A NOVEL REINFORCEMENT LEARNING-OPTIMIZATION APPROACH FOR INTEGRATING WIND ENERGY TO POWER SYSTEM WITH VEHICLE-TO-GRID TECHNOLOGY

A Dissertation

by

POUYA SHARIFI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Amarnath Banerjee
Committee Members,	Lewis Ntaimo
	Natarajan Gautam
	Bala Shetty
Head of Department,	Lewis Ntaimo

December 2020

Major Subject: Industrial Engineering

ABSTRACT

High integration of intermittent renewable energy sources (RES), specifically wind power, has created complexities in power system operations due to their limited controllability and predictability. In addition, large fleets of Electric Vehicles (EVs) are expected to have a large impact on electricity consumption, contributing to the volatility. In this study, a well-coordinated smart charging approach is developed that utilizes the flexibility of EV owners in a way where EVs are used as distributed energy storage units and flexible loads to absorb the fluctuations in the wind power output in a vehicle-to-grid (V2G) setup. Challenges for people participation in V2G, such as battery degradation and insecurity about unexpected trips, are also addressed by using an interactive mechanism in smart grid.

First, a static deterministic model is formulated using multi-objective mixed-integer quadratic programming (MIQP) assuming known parameters day ahead of time. Subsequently, a formulation for real-time dynamic schedule is provided using a rolling-horizon with expected value approximation. Simulation experiments demonstrate a significant increase in wind utilization and reduction in charging cost and battery degradation compared to an uncontrolled charging scenario.

Formulating the scheduling problem of the EV-wind integrated power system using conventional stochastic programming (SP) approaches is challenging due to the presence of many uncertain parameters with unknown underlying distributions, such as wind, price, and different commuting patterns of EV owners. To alleviate the problem, a model-free Reinforcement Learning (RL) algorithm integrated with deterministic optimization is proposed that can be applied on many multi-stage stochastic problems while mitigating some of the challenges of conventional SP methods (e.g., large scenario tree, computational complexity) as well as the challenges in model-free RL (e.g., slow convergence, unstable learning in dynamic environment). The simulation results of applying the combined approach on the EV scheduling problem demonstrate the effectiveness of the RL-Optimization method in solving the multi-stage EV charge/discharge scheduling problem. The proposed methods perform better than standard RL approaches (e.g., DDQN) in terms of

convergence speed and finding the global optima. Moreover, to address the curse of dimensionality issue in RL with large action-state space, a heuristic EV fleet charging/discharging scheme is used combined with RL-optimization approach to solve the EV scheduling problem for a large number of EVs.

DEDICATION

I would like to dedicate this work to my parents and my brothers whose dreams and supports for me have resulted in this achievement. Without their loving upbringing and nurturing, I would not have been where I am today and what I am today. Had it not been for my parents' unflinching insistence and support, my dreams of excelling in education would have remained mere dreams.

ACKNOWLEDGMENTS

In my journey towards this degree, I have found a teacher, an inspiration, a role model and a pillar of support in my guide, Prof. Amarnath Banerjee. He has been there providing his heartfelt support and guidance at all times and has given me invaluable guidance, inspiration and suggestions in my quest for knowledge. He has given me all the freedom to pursue my research, while silently and non-obtrusively ensuring that I stay on course and do not deviate from the core of my research. Without his able guidance, compiling this dissertation would not have been possible and I shall be eternally grateful to him for his assistance.

I would like to thank Prof. Ntaimo, prof. Gautam, and Dr. Shetty for agreeing to serve on my committee. And a special thanks to my friend Ellie Solhjou Khah and Dr. Mohammad Javad Feizollahi for providing help to find the topic of my research.

Finally, I would like to acknowledge and thank the Department of Industrial & Systems Engineering at Texas AM University for providing all the necessary resources for me to conduct my research and providing any assistance requested.

NOMENCLATURE

RL	Reinfrocement Learning
SP	Stochastic Programming
$MIQP$	Mixed Integer Quadratic Program
MG	Micro-Grid
T	Set of time/periods in scheduling with index t.
I	Set of all EVs with index i.
$I_{g2v}(I_{v2g})$	Set of vehicles participating in G2V (V2G).
B	Set of vehicles that arrive with charge level below SOC_{min}.
EV^t_{all}	Set of all vehicles plugged-in during time slot $[t, t+1]$.
EV^t_{v2g}	Set of vehicles in V2G plugged-in during time slot $[t, t+1]$.
$t_i^{arr}(t_i^{dep})$	Arrival (departure) time of EV i.
T_i^p	Plug-in period of EV i.
$pr^t(pr_f^t)$	(Forecasted) Electricity price in ¢/kWh.
D^t	Total charging demand in time slot $[t, t+1]$.
$W^t(W_f^t)$	Actual (forecasted) wind production in time slot $[t, t+1]$ in kWh.
$P_i^c(P_i^d)$	Maximum energy EV i can take (discharge) in Δt (kWh).
η_i^c, η_i^d	Charging and discharging efficiency of EV i.
SOC_i^t	State of battery charge for EV i at time t.
$SOC_{init,i}$	Initial state of battery for EV i in kWh.
$SOC_{cap,i}$	Battery capacity of EV i in kWh.
$SOC_{min,i}$	Minimum level of battery charge for EV i in kWh.
$SOC_{desired,i}$	Desired level of battery charge for EV i in kWh.

