to illustrate the wide variability of EV charging behavior among the example service territories. For scenario (a) with no time-of-use (TOU) rate structure, regular morning and evening peaks recur with commuting cycles to and from work. Also, there is a consistent midday charging. This unregulated EV charging leaves peak load to chance. For scenario (b) with static TOU rate, the variability of EV load is reduced but produces an unintentional peak at the same time. The unintentional peak of EV charging at 7 pm coincides with a shifted residential load such as cooking, cleaning, or HVAC. For Scenario (C) with an EV charging control program, the EV load can be evenly distributed along the day and aggregated within the period of a residential load valley between 12 - 3 am.

Therefore, there is a need for EV charging management. Also, the TOU rate alone is not enough as it may result in coincident loads with the other shifted load like cooking, cleaning, and laundry, which defeats the purpose of shifting this load in the first place.

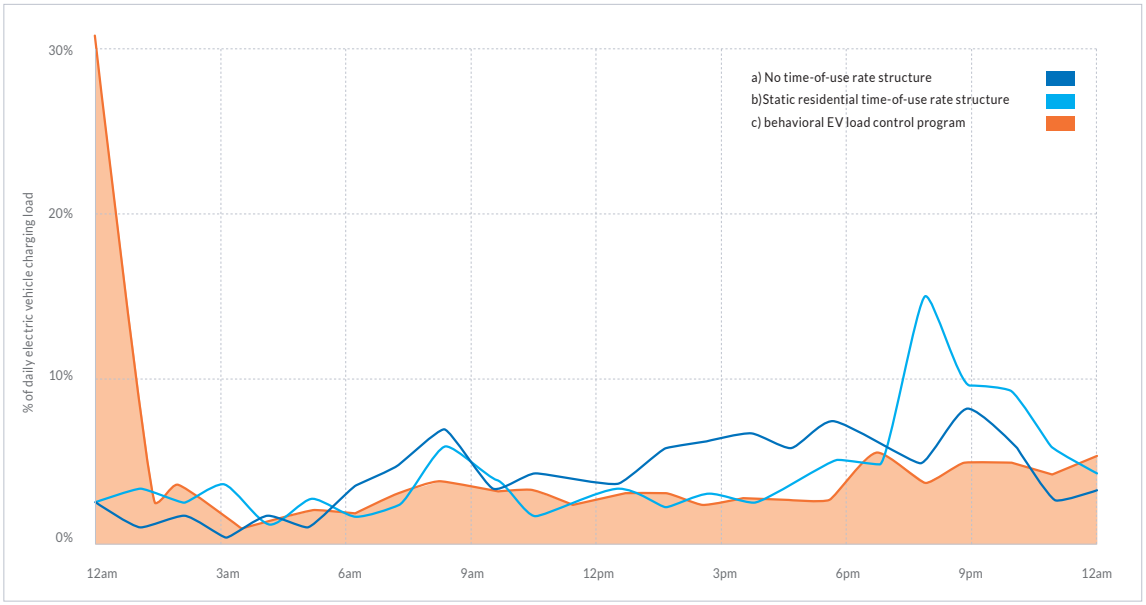

Figure 1.6: Three typical EV loads within a day [fle19]

EV charging scheduling methods have been studied through shifting EV load to off-peak and aligning with renewables and over-generation by solar. The results have shown the great potential to accommodate a large number of EVs and defer upgrading the electrical grid [XWC18, WWN17, KCC18, KCG19]. Also, from the perspective of an electrical grid, EVs offer an opportunity to provide ancillary services, such as DR, renewable generation integration, or providing emergency backup power. These benefits can be achieved by proper EV-grid integration. Therefore, EV has great potential to make an electrical grid more reliable. Fig. 1.7 shows the schematic of EV-grid integration.

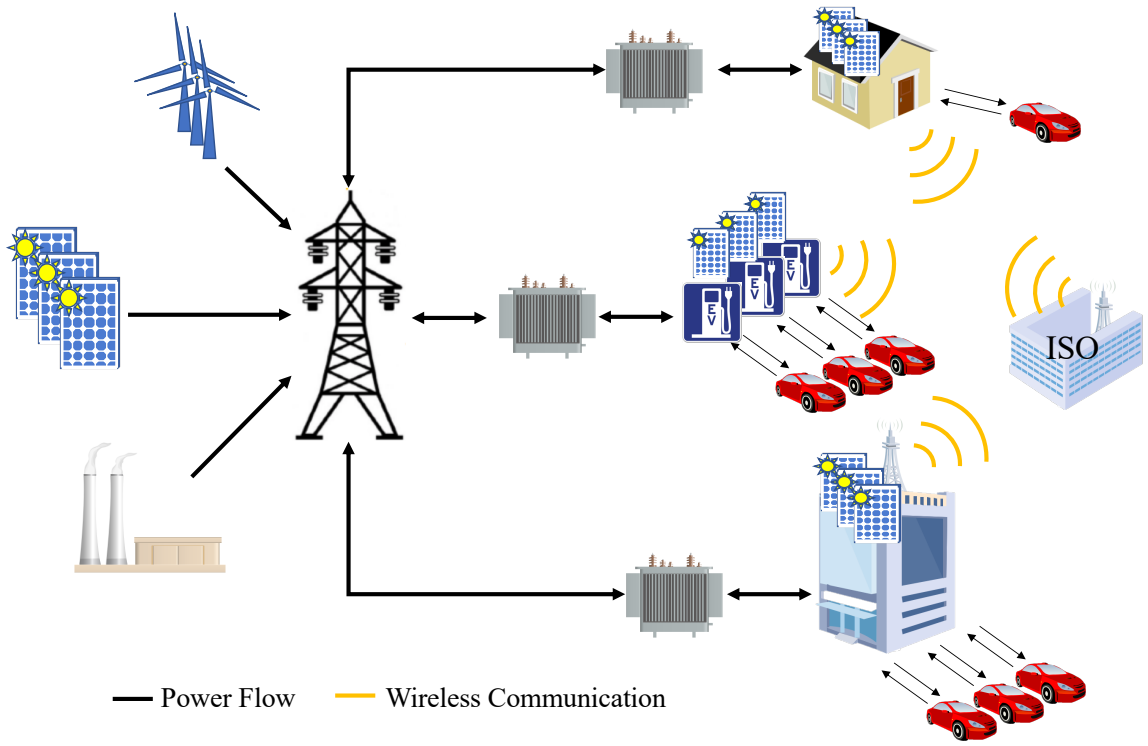

Figure 1.7: Schematic of EV-grid integration

Advanced communication scheme and vehicle-to-grid (V2G) are the key technologies to enable the EV-grid integration presented in Fig. 1.7. V2G technology enables bidirectional power flow between EV and electrical grid for EVs to support grid services. Monetary incentives may be applied to motivate EVs' participation so that utilities can achieve a certain level of grid management. Given that V2G with rapid charging/discharging may result in batter degradation, a study by NREL has shown that there is little to no impact on the battery life if operating with a proper V2G control [KMM16]. [UJW17] suggests that the optimal V2G control is able to reduce the EVs' battery pack capacity fade by up to 9.1% and power fade by up to 12.1%, namely the battery life can be extended with this V2G control compared to the scenario without V2G.

The dynamic electricity price is posted by an independent system operator (ISO) based on the system-wide energy consumption so that the end-users can adjust their load accordingly. Vehicle-to-Home (V2H) and vehicle-to-building (V2B) are also presented in Fig. 1.7. They

can be achieved by applying smart charging control. It manages the charging/discharging schedule to shave the load during peak hours, charge the EVs at low electricity price, or serve as a backup power source during outages. Distributed EV charging stations have great potential to facilitate renewable generation integration and enhance the smart grid's reliability. While RERs are largely deployed across the distribution grid, EVs can be utilized as a distributed energy storage system to support voltage and frequency regulation and even power compensation. These grid services provide benefits to the grid operator and the EV owner and charging network operators through lower or more predictable charging costs.

1.1.2 Smart EV Charging Framework

This section takes a close look at the smart EV charging system. Smart charging allows EV to be externally controlled for integration into the whole power system. Fig. 1.8 shows a smart charging framework with potential communication protocols to be used. A charging point operator(CPO) manages the EV charging points within the charging station(s) instructed by a smart charging service. A smart charging service has a close relationship with the distribution system operator (DSO) and coordinates other CPOs, smart buildings, and energy suppliers in the network. Smart charging service can be seen as an EV charging alliance that brings together the CPOs in the network to achieve the grid management objectives by optimizing the EV charging schedule. The more the CPOs participate in the alliance, the more significant result can be made. In this scenario, a scalable charging scheduling approach is needed to manage a large scale of EVs, such as a hierarchical distributed framework proposed by [KCG19].

The description of potential protocols presented in Fig. 1.8 is as follows. **Open Charging Point Protocol (OCPP)** is an open protocol provided by Open Charge Alliance. It is designed to standardize the communications between EV charging points and the control center. **Open Automated Demand Response Standard (OpenADR)** is an open protocol provided by OpanADR Alliance aiming at automating DR communication and support a system or device to the change power consumption. **Open Smart Charging**

Protocol (OSCP) is a protocol provided by Charge Alliance. The protocol communicates the forecast of the available capacity from the DSO to other systems. It is based on a budget system where CPOs can claim their energy consumption budget. Then the smart charging service will accommodate the energy demand within the boundaries of the available capacity. **Open Charge Point Interface (OCPI)** is an open and independent roaming protocol for EV that makes it easy to exchange data provided by NKL Nederland. **IEC 61850** is not a protocol but a document that defines communication protocols for intelligent devices and electrical substations. **ISO 15118** is an international standard between the communication between EV and the charging infrastructure. It supports V2G and allows the EV and charging station to dynamically exchange information based on which a proper charging schedule can be (re-)negotiated.

Common standards for EV charging management and the integration of charging stations and distribution networks is desirable due to the rise of EV penetration. Several norms were published at the global level by the International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and Society of Automotive Engineers (SAE). Main international norms related to EV in the aspects of connector, communication, safety, and charging topology are summarized in Fig. 1.9.

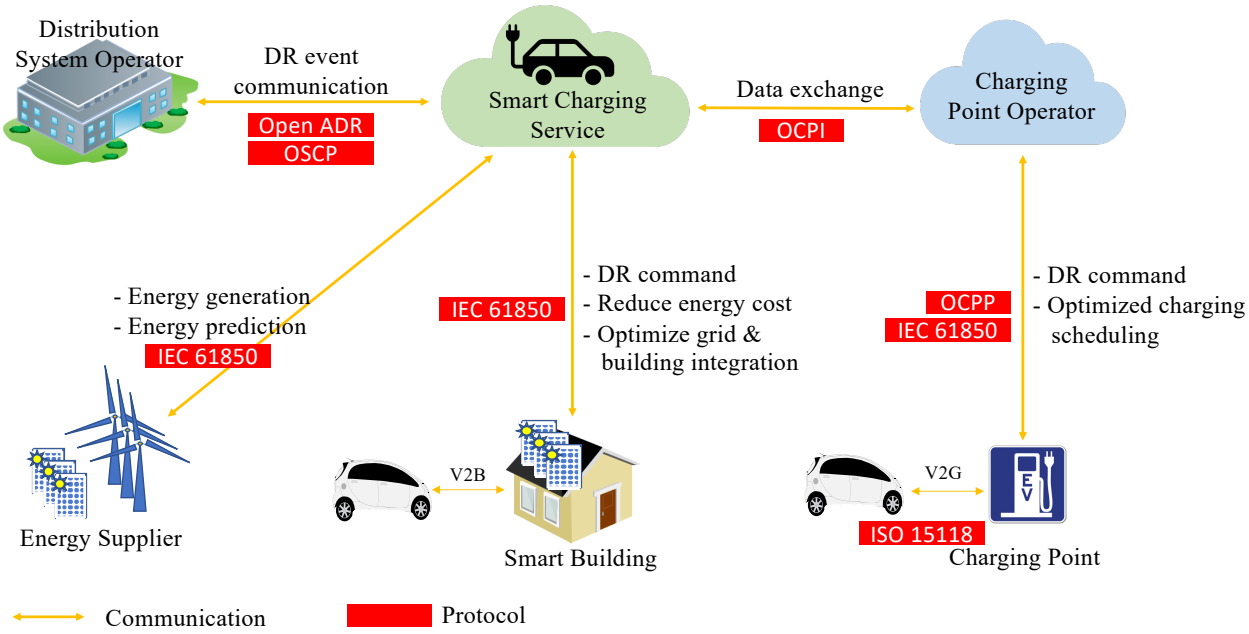

Figure 1.8: Smart EV charging system communication network

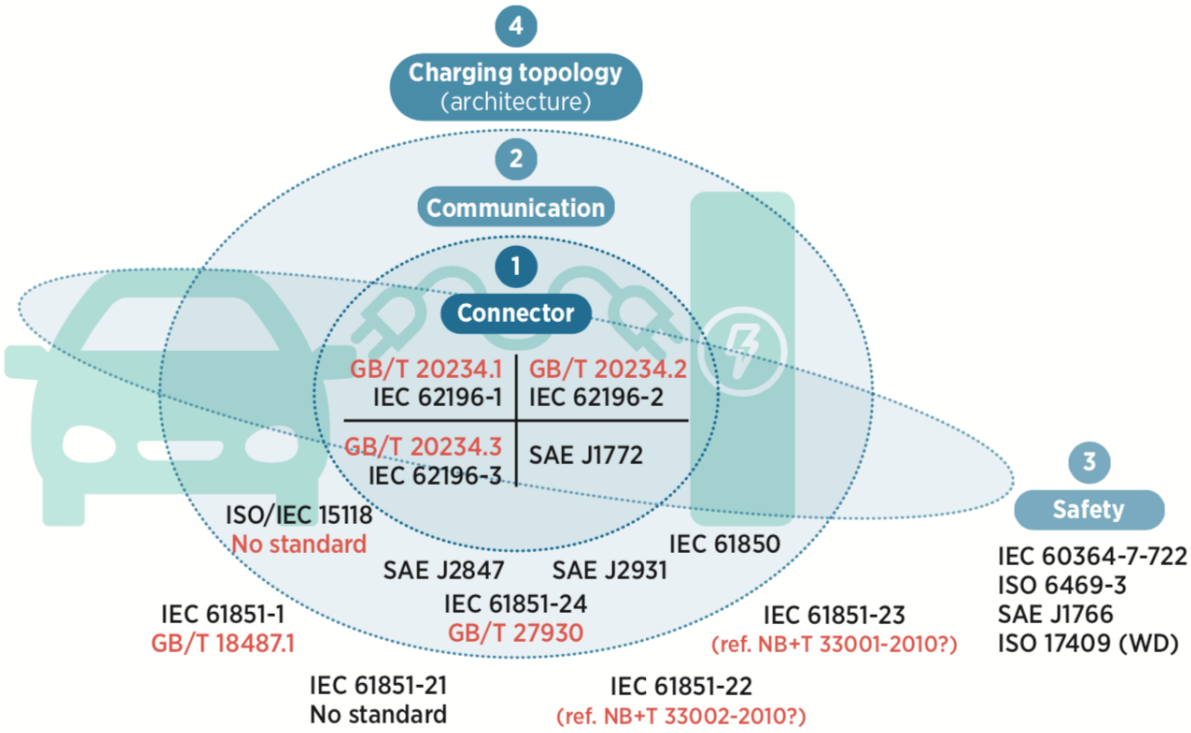

Figure 1.9: Overview of main international norms related to electric mobility [IRE19]

The expanded communication between EV and the charging network in the grid will improve the control of EVs and the integration of smart grids; however, at the expense of creating a larger attack surface. Smart charging development must take cybersecurity into consideration to provide a secure and reliable service.

1.2 Challenges and Contributions

It is shown that the electrical grid is undergoing tremendous change and in the transition to a smart grid. The changes from a centralized system to a decentralized and unidirectional power flow to bidirectional has made the power system more dynamic. As a result, advanced communication and control schemes should be further developed to manage the system and enhance its reliability. This dissertation focuses on the integration of EV into a smart grid, in response to the rapid growth of the EV population. The first challenge for EV-smart grid

integration is to accommodate the massive EV charging demand. To prevent the add up of EV charging power from forming a huge peak, charging scheduling is necessary to relocate the charging time according to each EV's availability and energy demand. However, EV's availability and energy demand are unknown and may be acquired by prediction, yet the user behavior uncertainty makes it challenging to be accurate. Therefore, the second challenge is to handle the EV user behavior uncertainty. There is no doubt that a reliable EV-smart grid integration requires smart charging. The advanced control and communication scheme for smart charging has turned the EV charging system into an information network. The enhanced control and monitoring of EV charging are at the cost of being more vulnerable to cyber-attack. Consequently, secure the EV charging system from potential cyber-attack becomes the third challenge.

The contributions of this dissertation are as follows.

1. A scalable smart charging algorithm is devised to manage EV charging with the objectives of reducing net load variance within a grid and lower the charging cost.

2. An ensemble machine learning approach is developed to address EV user behavior uncertainty and provide accurate predictions to leverage the smart charging's performance.

3. A codified methodology and taxonomy for assessing vulnerability and risk of cyber-physical attack on an EV charging network are carried out to create generalized and comprehensive solutions.

4. A novel anomaly detection approach that makes use of the property of invariant correlation under a controlled system is presented to protect an EV charging network.

1.3 Organization

The rest of the dissertation is organized as follows:

Chapter 2- In this chapter, real EV charging data is studied, and the different charging patterns are classified. Also, kernel density estimation, along with many commonly used machine learning prediction approaches, are discussed, aiming at developing a strategy to select the best prediction methods for different classes of charging patterns. An ensemble machine learning algorithm is proposed to leverage the performance of EV charging scheduling.

Chapter 3- In this chapter, a scalable and straightforward EV charging scheduling algorithm is presented, taking into consideration of dynamic electricity price, solar generation, and building load. The approach incorporates the prediction method introduced in Chapter 2 and minimizes the net load variance and charging cost for an EV charging network.

Chapter 4- In this chapter, comprehensive vulnerability analysis and risk assessment for the WinSmartEV^{TM} charging system on the UCLA campus is demonstrated. Several potential failure scenarios for the charging system were defined, and the impacts of potential cyber-physical attacks have been studied. Moreover, a codified methodology and taxonomy are provided for creating a generalizable and comprehensive solution.

Chapter 5- In this chapter, the concept of invariant correlation network is introduced, and the Greedy Gaussian Segmentation (GGS) method is applied to capture the system-wise correlations for the EV charging system. The anomaly detection method is then developed to detect the correlation changes that infer potential cyber-attack or malicious data injection.

Chapter 6- This chapter concludes the dissertation and discusses the future direction of method improvements.

CHAPTER 2

EV Load Modeling and Prediction

2.1 Overview

Electric vehicles (EVs) have received more and more attention since they became an essential part of a smart grid. This is not only because they are environmentally friendly but also because they provide an economical option to people, considering the high price of dwindling fossil fuels. According to *InsideEVs*' statistical report for 2018, around 361 thousand EVs sold in the US while 2 million in total worldwide, and the numbers almost doubled in comparison to that in 2016 [Lov19]. Currently, there are over 614 thousand EVs on the road in California[Vel19], spurred by the government's zero-emission vehicle mandate to achieve the goal of accommodating 1.5 million EVs by 2025 (California Executive Order B-16-2012). Therefore, a sharply increasing number of EVs on the road is foreseen. However, the increasing number of EVs also means that the rise of energy demand and is now becoming a challenge to the electrical grid. Based on the EV charging data collected on the University of California, Los Angeles (UCLA) campus, the average energy consumption is about 8 kWh per charge, which is similar to a daily household energy demand. EV charging load often shows two peaks in a day, one in the morning when people plug in the EV at the workplace and the other in the evening when people get home from work. Without proper energy management for EV charging, the huge power demand due to a large number of plugged-in EVs can stress the distribution grid, degrade the power quality [MWJ14, SIF15], and impact the wholesale electricity market [FTC13]. The AAA Foundation report reveals that US drivers spend only 0.8 hours in average behind the wheel everyday [Joh18] and mostly leave vehicles parked. This implies that EVs can have a great flexibility for charging and is it not necessary to

start charging right after plugged-in. Thus EV charging scheduling plays an important role in distributing and allocating the charging time according to the EVs' availability for overall load management. A proper EV load management not only mitigates the adverse effects of EV charging but also brings benefits to the grid such as load valley filling and peak shaving [GTL13]. Also EV as a mobile battery has a potential to participate in electricity market [KCM11]. Yet the stochasticity of EV user charging behaviors, including start time, stay duration, and energy demand, poses a significant challenge for the management of charging scheduling. Therefore, this chapter discusses and compares several commonly used prediction methods, aiming at developing an accurate predicting model for EV user behavior in order to improve energy management performance. In addition, since the predicting methods, such as regression or kernel density estimator (KDE) are based on the historical data and the historical charging patterns may be very different from each other, there is no one-size-fits-all predicting method for all different EV users. Thus, this chapter analyzes and classifies the different charging patterns, and uses different predicting algorithms accordingly.

2.2 Literature Review

Forecasting EV load and it's impact to a distribution grid has recently been brought to light by the development of smart grids and the growing number of EVs. However, due to limited access to real EV charging data, synthetic data from travel surveys are used for the majority of these studies. Gennaro et al. [GPS14] utilized the data collected from conventional fuel vehicles. Harris et al. [HW14] synthesized EV charging profile by using vehicle trip data from the National Household Travel Survey (NHTS). Wang et al.[WZO15] simulated EV energy consumption using car travel survey. In spite of the early stage of EV adoption, some utilities and aggregators have been collecting data from charging stations to gain insight into EV user behavior [Sma, XWC18]. EA Technology[Eat16] have conducted a three year project to collect data and investigate the impact of clusters of EVs on the electrical grid in the UK. There are two types of data can be used for the forecasting, which are station record and charging record. Station record comes directly from the measurement at the charging

outlets while charging record comes from the measurement of each user's charging session. In other words, station record is the aggregated load data over time and charging record is the data for a specific user during a charging session. In [MQC15], multiple methods including a k-nearest neighbor (KNN), a lazy-learning algorithm and a pattern sequence algorithm have been evaluated for aggregated EV load estimation. In [AKK16, WHQ15], an autoregressive integrated moving average (ARIMA) method has been proposed for aggregated EV load forecasting. In [XMC16], a data mining model was developed to predict EV charging demand for a geographical area. In [MQC14], modified pattern-based sequence forecasting (MPSF) was in comparison with KNN, support vector regression (SVR) and random forest (RF) algorithms and showed more accurate performance. Also, the aggregated data of EV loads may be used for coordinating the EV charging operation as in [XHS14]. However, to schedule the charging when EVs are plugged in, the charging parameters in each session are preferred instead of the aggregated load information. Furthermore, the prediction by aggregated load requires a large amount of EV charging data and currently the availability of the data is limited. The author in [MQC16] discussed EV charging load forecasting by using station records and charging records, and the results showed that charging record based prediction is faster and more accurate. The method such as Gaussian-based kernel density estimator (GKDE) has been applied to handle the uncertainties of user behaviors for each charging session in [WWN17, WSW17, WRW16]. But the use of optimal bandwidth selection for GKDE, a.k.a the normal reference rule [Sil86], usually leads to an over-smoothed probability and results in less accurate prediction. To overcome the deficiency of the normal reference rule, kernal density estimation via diffusion (DKDE) [BGK10], which provides a better bandwidth selection approach, has been used to improve the prediction accuracy of EV charging behavior[KCC18, CKC18]. By examining the performance of the algorithms applied to EV user behavior prediction, it is noted that the variances of the errors are usually large. This is because the EV charging patterns vary significantly and there is no unique algorithm that works for all. [CKC18] compares and discusses DKDE and GKDE, and the the result shows that DKDE has a higher accuracy for the users who charge their EVs regularly while GKDE works better for the irregulars. However, the overall performance for the prediction

still has room for improvement. To the best of the author's knowledge, there is no effective feature that can categorize different EV charging patterns associated with the most accurate predicting algorithms. Therefore, this chapter aims at classifying different charging patterns and uses the best approach to predict the charging behavior in each classification.

2.3 User Behavior Prediction Model

This section describes the method for EV user behavior prediction. The objective is to predict each specific EV user's stay duration and energy demand based on their historical charging data when they plug in their EVs. For each charging session, a 5-tuple of parameters is used to describe a charging behavior:

$$s \triangleq (u_{id}, t_s, t_d, d_w, e), \tag{2.1}$$

where u_{id} is the unique identifier (*user ID*) for each user in our system; t_s and t_d denote *start time* and *stay duration*, respectively; d_w denotes *day of week*; and e denotes *energy consumption*. Those charging parameters are of vital of importance for EV charging scheduling algorithms to determine an optimal solution. To be specific, once a user initiates a charging session, the predictions of *stay duration* and *energy consumption* are required for the scheduling services to determine energy allocation schedule. It is noted that *stay duration* is related to *start time* and *day of week* since users in our model may have their fixed weekly working schedules. Therefore, the prediction of *stay duration*($\widehat{t_d}$) can be expressed as follows:

$$\widehat{t_d} = f_d(t_s, d_w). \tag{2.2}$$

Also, *energy consumption* is related to *start time*, *day of week*, and *stay duration*, such that:

$$\widehat{e} = f_e(t_s, d_w, \widehat{t_d}). \tag{2.3}$$

As shown in (2.3), when predicting *energy consumption*, the *stay duration* is unknown and thus rely on its predicted value, $\widehat{t_d}$. The predicting procedure is illustrated in Fig. 2.1.

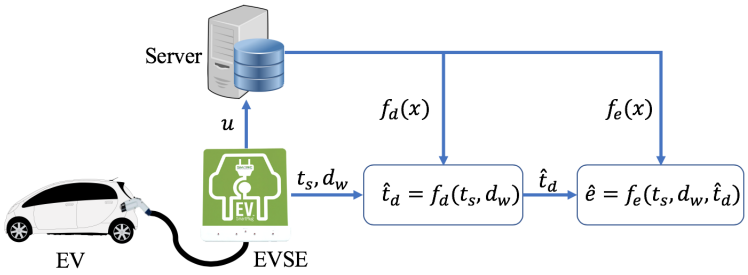

Figure 2.1: User behavior prediction

To evaluate the performances of different prediction algorithms, symmetric mean absolute percentage error (SMAPE) is chosen here based on the following reasons:

1. SMAPE is a unit free percentage error and it is easier to present the prediction accuracy with different data sets, which are *stay duration* and *energy consumption* in this paper.

2. Percentage error such as mean absolute percentage error (MAPE, defined as MAPE = mean $(|y - \hat{y}|/y)$) has a problem when y value becomes very small. This small value will result in a huge error that bias the overall accuracy. Therefore, SMAPE would be more accurate since it considers both y and $\hat{y}$ in the denominator, given that the data is strictly positive.

3. SMAPE is widely used in evaluating EV charging prediction accuracy, it would be easier for comparison.

For charging session i, the SMAPE is defined as:

$$SMAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{P}(i) - T(i)|}{\hat{P}(i) + T(i)}, \tag{2.4}$$

where N is the number of charging sessions, $\hat{P}$ is the prediction, and T is the corresponding true value.

2.4 Hybrid Kernel Density Estimator (HDKE)

This section presents a hybrid kernel density estimator (HKDE) that uses both Gaussian- and Diffusion-based KDE (GKDE and DKDE) to predict the stay duration and charging demand of electric vehicles (EVs), which are essential parameters for optimizing EV charging schedule. While DKDE has higher accuracy in general, GKDE tends to result in better estimation for users who charge the EV irregularly. Therefore, the HKDE evaluates and categorizes the charging pattern regularity of a user, and determines which KDE to use by a novelty detection method based on the users historical data. The estimations are then applied to an optimal EV charging algorithm to minimize load variance in an EV charging infrastructure and reduce EV charging cost.

2.4.1 Gaussian Kernel Density Estimator (GKDE)

KDE is widely used as a nonparametric distribution estimation method. Given an observed dataset $X = [X_1, X_2, ..., X_N]$, a probability density function can be estimated as follows[Cri16]:

$$\hat{P}_{\text{KDE}}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K(\frac{x - X_i}{h}) \tag{2.5}$$

where N is the size of X, h is the bandwidth of the Gaussian kernel $K(\cdot)$, and $K(\cdot)$ is defined as:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}u^2)} \tag{2.6}$$

Bandwidth h defines the shape of the kernel function and thus is a deterministic factor to the performance of the estimator. A large h oversmoothes the density function that masks the structure of data while a small h generates a spiky one that makes the interpretation difficult. It is desired to find a value of h that minimizes the error between the estimated density and the true density. However, there is a bias-variance trade-off for the bandwidth selection, which means a large bandwidth reduces the variance of $\hat{P}_{\text{KDE}}(x)$ but increases the bias with respect to the true density. On the other hand, a small bandwidth decreases the bias of $\hat{P}_{\text{KDE}}(x)$ at the expense of larger variance. Silverman's rule of thumb [Sil86], also

known as the normal reference rule, provides a simple solution for the optimal bandwidth, with the assumption that the true density has Gaussian normal distribution. The optimal bandwidth determined by the normal reference rule is as follows:

$$h^* \cong 1.06\sigma N^{-\frac{1}{5}}, \tag{2.7}$$

where σ is the sample standard deviation of N training examples.

However, this method usually leads to an over-smoothed result in multimodal models such as EV user charging behaviors.

2.4.2 Kernel Density Estimator via Diffusion (DKDE)

Different from the normal reference rule, the optimal bandwidth can be derived from the observed dataset X using an improved plug-in method introduced in[BGK10].

The kernel density (2.5) can be expressed in an alternative form:

$$\hat{f}_{\mathrm{X}}(x;y) = \frac{1}{N}\sum_{i=1}^{N}\phi(x, X_i; y), \tag{2.8}$$

where

$$\phi(x, X_i; y) = \frac{1}{\sqrt{2\pi y}}\exp(-\frac{(x-X_i)^2}{2y}), \tag{2.9}$$

in which $\sqrt{y} = h$ is defined as in (2.5).