$T_{min,i}$	Minimum number of periods to reach $SOC_{min,i}$.
Ψ_i	The battery degradation cost for EV i in ¢.
δ	Penalty for energy curtailment in ¢/kWh.
λ	Owners' level of tolerance for battery degradation in [0, 1].
P_G^{max}	Maximum transmission power between microgrid and the external grid in Δt period.
$X_{c,i}^t$	Charging rate for EV i in time slot $[t, t+1]$.
$X_{d,i}^t$	Discharging rate for EV i in time slot $[t, t+1]$.
$Y_{c,i}^t$	Binary variable, 1 if EV i is charging in interval $[t, t+1]$.
$Y_{d,i}^t$	Binary variable, 1 if EV i is discharging in interval $[t, t+1]$.
Ω^t	Wind curtailment in time slot $[t, t+1]$ in kWh.
G^t	Energy supplied from the external grid in time slot $[t, t+1]$.
J	Set of planning times in dynamic modeling with index j.
Δj	Planning intervals.
E^j	Set of all vehicles to be planned at planning time j.
E_{v2g}^j	Set of vehicles in V2G mode planned at planning time j.
E_{g2v}^j	Set of vehicles in G2V mode planned at planning time j.
E_i^{req}	Remaining required energy for EV i.
T_i^{rem}	Remaining parking periods for EV i.
τ_{max}^j	End of the planning (rolling) window planned at time j.
$LC_i^{\{j\}}(LD_i^{\{j\}})$	The (dis)charging rate in the period previous to time j.
R_{v2g}	Ratio of vehicles participating in V2G.
D_f^t	Charge demand for unknown future arrivals in time slot $[t, t+1]$.
$\hat{N}_j^t$	Estimated number of vehicles arrived after time j and plugged-in at time t.
$P_{a,b}^k$	The transition probability from wind state a to b in k time steps.
$\mathcal{S}$	State Space in MDP formulation.

$\mathcal{A}$	Action Space in MDP formulation.
$\mathcal{D}$	Memory buffer.
$\mathcal{F}$	Minibatch of transitions.
N_{max}	Maximum capacity of the parking lot.
$N_{ch,t}$	Number of EVs in charge mode at time t.
$N_{dc,t}$	Number of EVs in discharge mode at time t.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND OVERVIEW

In recent years, there has been a significant interest in integrating renewable energy sources (RES) into the power system motivated by their positive environmental impact, low carbon emission, and low production costs. However, high integration of these resources creates complexities in the operations of the power system due to the intermittent nature of these energy sources and could lead to energy curtailments. For instance, wind and solar energy account for more than 80% of total renewable installation in California as of February 2020 (1), and the level of wind and solar energy curtailments in February 2020 was more than 100,000 MWh.

Among all renewable resources, the focus of this dissertation is on wind power integration as wind energy is of special interest due to its recent improvements in technology, and its low generation cost. Wind power is intermittent in nature and has limited controllability and predictability, and hence, it poses challenges to the quality of power, reliability of the system and the power system scheduling. Without a well-defined control process, wind power may need to be curtailed to balance the supply and demand.

On the other hand, an increasing number of electric vehicles (EVs) are coming on the roads every year. Although the EVs have no tailpipe emissions, the electricity production has emissions if it is from conventional power generators. High penetration of EVs also creates challenges in power system operations as the peak-demand increases, and the EV charging demand and driver behaviors are unpredictable.

One approach to accommodate the high penetration of RES and to alleviate the unpredictability of EV charging demand is vehicle-to-grid (V2G) technology. V2G is classified into unidirectional and bidirectional V2G. Unidirectional V2G, which is also referred as Grid-to-Vehicle (G2V), refers to a situation when the EV can only charge from the electrical grid. In bidirectional V2G, the EV battery can not only charge but also discharge energy back to the grid. Advances in smart grids have enabled EV to be a good potential as a distributed energy storage unit and as a flexible dynamic load that can be managed by system operators. When connected to a charging station (plugged-

in mode), EV can charge power from the grid as well as send power back to the grid whenever needed. If the charging process is not controlled, it only adds an uncertain load element into the power system operation, but given the recent advances in smart grid and grid technologies, the question is how can the EVs potentially benefit their owners and the electrical grid together.

EV penetration is growing exponentially as it is projected that EVs are going to account for 35% of all new vehicle sales in the US by the year 2040. Vehicles are parked 95% of the time, but only require a few hours to recharge, so the remaining time can be used as the available flexibility to support renewable energies. For instance, if vehicles are parked during nighttime or at a workplace for 7-8 hours, the charging process can be shifted to the time when renewable production is high or when the demand (or electricity price) is low.

This dissertation focuses on real-time charge/discharge scheduling of the EVs to support the integration of wind energy in the power system. This problem is studied in the context of an aggregator who decides for all the EVs in the fleet. The aggregator receives information about available renewable energy, future wind and price forecasts, and the data related to the status of the EVs, and their owners' preferences. Then, the aggregator runs a scheduling algorithm and determines when to charge or discharge the EVs. The scheduling problem is modeled as a multi-objective optimization problem to minimize the wind curtailment as well as the charging cost for the owners while satisfying the owners' charging requirements. People participation in V2G has been reported to be limited as it is perceived that allowing system operators to modify the charging rate and frequent cycling of the battery will significantly reduce the lifespan of the battery. Another major impediment in participation is that the owners feel insecure for urgent needs when the battery might not have the desired energy level to reach the destination. Not considering the battery degradation cost and user discomfort will adversely impact the owners decision to participate in V2G. Motivated by these factors, the challenges of people participation in V2G are studied and addressed. The day-ahead scheduling problem is formulated as a deterministic problem with known EV demand, wind generation, and price signals. Then, the real-time scheduling problem is solved using a dynamic scheduling approach with mean-value optimization and rolling horizon.

In the dynamic model, the aggregator schedules the EV fleet charge/discharge strategy in real-time based on the availability of EVs as well as the wind and price forecasts in the future.

In the presence of uncertainty, this dynamic scheduling is a classic multi-stage stochastic problem (SP). However, conventional SP approaches may not be suitable for this problem due to the presence of large number of random variables. The other main challenge is to construct accurate models for the random variables and the system model. Unlike SP approaches that are model based, Reinforcement learning (RL), is potentially a model-free algorithm that does not require models for uncertain parameters and can be applied in complex stochastic environments. The RL agent learns the model dynamic explicitly through interacting with the environment in a trial-and-error fashion. Some RL methods are also capable of learning off-line from experiences. Thus, rather than finding a sufficiently accurate distributions for the uncertain elements, a simulation of the environment from historical data should provide the mechanism for the agent to learn the optimal strategy (i.e., policy). However, in large-scale dynamic environments with large state-action spaces, RL approaches suffer not only from slow convergence, but also they may approach a local optima rather than a global optimal solution.

With the curse of dimensionality of multi-stage SP approaches as well as the slow convergence and local-optimal trapping of RL methods, there is a need for a better approach to solve the scheduling problem. This motivated the development of an integrated RL-optimization algorithm that mitigates the slow convergence issue and increases the chance of reaching the global optima while learning from past experiences in a modified model-free RL approach.

In this dissertation, a comprehensive literature review on recent approaches to accommodate high integration of wind power generation is provided in Chapter 2. Next, Chapter 3 focuses on EV charge/discharge scheduling for the deterministic case. First, the case in which the wind profile and electricity price is determined ahead of schedule are explored. This problem is divided into two separate problems. In the first problem, the EV related information, such as arrival and departure time, the initial state of the battery, and battery capacity are assumed to be known. This is referred to as the deterministic static problem. In the deterministic static model, the EV owners provide

the arrival time, departure time, and the desired level of battery charge prior to the scheduling day. Then the aggregator determines the charge/discharge schedule for the planning day. In the second problem, a more realistic model referred to as the deterministic dynamic model is studied. In this model, the EV owners input their departure time and the desired level upon arrival. The aggregator batches all the vehicles that have arrived during a period t and plan for their charging schedule at the end of the period. With updated information regarding the arriving EVs, the grid electricity price, and the renewable energy generation, the aggregator runs the scheduling algorithm for the current set of EVs and assigns appropriate charging time to the EVs. For all the cases, the performance of the models for both unidirectional and bidirectional V2G are examined.

Chapter 4 considers the uncertainty of all the stochastic elements. First, a rolling-horizon mean-value optimization algorithm is proposed for the charge/discharge schedule. Then, a model-free RL algorithm combined with the deterministic optimization is proposed to formulate the charge/discharge scheduling of a small number of EVs without requiring any prior system model information. Finally, a heuristic scheme is used in combination with the RL-optimization method to solve the scheduling problem for a large number of EVs.

In summary, the remaining chapters of this dissertation are as follows. In chapter 2, a review of literature on the effects of high wind integration in the power system, and recent approaches used to accommodate these effects are provided. Chapter 3 focuses on the optimization of EV charge/discharge schedule in the deterministic case. In chapter 4, we expand the optimization problem to the stochastic case in which wind generation, drivers' behavior, and electricity price are unknown stochastic parameters. Figure 1.1 shows the hierarchy of the models that are investigated in this dissertation.

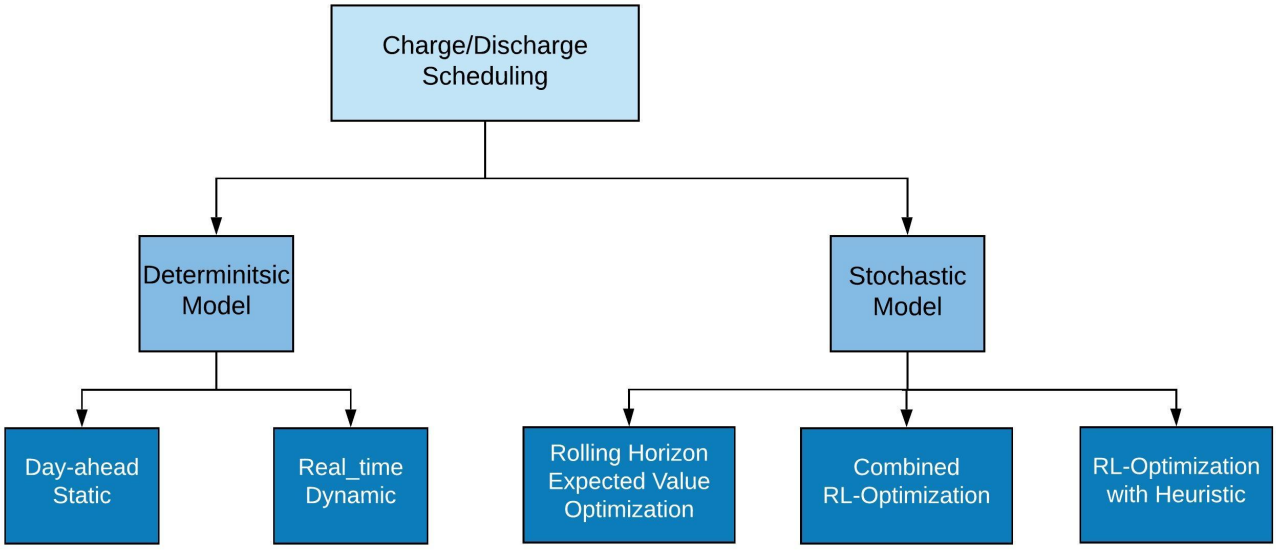


Figure 1.1: Schema hierarchy

2. HIGH WIND POWER INTEGRATION

2.1 Introduction

High penetration of renewable energy sources (RES), predominantly wind energy, into the grid has created complexity in the operation of power systems due to the intermittent nature of these sources. A large number of research studies has been conducted to accommodate the integration of wind energy in power systems, including the incorporation of energy storage devices, demand response tariffs, smart charging with Vehicle-to-Grid(V2G) technology, and stochastic unit commitment. This chapter provides a review of the effects caused by uncertainty and unpredictability of wind energy on power systems, and recent approaches to mitigate these effects. We also review the different storage systems incorporated with wind energy into the power system, and discuss the advantages and limitations of each system. Moreover, different demand response programs and their benefits and barriers are explored. A brief review of V2G technology for RES integration support is also provided. Finally, various optimization methods developed for optimal scheduling under wind power uncertainty are reviewed.