The main observation is that GKDE (2.8) is the unique solution to the Fourier heat equation as follows [BGK10]:

$$\frac{\partial}{\partial y}\hat{f}(x;y) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\hat{f}(x;y), \quad x \in \mathscr{X}, \quad y > 0, \tag{2.10}$$

with initial condition:

$$\hat{f}(x;0) = \frac{1}{N}\sum_{i=1}^{N}\delta(x - X_i), \tag{2.11}$$

where $\hat{f}(x;0)$ represents the empirical density of X, and $\delta(x - X_i)$ is the Dirac measure at X_i. The Neumann boundary condition to solve the diffusion equation (2.10) is as follows:

$$\frac{\partial}{\partial y}\ \hat{f}(x;y)\Big|_{x=1} = \frac{\partial}{\partial y}\ \hat{f}(x;y)\Big|_{x=0} = 0. \tag{2.12}$$

Exploiting the link between GKDE and Fourier heat equation, finding the optimal bandwidth of (2.8) is equivalent to finding the optimal mixing time y^* of the diffusion process governed by (2.10). Considering those conditions and the finite domain $[0, 1]$, the analytical solution of (2.10) is obtained by:

$$\hat{f}_{\text{diff}}(x; y^*) = \frac{1}{N} \sum_{i=1}^{N} \kappa(x, X_i; y^*), \quad x \in [0, 1], \tag{2.13}$$

in which the kernel function is given by:

$$\kappa(x, X_i; y^*) = \sum_{k=-\infty}^{\infty} \phi(x, 2k + X_i; y^*) + \phi(x, 2k - X_i; y^*), \quad x \in [0, 1], \tag{2.14}$$

and it is equivalent to:

$$\kappa(x, X_i; y) = \sum_{k=-\infty}^{\infty} e^{-k^2\pi^2 y/2} cos(\pi x) cos(\pi X_i). \tag{2.15}$$

Although both estimators (2.8) and (2.13) behave similarly in the interior of the domain [0,1] for a small bandwidth, (2.13) performs better near the boundaries where $x = 0, 1$. The reason is that DKDE is consistent with the true density while GKDE is inconsistent at the boundaries.

The optimal bandwidth can be expressed as follows:

$$y^* = \left(\frac{6\sqrt{2} - 3}{7} \right)^{2/5} \gamma^{[l]}(y), \tag{2.16}$$

in which

$$\gamma^{[l]}(y) = \gamma_1(\gamma_2(\cdots \gamma_l(y) \cdots)), \ l \leq 1, \tag{2.17}$$

and

$$\gamma_l(y) = \left(\frac{\left(1 + (1/2)^{(l+1/2)}\right)(1 \times 3 \times \cdots (2l-1))}{3N\sqrt{\pi/2} \parallel \hat{f}^{(2)}_{\text{diff}} \parallel^2} \right)^{2/(3+2l)}, \tag{2.18}$$

where $l = 5$ in this paper as recommended by [BGK10].

2.4.3 Implementation of HKDE

Here the one-class SVM for novelty detection according to Schölkopf[SWS00] is deployed to examine how different a test data is from the corresponding training data. Considering

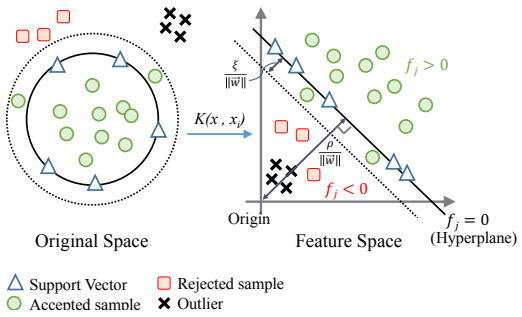

Figure 2.2: One-class SVM

a dataset $x = \{x_1, x_2, ..., x_n\}$, the kernel function maps all the data points into a feature space, and the one-class SVM generates a hyperplane, which serves as a decision boundary, to separate them from the origin while maximizing the distance from this hyperplane to the origin. This leads to a binary function f which returns $+1$ in a small region capturing most of the training data points, and -1 elsewhere. The schematic figure of one-class SVM is illustrated below in Fig. 2.2.

The following quadratic programming minimization function is used to separate the data set from the origin:

$$\min_{w,\xi_i,\rho} \frac{\|w\|^2}{2} + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho \tag{2.19a}$$

$$\text{s.t.}\,(w \cdot \phi(x_i)) \geq \rho - \xi_i \quad \& \quad \xi_i \geq 0,\ i = 1, ..., n \tag{2.19b}$$

where w is the normal vector to the hyperplane, $\phi(x_i)$ is the transformation function of x_i, $\rho \in R$ is a bias, ξ_i is the slack variable. $\nu \in [0, 1]$ characterizes the fractions of support vectors (SVs) and outliers. More specifically, ν sets an upper bound on the fraction of outliers and lower bounds on the number of training data points used as SV.

The minimization problem can be solved by Lagrange multipliers and the decision rule shown below:

$$f(x) = sgn\left((w \cdot \phi(x_i)) - \rho\right) = sgn\left(\sum_{i=1}^{n}\alpha_i K(x, x_i) - \rho\right) \tag{2.20}$$

where α_i is the Lagrange multiplier, and $K(x, x_i) = \phi(x)^T \phi(x_i)$ is the kernel function, and

in this paper, Gaussian radial basis function (RBF) is used:

$$K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}} \tag{2.21}$$

where $\sigma \in R$ is the kernel parameter.

To have high accuracy predictions, determining the parameter ν in (2.19) and the threshold for the HKDE is paramount. ν in the range $[0.1, 0.7]$ is examined to find the optimal one which minimizes prediction error. To determine the threshold, 20-month EV charging data collected from 55 users on UCLA campus as well as 52 users' charging data from EA Technology [Eat16] is analyzed, of which 60% is for training set, 20% for the validation set and 20% for the test set. For each user's data, both GKDE and DKDE are applied, and the errors are calculated using the validation set. Novelty detection using the Scikit-learn framework[PVG11] is utilized to find the out-of-class rate (OCR) for the training set and validation set, called TOCR and VOCR, respectively. OCR is defined as the number of data points that is out of the classification over the total number, according to the decision rule shown in (2.20). The difference between TOCR and VOCR, called DOCR, is defined in (2.22), which its optimal value is the threshold minimizing the HKDE error.

$$DOCR = \frac{VOCR - TOCR}{TOCR} \tag{2.22}$$

Once the optimal values of ν (ν^*) and threshold ($DOCR^*$) are determined by the procedure shown in Algorithm 1, the proposed HKDE is able to select an appropriate KDE for

each EV user in the system.

Algorithm 1: Calculation of $DOCR^*$ and ν^*

Initialization:

- Retrieve historical charging data
- Separate the data into training and validation set

for $\nu \in [0.1, 0.2, ..., 0.7]$ **do**
  STEP.1: Record the errors of GKDE and DKDE, and calculate DOCR
  for $UserID = 1 : \mathcal{N}$ **do**
    Calculate the errors of GKDE, DKDE
    Apply one-class SVM and calculate DOCR (2.22)
  end
  STEP.2: HKDE Test
  for $r \in [DOCR_{min}, ..., DOCR_{max}]$ **do**
    for $UserID = 1 : \mathcal{N}$ **do**
      if *user's* $DOCR \leq r$ **then**
        DKDE
      else
        GKDE
      end
    end
    Calculate HKDE error
  end
  $Threshold = DOCR^*$
  Evaluate the errors of GKDE, DKDE and HKDE
end
Determine ν^*

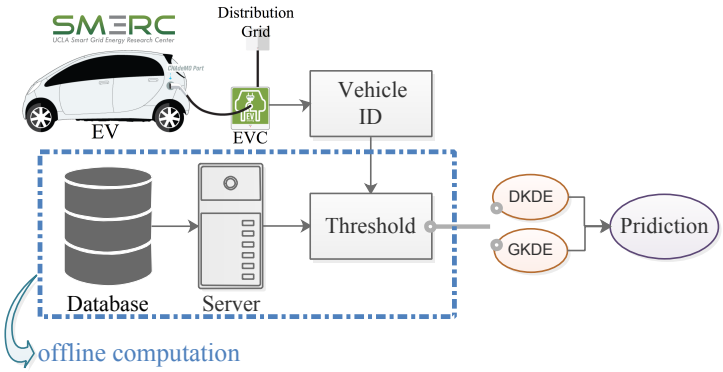

Figure 2.3: Schematic overview of the proposed HKDE.

As Fig. 2.3 shows, when a vehicle plugs in, the EV charging system recognizes the vehicle, and the proposed HKDE predicts its stay duration and energy consumption.

2.4.4 Results and Discussion

Regular and irregular charging patterns are identified by the novelty detection as shown in Fig. 2.4. Learned frontier is the boundary of the classification, which is the hyperplane in the feature space, and the color bar shows the mapping distance in the feature space to this hyperplane. The regular user starts charging the EV around 10 am to 12 pm or earlier at 7am regularly, and the stay duration is around 4 hours. In contrast, for the irregular users, the start time and stay duration vary along the axes, and most of the validation data points are located outside of the frontier.

The value of ν^* for equation (2.19) is determined by Algorithm 1, and the result shows that $\nu^* = 0.4$ as demonstrated in Fig. 2.5. Some charging events with small value of duration or energy consumption can produce very large errors and skew the overall error rate. Therefore, symmetric mean absolute percentage error (SMAPE) is used to mitigate the influence of those small values and evaluate the prediction errors. Since the number of charging sessions varies in each test day, SAMPE is modified to fit the use case as shown in the following:

$$SMAPE = \frac{1}{N_d}\sum_{i=1}^{N_d}\frac{1}{E_i}\sum_{j=1}^{E_i}\frac{|T_j^i - \hat{P}_j^i|}{T_j^i + \hat{P}_j^i} \times 100\% \tag{2.23}$$

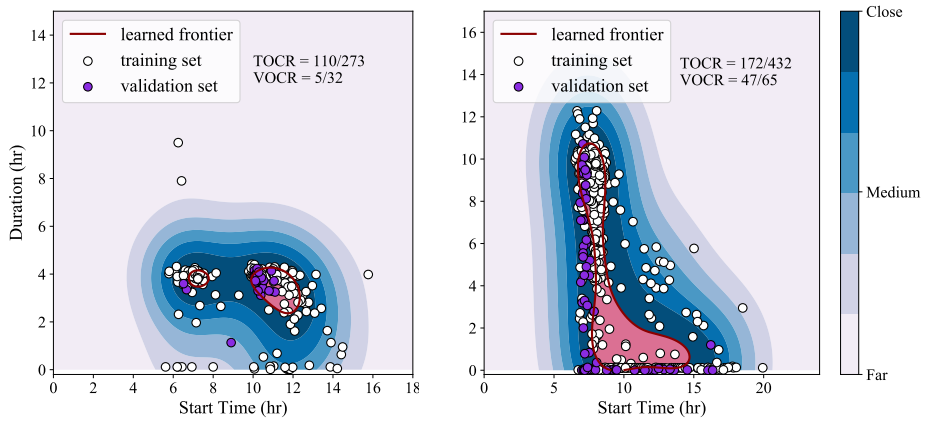

Figure 2.4: EV user behaviors under novelty detection analysis. **Left:** regular user; **:**Right: irregular user.

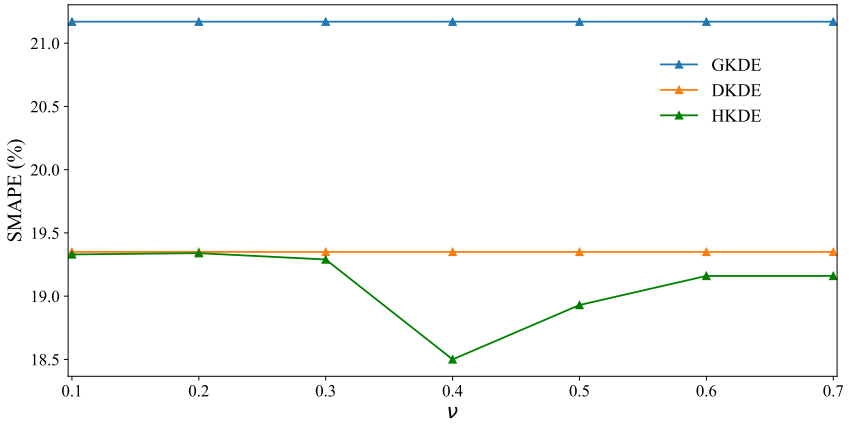

Figure 2.5: SMAPE of stay duration with different values of ν.

Where N_d is the total number of days that is evaluated, E_i is total number of charging sessions in the i-th day, $\hat{P}_j^i$ is the j-th prediction value of i-th day, and T_j^i is the corresponding actual value, i.e. the true value of stay duration or energy consumption.

The proposed HKDE predicts the charging parameters when an EV plugs in. Fig. 2.6 and Fig. 2.7 show the prediction errors of stay duration and energy consumption, respectively, using the test dataset for each user. The proposed HKDE selects the KDE with smaller error for predicting charging behavior in the most cases.

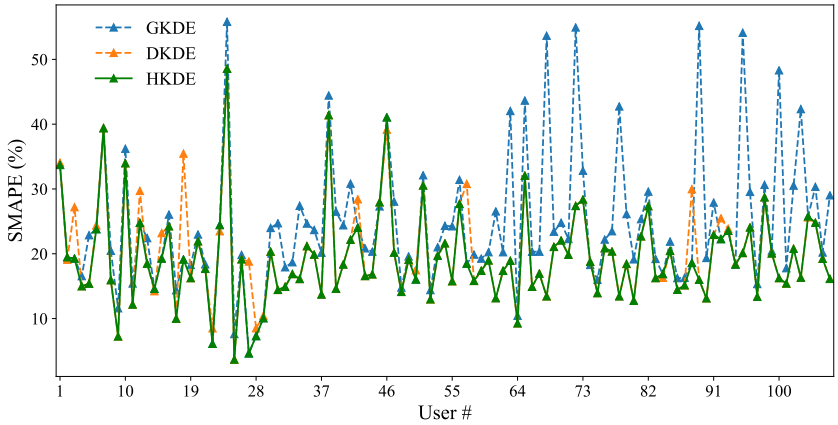

Figure 2.6: SMAPE of stay duration.

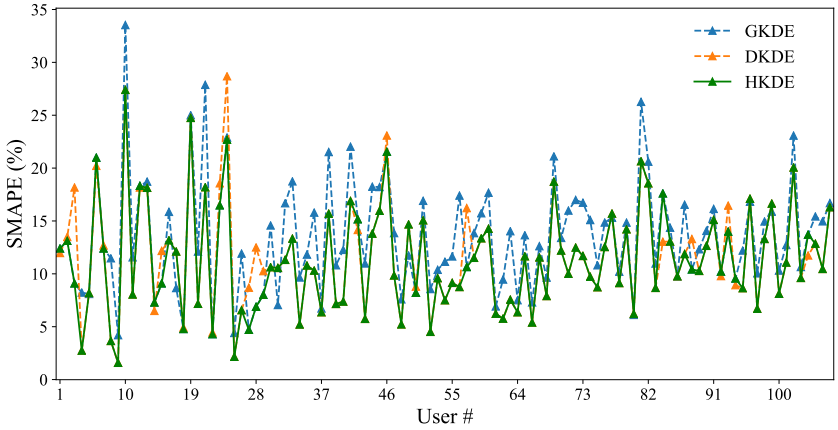

Figure 2.7: SMAPE of energy consumption.

Mean estimation deviation (MED) in (2.24) is also defined to assess the overall accuracy for the predictions.

$$MED = \frac{1}{N_{user}} \sum_{i=1}^{N_{user}} \left(\sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (\hat{P}(j) - T(j))^2} \right), \tag{2.24}$$

where N_{user} is the number of users, N_i is the number of charging sessions for the i-th user, $\hat{P}$ is the prediction, and T is the corresponding true value. For HKDE, the prediction MED for stay duration is 1.42 hour, and for energy consumption is 3.38 kWh. Table 2.1 shows the comparison of MED among GKDE, DKDE and HKDE.

Table 2.1: Comparison of prediction MED.

	Stay Duration (hr)	Energy Consumption (kWh)
GKDE	1.93	4.96
DKDE	1.44	3.84
HKDE	1.42	3.38

2.5 Ensemble Machine Learning Method

This section investigates electric vehicle (EV) charging behavior and aims to find the best method for its prediction in order to optimize the EV charging schedule. Here we discusses several commonly used machine learning algorithms to predict charging behavior, including stay duration and energy consumption based on historical charging records. It is noted that prediction error increases along with the rise of data entropy or the decrease of data sparsity. Thus, this paper accounts for both indicators by defining the entropy/sparsity ratio (R). When R is low, support vector regression (SVR) and random forest (RF) regression show better accuracy for stay duration and energy consumption predictions, respectively. While R is high, a diffusion-based kernel density estimator (DKDE) performs better for both predictions. The three methods are assembled as the proposed Ensemble Predicting Algorithm (EPA) to improve predicting performance by decreasing 11% of the duration and 22% of the energy consumption prediction errors.

2.5.1 Machine Learning Algorithms

Since the charging pattern varies from each other, there is no one-size-fits-all predictor for all EV users. Therefore, EV users charging patterns were classified, and eight different prediction algorithms were applied to those different classes for comparison to find the optimal solution. Eight prediction algorithms are reviewed here in the following sub-sections. By examining the EV charging data, the data can be roughly classified into four categories: linear, non-linear, clustered, and scattered patterns. Multiple linear regression is suitable

for a linear pattern. SVR can predict both linear and non-linear patterns and is not biased by outliers. Decision tree (DT), random forest (RF) regression are appropriate for clustered patterns. RF can be more accurate than DT since DT may easily lead to over-fitting. However, it is required to determine the proper number of trees for RF. KNN regression can also be applied for a clustered pattern. Scattered pattern is challenging for prediction. In this case, GKDE and DKDE are used to find the probability density function and make a prediction by calculating the expected value. A statistical method is applied here for comparison. These algorithms are compared, and their effectivenesses are evaluated for different EV charging patterns.

2.5.1.1 Statistical Method

Statistical method such as historical average are referred to as a naive approach, and it is a simple algorithms that used only for comparison with the other forecasting techniques. For the historical average algorithm, the prediction is the average of the past data.

2.5.1.2 Multiple Linear Regression(MLR)

MLR is used to describe the mathematical relationship between several explanatory variables and a response variable, and the goal is to make predictions about the response variable based on the known explanatory variables according to this relationship. For example, to predict a *stay duration* based on the *start time* and *day of week*. The model of MLR with k explanatory variables and n observations is as follows:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} ... + b_k x_{ik} + e_i \quad for \;\; i = 1, 2, ...n, \tag{2.25}$$

where y_i is the response variable, b_0 is the y-intercept term, $[b_1, b_2, ..., b_k]$ are the regression coefficients, $[x_{i1}, x_{i2}, ..., x_{ik}]$ are explanatory variables and e_i is the error term, which is also known as residual that is used to account for the difference between the actual outcome and the prediction. In this paper, for *stay duration* prediction, x_{i1} is *start time* and x_{i2} is *day of week* ($k = 2$). For *energy consumption*, x_{i1} is *start time*, x_{i2} is *day of week*, and x_{i3} is *stay du-*

ration($k = 3$). Here we use the Python package (*sklearn.linear_model.LinearRegression*) [BLB13] for the MLR model.

2.5.1.3 Support Vector Regression (SVR)

SVR is a type of support vector machine that supports linear and non-linear regression. Unlike general linear regression methods, which try to minimize the error between the prediction and data, SVR makes sure the errors do not exceed the threshold. Specifically, in ε-SVR[Vap95], the goal is to find a function $\hat{y}(x)$ that has at most ε deviation from the obtained targets y_i for the training data, ignoring the outliers that locate outside of the ε-tolerence band. Consider a training dataset $\{(x_1, y_1), ..(x_n, y_n)\} \subset \mathbb{X}$, where $\mathbb{X}$ denotes the space of the input data, SVR can be expressed as follows:

$$\hat{y}(x) = \langle \omega, x \rangle + b \quad with \quad \omega \subset \mathbb{X}, b \subset \mathbb{R}, \tag{2.26}$$

where ω an b are the solutions of the following optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|\omega\|^2 + \mathcal{C}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{subject to} \quad & \begin{cases} y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0. \end{cases} \end{aligned} \tag{2.27}$$

In equation(2.27), slack variables ξ_i, ξ_i^* are introduced to handle the problem of infeasible ε-precision constraints. The constant $\mathcal{C} > 0$ controls the trade-off between the flateness of $\hat{y}(x)$ (which is $\|\omega\|^2$) and the number of training data points that deviate larger than ε is tolerated. This optimization problem can be solved by Lagrange multipliers method and the solution is given by

$$\omega = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\Phi(x_i) \quad and \quad f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k(x_i, x) + b, \tag{2.28}$$

where α_i, α_i^* are Lagrange multipliers in which $\alpha_i, \alpha_i^* \in [0, C]$, $\Phi(x_i)$ is a transformation function, and $k(x_i, x) \triangleq \langle \Phi(x_i), \Phi(x) \rangle$ is a kernel function. The kernel function transforms

the data into a higher dimensional feature space to make it possible to perform the linear regression. The Gaussion radial basis function (RBF) is used here as a kernel function:

$$k\left(x, x_i\right) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}}, \tag{2.29}$$

where $\sigma \in \mathbb{R}$ is the kernel parameter. The detail of the SVR formulation can be found in [SS03]. Here we use the Python package ($sklearn.svm.svr$) [BLB13] for the SVR model.

2.5.1.4 Decision Tree (DT) Regression

Decision tree (DT) regression is a regression model in the form of a tree structure that breaks down a dataset into smaller classified subsets using each of the independent variables' split points. The average of the classified subset is the prediction value for the target with respect to its corresponding independent variable values. The classified subsets are called leaf nodes whereas the split points are decision nodes. For each decision node, mean square error (MSE) are compared across the independent variables and the variable/point rendering the lowest MSE is chosen as the root node/decision node. The process is recursively continued until the optimal split of the data is achieved, which is defined in terms of tree size constraints within the Python package ($sklearn.tree.DecisionTreeRegressor$)[BLB13] used here.

2.5.1.5 Random Forest (RF) Regression

Random forest (RF) regression is an ensemble learning method that combines and averages decisions from a sequence of DT models. Formally, RF regression can be expressed as follows:

$$g(x) = \frac{1}{N_{tree}} \sum_{i=1}^{N_{tree}} f_i(x), \tag{2.30}$$

where $g\left(x\right)$ is the RF model, $f_i\left(x\right)$ is the i^{th} DT model, and N_{tree} is the number of decision trees. Each $f_i\left(x\right)$ is built from a sample drawn with replacement from the training dataset. By using the average of the multiple DT models on the corresponding sub-samples of the dataset, the predictive accuracy can be improved.

Here we use the Python package (sklearn.ensemble.RandomForestRegressor) [BLB13] for the

RF model.

2.5.1.6 K-Nearest Neighbor (KNN) Regression

K-Nearest Neighbor (KNN) is a non-parametric method used for classification and regression [Alt92]. The regression model is used since the data labels are continuous instead of discrete variables. The model implements learning based on the k-nearest neighbors of each query point, where $k = 4$ is specified in this paper. The prediction of a query point is the average of its nearest neighbors, and it is assumed that each neighbor contributes uniformly to the classification of the query point.
Here we use the Python package (sklearn.neighbors.KNeighborsRegressor) [BLB13] for the KNN model.

Kernel density estimators including **GKDE** and **DKDE** are reviewed in Section 2.4.1 and 2.4.2, respectively.

2.5.2 Data Preparation

Two sources of EV charging data were applied to this research, including SMERC charging stations on the UCLA campus[Sma] as working space and real residential EV users' data in the UK that is available from the EA technology website[Eat16]. The data used from the UCLA charging stations was recorded from October 1, 2015 to December 31, 2017 and the data from the EV technology between February 16, 2014 and November 29, 2015. However, not every user in those datasets has a charging history that is long enough for data analysis and prediction. Therefore, we selected 50 users' data from UCLA and 202 form EV technology, which have at least 100 charging records, with 39,458 records in total. The data was split into 70% for the training set, 20% for the validation set, and 10% for the test set.

The statistics for charging *start time*, *stay duration*, and *energy consumption* are shown in the figures below. Fig. 2.8 shows two peaks for EV charging *start time*: one at 7:30 in the morning and the other at 17:30 in the evening. The average *stay duration* is 3 hours, and the average *energy consumption* is 10.63 kWh, as shown in Fig. 2.9 and Fig. 2.10, respectively.

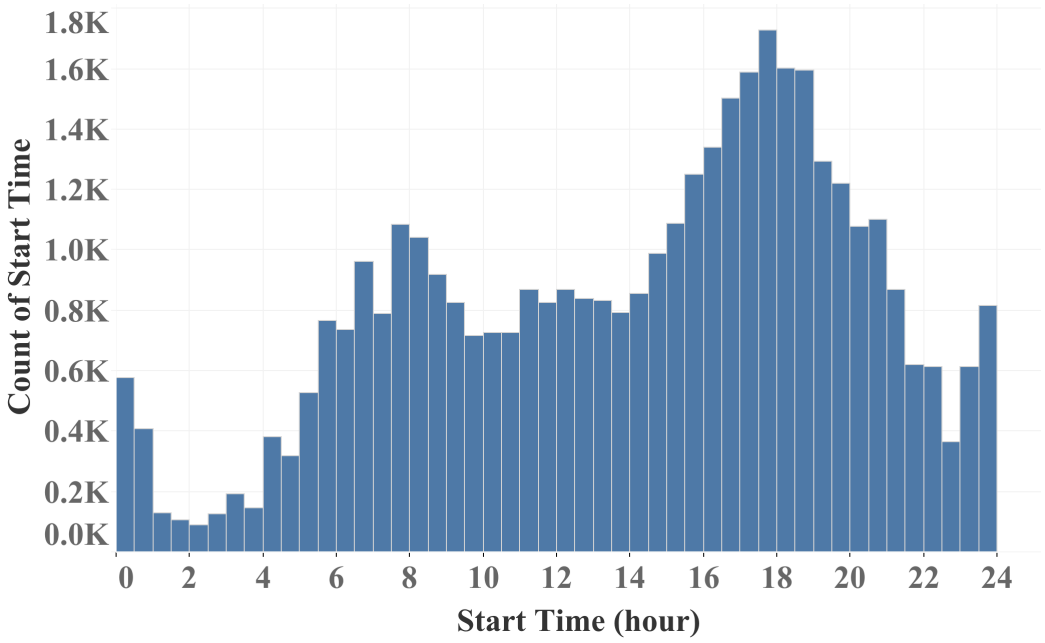

Figure 2.8: Statistics of EV charging *start time*

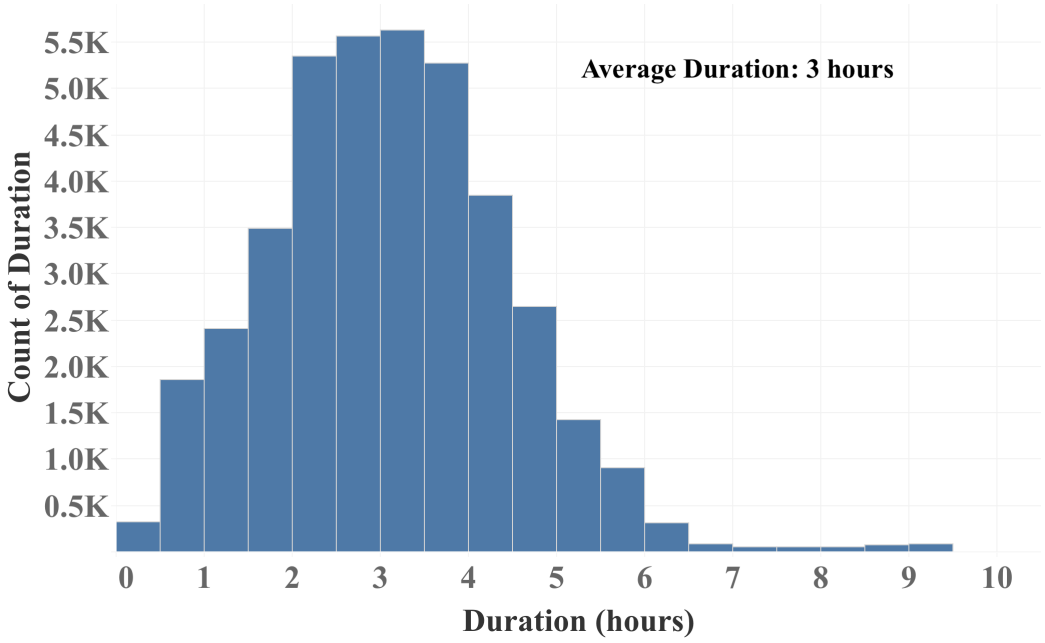

Figure 2.9: Statistics of EV *stay duration*

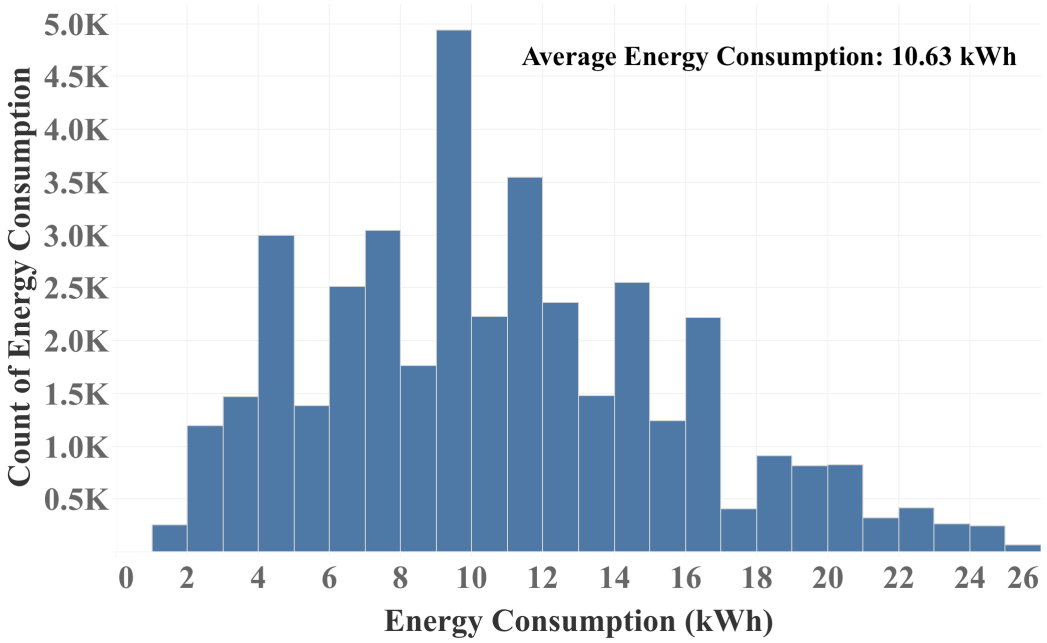

Figure 2.10: Statistics of EV *energy consumption* per charge

2.5.2.1 Data Preprocessing

The charging *start time* and *stay duration* were converted to hour. For instance, 13:15 will be noted as 13.25 hour. If a *stay duration* was smaller than 0.5 hour or an *energy consumption* was smaller than 1 kWh, the entire 5-tuple parameter for that charging session was removed from the dataset. Also, if an *energy consumption* was mistakenly recorded as more than the physical maximum of the charging device, the record value was replaced by the maximum value of its historical *energy consumption*.