With the ever increasing concern over global climate change and the need for environmental protection, renewable sources that are naturally replenished, such as power from wind energy and solar, have been given a significant amount of consideration owing to their zero greenhouse gas emissions (2; 3). Among all renewable resources, power from wind energy, the main focus of this chapter, is of special interest due to its recent improvements in technology, and its low generation cost (4). The United States aims to provide 20 percent of the electricity supply from wind generation by the year 2030 (5). However, high penetration of wind energy presents challenges to the power systems operations. Wind generation is inherently difficult to predict and causes problems to maintain system reliability (6). The unpredictability of renewable energy generations can be translated into more production volatility and an increase in the reserves (7). In the literature, various approaches have been proposed to accommodate the high integration of wind

energy, including the incorporation of energy storage devices, demand response programs (DRP), vehicle-to-grid technology, improving wind forecasting, and stochastic unit commitment (UC) (8). These approaches play an important role in providing support in integrating wind energy into the grid.

Wind power output is not fully controllable, and thus, it poses challenges to the quality of power, reliability of the system and the scheduling of a power system. Without any control process, in order to balance the supply and demand, wind power may need to be curtailed, meaning system operators may have to dispatch electricity power less than what is generated from wind turbines. One way to control the wind output is to install Energy Storage Systems (ESS) (9). The main function of ESS is to mitigate the intermittency of wind power output by storing or discharging electricity when there is an excess or shortfall of the actual output compared to the forecast wind power generation and load, respectively (10). By controlling wind power output, ESS can provide greater support in the penetration of wind generation into the power system. Ignoring the technical aspects of different storage systems, this chapter attempts to review the literature in finding the appropriate ESSs incorporated into the wind integrated grid. Technical features of various storage systems are discussed in (11; 12).

Another efficient approach in mitigating wind power uncertainty is to apply demand response (DR), which is also known as load management. DR can help balance power generation and load by managing end-use consumers' electricity consumption via price-based and incentive-based programs benefiting both retail customers and the market (13). When wind power output is higher than the forecast, DR can help consume excessive wind energy by lowering the price or by offering incentives. On the other hand, it can reduce the peak-load by increasing the price. Some papers such as (6; 14), have found that DR can accommodate wind power output uncertainty by reducing the amount of reserves.

Unit Commitment (UC) is the process of finding the optimal electricity generation schedule while satisfying a set of operating constraints. With respect to uncertainty in model parameters, UC is categorized into deterministic UC and stochastic UC models (15). Traditionally, the UC

uncertainty originates from the load forecast error. Failure in generators, transmission lines, and other devices are other sources of UC uncertainty, which are considered in security constrained UC (SCUC) models. In deterministic UC, load forecasts are treated as a known parameter defined by single point estimates. To manage uncertainty, recent UC optimization techniques tend to seek the optimal solution by taking into account the stochastic nature of wind power. With the increased level of wind farm integration, the necessity for stochastic UC has grown even stronger due to the high generation forecasting errors. Various optimization approaches have been developed to solve UC problems, including stochastic programming (SP), robust optimization (RO), chanced-constrained optimization, and stochastic dynamic programming (SDP). In this chapter, we focus only on mathematical formulations and numerical optimization algorithms. The reader is referred to (16) for other UC approaches in literature including numerical and heuristic methods. The recent optimization approaches to mitigate the negative effects of wind power forecasting errors (WPFE) in power system scheduling are provided.

The remainder of this chapter is organized as follows. First, a review of the impact of high wind penetration on system reliability and operations is provided in Section 2.2. In Section 2.3, various energy storage devices incorporated with RES are explored. Section 2.4 demonstrates the concepts behind a demand response program and how it mitigates the negative effects of wind power. In Section 2.5, vehicle-to-grid technology for RES support is reviewed. In section 2.6, different optimization methods for UC under uncertainty are reviewed. Finally, the summary of chapter 2 is provided in Section 2.7.

2.2 Effects of Wind Energy on Power System Operations

Identifying and analyzing the impacts of wind power on system operations is the first step to increase its level of penetration. The effects of wind power depend on two factors: penetration level and system flexibility. With the increased level of penetration, the impact on power systems operation will continue to increase. On the other hand, highly flexible systems tend to accommodate wind energy more easily (17). We elaborate on these two factors below.

2.2.1 Effects of High Wind-Integrated Grid

In a wind power integrated grid, system operators and planners encounter a new set of challenges due to the high volatility of wind power. Wind power fluctuates based on wind speed, which varies significantly both spatially and temporally. In other words, wind speed fluctuates depending on the geographical location, and also at a specific location, the amount of wind varies over time (18). Thus, forecasting the wind power generation is challenging and extensive research has been done on this topic (19; 20; 21; 22). Although significant advances have been made in forecasting wind power generation, there still exists a serious uncertainty with its forecast. In (23), the authors analyze the effects of different sources on the uncertainty of wind power forecast. These uncertainty sources include weather conditions, power curve, input data, and various prediction algorithms. It is concluded that probabilistic wind power forecasting compared to single-point estimate improves the solution quality of dispatch optimization problem in terms of reducing the operations cost and the reserve capacity; hence, improving the integration of wind power in the grid.

General impact of intermittent wind power on other conventional generators' efficiency, system reliability, transmission outages, voltage, and reactive powers are discussed in (18). Even though wind farms have low generation cost compared to the other fuel power plants, its impact on power systems and other conventional generators may cause increase in the operations cost at the system level (24). Fluctuating wind power causes the other conventional generators to perform in a sub optimal manner. A conventional power plant can adjust the power output to follow the demand, known as load following power plant. However, since the energy generated from RES does not follow the load and has a higher priority, conventional generators need to adjust their output based on both load and RES generation, and thus, will result in sub-optimal solutions for economic dispatch scheduling. Wind power may also cause transmission and distribution losses, increase in the amount of reserves, and discarded energy that the system cannot absorb (17). High penetration of wind energy requires expansion of more complex transmission networks resulting in more transmission losses. Furthermore, since wind farms are usually located in remote areas

far from cities, long-distance transmission lines are required. The long distance between the wind sites and the load can result in transmission congestion as well (10; 18).