2.5.2.2 Data Entropy

Joint entropy is used here to characterize the uncertainty of a set of variables. Two kinds of datasets were analyzed, which are *start time* vs. *duration* data and *duration* vs. *energy consumption* data. For calculation, *start time* and *duration* are rounded to the closest half hour, and the *energy consumption* is rounded to the closest integer. The values of *start time* and *duration* are then mapped into a set of integers $\in [0, 47]$, which represents

[0 : 00, 23 : 30].The formulation of a joint entropy is as follows:

$$E(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)log_2[P(x,y)], \tag{2.31}$$

where x and y are the two variables in dataset X and Y, respectively; $P(x,y)$ is the joint probability of the two variables.

2.5.2.3 Data Sparsity

Sparsity is defined as the number of zero entries devided by the total number of entries. Intuitively, if a sparsity is high the data is less variant because most entries are repeated. On the other hand, for low sparsity, the data is more scattered. As was the data entropy discussed in the previous section, *start time* vs. *duration* data and *duration* vs. *energy consumptuon* data are analyzed. The values of *start time*, *duration*, and *energy consumption* are rounded. Following are the examples of sparsity calculation:

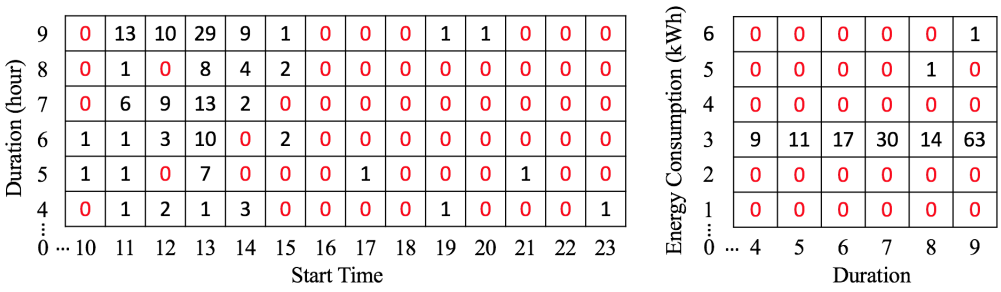

Figure 2.11: Sparsity of EV charging patterns. **Left:** *start time* vs. *duration*, **Right:** *duration* vs. *energy consumptuon*

In Fig. 2.11, the numbers in the cells are the number counts for the data points. For *start time* vs. *duration*, the *start time* ranges from 0 to 23 while the *duration* from 0 to 9. The number of non-zero entries is 31 and the total entries is 240, thus the sparsity is (240-31)/240 = 0.87. In the same manner, the sparsity for *duration* vs. *energy consumptuon* is 0.89.

2.5.3 Preliminary Result and Proposed Algorithm

2.5.3.1 Preliminary Results

Fig. 2.12 and Fig. 2.13 show the comparisons of eight algorithms' prediction errors with regard to data entropy, data sparsity and the ratio of entropy/sparsity (R). Generally, SMAPE positively correlates to data entropy and negatively correlates to data sparsity. Therefore, this paper takes into account both of the effects of entropy and sparsity by defining the ratio: $R = entropy/sparsity$. SD and DE represent the datasets of *Start time* vs. *Duration* and *Duration* vs. *Energy Consumption*, respectively. Dataset SD is used to predict *stay duration* while DE is utilized to predict *energy consumption*. Ratios of R_SD and R_DE are calculated using the training datasets' data entropy and sparsity. ρ is the correlation coefficient of SMAPE and R. P-value indicates the statistical significance of the trend (significant if P-value < 0.05).

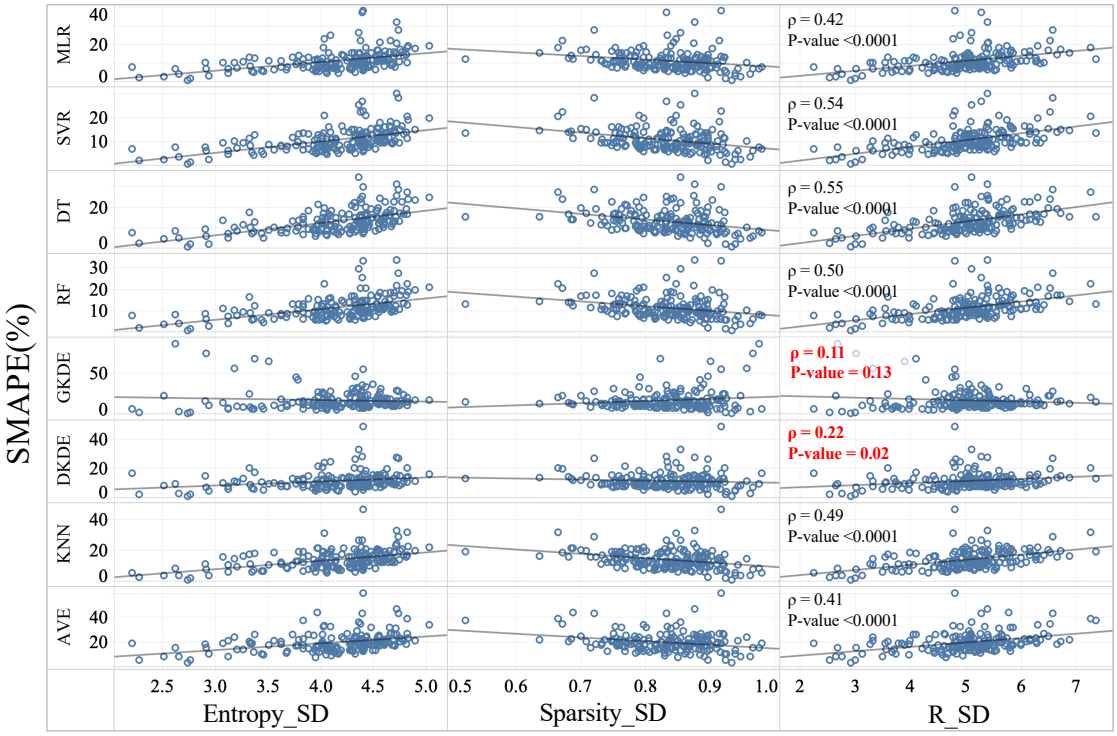

Figure 2.12: Comparisons of SMAPE(%) versus entropy, sparsity and R_SD (entropy/sparsity)

The SMAPEs of MLR, SVR, DT, RF, and KNN are compared with DKDE as shown in Fig. 2.14 and Fig. 2.15. Fig. 2.14 compares the SMAPEs of *duration*, and it shows that when R_SD is larger than 5.5, DKDE performs better. Likewise, Fig. 2.15 compares the SMAPEs of *energy consumption* predictions, and it shows that DKDE performs better when R_DE is larger than 4.

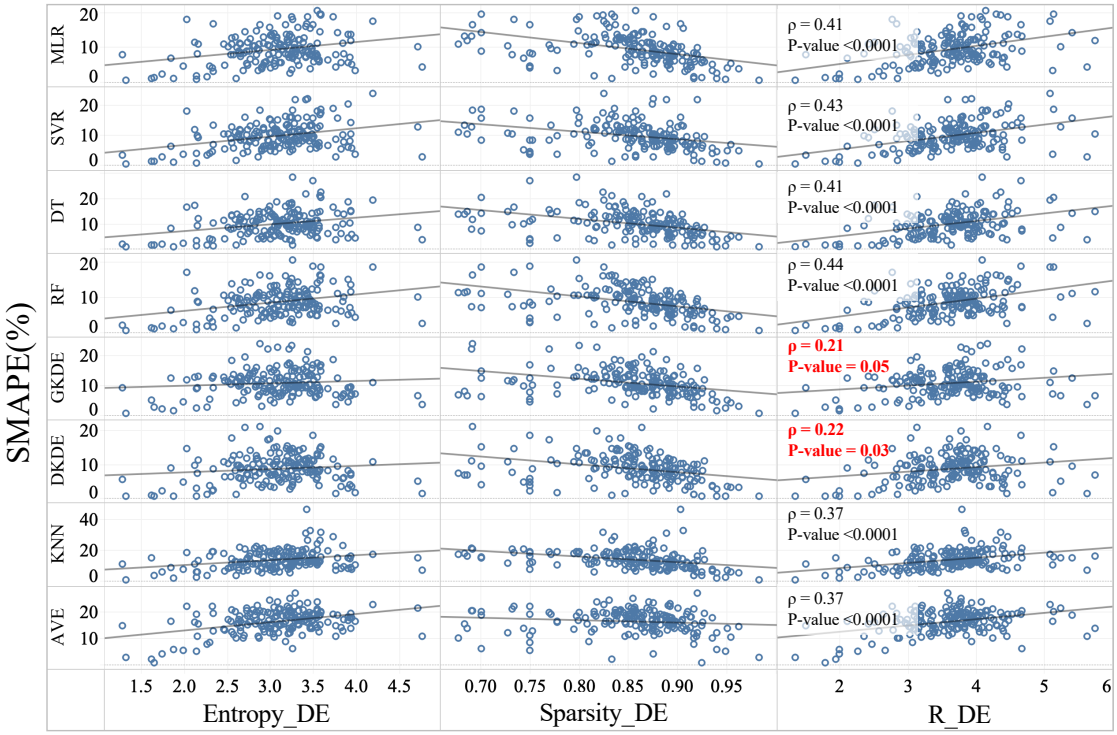

Figure 2.13: Comparisons of SMAPE(%) versus entropy, sparsity and R_DE (entropy/sparsity)

Table 2.2 shows the *duration* prediction results of the different algorithms. It indicates that SVR is most accurate overall, especially when R_SD $\leq$ 5.5. DKDE is the best when R_SD $>$ 5.5 and the SMAPE does not change significantly in different R_SD categories.

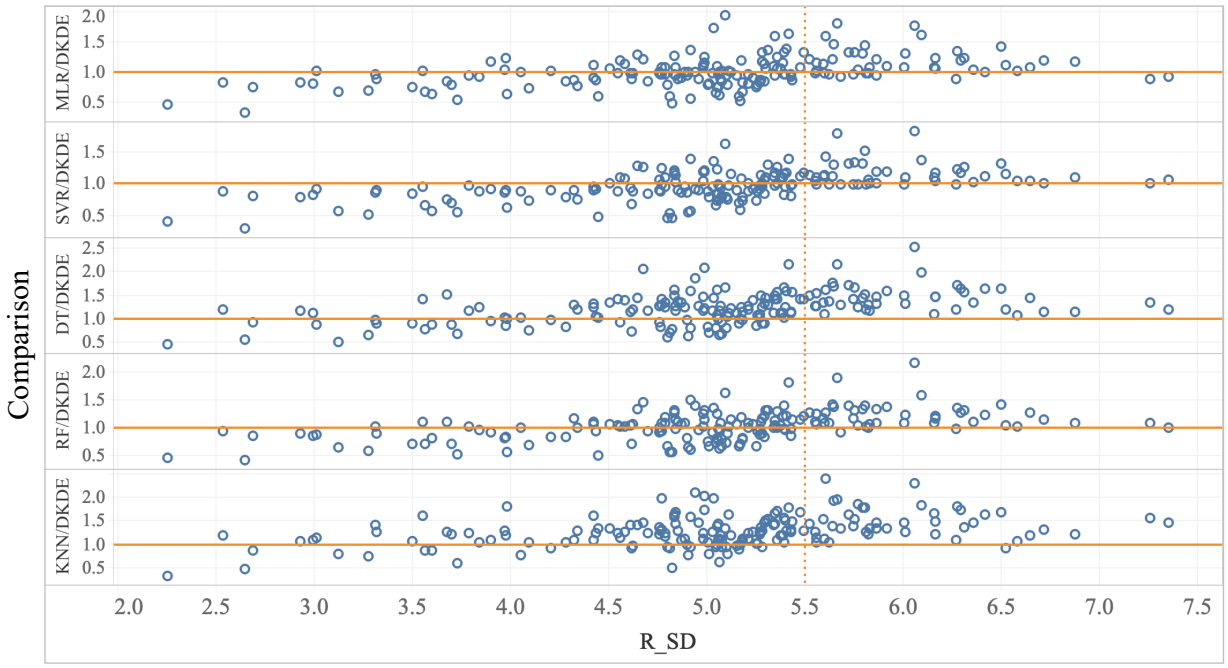

Figure 2.14: Comparisons of different algorithms with DKDE for duration prediction

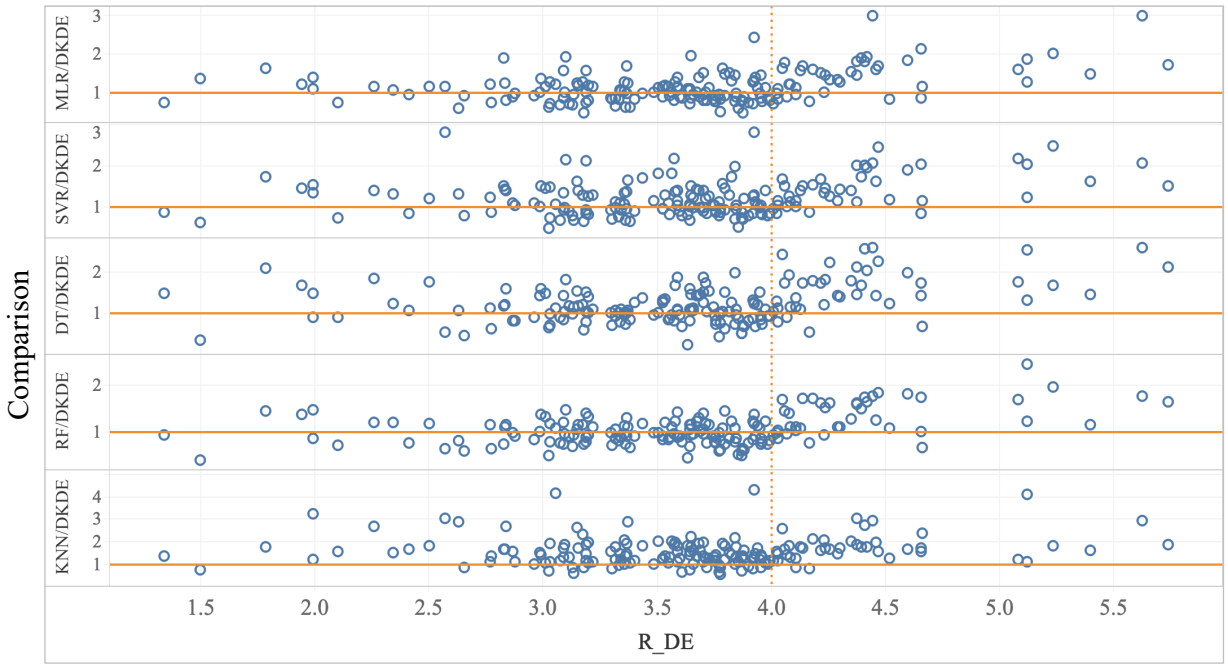

Figure 2.15: Comparisons of different algorithms with DKDE for energy consumption prediction

Table 2.2: Average and Standard deviation (in parentheses) for the SMAPE(%) of *duration* prediction

Ratio(R_SD)	SVR	MLR	DT	RF	DKDE	GKDE	KNN	AVE
R_SD≤ 5.5 (n=187)	9.54 (4.66)	10.19 (5.91)	11.84 (5.62)	10.21 (5.32)	10.96 (6.37)	17.39 (14.32)	12.65 (6.26)	18.43 (7.78)
R_SD>5.5 (n=65)	13.40 (4.40)	13.51 (4.20)	16.66 (4.78)	14.15 (4.33)	11.81 (4.15)	14.90 (5.84)	16.82(4.95)	23.47 (7.61)
Overall (n=252)	10.54 (4.89)	11.05 (5.69)	13.09 (5.80)	11.23 (5.36)	11.18 (5.88)	16.75 (12.72)	13.73(6.21)	19.73 (8.03)

Table 2.3: Average and Standard deviation (in parentheses) for the SMAPE(%) of *energy consumption* prediction

Ratio (R_DE)	SVR	MLR	DT	RF	DKDE	GKDE	KNN	AVE
R_DE ≤ 4 (n=204)	9.06 (4.25)	8.71 (4.14)	9.16 (4.55)	7.96 (3.68)	8.31 (4.20)	10.33 (4.23)	11.80 (5.38)	16.10 (4.64)
R_DE >4 (n=48)	12.91(4.11)	12.69 (4.32)	13.68 (5.52)	11.59 (4.04)	10.54 (3.76)	12.68 (4.80)	15.70 (5.54)	17.86 (3.21)
Overall (n=252)	9.79 (4.48)	9.46 (4.45)	10.01 (5.05)	8.65 (4.00)	8.73 (4.21)	10.78 (4.43)	12.54 (5.61)	16.43 (4.45)

Table 2.3 shows the *energy consumption* prediction results of the different algorithms. RF is shown to be the most accurate overall, especially when R_DE ≤ 4. Similarly, DKDE performs the best when R_DE > 4.

2.5.3.2 Proposed Algorithm

Based on the preliminary results, the combination of SVR, RF, and DKDE is proposed to form an ensemble algorithm, namely the EPA. The algorithm is depicted in Fig. 2.16 below.

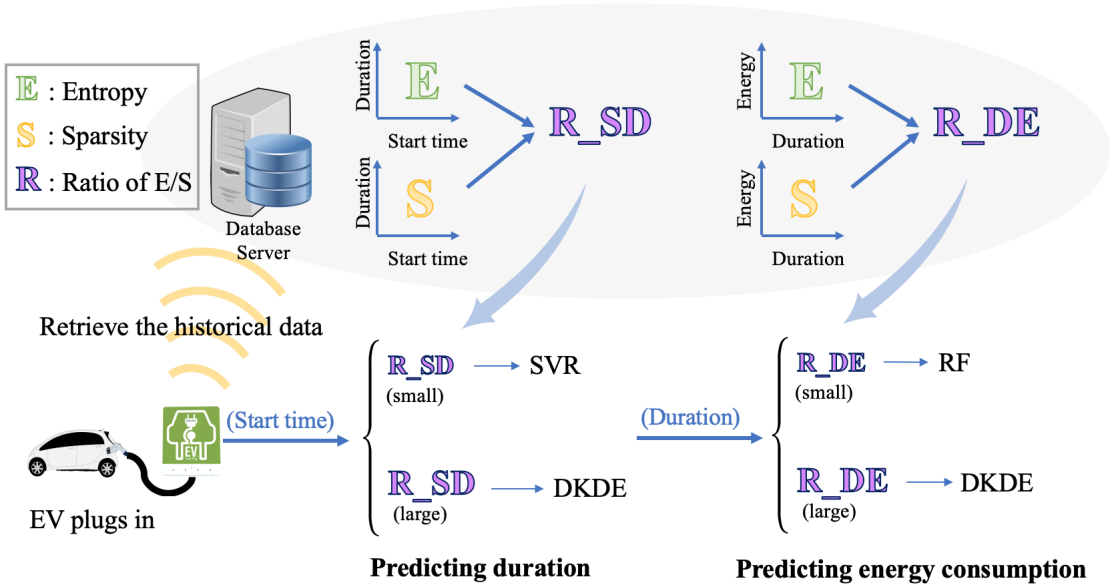

Figure 2.16: Flowchart of the ensemble predicting algorithm

Data entropy and sparsity are analyzed for all registered EV users in the system in order to calculate the R value. When an EV is plugged in, the user's R_SD is retrieved to determine either SVR or DKDE to be used for predicting *stay duration*. The predicted stay duration is then sent to the next step for *energy consumption* prediction. Similarly, RF or DKDE is applied depending on the value of R_DE. The threshold of R to switch the algorithms may need to update quarterly since user behaviors may change over time. The EPA is evaluated using a 10% test dataset. The prediction results along with the EV scheduling results are presented in the next section.

2.5.4 Results and Discussion

Fig. 2.17 and Fig. 2.18 show the SMAPE with regard to R for *duration* and *energy consumption* predictions, respectively. Fig. 2.17 illustrates the comparison between SVR and DKDE. As shown in the figure, the SMAPE of SVR is smaller when R_SD is smaller than 5.5, whereas the SMAPE of DKDE is smaller when R_SD is larger than 5.5. Fig. 2.18 demonstrates the comparison between RF and DKDE. As expected, RF is more accurate when

R_DE is smaller than 4, while DKDE performs better when R_DE is larger than 4.

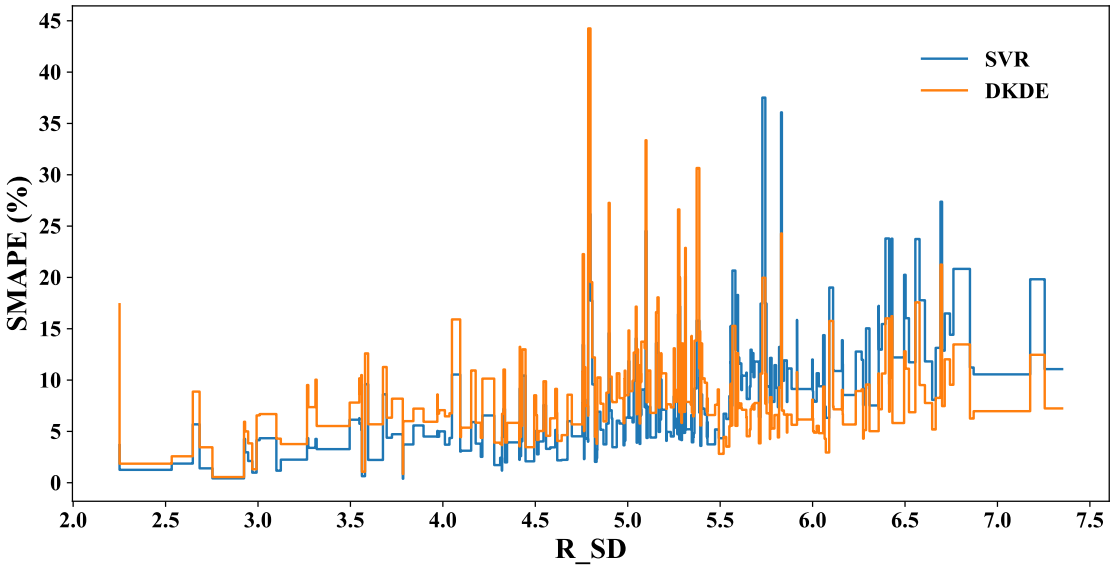

Figure 2.17: Average SMAPE vs. R_SD for SVR and DKDE

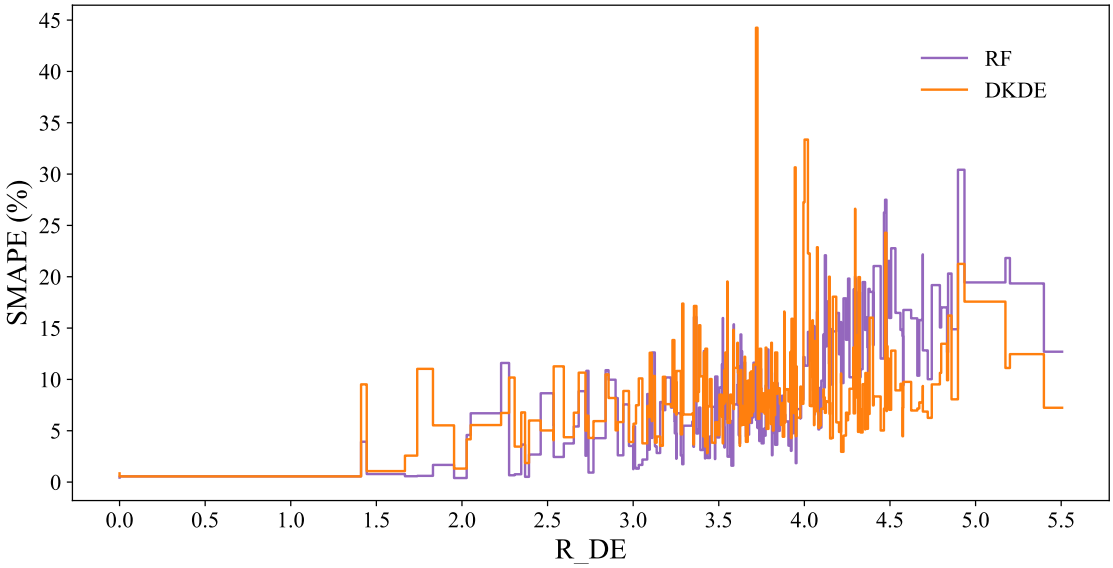

Figure 2.18: Average SMAPE vs. R_DE for RF and DKDE

Table 2.4 shows the average and standard deviation for the SMAPE of *duration* and *energy consumption* predictions. A pairwise T-test with the null hypothesis that the EPA has the same performance as the other algorithms is rejected by the small P-values ($p <$

0.05). The results show that EPA has decreased the errors significantly for *duration* and *energy consumption* predictions by around 11% and 22%, respectively.

Table 2.4: SMAPE(%), standard deviation (in parentheses), and the pairwise T-test result

	Duration Prediction			Energy Consumption Prediction		
Algorithm	SVR	DKDE	EPA	RF	DKDE	EPA
SMAPE (%)	11.53 (5.18)	11.67 (6.36)	10.40 (5.80)	9.69 (4.60)	9.56 (4.26)	7.54 (4.24)
P-value	0.00964409	0.00833339	-	0.00182736	2.00649×10^{-05}	-

Root mean squared error (RMSE) is also evaluated to show the effectiveness of EPA. Since each user has different number of charging records, here we calculate the mean of RMSE of all users, called mean estimation deviation (MED) as defined in Equation 2.24. Table 2.5 shows the comparison of MED among SVR, DKDE, RF, and EPA. For EPA, the prediction MED for *stay duration* is 1.16 hour and for *energy consumption* is 2.52 kWh.

Table 2.5: MED (Duration: hour; Energy: kWh), standard deviation (in parentheses), and the pairwise T-test result

	Duration Prediction			Energy Consumption Prediction		
Algorithm	SVR	DKDE	EPA	RF	DKDE	EPA
MED	1.36 (0.69)	1.38 (0.47)	1.16 (0.54)	2.94 (1.35)	2.65 (0.87)	2.52 (0.97)
P-value	0.00449328	1.3247×10^{-07}	-	1.49621×10^{-6}	0.0017243	-

2.6 Conclusion

Section 2.4 shows the comparison of GKDE and DKDE. While GKDE and DKDE have their strengths in predicting irregular and regular charging patterns, respectively, the novelty detection method exploits the synergy between them. Specifically, the novelty detection is utilized to determine charging pattern regularities, so that GKDE can be applied for irregular-pattern EV users while DKDE for regular patterns. Thus the proposed HKDE leads to more accurate predictions in comparison to using only either GKDE or DKDE. However, HKDE results in slightly higher prediction accuracy than DKDE. Therefore, Sec-

tion 2.5 examines more machine learning prediction approaches and investigates EV charging behaviors aiming to find the best method for its prediction. It is found that, in general, predicting SMAPEs positively correlate to data sparsity/entropy ratio (R) but this relationship for GKDE and DKDE is relatively weak. Therefore, the KDE method can be utilized to handle the high R data with lower prediction error. Based on this property and the analysis result, SVR, RF, and DKDE are selected to compose the EPA. The synergy of the three algorithms enhances the prediction performance where SVR is good at predicting EV stay duration, RF performs better on energy consumption estimation, and DKDE takes care of the prediction with the high R data. The estimations by HKDE and EPA are applied to the optimal EV charging scheduling algorithm for load variation and charging cost minimization in the next chapter.

CHAPTER 3

EV Charging Scheduling Model

3.1 Overview

The surge of EV has been observed in the past few years, and the global EV market continues to grow[MH14, SL19] because of the dwindling of the fossil fuel and the dedication to reducing carbon footprint emission worldwide. As in California, it is expected to have 1.5 million zero-emission EVs on the road by 2025 from the initiative from the government[Off12], and consequently, the increasing demand of the Electrical Vehicle Supply Equipment (EVSE) is foreseen. While the number of EV is increasing, without proper charging management, the uncoordinated power consumption on a local scale can stress the electrical grid and lead to grid problems that degrading power quality and reliability[CHD09, MWJ14, SIF15]. Therefore, more and more studies are focusing on EV charging coordination in order to accommodate the increasing number of EVs.

It is still a challenging task to manage a massive number of EV charging. First of all, there are several uncertainties on the demand side, such as start charging time, stay duration, and energy demand, as discussed in Section 2. Secondly, an uncoordinated EV charging may degrade the grid power quality or even damage the grid because it can produce a huge power demand that exceeds the grid capacity [LSA10]. Thirdly, the integration of renewable energy resources such as photovoltaic (PV) panel, requires proper control method for optimal energy utilization and PV intermittency alleviation [Cal16]. Lastly, the dynamic electricity price may significantly affect the EV charging cost. All of the above should be taken into consideration for an effective, real-time EV charging scheduling system.

3.2 Literature Review

A considerable number of studies have been made on EV charging management not only in the aspect of an economical implementation of EVSE but also in the reliability of a distribution grid, which is to alleviate the deteriorating impact of uncoordinated EV charging. [DTB12] and [LWL12] discuss the optimal sizing and location of EVSE while [LGZ18] further demonstrates the need of multi-types of charging facilities for optimal EVSE deployment. [TMN17] presents a charging scheduling algorithm to accommodate a high penetration of EVs and DERs. Also, [HC16] discusses EV charging scheduling with the consideration of vehicle-to-grid (V2G) capability. Studies of EV load scheduling fall into two approaches: centralized and distributed methods. Centralized means a central entity (CE) directly controls the EV; namely, a CE solves the optimization problem and broadcasts the results to the EVs [CHD09, BNE16]. The objectives of the optimization includes minimizing power loss [CHD09, SHM10], regulating load factor [SHM10] or maximizing supportable EV penetration [LSA10]. The centralized infrastructure requires to collect the information form all EVs and centrally optimize their charging schedules. Therefore, EV owners' privacy becomes an issue. Also, when EV penetration increases, the data is more difficult to manage, and the curse of dimensionality becomes a problem. On the other hand, a distributed approach is more suitable for managing a large scale of EV charging. In this method, CE coordinates the EV load demand through communication with the EV chargers [LBM16]. That is, instead of solving the scheduling problem, including many variables centrally, it is solved in a distributed and iterative manner between CE and EV charging agents. Over the past few years, a large number of articles have been devoted to the study of distributed EV charging scheduling approach. What seems to be lacking, however, is to consider user behavior stochasticity. Load flattening is achieved in [CF10], but the EV stay duration, energy demand are assumed to be known. [GTL13, MCH11] have optimally scheduled EV charging to achieve load valley filling, but user behavior uncertainties are not considered. [MCH11] assumed every EV has the same charging window and the same charging demand, which is not realistic.

In the following section, an EV charging scheduling algorithm incorporated with the EV user behavior prediction method introduced in chapter2 is presented, with the objectives of minimizing load variance and reducing charging cost.

3.3 Model Description

We consider an EVCI that is controlled and managed by a control entity (CE). The purpose of CE is twofold: minimizing the peak load, which is equivalent to load variance minimization [SHM10], as well as reducing total charging cost. We assume that EVCI (Fig. 3.1) has a 45 kWh BES and 35 kW PV panels, and it is supplied by an electrical feeder shared with an office building which its average net load demand is 250 kW.

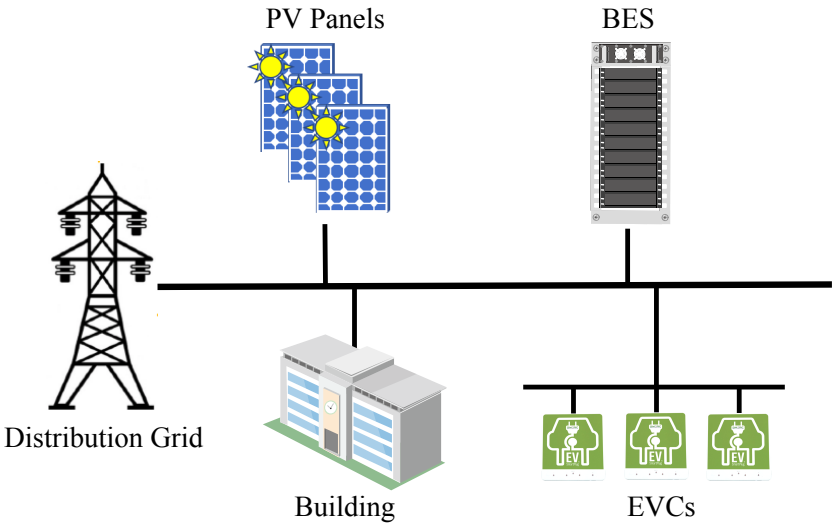

Figure 3.1: EVCI configuration.

The total number of EVs and EV chargers (EVCs) are denoted by $\mathcal{N}$ and $\mathcal{M}$, respectively, where $\mathcal{N}$, $\mathcal{M} \in \mathbb{N}$. In this configuration, each EVC has four charging outlets and can charge four EVs at the same time, so we use $\mathcal{N}_l$ to show the set of EVs supplied by EVC_l. The

model of EVCI and the building can be written as:

$$z_j(t+1) = z_j(t) + T_h u_j(t),\ j = 1, ..., \mathcal{N}, BES \tag{3.1a}$$

$$e_{CIB}(t) = d_B(t) - p_{PV}(t) + \sum_{j=1}^{\mathcal{N},BES} u_j(t) \tag{3.1b}$$

where $z_j(t) \in \mathbb{X}_j \in \mathbb{R}$, $u_j(t)$, $e_{CIB}(t)$, $d_B(t), p_{PV}(t) \in \mathbb{R}$ and $t \in \mathbb{N}$. $z_j(t)$ in [kWh] is the energy stored in EV_j or BES, $e_{CIB}(t)$ in [kW] is the total net load demand of the EVCI and the building, and $p_{PV}(t)$ in [kW] is the power generated by PV. It is assumed that EVs and EVCs have Vehicle-to-Grid (V2G) capability, so they can supply power to the grid. $d_B(t)$ in [kW] is the load demand of the building. In (3.1a), T_h (in hours [h]) is the discretization in time, e.g. $T_h = 0.5$ corresponds to 30 min. $u_j(t)$, which is the EV/BES charging (discharging) power, is introduced as the optimization variable. $u_j(t)$ is positive in the charging mode, and it is negative in discharging mode.

The constraints on the energy capacity of EVs and BES are:

$$\underline{C}_j(t) \leqslant z_j(t) \leqslant \overline{C}_j(t),\ j = 1, ..., \mathcal{N}, BES \tag{3.2}$$

where $\underline{C}_j(t)$ and $\overline{C}_j(t)$ are the bounds on the energy stored in EVs and BES. The constraints on charging/discharging power of EVs and their corresponding EVCs as well as BES are:

$$\underline{u}_j \leqslant u_j(t) \leqslant \overline{u}_j,\ j = 1, ..., \mathcal{N}, BES \tag{3.3a}$$

$$\underline{g}_l(t) \leqslant \sum_{i=1}^{\mathcal{N}_l} u_{li}(t) \leqslant \overline{g}_l(t),\ l = 1, ..., \mathcal{M} \tag{3.3b}$$

where $\underline{u}_j$ and $\overline{u}_j$ are the minimum and maximum power ratings of the EV and BES chargers, and $\underline{g}_l$ and $\overline{g}_l$ are the minimum and maximum power ratings of the EVCs.

The bounds in (3.2) for EVs are time-varying and defined as follows; if $EV_j, j = 1, ..., \mathcal{N}$ is:

- not plugged in EVC, $\underline{C}_j(t) = \overline{C}_j(t) = 0$
- plugged in EVC, but it is in idle mode, $\underline{C}_j(t) = 0$ & $\overline{C}_j(t) = C_j$
- plugged in EVC, and it is needed by time t, $\underline{C}_j(t) = \overline{C}_j(t) = C_j$

where C_j is the maximum capacity of EV_j's battery.

3.4 Problem Formulation

For the given time index t and prediction horizon $N \in \mathbb{N}$, let's denote the vector notation $\mathbf{u}_j = \big(u_j(t), u_j(t+1), \ldots, u_j(t+N-1)\big)^T, \mathbf{u}_j(t) \in \mathbb{R}^N$, which is used for all other variables as well. To formulate the objective function, we define total net load demand at time t by (3.4):

$$\mathbf{\Omega} := \mathbf{d}_B - \mathbf{p}_{PV} \tag{3.4}$$

Also, the average net load demand at time t over time horizon $N \in \mathbb{N}$ is:

$$\overline{\mathbf{\Omega}} := \frac{1}{N} \sum_{t=k}^{k+N-1} \mathbf{\Omega} \tag{3.5}$$

Accordingly, the twofold objective function of EV charging coordination (CC) is written as:

$$\begin{aligned} V := \min_{\mathbf{u}} \{ & \sum_{j=1}^{\mathcal{N},BES} \alpha\Big(\mathbf{\Pi}^T \mathbf{u}_j\Big) + \sum_{t=k}^{k+N-1} \Big(\overline{\mathbf{\Omega}} - (\mathbf{\Omega} + \sum_{j=1}^{\mathcal{N},BES} \mathbf{u}_j)\Big)^2 \} \\ & \text{s.t.} \quad (3.1) - (3.3) \end{aligned} \tag{3.6}$$

where α is a weighting factor, and $\mathbf{\Pi} \in \mathbb{R}^N$ is time of use (TOU) price vector [Cal]. The first part in (3.6) reduces charging cost, while the second part minimizes the total load variance.

3.5 EV Scheduling Results and Discussion

3.5.1 EV Scheduling using HKDE prediction

Using HKDE prediction results, we run CC for the EVCI including 107 EVs and 27 EVCs. Dynamic electricity price used in our numerical simulation is shown in Fig. 3.2. Total charging profiles of the EVCI for uncoordinated CC (uCC) and CC using real and HKDE data are shown in Fig. 3.3. As it is clear, CC flattens the total load profile which results in peak load shaving and valley filling. Also, the difference between load profile using HKDE and real data is negligible during most of the time intervals. Comparing BES profile (Fig. 3.4)

and aggregated EV load profile (Fig. 3.5) with dynamic price (Fig. 3.2), when energy price is high (7:00 and 20:00), BES is discharged in order to supply EV load, and it is charged while the energy price is low.

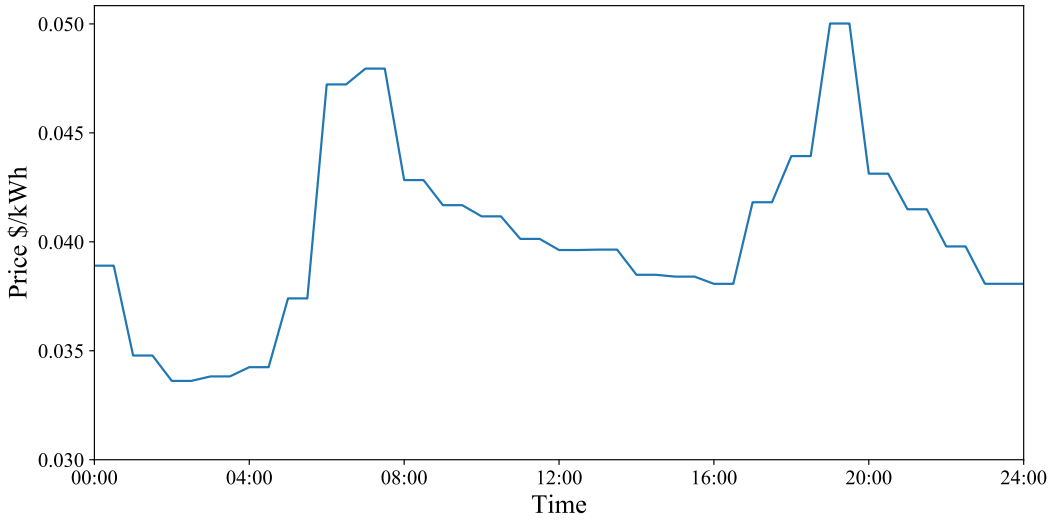

Figure 3.2: Dynamic price for the EV charging scheduling simulation [Cal].

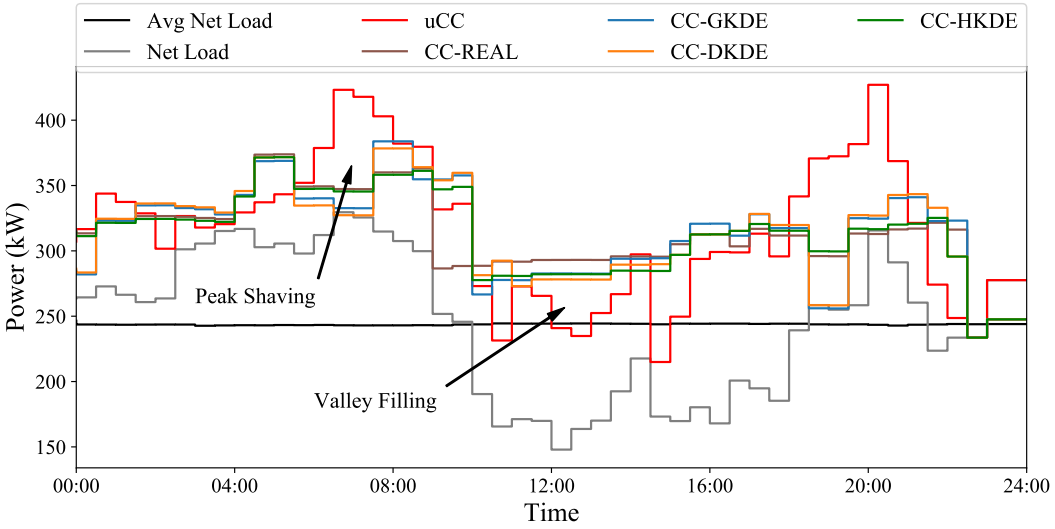

Figure 3.3: Load profile using uCC and CC based on real data and GKDE, DKDE and HKDE estimations.

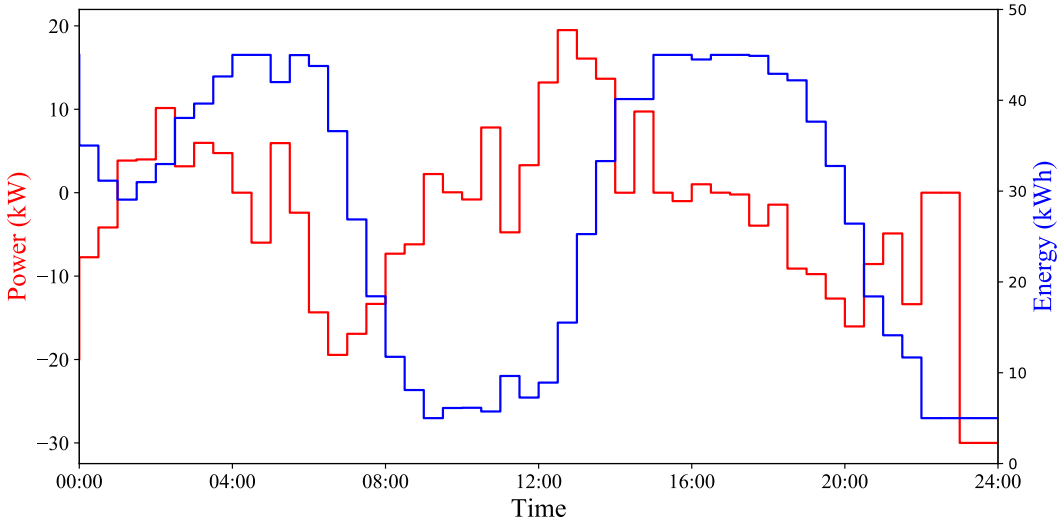

Figure 3.4: BES power and energy. Positive power: BES charging; Negative power: BES discharging.

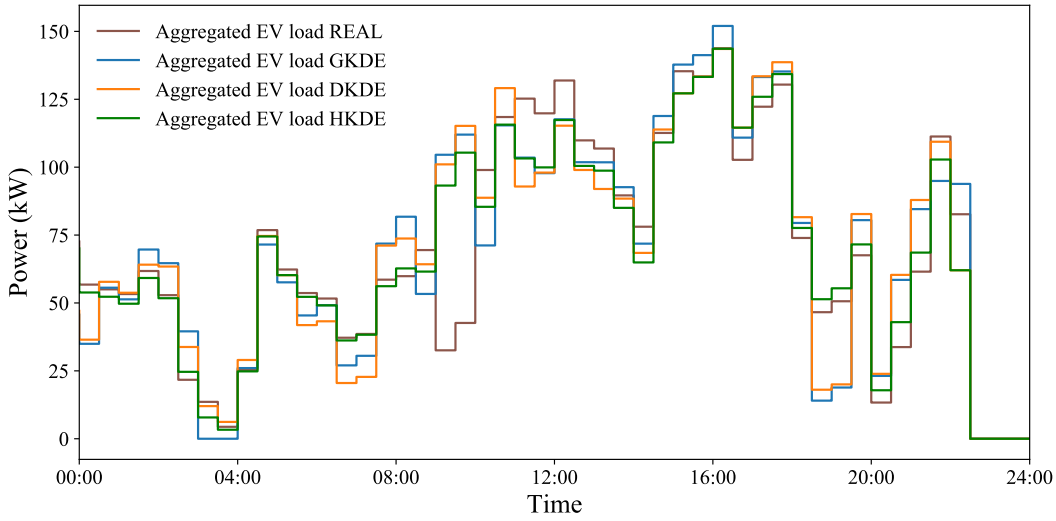

Figure 3.5: The comparison of aggregated EV load using real data and GKDE, DKDE and HKDE estimations.

Table 3.1: Comparison of RMS errors of aggregated EV loads.

	RMS Error (%)
GKDE	20.24
DKDE	20.15
HKDE	17.72

The comparison of RMS errors among the aggregated loads of GKDE, DKDE and HKDE

is shown in Table 3.1. Together with the aggregated EV load profiles illustrated in Fig. 3.5, HKDE demonstrates a better prediction accuracy over GKDE and DKDE.

Depending on the α values in (3.6), total charging cost, peak-to-peak (PTP) and root-mean-square (RMS) of the total load profile vary. By increasing α, PTP and RMS increase and total charging cost decrease and vice versa. For comparison, numerical simulation results for different α values are shown in Table 3.2. As it is seen, by increasing α from 10 to 100 using real data (CC-Real), the total charging cost decreases from \$111.70 to \$107.68 at the expense of PTP and RMS degradation. As it is shown in the table, the PTP with $\alpha = 100$ is even worse than that with uCC. Therefore it is important to consider PTP and RMS while minimizing the charging cost.

Table 3.2: Comparison between uCC with CC using real data and HKDE

Charging Type	α Value	PTP (kW)	RMS (kW)	\$Charging Cost
uCC	–	212.11	91.07	120.66
CC-Real	10	140.26	76.74	111.70
	50	166.44	80.61	109.73
	100	255.46	89.68	107.68
CC-HKDE	10	138.08	77.36	111.93
	50	156.16	80.49	110.32
	100	251.54	89.13	108.43

3.5.2 EV scheduling using EPA

Using EPA prediction results, we run CC for the EVCI including 252 EVs and 63 EVCs. Dynamic electricity price used in our numerical simulation is shown in Fig. 3.6. Total charging profiles of the EVCI for uncoordinated charging (uCC) and CC using real and prediction data are shown in Fig. 3.7. As it is clear, CC flattens the total load profile which results in peak load shaving when dynamic price is high and valley filling when the price is low. Also, the difference between load profile using EPA and real data is negligible during most of the time intervals. However, there appears a valley during 23:00 and 4:00 in Fig. 3.7

and is not filled. This is because of less availability of EVs according to our dataset, in which the "end time" refers to the "end charging time" instead of "un-plugging time," and therefore further restrains the EVs' time flexibility for charging.

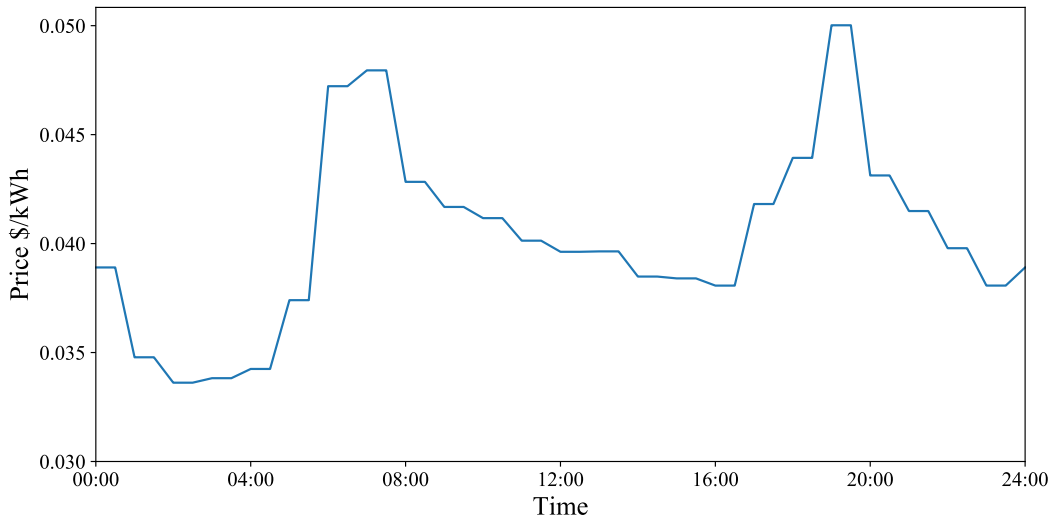

Figure 3.6: Dynamic price for the EV charging scheduling simulation [Cal]

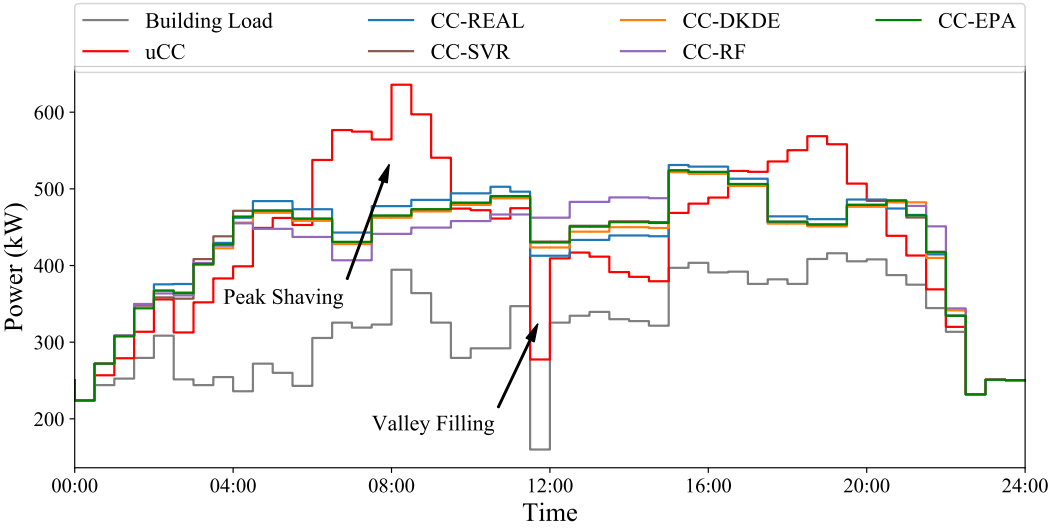

Figure 3.7: Load profile using uCC and CC algorithms based on real data

The coordinated and uncoordinated EV charging scheduling results with respect to peak-to-peak (PTP) and root-mean-square (RMS) of the load profile, and charging cost are shown in Table 3.3. The result of CC_EPA aligns well with CC_Real, and reduces 27% peak load, 10% load variation, and 4% charging cost from that of uCC's.

Table 3.3: Comparison between uCC with CC using real data and EPA

Algorithm	PTP (kW)	RMS (kW)	Total Cost ($)
CC_Real	307.13	139.59	219.94
CC_EPA	300.50	139.24	219.25
uCC	411.83	156.44	229.13

Although the result shows only $10 can be saved per day by scheduling 252 EVs' charging comparing to uncoordinated charging, with a large number of EVs, the saving can be significant. Furthermore, according to [Key16], the energy unit cost (EUC) negatively related to a load factor (LF) along with a hyperbolic function ($EUC \propto 1/LF$), where LF is defined as follows:

$$LF = \frac{AveragePower}{PeakPower} * 100\%. \tag{3.7}$$

The price will approach the minimum when LF close to 1. As shown in Fig. 2.8, people tend to plug in EVs in the morning when they arrive at work, and in the evening when they get home. If the energy peak produced by EV charging that drastically lower the LF, the energy price will increase sharply. Therefore, EV charging control is necessary to accommodate such larger amount of EVs within the electrical grid.

The results show that the EPA model fits the true densities, including start time, stay duration, and energy consumption, better than the other algorithms. Therefore, the control entity (CE) can schedule EV charging optimally in terms of minimizing peak loads and reducing charging cost. For scale-up, a considerable amount of EVs can be utilized to mitigate the renewable energy intermittency issue such as solar duck-curve problem. Since the EPA algorithm can predict the EVs' availability very well, in combination with vehicle-to-grid (V2G) technology, the charging CE can manage to charge EVs during the midday when solar power is ample and discharge during the evening to reduce the peak load.

3.6 Conclusion

As can be seen in the prediction result in Section 2.4.4, there is a slight improvement of HKDE over DKDE. This is because of DKDE's over-fitting issue for the irregular charging patterns, instead of an accurate prediction generated by GKDE. In other words, the prediction error for GKDE is still significant, but it is even larger for DKDE in the case of over-fitting. This can be due to a short charging history of a user or the user's irregular charging behavior in nature. However, this case is rare for most EV users. DKDE can provide more accurate predictions in general, thanks to its capability to model the charging record distribution accurately. The comparison between GKDE and DKDE is made in Fig. 3.8, and it shows that DKDE aligns with the empirical data better than GKDE. Therefore, DKDE is used in the EPA.

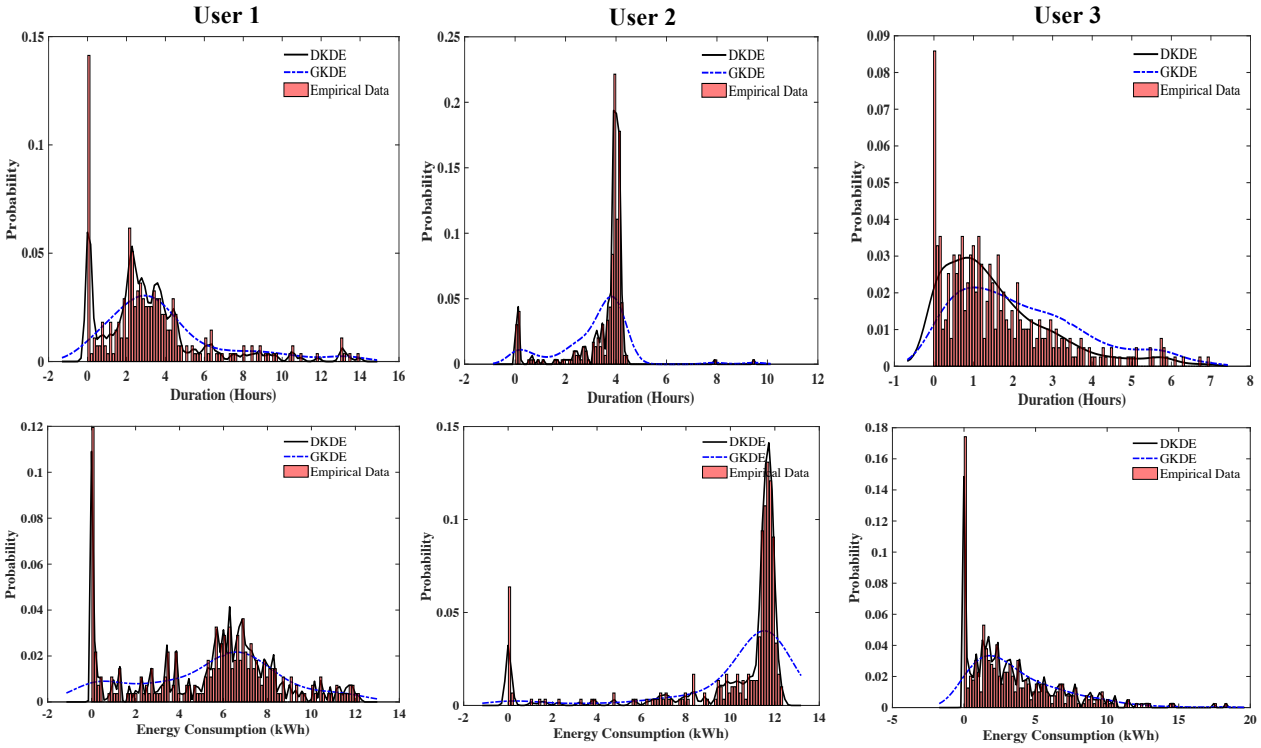

Figure 3.8: The comparison of modeling EV users' charging behavior between GKDE and DKDE. (textbfTop: stay duration; textbfBottom: energy consumption)

It is founded that, in general, predicting SMAPEs positively correlate to data sparsity/entropy ratio (R), but this relationship for GKDE and DKDE is relatively weak. There-

fore, the KDE method can be utilized to handle the high R data with lower prediction error. Based on this property and the analysis result, SVR, RF, and DKDE are selected to compose the EPA. The synergy of the three algorithms enhances the prediction performance where SVR is good at predicting EV stay duration, RF performs better on energy consumption estimation, and DKDE takes care of the prediction with the high R data. The estimations by EPA are then applied to the optimal EV charging scheduling algorithm for load variation and charging cost minimization. Owing to the increased accuracy of the prediction, the scheduling algorithm can provide better EV charging load management in terms of reducing load variation and charging cost. Real data is employed for a numerical simulation to demonstrate the improved prediction accuracy of EPA and validate the effectiveness of the EV charging scheduling algorithm.

The proposed EPA algorithm can be applied to any scale of charging station, with an assumption that EVs' charging records are known. However, to reach optimal scheduling within a distribution grid, the connection between each charging station is required.

CHAPTER 4

Vulnerability and Risk for EV Charging System

4.1 Overview

The electricity distribution system has become more complex and dynamic because of the increasing deployment of renewable energy resources such as wind and solar energy, and the surge of EVs within electrical grids. These changes have brought about significant challenges for the distribution system operator (DSO) who manages the grid. Therefore, new technologies such as smart controllers, smart meters, or demand response incorporating the Internet of Things (IoT) technology have been developed to cope with the challenges of managing this complex grid, rendering the digitalization of the power grid, namely, the smart grid. Improved sensing, communication, and control capability enhance the performance of smart grid operation, but at the expense of increased vulnerabilities to deliberate attacks and accidental failures, threatening the grids functionality and reliability.

Although an increasing number of studies have been made on cybersecurity for the power system, there is a lack of a consistent cyberattack assessment in EV networks. Each of the new layers of data integration and control that are added to electric distribution systems can create new attack surfaces and potential privacy breaches. The rapid pace of electrification in the transportation sector, realized through plug-in EV (PEV) integration, necessitates the smart charging infrastructures. These systems built on real-time data collection and decision making coordinate the charging demand to facilitate high penetration of PEVs in the power grids. Accordingly, the inherent cyber-physical characteristic of smart charging networks makes them susceptible to cyber-physical threats.

This chapter presents the vulnerability analysis and risk assessment for the smart charging

infrastructures. To this end, several potential failure scenarios for the WinSmartEV^{TM} charging system on the UCLA campus were defined, and the impacts of potential cyber-physical attacks have been studied. Moreover, a codified methodology and taxonomy were provided for assessing vulnerability and risk of cyber-physical attacks on the EV charging networks in order to create a generalizable and comprehensive solution. The outcome is a framework to prioritize the degree of the vulnerabilities and risks in the EV networks and to develop effective countermeasures.