One of the main issues that system operators are facing is the provision of system reserves. Several research studies, such as in (25; 26; 27), confirm that as wind power integration increases, the amount of reserve needs to be increased over longer periods of time in order to maintain system reliability. However, in (25), the authors propose that with fast acting reserves, the impact of wind power over short time frames is trivial due to the small variation of wind power in a short time scale.

Previous studies have concluded that voltage stability is more jeopardized in systems with wind power integration (28; 29). Also, with high integration of wind energy, maintaining frequency stability of power systems is also a challenge. This is due to the fact that maintaining the balance of supply and load demand is more difficult in case of high volatility of supply, and hence, leads to frequency deviation from the normal range (30). Moreover, increasing penetration of wind power leads to utilizing fewer number of flexible conventional generators, which are used to adjust the frequency of the power system (31).

Wind farms aggregation has also been analyzed by many researchers. In an international collaboration study (32), it was concluded that geographically-spread wind farms could result in lower forecasting errors. The study in (18) also confirms that increased number of wind turbines in a farm and spatial spread of wind farms can help reduce the volatility of wind power. In (33), the authors present a method for an optimal allocation of wind farms in a given region that reduces the wind power fluctuations significantly.

Traditionally, with all electricity supply from thermal units, or in other words, with low penetration of wind farms into the grid, the variation mainly results from load forecast errors. Some researchers have developed methods to solve unit commitment under the uncertainty of load (34; 35). However, with wind power integration, a significant stochastic element of power generation is added to the system. Thus, finding an optimal schedule under uncertainty is more challenging. Numerous attempts have been devoted to solve unit commitment considering wind power uncertainty, which

is further discussed in Section 2.6.

2.2.2 System Flexibility

A general approach to accommodate the impacts of wind energy on system operations is to build a flexible system. The flexibility of a system can be enhanced through generation flexibility (36), incorporating demand response programs (37), usage of storage systems (38), exploiting vehicle-to-grid (V2G) technology and system scheduling for near real-time planning horizons (32). Flexible or fast-responding generators can follow sudden changes in demand and wind supply, thus mitigating the negative effects of wind and load volatility. Storage devices are used to mitigate the fluctuation in supply by charging excessive power that would be curtailed in the absence of storage, and discharge the power back to the grid when needed. Demand Response, on the other hand, aims to change the consumer electricity usage to smooth out the load fluctuation. With a large fleet of electric vehicles (EV), V2G technology enables the system operators to exploit the vehicles as storage devices as well as flexible loads that can smooth out the electricity demand. Storage system, demand response, and V2G are discussed in the following sections. Moreover, the study in (32) has shown that the system operation with intra-day or intra-hour planning horizon helps to reduce the effects of forecast errors substantially compared to day-ahead planning.

2.3 Storage Systems

Energy Storage Systems (ESS) are considered a critical factor to accommodate variation of wind power generation by controlling the power generators output, and thus, empowering the integration of wind energy into the grid. The capability of ESS to smooth out wind farm output results from its ability to charge and discharge electricity when power supply differs from the electricity demand (10).

Both technical and economic benefits of energy storage have been studied extensively (15; 39; 40). The study in (15) discusses the potential benefits of bulk storage by considering an ideal storage system, in which there are no energy losses, no up and down ramps, and there is instant power charge/discharge. Ideal storage units can lead to a significant integration of RES, reduction

in power curtailment, improved operation of the conventional generators by reducing the number of shut-downs and start-ups, and improving the system reliability by providing real-time support to balance the supply and demand. By considering more realistic features of ESS, the study in (39) lists the benefits of ESSs and their desired characteristics. The paper also addresses how different set of characteristics are used to handle the issues imposed by the volatile generation of RES, such as, power quality, unit commitment, and load following. ESS can also provide economic benefits by reducing the system reserves and mitigating power failure costs associated with electricity outages and short-term fluctuations (40).

Despite the apparent advantages of storage technologies, there is a need to improve energy storage technologies to overcome their barriers and limitations. The primary barrier is the actual initial cost per kW of storage devices. For instance, in 2016, the installation cost of compressed-air energy storage (CAES), Lithium titanate, and other Li-ion batteries was estimated at 53 USD/kWh, 473-1,260 USD/kWh, and 200-840 USD/kWh, respectively (41). Besides the high capital cost, there are other barriers to the broad implementation of storage units in the grid. The study in (39) discusses external factors, such as mineral availability and geographical restrictions that may affect high deployment of some storage technologies. A 2013 Sandia National Laboratories report (42) discusses market and policy barriers to widespread usage of storage technologies.

Various energy storage technologies are used in the wind integrated power system. Pumped hydro energy storage (PHES) (43) and CAES (44) are the two most widely implemented technologies for large-scale electricity storage. The former stores energy in the form of water by pumping water between upper and lower reservoirs, while the latter stores energy as compressed air. There are many other storage technologies, including batteries (45), flow battery (46; 47), superconducting magnetic energy storage (SMES), flywheel energy storage (FES) (48; 49), and supercapacitors (50; 51). Each of these energy storage technologies has its own characteristics. Some of the main features of ESSs, which are the critical factors in determining proper storage technologies for wind integration, are described below.

Energy Efficiency or round-trip energy efficiency is the ratio of energy retrieved (in MWh)

from a storage device to the energy stored (MWh). *Response time* is the time needed for the storage device to provide the power output. *Power density (W/L)* is the amount of power per unit volume. *Energy storage capacity (kWh)* is the maximum amount of energy that can be stored by a device. *Charge/discharge duration* is the time needed to fully charge/discharge.