4.2 Literature Review

Automotive manufacturers are expanding their electric vehicle (EV) offerings, and the charging infrastructure is rapidly following. As of November of 2017, there were about fifty thousand level 2 and DC fast-charging stations throughout the United states[WRM17], and as of May 2019, the number rose to more than sixty-eight thousand [Gre19]. As EV charging becomes a significant power consumer within a distribution grid, smart charging control is essential for charging regulation. However, while charging stations become smart, they become more susceptible to cyber-physical attacks. Nonetheless, there is a lack of studies in potential cybersecurity problems for EV charging systems. Only a few attempts have so far been made at different aspects of cybersecurity issues for EV.[CB12] discussed the potential security vulnerability for EV infrastructure and the security issue for the communication between EV and EVSE. [AAJ15] examined the cybersecurity issues for EV and smart grid integration, and reviewed the state-of-the-art methods for cyber-attack detection. [FAT18] justified the research need for Internet of EV (IoEV) security as it is a complex system that involves vehicles, humans, sensors, road infrastructure and charging stations that are vulnerable to cyber-physical security threats. What seems to be lacking, however, is a comprehensive vulnerability analysis and risk assessment for an EV charging network.

Currently, a method for assessing the risk and impact of successful attacks against plug-in EV (PEV) charging networks does not exist. As a result, this chapter aims to present a method by working through several case studies regarding potential mock cyber-physical

attacks against the UCLA WinSmartEV^{TM} charging platform. A survey has been conducted to understand the attack feasibility and investigate the potential impact and risk of an individual carrying out such an attack. Also, it helps to create the method outline for assessing the risk and impact. In the next sections, we outline the UCLA charging network structure, conduct vulnerability analysis of current systems, and provide a series of case study topics of cyber-physical attacks and the several corresponding electric transportation (ET) impact failure scenarios.

4.3 Vulnerability Analysis and Risk Assessment

4.3.1 EV Charging Network

There are two types of charging networks with varying levels of complexity: conventional and smart grid. Conventional charging networks are connected to the primary power grid infrastructure. However, the existing power grid suffers from a tight coupling that exposes it to single points of failure for power distribution. Smart grids, localized power grids consisting of smart devices that can connect to a larger power grid or generate, store, and distribute electricity from within, alleviate some of the single points of failures in the network as shown in Fig.4.1, UCLA WinSmartEV^{TM}.

The UCLA WinSmartEV^{TM} Network is a smart grid EV charging network. It generates power from rooftop photovoltaic solar panels, stores the energy locally, and can send energy to EV charging terminals for users to charge their EVs. The smart charging network monitors the charging, schedules optimized aggregated charging sequences, and executes the schedule via the control network. Critical data, including energy consumption and various power-quality related variables, are recorded and uploaded to a centralized database managed by the control center. The EV users can access the charging status via the UCLA WinSmartEV^{TM} mobile application.

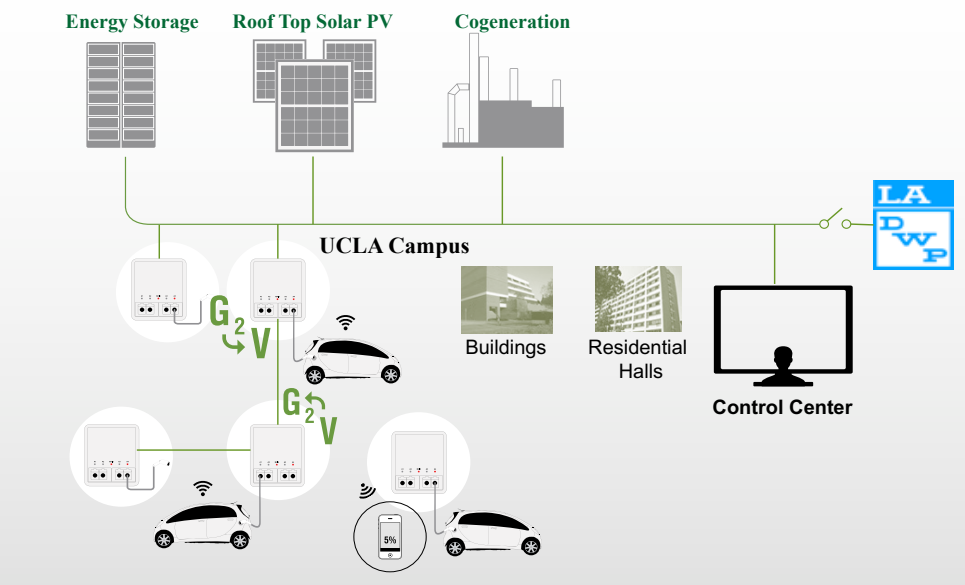

Figure 4.1: UCLA EV WinSmart™

4.3.2 Potential Attacks and System Vulnerability

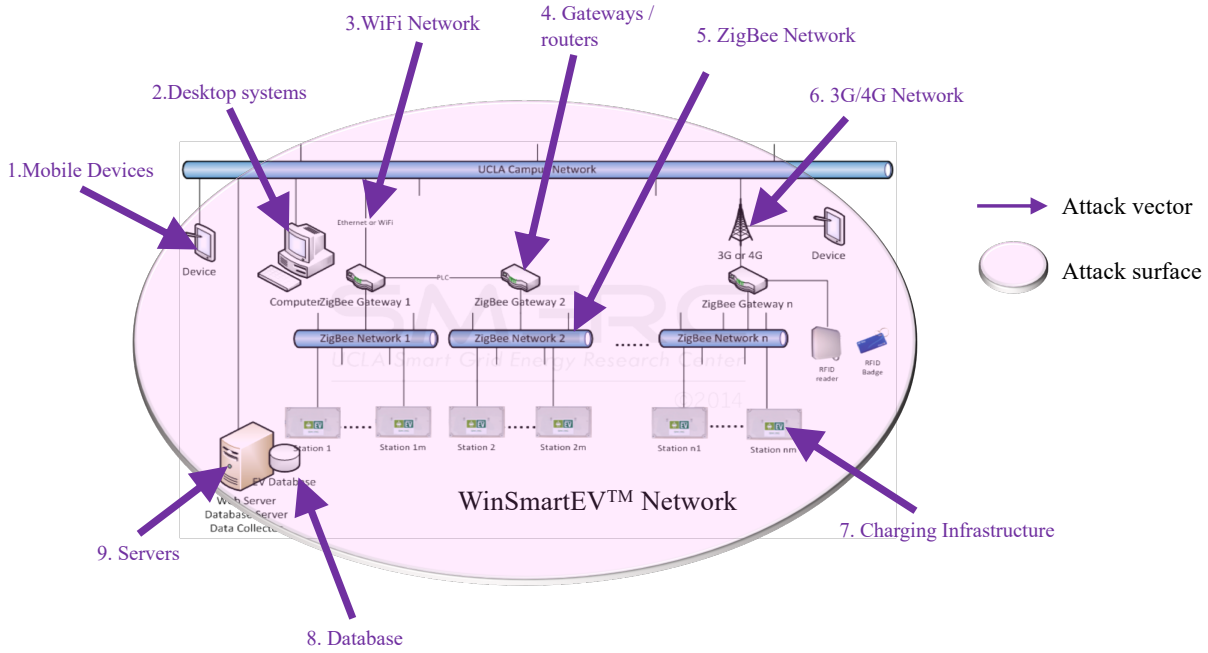

Figure 4.2: Attack vectors and the attack surface of UCLA EV WinSmartEV™ network

We define vulnerability as a weakness in an information system, system security procedures, internal controls, or implementation that could be exploited or triggered by a threat source [Nat10]. Since 2013, an estimated 14 billion data record records have been lost or stolen [Bre]. In the PEV network, cyber and physical vulnerabilities lie in the components shown in Fig.4.2. Hackers have the advantage of choosing the time of the attack and the vulnerability to exploit. In terms of cyber-physical compromise, both attacks and the impacts can be cyber or physical domain. Table 4.1 shows the causality of some common cyber-physical attacks [MKB12].

Table 4.1: The impact of cyber-physical device compromise[MKB12]

Attack \Impact	Cyber	Physical
Cyber	OpenSSL heartbleed bug - Eavesdropping of private information	Stuxnet, WannaCry virus
Physical	Meter bypassing	Instability due to physical destruction

As shown in the Table.4.1, the impacts of the cyber-physical attacks can be categorized into four classes: Cyber-attack-Cyber-impact (CC); Cyber-attack-Physical-impact (CP); Physical-attack-Cyber-impact (PC); and Physical-attack-Physical-impact (PP). Understand the nature of the attacks would be helpful to come up with the solution for remedy and the corresponding protecting action.

This section will explore several typical attack vectors on the UCLA WinSmartEV^{TM} network components. It should be noted that while we cover five typical types of attacks, we considered a multitude of others, and not all will be listed, nor could all be sufficiently covered.

Following are the potential attacks:

4.3.2.1 Man-in-the-Middle Attack

The major risk of an attack comes from the router. The man-in-the-middle (MITM) attack refers to the attacker secretly replaying and possibly altering the communication between two parties by placing himself in the middle of communication [Rah17, CCO18]. As in Fig.4.3, an attacker can intercept communication between the EV charging control center and drop, modify, or add data transmissions. This can lead to simultaneous fast charges that can cause a transformer overload.

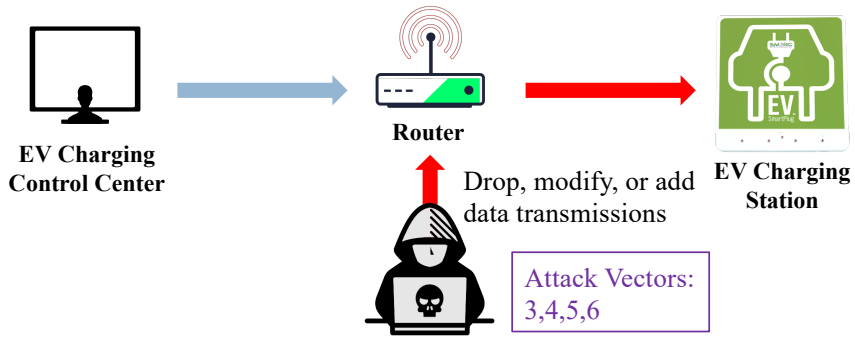

Figure 4.3: Man-in-the-Middle Attack

4.3.2.2 Denial-of-Service Attack

A denial-of-service (DoS) attack occurs when an attacker takes action intending to overload and flood the network, so that a network service is unavailable to its intended users [CCO18, QLS18]. In this scenario, the hacker can attack via the server and block an EV user from the charging station, as shown in Fig.4.4. For example, unavailable communication blocks the customer's use of EV preferential rates. This can lead to a delay for high priority vehicles such as ambulance and firetruck. There is an advanced DoS called distributed DOS(DDoS), which can lead to a more severe outcome. While The DoS attack typically uses one computer and one Internet connection to flood a targeted system or resource. The DDoS attack uses multiple computers and Internet connections to flood the targeted resource. DDoS attacks are often global attacks, distributed via botnets.

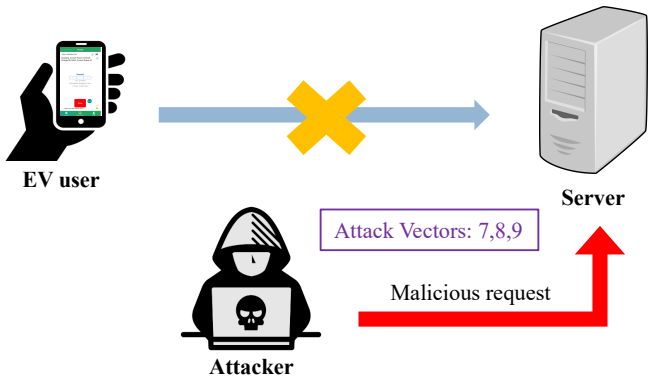

Figure 4.4: Denial-of-Service Attack

4.3.2.3 Packet Replay Attack and Eavesdropping

As illustrated in Fig. 4.5, packet replay attack or eavesdropping occurs when an attacker intercepts a request from an EV user. The action captures and repeats or delays valid data transmissions, resulting in modified messages or spoofed on demand response automation system(DARS) communication channels, or collect private EV user information [CCO18, CMA17, BGS17]. This can lead to EV registration ID theft to falsifying credentials to access the preferential rate of high priority EV users.

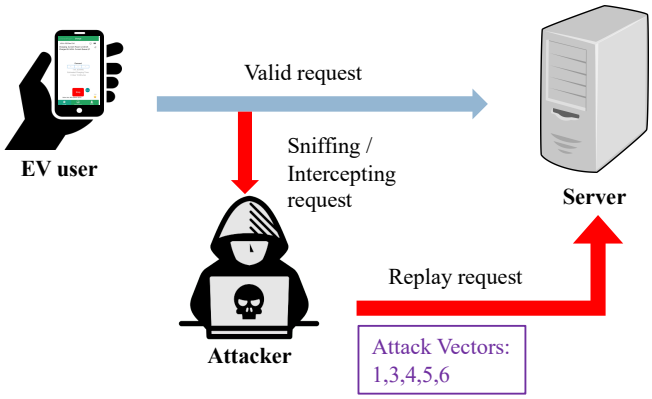

Figure 4.5: Packet Replay Attack and Eavesdropping

4.3.2.4 Address Resolution Protocol Spoofing

Address Resolution Protocol (ARP) spoofing attack occurs when an attacker sends a falsified ARP message over a local area network, resulting in the linking of an attackers MAC address with the IP address of a legitimate computer or server on the network. Therefore, the attacker will be able to receive any data that is intended for that IP address[BGS17]. The ARP spoofing is illustrated in Fig.4.6.

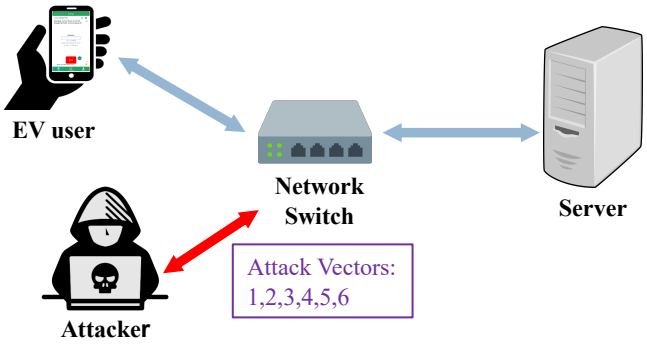

Figure 4.6: ARP Spoofing

4.3.2.5 Insider Attack

While an attacker tries to break into a network, an insider is much dangerous and unpreventable. Even the most secure firewall does not stop an insider, as shown in Fig.4.7. An insider can be employees, contractors, or an insider from outside [CMA17].

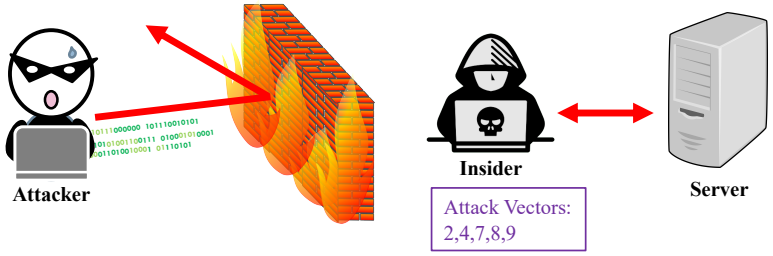

Figure 4.7: Insider Attack

Table 4.2: Electric Transportation (ET) Failure Scenarios (**I: Impact C: Cost R: Ratio**)

Scenarios	Description	I	C	R	Ranking	Class
ET1	Custom Malware causes EV Overcharge and Explosion	3	3	1	Low	CP
ET2	Simultaneous Fast Charges cause Transformer Overload	9	9	1	Low	CP
ET3	Virus Propagated between EVs and EV Service Equipment (EVSE)	9	3	3	Medium	CP
ET4	EV Charging Locations Disclosed via Utility Database	1	1	1	Low	CC
ET5	Compromised Protocol Translation Module Enables Control of EVs	3	3	1	Low	CP
ET6	EVSE Connects Wirelessly to Wrong Meter and Compromises Billing	3	3	1	Low	CC
ET7	Private Information Disclosed in Transit between EV and EVSE	3	3	1	Low	CC
ET8	Customer Misuses their EV Registration ID to Obtain Preferential Rate	0	0.1	0	Negligible	CC
ET9	EV Registration ID Stolen to Obtain Preferential Rate	0	0.1	0	Negligible	CC
ET10	High Priority EV Registration Identity Misused to Obtain Faster Charging	0	1	0	Negligible	CC
ET11	All EV Registration IDs Stolen from Utility	3	1	3	Medium	CC
ET12	Unavailable Communication Blocks Customer Use of EV Preferential Rate	1	3	0.33	Low	CC
ET13	Invalidated EV Registration ID Blocks Customer use of Preferential Rate	1	3	0.33	Low	CC
ET14	EV Charging Process Slowed by Validation Delay of EV Registration ID	1	3	0.33	Low	CC
ET15	Malware Causes Discharge of EV to the Grid	3	0.1	30	High	CP
ET16	An EV is Exploited to Threaten Transformer or Substation	9	9	1	Low	CP
ET17	EVSE Meter Bypassing Result in Wrong Billing	3	1	3	Medium	PC
ET18	EVSE Destruction Result in Unavailability of Charging Service	1	1	1	Low	PP

4.3.3 Risk Assessment

When investigating the cyber-physical attacks on EV networks, we found that there exist two broad categories: (1) impact (2) cost. The impact is the impact and effects on the likelihood and opportunity of a successful attack. The cost refers to the cost or resources necessary for the attacker to be successful [Lee14]. To capture the relationship of impact and cost, we defined the risk as the ratio of impact to cost.

$$Risk = \frac{Impact}{Cost}. \tag{4.1}$$

The impact can be quantified as a 0, 1, 3, or 9 that represents the severity of a specific failure scenario, in which 0 is least significant and 9 most significant. Similarly, the cost can take on the values 0.1, 1, 3, or 9. Higher values of risk indicate risker systems. The potential parameters for each variable were chosen to make calculating risk easier and understandable.

The distance between the values also makes it easier for users to understand the different levels of risk, impact, and cost. These values are obtained by surveying individuals knowledgeable on cyber-physical system attacks. This formula highlights the areas of highest risk and provides a ranking system that prioritizes remediation effort.

A NESCOR member company has successfully used this approach in the past [Lee14]. Using NESCOR as a reference, we quantified the impact of a failure scenario as an impact score, which can take on the value 0, 1, 3, or 9. The values represent increasing severity of impact. For example, impact scores could be:

- 0: one customer out of power for 15 minutes, petty cash expenses,
- 1: small generation plant offline,
- 3: 20% of customers experience defect from smart meter deployment,
- 9: large transformer destroyed and major city out of power for a week.

Additionally, we created a cost score that represents the cost and difficulty to the threat agent to carry out the failure scenario, which can take on values 0.1, 1, 3, or 9. For example, cost scores could be:

- 0.1: It is easy to trigger the failure scenario, almost no cost,
- 1: a bit of expertise and planning needed, such as capture keys off unencrypted smart meter bus
- 3: serious expertise and planning needed to carry out scenario,
- 9: probably needs nation-state resources to carry out scenario (e.g., Stuxnet).

These scores are collected via a survey given to researchers familiar with the resources required to carry out such cyber-physical attacks. In both cases, the scores increase in severity as the number assigned increases. In cases, where scores are not the same values, we proposed using equation(4.1) to calculate risk, since the likelihood of the impact of a

cyber-physical attack and the means of carrying one our are directly proportional. In other words, as the potential impact of a cyber-physical attack increases, the amount of resources necessary to carry one out also increases. Thus, a higher ratio means a higher level of urgency. The risk assessment for 18 scenarios for the EV charging system is summarized in Table 4.2. The following subsections present three case studies for low- medium-, and high-risk scenarios.

4.3.3.1 Case Study I: Low Risk

A possible vulnerability could exist in a protocol translation module where unauthorized changes can be made. A successful attempt to exploit this may enable unauthorized control of EVs such as ET 5 shown in Table 4.2. The resources to accomplish this would require expertise and planning so that the cost score would be a three. The impact scenario could be altering charging levels for a large number of vehicles within a short period, which can have varied impacts ranging from inconveniencing customers. The impact value is also three because it primarily targets inconvenience to consumers, but it can have an impact on multiple consumers at the same time. Since the impact and cost are both three, the risk ration is one that is not high, but it is essential to know to figure out if an attacker would target this vulnerability.

4.3.3.2 Case Study II: Medium Risk

A possible vulnerability is the installation of malware in an EV. An attacker can propagate a virus between EVs and EV Service Equipment (EVSE) such as ET 3 shown in Table 4.2. Malware could affect driving mechanisms that could result in severe injury or loss of life. The impact would be severe as it affects multiple EV drivers, so the impact score is 9. The resources and cost to the attacker would require installing a virus, so the cost score is 3. This scenario has a risk ratio of three and can be used to prioritize this issue above the previous.

4.3.3.3 Case Study III: High Risk

A possible scenario is a malware causing the discharge of the electric vehicle to the microgrid, such as ET 15 shown in Table 4.2. Relevant vulnerabilities in the system will be changes to code in the charging station management system and protocol translation module or design, implementation, or maintenance permits system to enter a hazardous state by overloading of the distribution transformer if many EVs are discharged. The impact of such an attack could be Critical damage to electric vehicles and associated costs, violation of customer contracts and loss of customer confidence, or even sudden discharges that damage a transformer. In this scenario, the impact could be assigned a 3, and the cost could be considered a 0.1, which results in a risk score of 30.

4.3.4 Cybersecurity Survey of EV Users

A cybersecurity risk assessment is conducted via a survey on EV users' concerns on the EV charging system security. The purpose of this survey is to observe how consumers rank the importance of the effects of cyber-physical attacks on EV charging networks. The consumer is defined as any stakeholder whose product uses or is concerned with EV charging networks, such as EV car manufacturers or electric utility providers, and any individual who is an electric vehicle operator and who uses commercial PEV charging networks.

56 EV users have participated in the survey, and 50 of them had driven EV for at least six months (Fig. 4.8). 52 users have the experience of using commercial plug-in charging stations, which is defined as any charging station that does not use a charger plugging into a wall outlet. The frequency of using commercial plug-in charging stations is shown in Fig. 4.9.

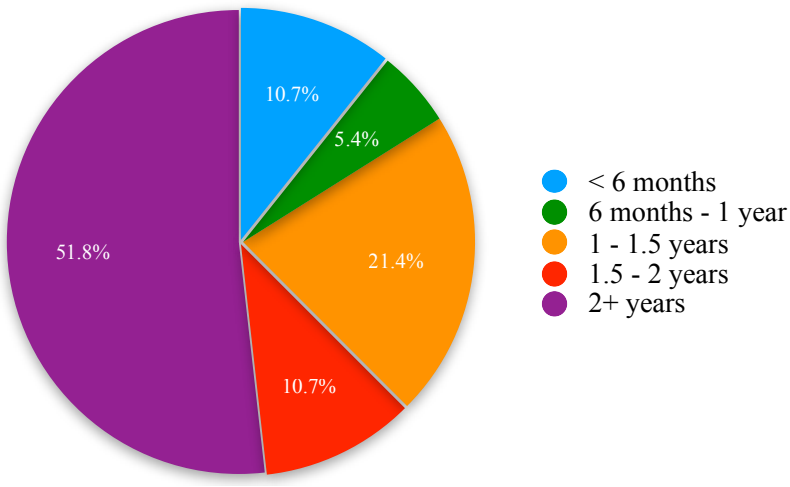

Figure 4.8: Time of having EV

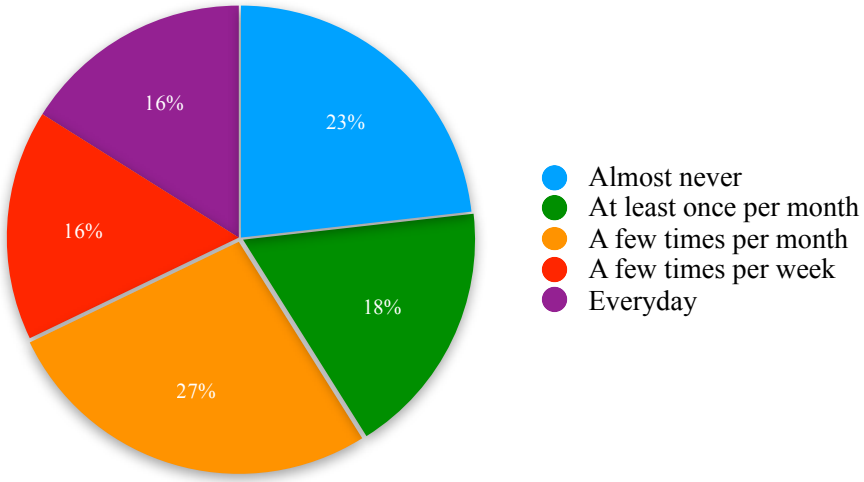

Figure 4.9: The frequency of using commercial plug in charging stations

There are two parts of the survey, the perception of inconvenience scenarios that EV users may experience when using commercial charging station and the concerning level for the private information compromise, respectively. Table 4.3 shows the rank of the concerning level of inconvenience.

Table 4.3: Rank of the concerning level of inconvenience

Scenarios	Rank
Charging station failure to charge electric vehicle	1
Car charges at slower than normal rate	2
Price fluctuation during car charging	3
Small inaccuracies in billing ranging $1-2	4

Table 4.4 shows the rank of the concerning level of private information compromise.

Table 4.4: Rank of the concerning level of private information compromise

Private information being compromised	Rank
Credit or Debit card information	1
Account login information	2
Home address	3
Phone number	4
Transaction history at charging station	5
Popular routes,taken by your car	6
Frequently visited destinations	7
Name	8

4.4 Discussion

Table 4.2 shows the impact scenarios that can occur as a result of different attack types on EV networks. Table 4.5 maps each scenario to the potential attacks.

The goal of the scoring mechanism is to rank risk in order to highlight areas of highest risk and prioritize remediation effort. The mapping attacks is to identify the nature of the attacks, thereby helps to find the corresponding solution. It is noted that a high ranking does not necessarily have the highest impact. For ET15, malware may be injected by an angry worker form EV maintenance service or anyone who has access to the EV. An EV

Table 4.5: Mapping of potential ET impact scenarios of listed attack types.

Attack Type	**ET Failure Scenarios**
Man in the Middle	ET2, ET5, ET6, ET15, ET16
Denial of Service	ET12, ET14
Packet Replay	ET14
Eavesdropping	ET4, ET7, ET9, ET10
ARP Spoofing	ET4, ET7, ET9, ET10, ET13
Insider	ET1, ET2, ET3, ET6, ET15, ET16

without malware detection can affect the EVSE when plugging in. On the other hand, low ranking can also result in high impacts as ET2, ET3, and ET16. For those cases, the attackers require a higher level of computer skill to compromise the system and thus increase the cost of the attacks. The nature of the attack-impact causality is marked in the "Class" column. Generally, preventing cyber attacks relies on a stronger authentication process, and avoiding cascading of physical impacts requires physical protection mechanisms such as circuit breakers. Fault detection is essential for both cyber and physical consequences. Physical attack, which is relatively rare, can be avoided by physically secure the access to the infrastructures. The strategy of mitigation for each scenario is summarized in Table 4.6.

Table 4.6: Mitigation action for each ET scenario

Scenarios	**Mitigation Act**
ET1	Overcharge-prevention hardware for EV battery[LSK15]; A stronger authentication mechanism for modifying EV firmware [ZL12].

Continued on next page

Table 4.6 – *Continued from previous page*

Scenarios	**Mitigation Act**
ET2	A stronger authentication mechanism for configuring fast charging management system[ZL12]; Fault-detection scheme for an unusual fast charging load[HST18]; Set an upper limit of EVs that can charge simultaneously; Deploy a circuit breaker to protect distribution transformer.
ET3	Anti-virus program in charging system to detect unauthorized software; Fault-detection scheme to detect abnormal events or functionality[HST18].
ET4	Enforcement of user password rule; Improve data encryption method[ALF14]; A stronger authentication process to access the database[ZL12].
ET5	Strengthen the integrity protections for translation modules;
ET6	A stronger authentication check between EVSE and the smart meter[MXD13]; A stronger authentication process to pair smart meter and EVSE configuration[MXD13].
ET7	Improve data communication encryption method between EV and EVSE [YYL11].
ET8	Deploy a power usage monitoring program to recognize EV charging pattern and identify abnormal usage pattern[HST18]; A stronger authentication process to verify the EV identity[MXD13].
ET9	Use multisignature method to authorize EV charging;[Bol03].
ET10	Use multisignature method to authorize EV charging;[Bol03].

Continued on next page

Table 4.6 – *Continued from previous page*

Scenarios	**Mitigation Act**
ET11	A stronger authentication process to access the database[ZL12]; Improve data encryption method[ALF14]; Use multisignature method to authorize EV charging;[Bol03]; Enable user to dispute the abnormal charging event and re-issue an EV ID.
ET12	Design resilient communication paths for EV identity verification.
ET13	Design resilient communication paths for EV identity verification; Use an alternative authentication method to verify EV identity.
ET14	Design resilient communication paths for EV identity verification; Use an alternative authentication method to verify EV identity.
ET15	A stronger authentication mechanism for configuring charging management system [LNZ14]; Require EV users' authorization for discharging; Deploy a circuit breaker to avoid over reverse power flow to the grid.
ET16	A stronger authentication mechanism for configuring charging management system[LNZ14]; Fault-detection scheme for an unusual charging load[HST18]; Set an upper limit of EV charging load; Deploy a circuit breaker to protect distribution transformer.
ET17	Secure the access to the EVSE.
ET18	Secure the access to the EVSE.

4.5 Conclusion

This chapter presents a comprehensive cyber-physical system vulnerabilities analysis for the EV charging domain. We analyze the UCLA WinSmartEV^{TM} charging network and identify the potential attack vectors and its attack surface. Since the cyber-security issue is an unfair advantage for hackers as they can choose the time and place of battle and attack only a single weak point of the system, understand the weakness and strengthen the protection scheme is vital of importance to secure the network. Therefore, we review the potential attack types to the system weakness and discuss their impacts. Eighteen ET failure scenarios are presented and categorized based on the attack-impact causality. We also conduct a risk assessment for each scenario and rank them in order to prioritize the remediation effort and allocate security resources accordingly. The major challenge of the security and privacy of an EV charging network would be EV authentication, user authorization, and communication encryption. The best strategy to secure the system is to increase the attacking cost until it outweighs the value of the attack. However, the potential attacks are included but not limited to the above mentioned eighteen scenarios. While we try to shorten the attack surface, the attackers may still strike the system via the weakness. In addition to improving system security, resiliency is critical when the system is compromised. Therefore, more research is needed for cyber-physical attack detection.

CHAPTER 5

Anomaly Detection for EV Charging Network

5.1 Overview

With the rapid growth in computer networks and the IoT technology deployment, the information network and its services are becoming much more complex and vulnerable to cyber-attacks. Even if a system has been hardened concerning all potential attack vectors, it is still inevitable that security has flaws. Motivated and well-resourced attackers will always breach it at some points. Moreover, attackers have the advantages of choosing the time and the vulnerability to exploit. System intrusion, data breaches, and privacy compromise have become the major concerns of such attacks. Cybersecurity and anomaly detection have been widely studied for computer networks but not as much as that for power systems. Studies have shown that the current US power system is at risk of a major cyberattack that could possibly result in devastating outcome[NE12, HHT14, Kop15]. EV charging infrastructures have been widely deployed in the power system to meet the enormously increasing energy demand [EVA18a]. To properly manage the power consumption, smart charging technology is currently under research and development and turns the charging network into an information-interconnected network. Therefore, the ability to identify potential attacks is imperative.

Anomaly detection is the process of identifying suspicious events or observations that do not conform to the typical behavior in the majority of the data. Anomalous data infers some problems or rare events such as structural defects, equipment malfunctions, or system intrusions. The connection between anomaly and the causal factors is valuable as it can diagnose a system condition and identify system faults followed by remedial responses. A

simple and common way to identify anomalies is to introduce a range of values for a normal condition. If an observation falls outside this range, it is considered as an anomaly. However, this hardcoded range values may result in a large number of false alarms or missing alarms, without considering other system parameters. For example, a false alarm may arise because of an abnormal high EV charging load, without knowing that the electricity price at that moment is extremely low. On the other hand, a typical EV charging load does not necessarily mean that there is no problem in the system, such as a spoofed electricity price suggesting there is no need to reduce the load.

EV charging management system controls and schedules EV load according to the measurements of local building load, solar generation, and dynamic electricity price. Within this information network, any replaced and modified data by an attacker will disrupt the EV charging schedule or even cause damage to the electrical grid. Those measurements are correlated under genuine circumstances, while compromised measurement disturbs the correlations. This chapter discusses the correlation of pair-wise measures within the EV charging network and analyzes the differences in normal and abnormal circumstances.

5.2 Literature Review

Anomaly detection has been widely studied in statistics and machine learning due to an increasing concern on information network security. Anomaly detection techniques can be classified into four categories, which are classification, clustering, statistical, and information theory [AMH16]. The application fields range from intrusion detection, fault detection, fraud detection, system condition monitoring, and event detection in sensor networks. Each class of the approaches has the strength based on the need of the application.

Surprisingly few studies have so far been made at anomaly detection for smart EV charging network. To ensure a smart EV charging system's reliability and guarantee the quality of the service output, a monitoring system that can detect anomalies and diagnose system faults is desirable. [Ger17] proposed a method to monitor the system status and a thresholding scheme to detect anomalies. [HLV03] introduced a Robost SVM (RSVM) method to

improve the detection accuracy by filtering the false-positive incidents. [KTK02] presented a statistical-based approach to detect anomalous network traffic. The above methods require to learn a universal and reliable threshold to raise alarms for anomalies, which may be difficult for a complex information network. Information theory provides an approach to evaluate and characterize the property of a system data, such as information gain [LLL19] and correlation analysis [CZC16, IML16]. A system profile is then built based on those properties, and anomalies can be identified by analyzing the pattern changes of the profile. For example, [CZC16, IML16] discovered the stable and significant correlations within a system under genuine circumstances so that the compromised measurements, which disrupt the correlations, can be captured. To efficiently analyze multiple pair-wise correlations at the same time, partitioning the time-series measurement data into segments can be helpful [RR06]. Based on the time-series partitioning approach, [HNB19] has developed a powerful and efficient way to analyze system-wise correlations, which is called Greedy Gaussian Segmentation (GGS) algorithm. The algorithm is formulated as a covariance-regularized maximum likelihood problem, dividing the multivariate time-series over which the data is well explained as independent samples. It is assumed that in each segment, the mean and covariance are constant and unrelated to information from other segments. Therefore, the anomalous data, in which the mean and covariance are peculiar, can be identified and marked out as a segment.

The smart charging here has taken the dynamic electricity price, building load, and solar generation into consideration to minimize the EV charging cost and net load variation. In this chapter, an information theory-based anomaly detection method is proposed here, and the GGS algorithm is applied to analyze the pair-wise correlations among EV charging load, building load, dynamic electricity price, and solar generation.

5.3 EV Charging Invariant Network and Anomaly Detection

5.3.1 System Overview

As shown in Fig. 5.1, the EV control center schedules the charging according to the meter data from the building and the solar power system as well as the dynamic electricity price from a utility. The optimization function of the scheduling is formatted as a quadratic program, and the objectives are to reduce charging costs and total load variance, as described in Chapter 3. To recap, there is a weighting factor α in (3.6) that controls the trade-off between mitigating net load variance and charging cost reduction. Table 5.1 demonstrate the impact of α to the charging scheduling algorithm.

Table 5.1: The impact of α to the EV charging scheduling algorithm

α Value	% Peak Reduction	% Cost Reduction
5	20.4	3.7
10	20.5	1.1
20	21.62	3.05
50	17.64	0
100	18.17	2.4
500	19.55	7.80
1000	19.30	10.60
5000	12.37	13.91
10000	4.53	18.12

For the value smaller than 50, α does not have a significant impact to the algorithm. For $\alpha > 50$ while increasing the values, the percentage of the charging cost reduces at the expense of increasing the net load variation. Here $\alpha = 1000$ is selected for the EV charging system.

Under this framework, there exists a set of pair-wise correlations among the components which is shown by the yellow dotted lines. The goal is to characterize the EV charging

correlation network under regular operation so that one can identify the abnormal state which disturbs these correlations.

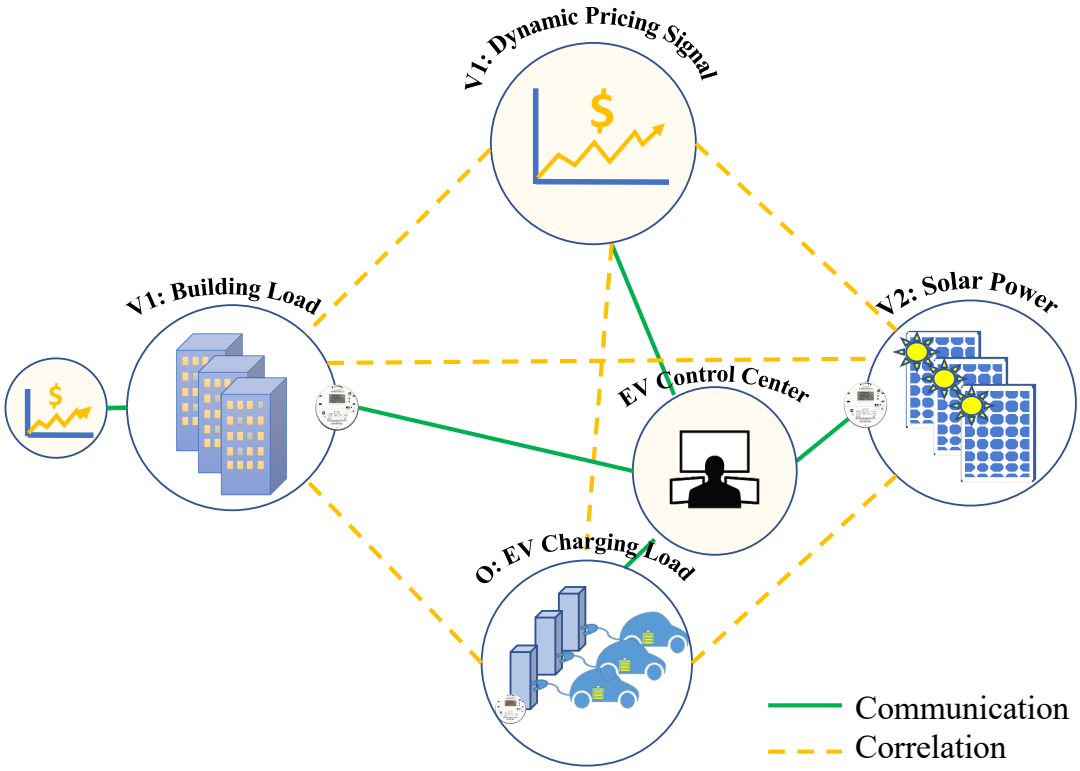

Figure 5.1: EV Charging Network

Real EV charging data with 7,994 records collected from UCLA campus was used here. The EV charging station is consisted with 7 level-2 EVCIs with a maximum charging rate 7 kW for each. We also implemented the real data of building load and solar generation from Cornell University [EM19], and the dynamic electricity price from ISO New England [Eng19]. To fit the charging station model, the building load and solar power were scaled to 25 kW in average and 10 kW, respectively. Unfortunately, smart building load in response to dynamic electricity prices was not available. To obtain this data we simulated an intelligent building load based on the data retrieved from Cornell University, with an assumption that 20% of the load are controllable and can be reallocated to a different point in time according to the dynamic pricing signal. The formulation of GGS algorithm is described in the next section.

5.3.2 Greedy Gaussian Segmentation (GGS) Algorithm

This section provides a brief review of the GGS algorithm. The details can be find in the reference [HNB19]. As illustrated in Fig. 5.2, considering a given time series $x_1, ..., x_T \in R^n$, the goal is find K break points $(b_1, ..., b_K)$ to divide the time series into $K+1$ segments, with the means and covariances

$$\mu^{(1)},..., \mu^{(K+1)}, \qquad \Sigma^{(1)},..., \Sigma^{(K+1)}$$

in each segment between the breakpoints. It is assumed that the mean and covariance in each segment are constant and unrelated to that in all other segments. In other words, the $x_t's$ are independent samples with $x_t \sim \mathcal{N}(\mu_{(t)}, \Sigma_{(t)})$, where the mean $\mu_{(t)}$ and covariance $\Sigma_{(t)}$ only change at the break points.

The breakpoints must satisfy

$$1 = b_0 < b_1 < \cdots < b_K < b_{K+1} = T + 1,$$

and the means and covariances are given by

$$(\mu_t, \Sigma_t) = (\mu^{(i)}, \Sigma^{(i)}), \qquad b_{i-1} \leq b_i, \qquad i = 1, ..., K$$

where t and i denote time t and segment i, respectively.

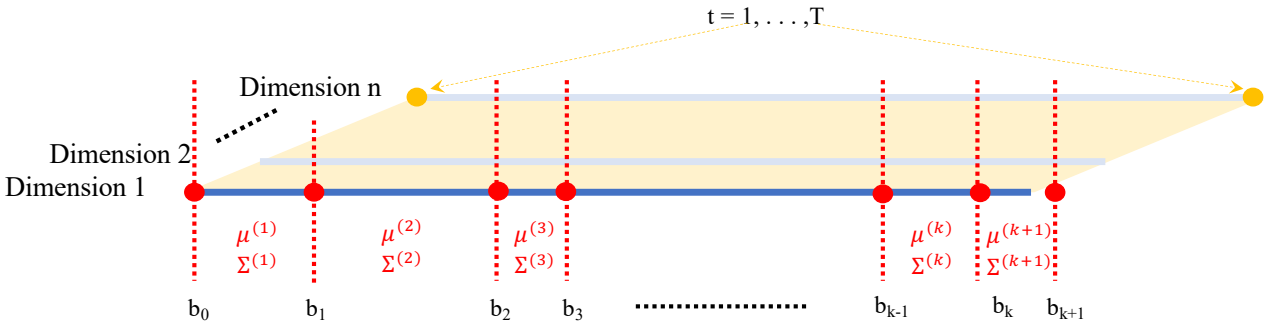

Figure 5.2: Segmented Gaussian model

The log-likelihood of the data $x_1, ..., x_T$ under the segmented Gaussian model (SGM) is formulated as

$$\ell(b, \mu, \Sigma) = \sum_{i=1}^{K+1} \ell^{(i)} \left(b_{i-1}, b_i, \mu^{(i),\Sigma^{(i)}} \right), \tag{5.1}$$

where $b = (b_1, ..., b_k)$, $\mu = (\mu^{(1)}, ...\mu^{(K+1)})$, $\Sigma = (\Sigma^{(1)}, ..., \Sigma^{(K+1)})$, and

$$\ell^{(i)} \left(b_{i-1}, b_i, \mu^{(i),\Sigma^{(i)}} \right) = -\frac{1}{2} \sum_{t=b_{i-1}}^{b_i-1} (x_t - \mu^{(i)})^T (\Sigma^{(i)})^{-1} (x_t - \mu^{(i)}) \\ - \frac{b_i - b_{i-1}}{2} \left(\log(\det \Sigma^{(i)}) + n \log(2\pi) \right).$$

Note that $b_i - b_{i-1}$ is the length of the ith segment.

To avoid the errors due to more dimension than samples in a segment, a covariance-regularized log-likelihood is formulated as

$$\begin{aligned} \phi(b, \mu, \Sigma) &= \ell(b, \mu, \Sigma) - \lambda \sum_{i=1}^{K+1} \mathrm{Tr}(\Sigma^{(i)})^{-1} \\ &= \sum_{i=1}^{K+1} \left(\ell^{(i)} \left(b_{i-1}, b_i, \mu^{(i),\Sigma^{(i)}} \right) - \lambda \sum_{i=1}^{K+1} \mathrm{Tr}(\Sigma^{(i)})^{-1} \right), \end{aligned} \tag{5.2}$$

where $\lambda \leq 0$ is a regularization parameter and K is fixed. The analytical solutions for μ and Σ are

$$\mu^{(i)} = \frac{1}{b_i - b_{i-1}} \sum_{t=b_{i-1}}^{b_i-1} x_t \tag{5.3}$$

$$\Sigma^{(i)} = S^{(i)} + \frac{\lambda}{b_i - b_{i-1}} I, \tag{5.4}$$

where

$$S^{(i)} = \frac{1}{b_i - b_{i-1}} \sum_{t=b_{i-1}}^{b_i-1} (x_t - \mu^{(i)})(x_t - \mu^{(i)})^T.$$

Therefore, the maximum covariance-regularized log-likelihood of (5.2) can be expressed as

$$maximize \quad -\frac{1}{2} \sum_{i=1}^{K+1} \psi(b_{i-1}, b_i), \tag{5.5}$$

where

$$\psi(b_{i-1}, b_i) = \left((b_i - b_{i-1}) log \left(det \left(S^{(i)} - \frac{\lambda}{b_i - b_{i-1}} I \right) \right) - \lambda \, \mathrm{Tr} \left(S^{(i)} - \frac{\lambda}{b_i - b_{i-1}} I \right)^{-1} \right),$$

and the variable set of $b = (b_1, ..., b_k)$ is to be chosen to maximize the objective function (5.5).

To find the optimal break points (b_i), GGS algorithm along with a split subroutine $Split(b_{i-1}, b_i)$ are implemented. $Split(b_{i-1}, b_i)$ screens the values between the segment b_{i-1} and b_i, and determines the optimal t that optimize $\psi(b_{i-1}, t) + \psi(t, b_i)$. The GGS algorithm is illustrated in Fig. 5.3. The alogrithm loops over from $i = 1$ to $i = K + 1$ and adds the optimal break point in each loop. Adding the point is followed by relabeling the point and points adjustment until all K break points are acquired.

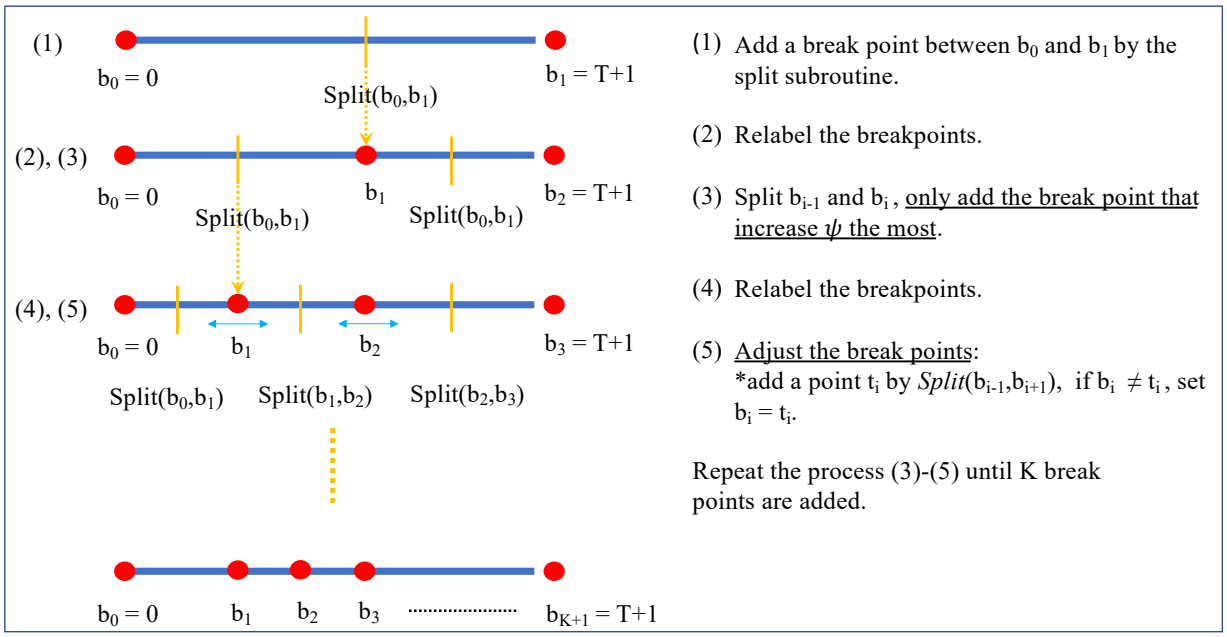

Figure 5.3: GGS algorithm

5.3.3 Validation and Parameters Selection

The λ and K in (5.5) is determined by running a 10-fold cross-validation. For each fold, 10% of the samples are selected randomly as the test set, and the remaining 90% are the training set. The averaged log-likelihood results versus different K values by the GGS algorithm for

test and training sets were compared. The comparisons with different λ values are presented in Fig. 5.4. When $\lambda = 10^{-3}$ and $\lambda = 10^{-2}$, the log-likelihood values drop significantly after $K = 2$ and $K = 1$, respectively. The divergence of the log-likelihood values indicates a overfitting of model. For $\lambda = 1$, the curve stops at $K = 5$ because there is no breakpoint can be found to increase the log-likelihood. For $\lambda = 10^{-1}$, the log-likelihood remains the same after $K = 5$. For GGS to model time-series data, small K and large λ are preferable because they make the model simpler and less sensitive to noise. Therefore, $K = 5$, and $\lambda = 10^{-1}$ would be reasonable choices for the purpose.

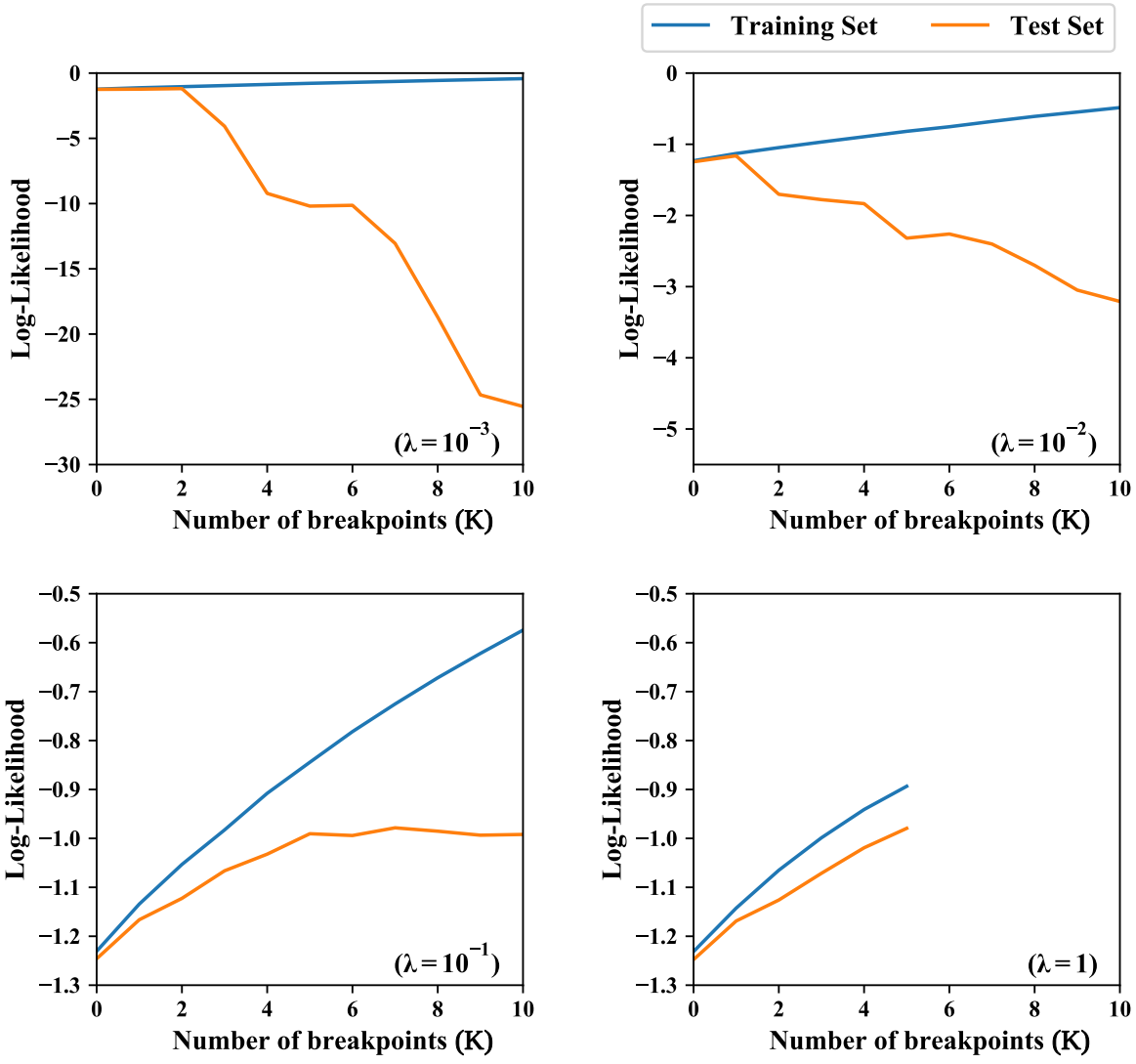

Figure 5.4: 10-fold cross validation with different K and λ

5.4 Result and Discussion

Smart EV charging control enhances the correlation among the EV charging load, building load, solar generation, and dynamic electricity price. The following figures present the comparison between coordinated and uncoordinated EV charging control, in terms of system-wise correlation.

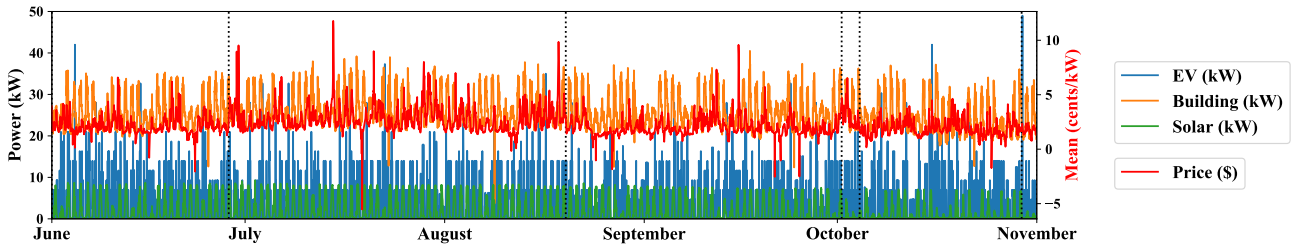

Figure 5.5: Time series for uncorrdinated EV charging

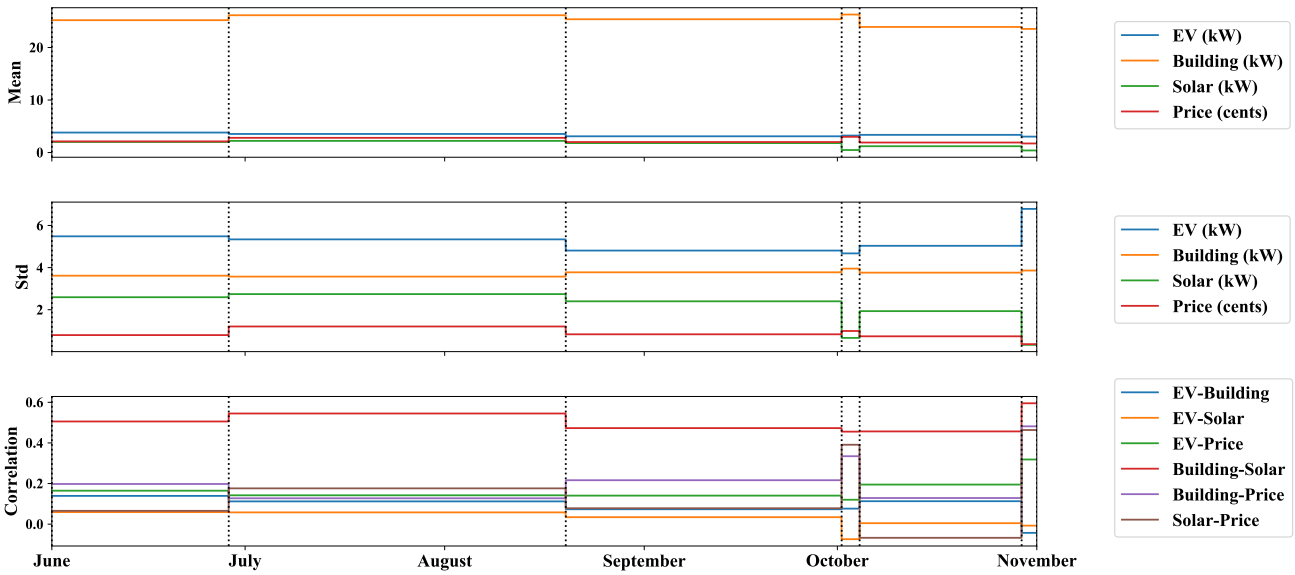

Figure 5.6: Mean, standard deviation(Std), and correlation for uncorrdinated EV charging

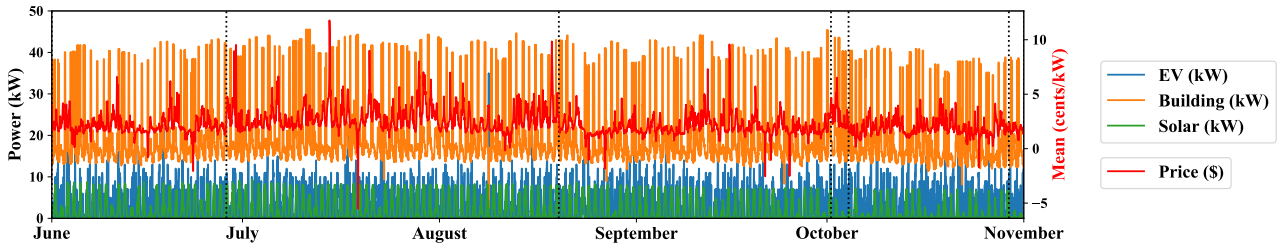

Figure 5.7: Time series for corrdinated EV charging

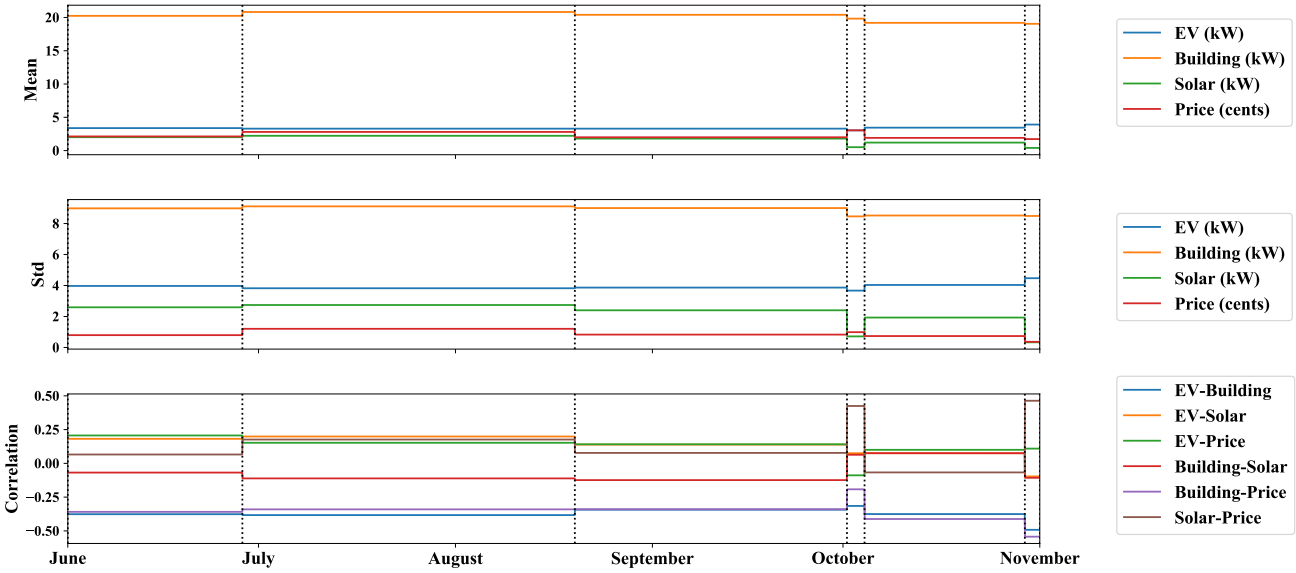

Figure 5.8: Mean, standard deviation(Std), and correlation for corrdinated EV charging

The correlations shown in Fig. 5.6 and Fig. 5.8, following changes are noticed:

- **EV-Building** becomes negative correlated (0.15 to -0.4).
- **EV-Solar** becomes slightly positive correlated (0.05 to 0.2).
- **EV-Price** correlation has little to no different change (0.1 to 0.2).
- **Building-Solar** becomes slightly negative to no correlated (0.5 to -0.1).