The characteristics of storage technologies have been investigated for wind power integration support (9; 39; 40; 52). In these works, the authors reviewed the state-of-the-art ESS technologies and discussed different applications and characteristics in selecting ESS technology for a wind-connected grid. The study in (53) presents a comparative economic analysis of different technologies. It is reported that CAES and PHES can provide relatively low energy costs due to their cheap storage media, while SMES and FES pose high energy costs. However, PHES and CAES technologies cannot be widely used due to their geographical limitations (6). In a similar study (54), the authors provide a technical comparison between CAES, SMES, FES, and hydrogen-based energy storage systems (HESS), and their impact on the grid are analyzed separately. It is concluded that all these storage devices are capable of smoothing out the wind power output. However, FES and SMES are designed to improve the power quality due to their fast response and low discharge duration, while CAES has higher storage capabilities compared to others. The study in (11), presents a comparison between PHES, different types of batteries and fuel cells considering different factors in selecting ESS for RES integration such as economic viability, efficiency, and life span. It is reported that PHES is a fully matured technology while it is highly limited to the areas with available hydro resources. It also has the largest lifetime close to 50 years compared to the fuel cells and batteries, which have lifespans ranging from 5 to 15 years. Moreover, fuel cells pose the highest investment cost while having low efficiency, which makes it economically unsuitable for wind power application.

There is a large body of literature trying to study the potential applications of ESSs in wind power integrated grid (9; 53; 55). Load following, power quality, spinning reserves, peak shaving, and voltage control are among the wind power related applications. *Load following* is simply regarded as adjusting electricity generation in order to meet the actual demand throughout the

day. *Power quality* refers to specific features such as sags, spikes, voltage/frequency of electricity supply. *Peak shaving* is the capability to store electricity in off-peak demand (overnight) and provide power during on-peak demand. *Voltage support* provided by storage devices is vital, since a sudden reduction in a wind farm output causes a substantial deviation from normal designed frequency/voltage of an equipment, which can result in equipment damage. Different energy storage technologies perform differently in each of these applications; for instance, HESS is well suited in the load following applications. SMES, FES, batteries, and super-capacitors enhance the power quality supplied into the system by charging the excessive power generated by wind turbines and discharging at low wind speed (40). FES, SMES, and batteries are suitable for frequency/voltage support owing to their fast response to the unpredictable changes in wind speed (53).

Table 2.1 (55) summarizes a comprehensive comparison between different storage technologies, their main characteristics and applications. As seen in Table 2.1, and as a common conclusion among papers cited here, there is no single technology that outperforms others in all applications and meets all the requirements. However, technologies with high power density, fast response and high energy efficiency, such as L/A batteries, SMES, and FES have better potential in dealing with volatile wind power output; while, systems like PHSS and CAES posses low cost. Hence, it is strongly believed that further developments are needed in storage systems to achieve systems with low costs and high efficiency.

The optimal location of storage devices are studied by several researchers (56; 57; 58). In these works, optimal ESS site is obtained with the objective of minimizing power losses and generation costs while considering the randomness of wind power output. In another thread of research, optimal operation and sizing of storage systems are analyzed (52; 59; 60; 61; 62). In (59; 60), optimal operational strategy of a combined wind-hydro pumping storage is obtained. In more recent studies (63; 64), optimal operations of energy storage systems is obtained by considering ESS as an independent unit, which is not co-located with wind farms.

Table 2.1: Characteristics and Applications of Storage Technologies. Reprinted with permission from (55)

Characteristics	**L/A battery**	**Li-ion battery**	**NaS battery**	**VRB flow battery**	**Super Capacitors**	**SMES**	**FES**	**PHES**	**CAES**
Energy storage capacity (kWh)	≤ 100	≤ 10	≤ 100	20-50	≤ 10	≤ 10	1-25	≥ 150	≥ 10
Energy density (Wh/L)	50-80	200-500	150-250	16-33	2-10	0.2-2.5	20-80	0.5-1.5	3-6
Power density (W/L)	10-400	0	0	0	100,000	1000-4000	1000-2000	0	0.5-2
Discharge duration	Hours	Minutes-hours	Hours	2-8 h	Seconds	Hours	Seconds-minutes	Several hours	Hours
Charge duration	Hours	Minutes-hours	Hours	2-8 h	Seconds	$\leq$Seconds	15 minutes	Several hours	Hours
Response time	$<$Seconds	seconds	Milliseconds	$<$Seconds	Seconds	$<$Milliseconds	Seconds	Seconds-minutes	Minutes
Lifetime (years)	3-10	10-15	15	5-20+	5-20	5-20	20	25+	20+
Lifetime (cycles)	500-800	2000-3000	4000-40,000	1500-15,000	50000	$> 50,000$	100,000	$> 50,000$	$> 10,000$
Round-trip efficiency	70-90	85-95	80-90	70-85	90	> 90	85-95	75-85	45-60
Capital cost per discharge ($/kW)	300-800	400-1000	1000-2000	1200-2000	1500-2500	2000-13,000	2000-4000	1000-4000	800-1000
Power quality	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No
Transient stability	Yes	No	No	No	Yes	Yes	Yes	No	No
Regulation	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Spinning reserves	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
Voltage control	No	Yes	Yes	Yes	Yes	No	Yes	No	No
Load following	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Firm capacity	No	No	No	No	No	No	Yes	Yes	No
Congestion relief	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes
Advantages	Low cost, High recycled content	High efficiency, high energy density	High energy density, quick response, efficient cycles	High depth of discharge, high cycling tolerance	Rapid response time, high power density	High efficiency	Rapid response time, low maintenance, high cycles	Rapid response time, large capacity	Rapid response time
Disadvantages	Low energy density, large footprint, limited discharge depth	Cost prohibitive, overheats, limited discharge depth	Safety issues	Low energy density, low efficiency	Cost prohibitive	Cost prohibitive, low energy density	Cost prohibitive, tensile strength limitations	Geographically constrained	Low efficiency, geographically constrained

2.4 Demand Response

Reliable operation of the power system requires a steady balance between supply and demand. With the increasing penetration of renewable energy, maintaining the system reliability and balance has been even harder, since both supply and demand are volatile and uncertain. Although storage can reduce this imbalance, high investment cost is a serious barrier in grid-level storage applicability. Traditionally, the power supply follows the power demand using the fast-responding, load-following power plants. However, in the last two decades, power grid operators have attempted to find a way that the power demand follows the supply, so that the demand can chase after the more economical energies of wind and solar generations. This is where DR comes into play to reduce the need for expensive flexible generators like gas fired plants.