- **Building-Price** becomes negative correlated 0.2 to -0.4.
- **Solar-Price** remains unchanged around 0.1.

The correlation of EV-Building changes more than EV-Price because the charging scheduling algorithm weighs more on reducing load variance than the price effect. Building-Solar becomes uncorrelated because building control only considers the price. The positive correlation found for the uncoordinated load may be because people tend to use more air-conditioners when it is hot during the summer.

To simulate anomalous data, the data of dynamic pricing, building load, EV load, and solar generation are falsified, with some fake data points inserted randomly. The following subsections present the result of anomaly detection by running the GGS algorithm.

5.4.1 Detecting False Pricing Data

Since the EV charging scheduling and building control consider dynamic electricity price, false pricing data injection may affect the EV charging schedule as well as building load. As Fig. 5.9 and Fig. 5.10shown, two fake pricing data inserted were identified (Anomaly 1 and anomaly 3). However, there is a false alarm (Anomaly 2) being raised due to a normal drop in the price. The correlation changes comparison among the three anomalies is shown in Fig. 5.11. As expected, Building-Price correlation changes because the building is controlled according to untampered price. For EV-Building, the correlations change because EV charging is encouraged due to the fake low price for Anomaly 1 and Anomaly 3. Anomaly 2 is identified because there is no EV charging at that period, and it happens to have a price drop. With a normal decrease in price, EV and building load would both increase and thus weaken the negative correlation. But a false decrease price further strengthens the negative correlation because building load does not respond to the false price. The increase in EV load due to the false price makes a negative EV-Solar correlation, while the normal price drop does not affect this correlation. EV-Price correlation should be no change since EV is controlled based on the false price in the same way as the untampered price. However, over-induced EV charging can make a significant effect, as shown in EV-Price correlation

change for Anomaly 1. Building-Solar correlation is irrelevant to this scenario.

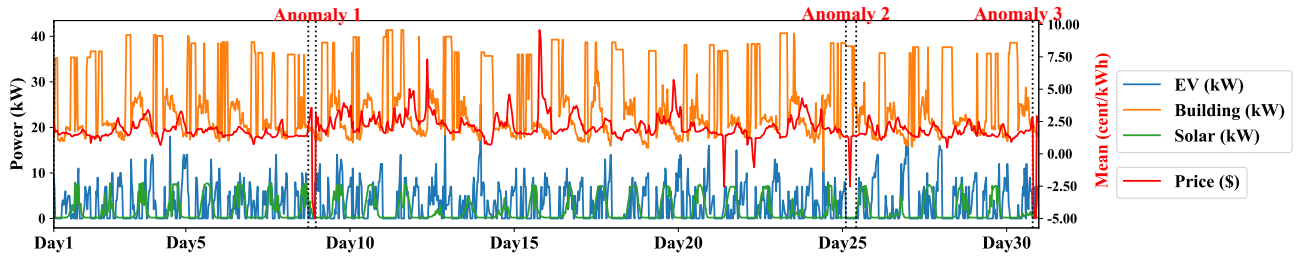

Figure 5.9: Time series with some false pricing data inserted

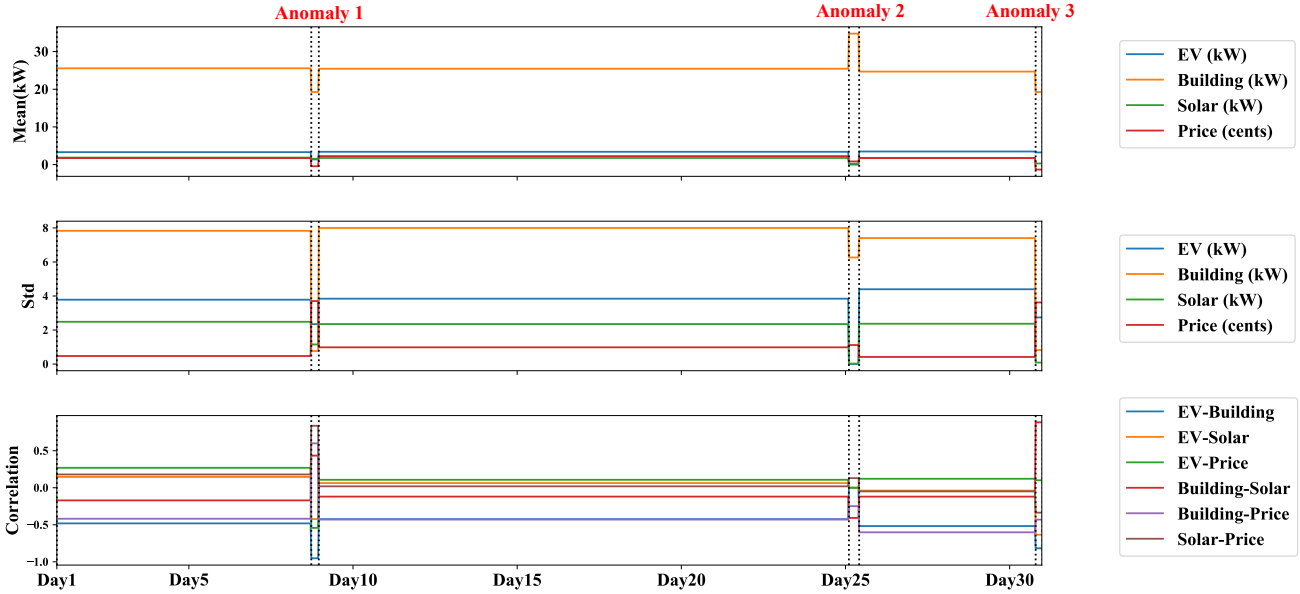

Figure 5.10: Detecting correlation changes due to false price

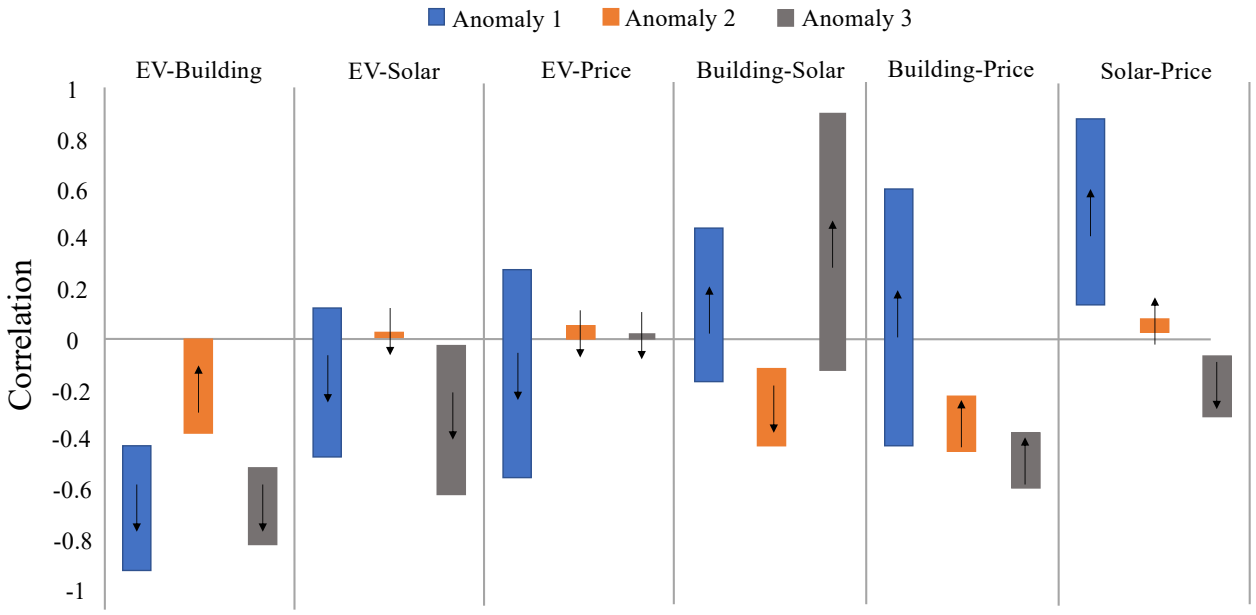

Figure 5.11: Correlation values change from the previous segment for the three identified anomalies

5.4.2 Detecting False Building Load Data

Falsified building load will affect the EV charging schedule heavily and result in the unexpected peak of EV load. Therefore it is important to identify any tampered building load data.Fig. 5.12 and Fig. 5.13 shows the detection of two anomalies. Anomaly 1 is a natural system anomaly, while Anomaly 2 is due to a tampered building load. The comparison of the correlation changes for the two anomalies is illustrated in Fig. 5.14. As shown in the figure, the changes of the two anomalies behaved differently. While the natural anomaly strengthens its original correlations, the tampered data induced anomalies change the direction of the correlations for most of the pair-wise relationships. For example, the EV-Building relationship for Anomaly 2 changes from a steady negative correlation to a weak positive correlation. Also, the fake building load makes it more correlated to solar generation, which does not make sense since the building control does not consider solar power.

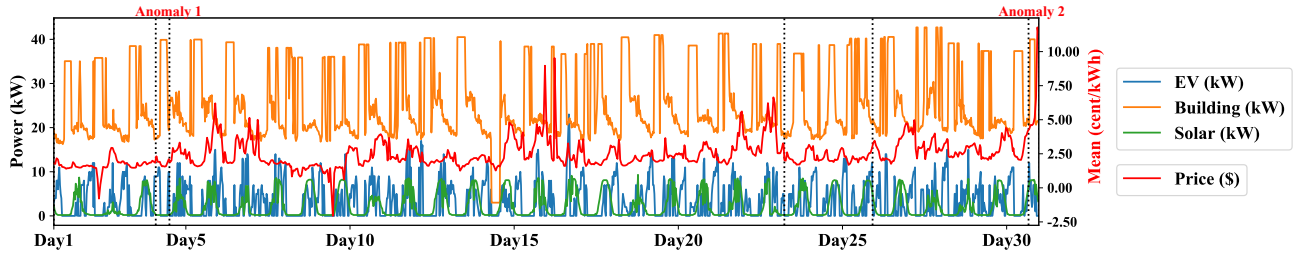

Figure 5.12: Time series with some false building load data inserted

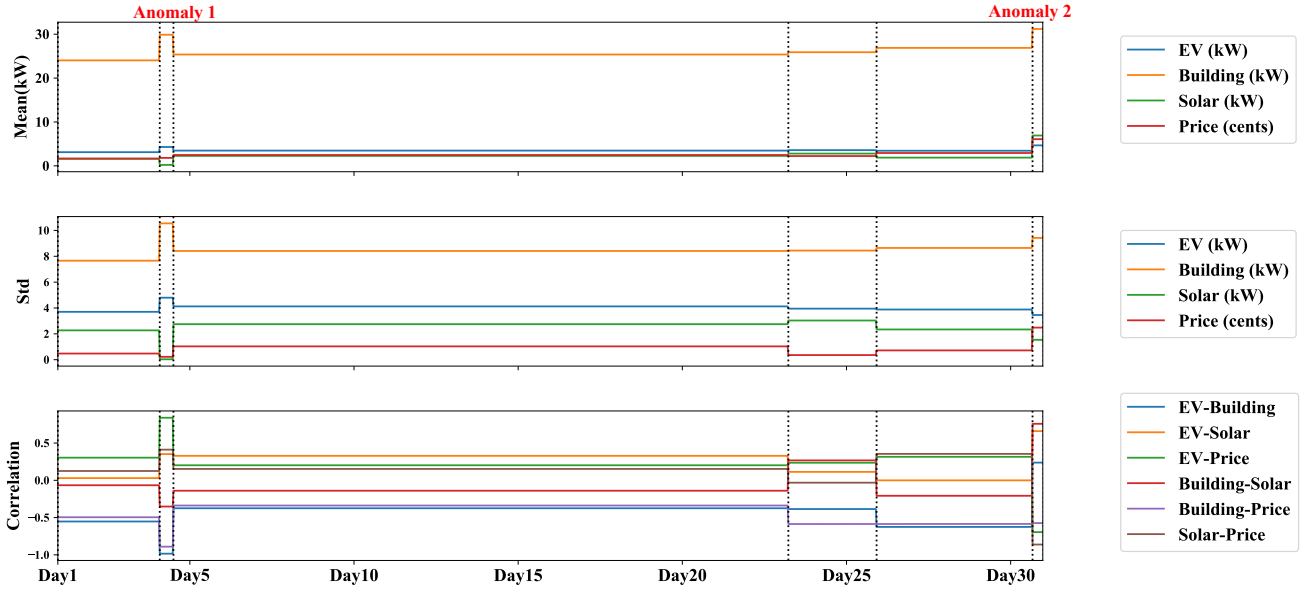

Figure 5.13: Detecting correlation changes due to false building load data

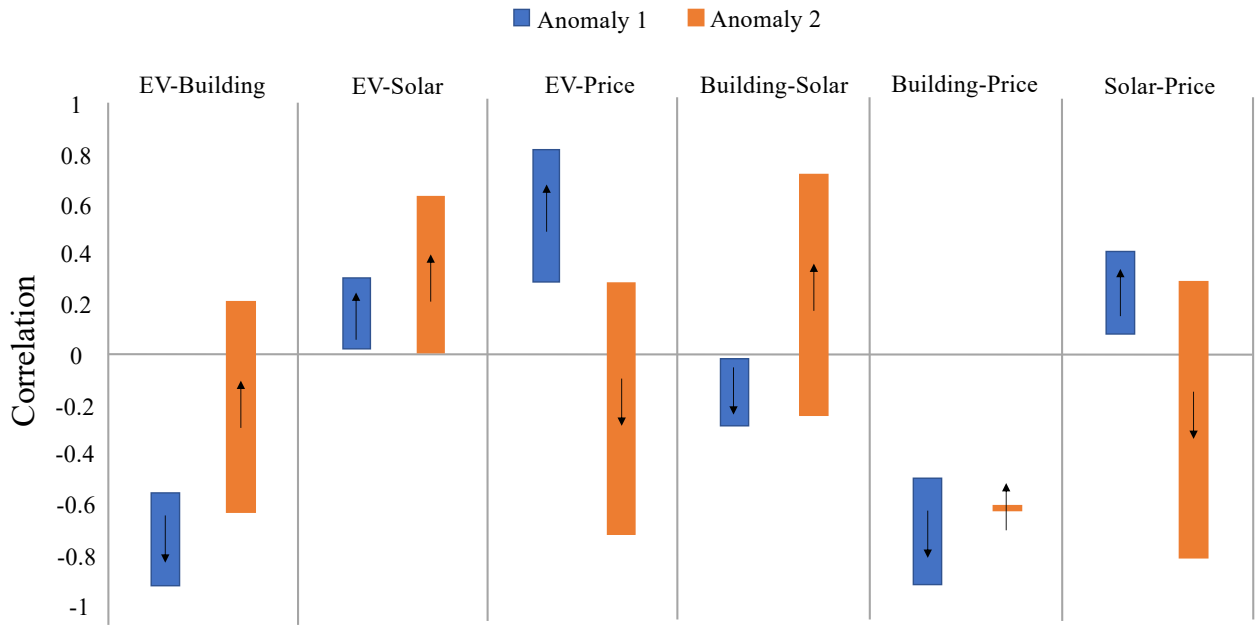

Figure 5.14: Correlation values change from the previous segment for the two identified anomalies

5.4.3 Detecting False EV Charging Load

While an EV charging schedule can be misled by false electricity price and false building load, EV charging load could even be altered by hacking into the system. Since under such attack, the EV load is unlikely to correlate to the other time-series data, the proposed detection method using the concept of the invariant network has the potential to identify this system intrusion. The correlations that are related to EV should be a focus. Fig. 5.15 and Fig. 5.16 show three identified anomalies. In fact, Anomaly 1 and Anomaly 3 were because of the tempered EV charging loads, and Anomaly 2 was a natural anomaly, which was the result of a drop in electricity price, and no EV charging happened at the same time. The analysis of the correlations change for the three anomalies are shown in Fig. 5.17. For EV-Building, both Anomaly 1 and Anomaly 3 change from strong negative correlations to strong positive correlations. This is because the tempered EV loads increase while the building load is increasing. For EV-Solar, Anomaly 1 drops to zero because the false EV charging occurs

at the time without solar generation; Anomaly 3 increases to positive correlation because the false EV charging increases while solar generation is increasing. For EV-Price, Anomaly 1 remains at the same level of the correlation, while Anomaly 3 drops to a strong negative correlation because of the same trend of the pricing and EV load in that period. The three correlations discussed above for Anomaly 2 all changes to zero because there is no EV charging at the time of a significant drop in price. The rest correlations are less relevant to false EV loads detection.

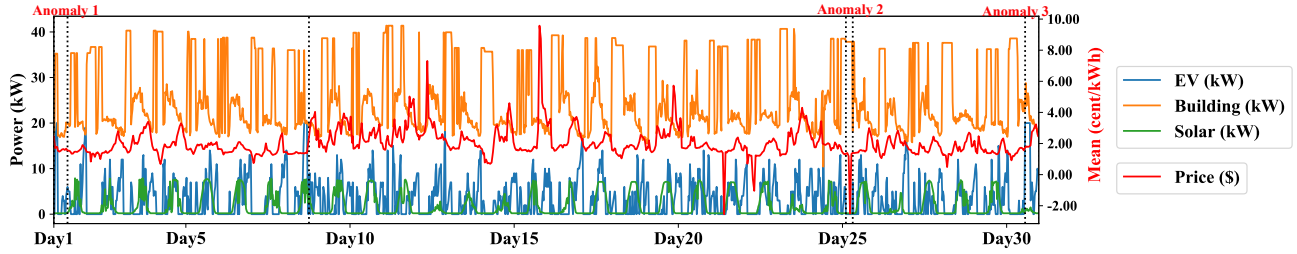

Figure 5.15: Time series with EV charging load being altered

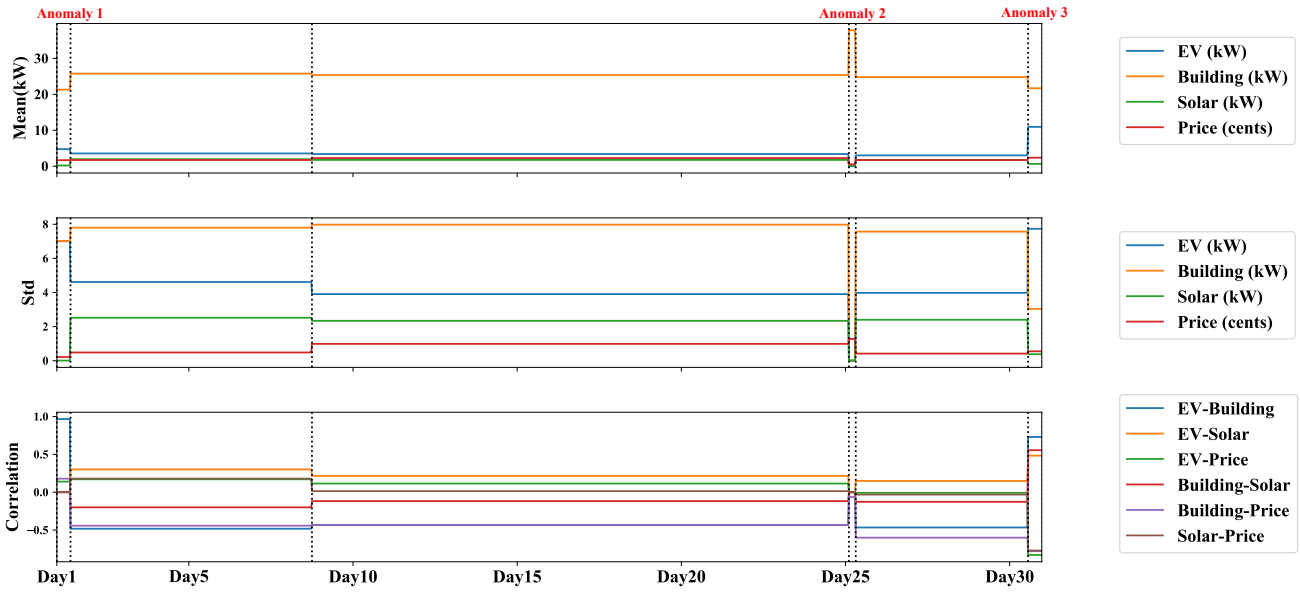

Figure 5.16: Detecting correlation changes due to anomalous EV charging events

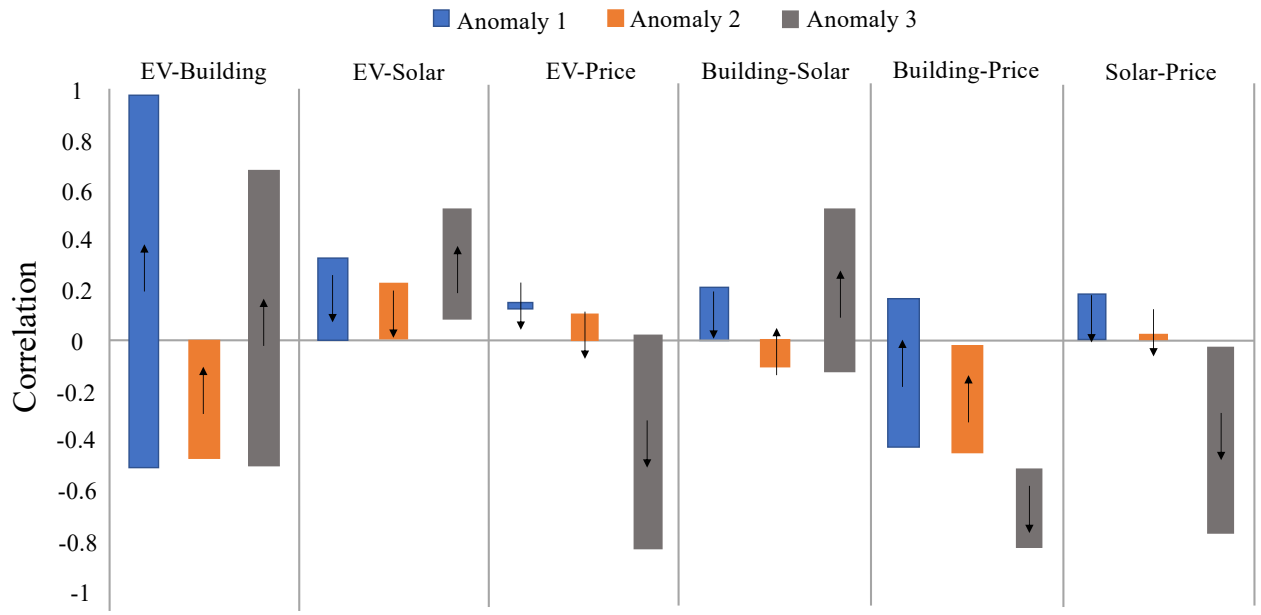

Figure 5.17: Correlation values change from the previous segment for the three identified anomalies

The results have demonstrated the capability to identify false pricing and building load data injection, and EV load manipulation, based on the steady correlations among the time series. 123 compromised events were issued and 112 were identified. Precision/Recall metric is used to evaluate this model, as shown in Fig. 5.18 below.

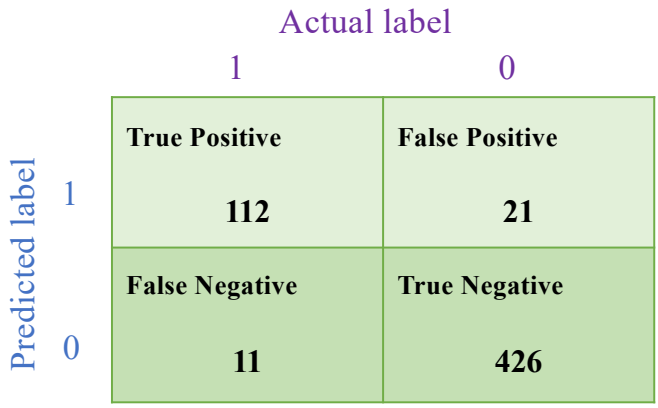

Figure 5.18: Precision/Recall Metric

Precision refers to the percentage of the predicted positives that are true positives, and recall refers to the percentage of real positives being identified correctly. The definitions precision, recall, and accuracy are defined as follows:

$$\textbf{Precision} = \frac{True\,Positives}{Total\,\#\,of\,predicted\,positives} = \frac{True\,Positives}{True\,Positives + False\,Positives},$$

$$\textbf{Recall} = \frac{True\,Positives}{Total\,\#\,of\,actual\,positives} = \frac{True\,Positives}{True\,Positives + False\,Negatives},$$

$$\textbf{Accuracy} = \frac{True\,Positives + True\,Negatives}{Total\,Population}.$$

The precision, recall, and accuracy for the detection model are 84%, 91.1%, and 94.3%, respectively.

However, because solar-related correlations are relatively weak in this experiment setup, the detection of false solar generation data is not significant. This is due to a data limitation. First off, solar and electricity price is not correlated in the region of New England, where the solar penetration is not yet to impact the electricity price. Solar-electricity price correlation is more significant in California. According to CAISO, the higher penetration of solar results in a stronger correlation between solar generation and electricity price ($\sim$ -0.5). Secondly, Smart building load is controlled according to electricity price only and thus is not correlated to solar generation. More solar-related correlations can be found in a more extensive and more complex EV charging network, and they can be utilized to detect the anomalies due to tampered solar data.

5.4.4 Identifying the Sources of Anomalies

As seen in the previous subsections for the detection, natural changes in correlation are sometimes identified as anomalies because they rarely happen. There is a need for a supervised machine learning approach to distinguish natural and malicious cases by using the correlation set as features, where the correlation set $\mathcal{C} \triangleq$ [EV-Building, EV-Solar, EV-Price,

Building-Solar, Building-Price, Solar-Price]^T. In addition, based on the observation, the correlation changes of the anomalies due to tampered measurement are much larger than that for natural anomalies as shown in the Fig. 5.11, Fig. 5.14, and Fig. 5.17. Therefore, a Euclidean distance ($\mathcal{D}$) between a correlation set ($\mathcal{C}$) to the averaged correlation set under normal operation ($\mathcal{C}_{mean}$) is taken as an extra feature, where $\mathcal{D}$ is defined as:

$$\mathcal{D} \triangleq \|\mathcal{C} - \mathcal{C}_{mean}\| = \sqrt{\|\mathcal{C}\|^2 + \|\mathcal{C}_{mean}\|^2 - 2\mathcal{C} \cdot \mathcal{C}_{mean}}.$$

Each segment by the GGS model can be described as an 8-tuple: ($\mathcal{C}$,$\mathcal{D}$, class), where $\mathcal{C}$ has six elements, $\mathcal{D}$ has one element, and class is the label of the source of the attack (class $\in$ [Normal, tampered price, tampered building load, tampered EV load]). 570 segments were labeled, including 470 normal segments, 37 segments due to tampered price, 34 segments due to tampered building load, and 41 segments due to tampered EV load. Because of the limited labeled dataset, a cross-validation testing method is conducted. For each cross-validation trial, 10% of the labeled data was selected randomly as a test set, and the remaining 90% as a training set.

A weighted k-nearest neighbor (kNN) classifier (weight = $\frac{1}{distance}$) is introduced to classify the sources of attacks. For a new sample, the kNN classifier finds its closest k labeled samples and classifies the sample by the majority vote of these k nearest neighbors. The closer neighbor has a higher weight on its vote. Fig. 5.19 shows the intuition for kNN classification. $k = 2$ is chosen because it has the best classification accuracy in comparison to the other numbers of k under 30-trials of cross-validation testing, as shown in Fig. 5.20. The curves of $k = 1$ and $k = 2$ overlap because they are equivalent.

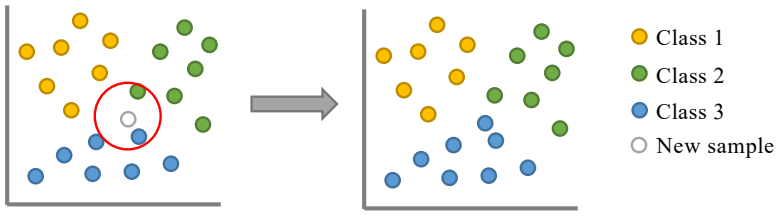

Figure 5.19: Intuition for kNN classification with k=2

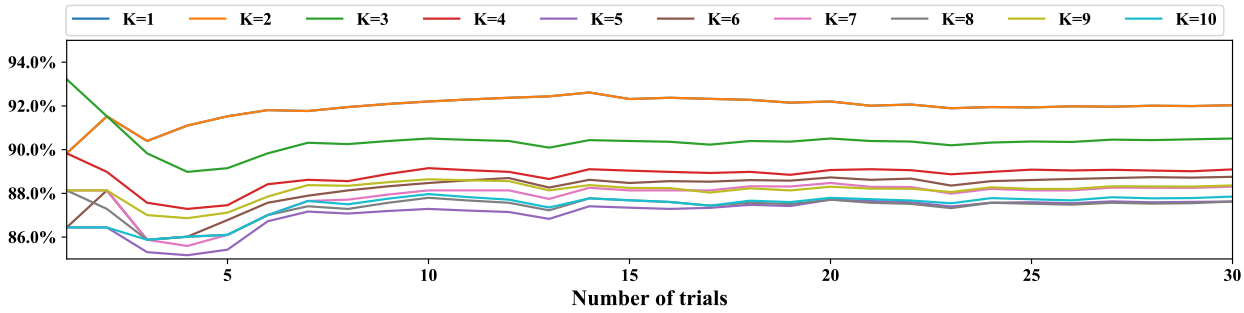

Figure 5.20: The comparison of classification accuracy with different numbers of k

Fig. 5.21 shows the result of kNN classification for one of the eighty trials. The numbers on the diagonal are the number of correctly labeled incidents. There are still a few misclassified cases. For example, the figure shows one attack due to a fake EV load was labeled as a fake price. The misclassification rate for the detected anomalies is 0.17 on average for the eighty-trials testing. The result for the cross-validation testing is shown in Fig. 5.22. The averaged Precision, Recall, and Accuracy are 96.8%, 77.2%, and 94.4%, respectively.

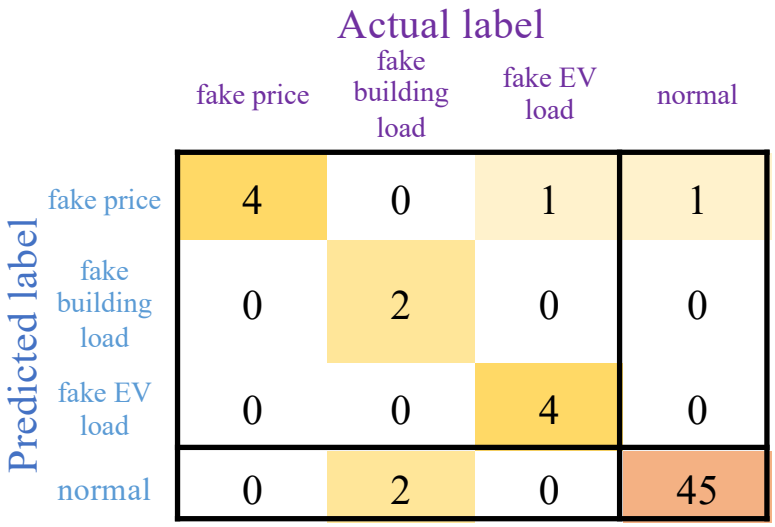

Figure 5.21: Confusion matrix for the kNN classification

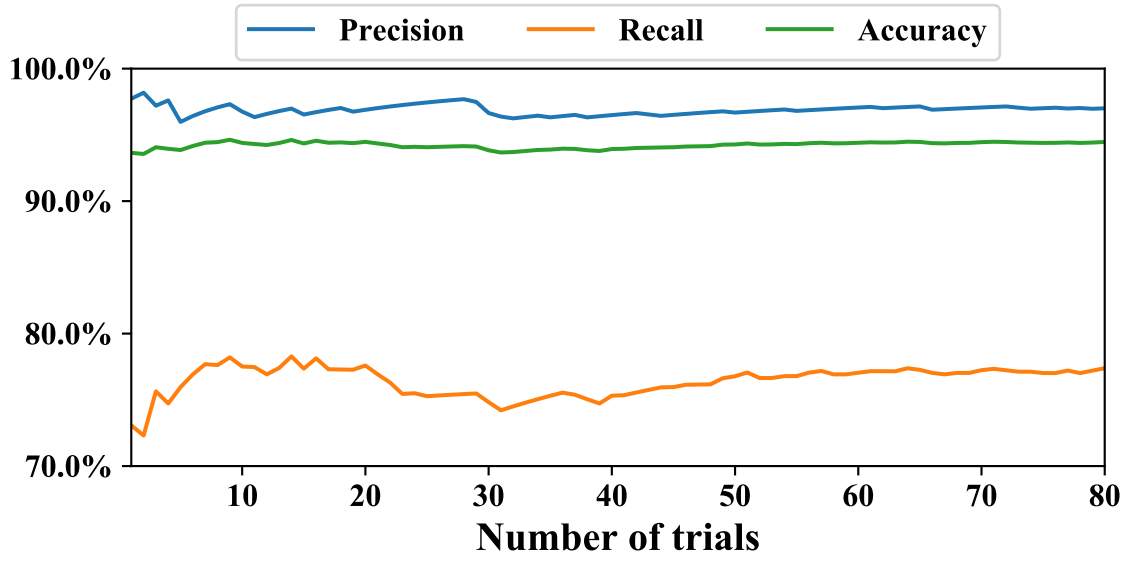

Figure 5.22: Mean prediction Precision, Recall, and Accuracy for 80-trials cross-validation testing

5.5 Conclusion

This section presents a novel strategy to detect system anomalies. Usually, the conventional detection methods characterize the measurements according to their typical behavior and identify the one that behaves anomalously. This kind of approach defines a hard-coded range based on the empirical rule and considers the data points lying outside of this range as anomalies. However, very often, the data points located within the normal range can be anomalous, and the points locate outside can be healthy. Therefore, using the conventional approach can result in many false-negative and false-positive cases.

The GGS algorithm examines all system measurements concurrently, take them as a whole entity by formulating it as a multivariate maximum log-likelihood problem. The method divides the time-series data into several segments, intending to isolate anomalous data. Under typical operation, each measurement has steady correlations with the other measurements. Any unusual event that disturbs the correlations will be detected as an anomaly. This can be more accurate because anomalies are identified by the result of many

correlation changes instead of a single determinant. In addition, a kNN classifier is applied to find the source of an anomaly upon detection. While different tampered data results in different correlation changes, the value of correlations can provide insight for tracing the source of the intrusion.

In this information interconnected era, any cyber-attack can result in adverse consequences, such as system malfunction, property damage, privacy compromise, or even threatening lives. Cybersecurity becomes paramount that should be taken seriously. Although an enlarger information network makes it more vulnerable to cyber threats, while the interconnected information working collaboratively, making use of their correlation can be a great idea to defense the cyber threats.

CHAPTER 6

Conclusion & Future Work

The current electrical grid is undergoing a revolutionary transition toward an environmental-friendly, energy-efficient, and information interconnected smart network. In response to the critical energy and environmental challenges, humongous advanced technologies have been developed to support this transition. Utilities, government, and regulators are working collaboratively to achieve the maximum benefits. The behavior of the new energy users is the most critical factor, with the vision that customers will be able to manage the energy usage accordingly. Therefore, customers can save energy costs, the government can reach energy reduction, and all of us can benefit from lower carbon emissions and a cleaner environment. The surge of EV to the world and the integration to the distribution grid is an excellent example of smart energy utilization since EVs have a great potential to provide grid services and stabilize the electrical grid. Yet, many technical challenges need to be resolved for EV-grid integration. As EVs are powered by electricity, a massive charging demand may add up the power consumption an result in many grid problems. Accordingly, smart charging scheduling strategies are developed, intending to reallocate energy consumption. However, the biggest challenge to reach the optimal charging schedule is the stochasticity of the charging behavior, including the time to charge and the energy demand. Therefore, Chapter 2 provides an improved prediction method, namely the ensemble predicting algorithm (EPA), to tackle the uncertainty of EV users' behavior and leverage the performance of EV charging scheduling. In this chapter, data sparsity/entropy (R) as an information property taken as is defined to classify different charging patterns. The EPA can thus select the proper methods according to R and make a more accurate prediction. Nonetheless, with the capability of EV scheduling, charging stations that are working independently may not achieve the maximum

benefit of what EVs can provide. Therefore, Chapter 3 presents a scalable and straightforward EV charging scheduling algorithm that can be easily adapted to a distributed control scheme. In this manner, more EVs within the distribution grid can work together to make a significant impact. For future work, a smart charging strategy for a distributed control scheme will be studied to manage the collaboration of different charging stations. Also, the extra EV charging load will broadly impact the original dynamic electricity pricing mechanism, which is taken as the input of the charging algorithm in this dissertation. This should be further studied for better EV-grid integration.

The smart charging control allows EV to be externally controlled for integration into the whole power system and provide grid services. However, the expanded communication network of EV charging systems become more vulnerable to potential cyber threats. Chapter 4 discusses the possible cyberattacks against an EV charging system and presents the result of the vulnerability analysis and risk assessment, intending to create a generalizable and comprehensive solution. Inevitably, there always exists a system weakness that can be exploited for a cyberattack. Thus, anomaly detection is critical to protect the system by handling the attacks quickly. For this reason, Chapter 5 demonstrates a novel approach to detect the EV charging system anomalies. The proposed method analyzes the system-wise correlations, characterizes the invariant-correlation network, and discovers the unusual correlation changes. The method can provide more accurate detection because anomalies are identified by the result of many correlation changes instead of a single determinant factor.

Furthermore, a weighted kNN classifier is applied to distinguish natural and malicious cases, making use of the correlation values as critical features. The results also show the potential to classify and identify the source of the attack by the kNN classifier. However, due to the data limitation, such as a relatively small scale of the labeled dataset and weak solar-related correlation, the current experimental setup does not detect tampered solar attacks well. For future work, an expanded EV charging network that involves more components, as well as more solar-related correlations, will be examined to verify the detection functionality and accuracy. Also, with a larger labeled dataset, neural network can be trained to improve anomaly classification performance.

In summary, this dissertation presents the strategy for EV-smart grid integration, in the aspects of EV load modeling and prediction, charging management and scheduling, and the cybersecurity discussion and anomaly detection. As we have witnessed the evolution of the electrical grid, the rapidly growing number of EVs, and the related-technology advancement, this work contributes to addressing the current technical issues and paving the essential step into the future.

REFERENCES

[AAJ15] Sajjad Abedi, Ata Arvani, and Reza Jamalzadeh. "Cyber security of plug-in electric vehicles in smart grids: application of intrusion detection methods." In *Plug In Electric Vehicles in Smart Grids*, pp. 129–147. Springer, 2015.

[AKK16] M. Hadi Amini, Amin Kargarian, and Orkun Karabasoglu. "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation." *Electric Power Systems Research*, **11**:378 – 390, 2016.

[ALF14] Man Ho Au, Joseph K Liu, Junbin Fang, Zoe L Jiang, Willy Susilo, and Jianying Zhou. "A new payment system for enhancing location privacy of electric vehicles." *IEEE transactions on vehicular technology*, **63**(1):3–18, 2014.

[Alt92] N. S. Altman. "An introduction to kernel and nearest-neighbor nonparametric regression." *The American Statistician*, **46**(3):175–185, 1992.

[AMH16] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. "A survey of network anomaly detection techniques." *Journal of Network and Computer Applications*, **60**:19–31, 2016.

[BGK10] Z. I. Botev, J. F. Grotowski, and D. P. Kroese. "Kernel density estimation via diffusion." *The annals of Statistics*, **38**(5):2916–2957, 2010.

[BGS17] Rajneetkaur Bijral, Alka Gupta, and Lalit Sen Sharma. "Study of Vulnerabilities of ARP Spoofing and its detection using SNORT." *International Journal of Advanced Research in Computer Science*, **8**(5), 2017.

[BLB13] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. "API design for machine learning software: experiences from the scikit-learn project." In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122, 2013.

[BNE16] Abdoulmenim Bilh, Kshirasagar Naik, and Ramadan El-Shatshat. "A novel online charging algorithm for electric vehicles under stochastic net-load." *IEEE Transactions on Smart Grid*, **9**(3):1787–1799, 2016.

[Bol03] Alexandra Boldyreva. "Threshold signatures, multisignatures and blind signatures based on the gap-Diffie-Hellman-group signature scheme." In *International Workshop on Public Key Cryptography*, pp. 31–46. Springer, 2003.

[Bre] Breach Level Index. "DATA BREACH STATISTICS - DATA RECORDS LOST OR STOLEN SINCE 2013." `https://breachlevelindex.com`. Accessed: 2019-04-20.

[Cal] California Independent System Operator (CAISO). "Locational marginal price [Online] (2018)." http://oasis.caiso.com. (accessed: 01.05.2018).

[Cal16] California Independent System Operator. "What the duck curve tells us about managing a green grid." https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf, 2016.

[Cal19] California Energy Commission (CEC). "Total System Electric Generation." https://ww2.energy.ca.gov/almanac/electricity_data/total_system_power.html, 2019.

[CB12] Hina Chaudhry and Theodore Bohn. "Security concerns of a plug-in vehicle." In *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, pp. 1–6. IEEE, 2012.

[CCO18] Cedric Carter, Patricia G Cordeiro, Ifeoma Onunkwo, and Jay Tillay Johnson. "Cyber Assessment of Distributed Energy Resources." Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2018.

[CF10] Michael C Caramanis and Justin M Foster. "Coupling of day ahead and real-time power markets for energy and reserves incorporating local distribution network costs and congestion." In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 42–49. IEEE, 2010.

[CHD09] Kristien Clement-Nyns, Edwin Haesen, and Johan Driesen. "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid." *IEEE Transactions on power systems*, **25**(1):371–380, 2009.

[CKC18] Yu-Wei Chung, Behnam Khaki, Chicheng Chu, and Rajit Gadh. "Electric vehicle user behavior prediction using hybrid kernel density estimator." In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2018)*, pp. 1–6, 2018.

[CMA17] Mehmet Hazar Cintuglu, Osama A Mohammed, Kemal Akkaya, and A Selcuk Uluagac. "A survey on smart grid cyber-physical system testbeds." *IEEE Communications Surveys & Tutorials*, **19**(1):446–464, 2017.

[Cri16] M. B. Cristopher. *Pattern recognition and machine learning*. Springer-Verlag, 2016.

[CZC16] Wei Cheng, Kai Zhang, Haifeng Chen, Guofei Jiang, Zhengzhang Chen, and Wei Wang. "Ranking causal anomalies via temporal and dynamical analysis on vanishing correlations." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 805–814, 2016.

[DTB12] Martin Densing, Hal Turton, and Georg Bäuml. "Conditions for the successful deployment of electric vehicles–a global energy system perspective." *Energy*, **47**(1):137–149, 2012.

[Eat16] Eatechnology.com. "My electric avenue data [Online] (2016)." https://www.eatechnology.com/americas/projects/my-electric-avenue/, 2016. (accessed: 03.01.2018).

[EM19] Cornell University Energy Management Control System (EMCS). "Real Time Building Utility Use Data." https://portal.emcs.cornell.edu/d/2/dashboard-list?orgId=2, 2019.

[Eng19] ISO New England. "Pricing Report - Preliminary Real-Time Hourly LMPs." https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/lmps-rt-hourly-prelim, 2019.

[EVA18a] EVAdoption. "EV Charging Statistics." https://evadoption.com/ev-charging-stations-statistics/, 2018.

[EVA18b] EVAdoption. "New Electric Vehicle Sales Market Share for 2016, 2017 and Forecast for 2018 (full year) for Selected Markets." https://evadoption.com/ev-market-share/, 2018.

[FAT18] Yosra Fraiji, Lamia Ben Azzouz, Wassim Trojet, and Leila Azouz Saidane. "Cyber security issues of Internet of electric vehicles." In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6. IEEE, 2018.

[fle19] fleetcarma. "Finding the value of managing your electric vehicle charging load." https://www.fleetcarma.com/finding-the-value-of-managing-your-electric-vehicle-charging-load/, 2019.

[FTC13] A. Foley, B. Tyther, P. Calnan, and B. . Gallachir. "Impacts of electric vehicle charging under electricity market operations." *Appl Energy*, **101**:93–102, 2013.

[Ger17] Janos Gertler. *Fault detection and diagnosis in engineering systems*. Routledge, 2017.

[GPS14] Michele De Gennaro, Elena Paffumi, Harald Scholz, and Giorgio Martini. "GIS-driven analysis of e-mobility in urban areas: An evaluation of the impact on the electric energy grid." *Appl Energy*, **124**:94 – 116, 2014.

[Gre19] Green Car Congress ©2004-2020 BioAge Group, LLC. "US has more than 68,800 electric vehicle charging units." https://www.greencarcongress.com/2019/07/20190709-fotw.html, 2019. (accessed: 01.28.2019).

[GTL13] L. Gan, U. Topcu, and S.H. Low. "Optimal decentralized protocol for electric vehicle charging." *IEEE Trans Power Syst*, **28**(2):940–951, 2013.

[HC16] Christopher G Hoehne and Mikhail V Chester. "Optimizing plug-in electric vehicle and vehicle-to-grid charge scheduling to minimize carbon emissions." *Energy*, **115**:646–657, 2016.

[HHT14] Michael Hayden, Curt Hébert, and Susan Tierney. "Cybersecurity and the North American electric grid: New policy approaches to address an evolving threat." *Bipartisan Policy Center, Electric Grid Cybersecurity Initiative*, 2014.

[HLV03] Wenjie Hu, Yihua Liao, and V Rao Vemuri. "Robust anomaly detection using support vector machines." In *Proceedings of the international conference on machine learning*, pp. 282–289. Citeseer, 2003.

[HNB19] David Hallac, Peter Nystrup, and Stephen Boyd. "Greedy Gaussian segmentation of multivariate time series." *Advances in Data Analysis and Classification*, **13**(3):727–751, 2019.

[HST18] Fouzi Harrou, Ying Sun, Bilal Taghezouit, Ahmed Saidi, and Mohamed-Elkarim Hamlati. "Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches." *Renewable energy*, **116**:22–37, 2018.

[HW14] Chioke B. Harris and Michael E. Webber. "An empirically-validated methodology to simulate electricity demand for electric vehicle charging." *Appl Energy*, **126**:172 – 181, 2014.

[IML16] Vittorio P Illiano, Luis Munoz-Gonzalez, and Emil C Lupu. "Don't fool me!: Detection, characterisation and diagnosis of spoofed and masked events in wireless sensor networks." *IEEE Transactions on Dependable and Secure Computing*, **14**(3):279–293, 2016.

[IRE19] IRENA. "Innovation outlook: Smart charging for electric vehicles." `https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/May/IRENA_Innovation_Outlook_EV_smart_charging_2019.pdf`, 2019.

[Joh18] Tamra Johnson. "Americans spend an average of 17,600 minutes driving each year." `https://newsroom.aaa.com/tag/american-driving-survey/`, 2018. (accessed: 11.10.2018).

[KCC18] Behnam Khaki, Yu-Wei Chung, Chicheng Chu, and Rajit Gadh. "Nonparametric user behavior prediction for distributed EV charging scheduling." In *2018 IEEE Power and Energy Society General Meeting Conf. (PESGM 2018)*, 2018.

[KCG19] Behnam Khaki, Chicheng Chu, and Rajit Gadh. "Hierarchical distributed framework for EV charging scheduling using exchange problem." *Applied energy*, **241**:461–471, 2019.

[KCM11] Trine Krogh Kristoffersen, Karsten Capion, and Peter Meibom. "Optimal charging of electric drive vehicles in a market environment." *Appl Energy*, **88**(5):1940 – 1948, 2011.

[Key16] Ali Keyhani. *Design of Smart Power Grid Renewable Energy Systems, 2nd Edition.* Wiley-IEEE Press, 2016.

[KMM16] Mithat Kisacikoglu, Tony Markel, Andrew Meintz, Jiucai Zhang, and Myungsoo Jun. "EV-Grid Integration EVGI Control and System Implementation." `https://www.nrel.gov/docs/fy16osti/65861.pdf`, 2016.

[Kop15] Ted Koppel. *Lights out: a cyberattack, a nation unprepared, surviving the aftermath*. Broadway Books, 2015.

[KTK02] Christopher Krügel, Thomas Toth, and Engin Kirda. "Service specific anomaly detection for network intrusion detection." In *Proceedings of the 2002 ACM symposium on Applied computing*, pp. 201–208, 2002.

[LBM16] Caroline Le Floch, Francois Belletti, and Scott Moura. "Optimal charging of electric vehicles for load shaping: A dual-splitting framework with explicit convergence bounds." *IEEE Transactions on Transportation Electrification*, **2**(2):190–199, 2016.

[Lee14] Annabelle Lee. "National Electric Sector Cybersecurity Organization Resource (NESCOR)." Technical report, Electric Power Research Institute (EPRI), Incorporated, 2014.

[LGZ18] Lizi Luo, Wei Gu, Suyang Zhou, He Huang, Song Gao, Jun Han, Zhi Wu, and Xiaobo Dou. "Optimal planning of electric vehicle charging stations comprising multi-types of charging facilities." *Applied Energy*, **226**:1087–1099, 2018.

[LL19] Lawrence Livermore National Laboratory (LLNL). "Energy Flow Charts: Charting the Complex Relationships among Energy, Water, and Carbon." `https://flowcharts.llnl.gov`, 2019.

[LLL19] Xianglin Lu, Pengju Liu, and Jiayi Lin. "Network Traffic Anomaly Detection Based on Information Gain and Deep Learning." In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, pp. 11–15, 2019.

[LNZ14] Hong Liu, Huansheng Ning, Yan Zhang, Qingxu Xiong, and Laurence T Yang. "Role-dependent privacy preservation for secure V2G networks in the smart grid." *IEEE Transactions on Information Forensics and Security*, **9**(2):208–220, 2014.

[Lov19] Steven Loveday. "June 2019 Plug-In electric vehicle sales report card." `https://insideevs.com/monthly-plug-in-sales-scorecard/`, 2019. accessed: 07.25.2019.

[LSA10] João A Peças Lopes, Filipe Joel Soares, and Pedro M Rocha Almeida. "Integration of electric vehicles in the electric power system." *Proceedings of the IEEE*, **99**(1):168–183, 2010.

[LSK15] Hae Kyu Lim, Jeong Hun Seo, Suk Hyung Kim, Yoon Cheol Jeon, Jun Seok Choi, and Eun Kyung Kim. "Overcharge prevention device of battery.", September 29 2015. US Patent 9,147,872.

[LWL12] Zhipeng Liu, Fushuan Wen, and Gerard Ledwich. "Optimal planning of electric-vehicle charging stations in distribution systems." *IEEE Transactions on Power Delivery*, **28**(1):102–110, 2012.

[MCH11] Zhongjing Ma, Duncan S Callaway, and Ian A Hiskens. "Decentralized charging control of large populations of plug-in electric vehicles." *IEEE Transactions on control systems technology*, **21**(1):67–78, 2011.

[MH14] Marc Melaina and Michael S Helwig. "California statewide plug-in electric vehicle infrastructure assessment." `https://lib.dr.iastate.edu/imse_reports/1/`, 2014.

[MKB12] Y. Mo, T. H. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. "CyberPhysical Security of a Smart Grid Infrastructure." *Proceedings of the IEEE*, **100**(1):195–209, Jan 2012.

[MQC14] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota. "A novel forecasting algorithm for electric vehicle charging stations." In *2014 International Conference on Connected Vehicles and Expo (ICCVE)*, pp. 1035–1040, 2014.

[MQC15] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R Pota. "Fast prediction for sparse time series: Demand forecast of EV charging stations for cell phone applications." *IEEE Trans. Ind. Informat.*, **140**(1):242–250, Feb. 2015.

[MQC16] Mostafa Majidpour, Charlie Qiu, Peter Chu, Hemanshu R. Pota, and Rajit Gadh. "Forecasting the EV charging load based on customer profile or station measurement?" *Appl Energy*, **163**:134 – 141, 2016.

[MWJ14] Y. Mu, J. Wu, N. Jenkins, H. Jia, and C. Wang. "A spatialtemporal model for grid impact analysis of plug-in electric vehicles." *Appl Energy*, **114**:456–465, 2014.

[MXD13] Leszek Mazur, Jianhui Xie, Sean D Daniel, and Cesare John Saretto. "Authenticating using cloud authentication.", November 12 2013. US Patent 8,584,221.

[Nat10] National Institute of Standards and Technology (NIST). "Guidelines for Smart Grid Cyber Security." Technical report, U. S. Department of Commerce, 2010.

[NCP17] Hamidreza Nazaripouya, Chi-Cheng Chu, Hemanshu Roy Pota, and Rajit Gadh. "Battery energy storage system control for intermittency smoothing using an optimized two-stage filter." *IEEE Transactions on Sustainable Energy*, **9**(2):664–675, 2017.

[NE12] North American Electric Reliability Corporation (NERC). "Cyber Attack Task Force - Final Report." `https://www.yumpu.com/en/document/view/11702317/cyber-attack-task-force-final-report-nerc`, 2012.

[Off12] Office of Governor Edmund G. Brown Jr. "EXECUTIVE ORDER B-16-2012." `https://www.ca.gov/archive/gov39/2012/03/23/news17472/index.html`, 2012.

[Ort18] Mario Roberto Durn Ortiz. "Top-selling light-duty plug-in electric vehicle global markets by country or region as of December 2018." https://commons.wikimedia.org/w/index.php?curid=66102727, 2018.

[PVG11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, **12**:2825–2830, 2011.

[QLS18] Jiahu Qin, Menglin Li, Ling Shi, and Xinghuo Yu. "Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks." *IEEE Transactions on Automatic Control*, **63**(6):1648–1663, 2018.

[Rah17] Robbi Rahim. "Man-in-the-middle-attack prevention using interlock protocol method." *ARPN J. Eng. Appl. Sci*, **12**(22):6483–6487, 2017.

[RR06] Venkatesh Rajagopalan and Asok Ray. "Symbolic time series analysis via wavelet-based partitioning." *Signal processing*, **86**(11):3309–3320, 2006.

[SBC04] Greg C Stone, Edward A Boulter, Ian Culbert, and Hussein Dhirani. *Electrical insulation for rotating machines: design, evaluation, aging, testing, and repair*, volume 21. John Wiley & Sons, 2004.

[SHM10] Eric Sortomme, Mohammad M Hindi, SD James MacPherson, and SS Venkata. "Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses." *IEEE transactions on smart grid*, **2**(1):198–205, 2010.

[SIF15] Florian Salah, Jens P. Ilg, Christoph M. Flath, Hauke Basse, and Clemens van Dinther. "Impact of electric vehicles on distribution substations: A Swiss case study." *Appl Energy*, **137**:88 – 96, 2015.

[Sil86] B. W. Silverman. *Density estimation for statistics and data analysis*. CRC press, 1986.

[SL19] P Slowik and N Lutsey. "The surge of electric vehicles in United States cities." https://theicct.org/publications/surge-EVs-US-cities-2019, 2019.

[SLC15] Wenbo Shi, Na Li, Chi-Cheng Chu, and Rajit Gadh. "Real-time energy management in microgrids." *IEEE Transactions on Smart Grid*, **8**(1):228–238, 2015.

[Sma] Smart Grid Energy Research Center (SMERC), UCLA. "Smart grid project - smart EV charging station." https://evsmartplug.net/smartgrid/ChargingRecord/. (accessed: 03.01.2018).

[Sol19] Solar Energy Industries Association (SEIA). "California Solar." https://www.seia.org/state-solar-policy/california-solar, 2019.

[Sol20] Solar Energy Industries Association (SEIA). "Solar Industry Research Data." https://www.seia.org/solar-industry-research-data, 2020.

[SS03] Alex J. Smola and Bernhard Schlkopf. "A tutorial on support vector regression." Technical report, Statistics and Computing, 2003.

[Sto18] Bartz Stockmar. "Staying big or getting smaller." `https://commons.wikimedia.org/wiki/File:Staying_big_or_getting_smaller.jpg`, 2018.

[SWS00] Bernhard Schölkopf, Robert C Williamson, Alex J Smola, John Shawe-Taylor, and John C Platt. "Support vector method for novelty detection." In *Advances in neural information processing systems*, pp. 582–588, 2000.

[SXC14] Wenbo Shi, Xiaorong Xie, Chi-Cheng Chu, and Rajit Gadh. "Distributed optimal energy management in microgrids." *IEEE Transactions on Smart Grid*, **6**(3):1137–1146, 2014.

[TMN17] Sajad Tabatabaee, Seyed Saeedallah Mortazavi, and Taher Niknam. "Stochastic scheduling of local distribution systems considering high penetration of plug-in electric vehicles and renewable energy sources." *Energy*, **121**:480–490, 2017.

[UJW17] Kotub Uddin, Tim Jackson, Widanalage D Widanage, Gael Chouchelamane, Paul A Jennings, and James Marco. "On the possibility of extending the lifetime of lithium-ion batteries through optimal V2G facilitated by an integrated vehicle and smart-grid system." *Energy*, **133**:710–722, 2017.

[US19] U.S. Energy Information Administration (EIA). "U.S. energy facts explained." `https://www.eia.gov/energyexplained/us-energy-facts/`, 2019.

[Vap95] V. Vapnik. *The nature of statistical learning theory*. Springer, 1995.

[Vel19] Veloz. "Sales Dashboard." `https://www.veloz.org/sales-dashboard/`, 2019. accessed: 07.25.2019.

[WHQ15] B. Wang, B. Hu, C. Qiu, P. Chu, and R. Gadh. "EV charging algorithm implementation with user price preference." In *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2015.

[WRM17] Eric W. Wood, Clement L. Rames, Matteo Muratori, Seshadri Srinivasa Raghavan, and Marc W. Melaina. "National Plug-In Electric Vehicle Infrastructure Analysis." Technical report, National Renewable Energy Lab. (NREL), Golden, CO (United States), 2017.

[WRW16] B. Wang, R.Huang, Y. Wang, H. Nazaripouya, C. Qiu, C. Chu, and R. Gadh. "Predictive scheduling for Electric Vehicles considering uncertainty of load and user behaviors." In *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D 2016)*, pp. 1–5, 2016.

[WSW17] Yubo Wang, Wenbo Shi, Bin Wang, Chi-Cheng Chu, and Rajit Gadh. "Optimal operation of stationary and mobile batteries in distribution grids." *Appl Energy*, **190**:1289 – 1301, 2017.

[WWN17] B. Wang, Y. Wang, H. Nazaripouya, C. Qiu, C. Chu, and R. Gadh. "Predictive scheduling framework for electric vehicles with uncertainties of user behaviors." *IEEE Internet of Things Journal*, **4**(1):52–63, Feb. 2017.

[WZO15] Hewu Wang, Xiaobin Zhang, and Minggao Ouyang. "Energy consumption of electric vehicles based on real-world driving patterns: A case study of Beijing." *Appl Energy*, **157**:710 – 719, 2015.

[XHS14] Zhiwei Xu, Zechun Hu, Yonghua Song, Wei Zhao, and Yongwang Zhang. "Coordination of PEVs charging across multiple aggregators." *Appl Energy*, **136**:582 – 589, 2014.

[XMC16] Erotokritos Xydas, Charalampos Marmaras, Liana M Cipcigan, Nick Jenkins, Steve Carroll, and Myles Barker. "A data-driven approach for characterising the charging demand of electric vehicles: A UK case study." *Applied energy*, **162**:763–771, 2016.

[XWC18] Yingqi Xiong, Bin Wang, Chi cheng Chu, and Rajit Gadh. "Vehicle grid integration for demand response with mixture user model and decentralized optimization." *Appl Energy*, **231**:481 – 493, 2018.

[YYL11] Zhenyu Yang, Shucheng Yu, Wenjing Lou, and Cong Liu. "P^2: Privacy-preserving communication and precise reward architecture for V2G networks in smart grid." *IEEE Transactions on Smart Grid*, **2**(4):697–706, 2011.

[ZL12] Dimitrios Zissis and Dimitrios Lekkas. "Addressing cloud computing security issues." *Future Generation computer systems*, **28**(3):583–592, 2012.