Incorporating DR into the grid can accommodate high penetration of wind energy by managing the end-use customer's electricity consumption. Unlike storage devices, which can provide support in mitigating fluctuation from the supply side, DR is designed to change the customer electricity consumption pattern to minimize the demand fluctuation and the system imbalance.

DR can be defined as changes in electricity consumption by end-use customers in response to price changes over time or to incentivize payments designed to reduce power consumption during peak-demand periods (65). DR approaches are classified into two main programs: price-based programs (PBP) and incentive-based programs (IBP) (13). The former provides dynamic price rates throughout the day with the objective to equalize the demand curve by raising the electricity price during peak periods and lowering price in off-peak periods, while the latter encourages customers to participate in DR by providing incentives to reduce their load at critical periods. Some examples of PBPs include Time of Use (ToU) (66), Critical Peak pricing (CPP) (67), and Real-Time pricing (RTP) programs. In the most basic DRP, the ToU program, price rates vary in different blocks of time; most commonly, it has two blocks; peak and off-peak demand. CPP is very similar to ToU, but the only difference is that it is only applied on some specific days called 'event days'. In the more complicated program, RTP implements smart metering to provide real-time pricing based on the real cost of electricity delivery for customers to reduce their use at peak-

demand (68; 69). RTP is considered the most economically efficient program for a competitive electricity market (70). Further classification of PBPs and IBPs is discussed in (68).

If implemented properly, DR programs can provide a vast amount of benefits to the system. Earlier works presenting benefits of DR are available in (65; 68; 69). Some of DR benefits highlighted in these works include customers bill savings, market price reduction, environmental benefits, market efficiency impacts, and reliability boosts. By diminishing electricity usage during peak-demand periods, DR can reduce the need for generator cycling or building new power plants to balance wind power fluctuation, which also leads to environmental benefits. According to the Federal Utility Regulatory Commission, DR can lessen the peak demand by 20 percent in the US (71). It is suggested that the CPP program is the most effective program in reducing peak load for residential electricity use in North America by utilizing the technology to curtail loads on special days (72). Cutting down the electricity demand leads to high market price reduction. This is due to the fact that with the production near maximum generation capacity, which is likely to occur in peak periods, the generation cost rises exponentially; thus, even a small reduction in electricity demand can lead to a considerable decrease in price.

In (73), DR benefits are categorized as operating, planning, and economic benefits. The operating benefits of DR mainly result from the ability to reduce reserve needs which leads to reducing the use of conventional generators. Moreover, the system operators will have more options and resources to match supply and demand; thus, maintaining system reliability in a wind-connected grid. Concerning the planning, load shifting ability of DR that can balance wind power fluctuation will result in reducing the need for high investment in flexible power plants. Economic benefits from DR programs include electricity bill savings for customers and reduction in price volatility.

There are numerous studies dedicated to investigating the effect of DR in a wind-connected grid (66; 69; 74; 75; 76; 77). In the study (69), different time interval uncertainties posed by wind power and the ancillary services required for DR programs to accommodate these uncertainties are discussed. For instance, frequency deviation requires an immediate response in demand, while

for power regulation and to prevent unplanned outages, demand should be managed for the 5-10 min horizon. In such emergency cases, customers enrolled in DR programs proactively reduce their electricity consumption for rewards of financial incentives, or the system operator shut downs some preselected electric devices with the customers' permission. In (66), the impact of RTP program on wind integration is evaluated and it is shown that by implementing real-time pricing program, the wind utilization increases up to 11% and as a result, reduces the amount of wind that has to be curtailed, since it allows the demand to increase in high wind generation days. Besides, RTP may lower the use of expensive transmission systems. Other PBPs can also mitigate the wind generation uncertainty and can reduce the operational cost by reducing the amount of reserve provided by conventional generating units (77).

The load control can be done in both direct and indirect manners. In the former, system operators or utilities directly modify the customer's energy consumption by, for instance, turning off their devices during high-load periods. In contrast, the latter allows the customers to decide on their reaction to the price changes. The work in (74) evaluates the effects of both direct and indirect controls for demand management, and different DR programs are compared. The total operating cost reduction to the energy market by both options is highlighted. That is because fewer thermal units are used for reserve. Besides, with high wind generation during low demand periods, DR could cause no wind curtailment. DR programs can also provide ancillary services to the system such as voltage control, frequency support (78; 79) and spinning reserve (77; 80).

Besides the benefits offered by DR, there are some challenges in implementing wide-use DR. Some of the main barriers for DR are discussed in (65; 73) including lack of knowledge of regulators and policy makers regarding the benefits of DR, improper market or regulatory structure, uncertainty about the economic and performance of new technologies and tariffs, and limited and uncertain response to the price changes. Moreover, the lack of experience in this field leads to challenges in modeling and assessing DR approaches. A 2009 study of DR experience in Europe (81) concludes that lack of knowledge of DR potentials in power system and high-cost estimates for DR technologies and frameworks are the reasons behind the slow development of DR in European

electricity markets. Moreover, the lack of real-time price information to customers and relatively small incentives are highlighted as reasons to limited response to the electricity demand. However, recent advances in smart grid have enabled better communication between the operators and users; thus, more real-time cost information is available to customers. A recent study in 2018 (82) shows that with smart grid technology, DR limitations tend to diminish, enabling large-scale wind power integration.

2.5 Electric Vehicles and Vehicle-to-Grid

As discussed in sections 2.3 and 2.4, high capital cost of ESSs as well as challenges in DR limit the incorporation of these two techniques in the power system. Advances in smart grids have enabled electric vehicle (EV) to be a good potential option as a distributed energy storage unit and as a flexible dynamic load that can be managed by system operators (83). When connected to a charging station (plugged-in), an EV can charge using power from the network as well as send back power to the grid whenever needed. The exchange of energy between EV and the grid known as Vehicle to Grid (V2G) technology, can improve the power system operations in many ways. In this dissertation chapter, we only focus on the benefits of V2G in support of RES integration. The reader is referred to (84) for an extensive review of the benefits and challenges of V2G technology.

V2G can be classified into two groups: Unidirectional and bidirectional V2G. In the former, the flow of energy is in one direction in which EVs can only be charged from the grid without any discharge capabilities. The latter provides a two-way power exchange between the grid and EV. Both techniques are able to absorb the fluctuation in the output of the RES.

System operators can utilize unidirectional V2G to smooth out the demand volatility by charging EVs on off-peak demand periods. The studies in (85; 86) have considered the load shifting and peak shaving potentials of the V2G technology. In (85), it is shown that V2G can reduce the peak load by 30% with EV adoption rate of 15%. The authors in (87) study demand-side management approaches for different scheduling models of EV charging that lead to lower demand in peak-load periods and an increase in demand for RES. With the focus on maximizing RES integration, many researchers have attempted to find the optimal scheduling for EV charging (88; 89; 90). The

authors in (88) have found that by utilizing the flexibility of EV drivers in charging their vehicles, V2G can lead to a high integration of wind power. An optimization study in (90) proposes that with the overnight charging of a large number of plug-in hybrid electric vehicles (PHEVs), the utilization of wind energy can increase significantly. The work in (86) shows that smart charging of EVs can shift its load toward the high RES generation periods, thus, increasing the share of RES.

On the other hand, bidirectional V2G provides additional benefits for RES integration, as it can absorb the excessive power supplied by wind power or other RES with charging while discharging electricity back to the electric grid during peak demand or low wind generation periods. EV drivers that participate in V2G service can sell their excessive power to the grid with a price higher than the purchase price in high-peak demand periods when electricity price is high. Feeling insecure for urgent needs and the battery degradation of their vehicles that frequently charge and discharge in bidirectional V2G technology, many EV owners are reluctant to participate in bidirectional V2G. Therefore, system operators must provide financial incentives, and the incentives need to be clearly and sufficiently stated to encourage their participation. Compared to the unidirectional V2G, there are fewer number of research studies devoted to finding the optimal charging/discharging scheduling (bidirectional V2G). The study in (91) evaluates the impact of V2G service on wind integration for different scenarios. It is shown that wind utilization can reach 89% with a coordinating charging and discharging strategy under the assumption of an EV for each household. The work in (92) shows that a large fleet of EVs can be used as distributed storage units that help to integrate the RES generation by providing the load following. The authors in (93) propose a controlled charging/discharging model for large number of EVs in an intelligent parking lot that provides financial incentives for the EV owners as well as satisfying technical operation goals.

2.6 Unit Commitment under Wind Uncertainty

Unit commitment (UC) is known as the decision-making process of finding an optimal schedule and operation of power generation units in a given set of operating and security constraints. UC decisions contain both the day-ahead decision of commitment status of all generation units as

well as the real-time economic dispatch (ED) decision. The objective of a generic UC is to minimize operational cost while satisfying the ramping limits, minimum up/down time constraints, power balance, and system reliability constraints (94; 95; 96). The on-off status of units are formulated as binary variables, the generation levels are considered as continuous variables, and energy generation cost curves are non-linear. Hence, the UC problem is often formulated as a mixed-integer nonlinear optimization problem (97).

The literature in this field can be categorized into basic UC and security-constrained UC, in which the latter considers the security constraints such as the transmission line and generator outages, while the former ignores these constraints. For addressing the data uncertainty, UC is divided into deterministic and stochastic approaches. To compare the two approaches, we first need to identify different sources of uncertainties.

2.6.1 Uncertainties in Power System

Uncertainties in the power system can be categorized into two classes: forecasting errors and component failures (7). Forecasting errors are simply the difference between the prediction and the actual value of an uncertain parameter. The two main forecasting errors in power system operations are associated with load and generation. With all the power supplied from conventional generators, the production is assumed to be fixed, and the load forecast error is the primary source of uncertainty (34; 35). However, the presence of wind generation adds another source of uncertainty to the system, which is known to be much greater than the load uncertainty (98). The other sources of uncertainty resulting from equipment failures include transmission and generator outages.

With no power generated from RES, the supply of conventional generators can be treated as a deterministic element, and the load forecast errors are captured by the downward and upside reserve (99). Traditionally, the deterministic UC considers the point forecast of the load and supply as the actual parameters in finding the optimal schedule. In this technique, the primary goal is to determine the optimal reserve level to balance the supply and demand. In the deterministic approach, the future situation is assumed to be precisely known (single point prediction value for

wind power output and load), and it attempts to find the level of reserve requirements that protects the power system from unpredictable changes in demand or generator outages. So deterministic approaches differ in the way they formulate the reserve constraint. For instance, a traditional UC expresses the reserve constraint in a way that the minimum amount of spinning reserve is constrained to be at least the capacity of the largest generator (100).

However, with the increased penetration of wind power, the deterministic approach suffers from low efficiency and high cost, due to the inability to capture the variation and sudden changes in wind power output (101). Thus, with the volatile supply of wind farms, the need to improve forecasting tools as well as UC approaches to accommodate the uncertainty has increased significantly.

2.6.2 Wind-Forecasting Tools

Traditional forecasting tools predict the wind power generation as a single point estimator or a conditional expectation of wind power output, which do not take the prediction uncertainty into account (102). There are several methods developed to predict the wind speed as a probabilistic framework. The recent forecasting techniques and the challenges associated with wind prediction are reviewed in (19; 103). The prediction approaches include numeric weather predictors (NWP), statistical or machine learning methods (e.g. artificial neural network (ANN) (104; 105) and time-series models (106)), and hybrid models(107; 108).

Short-term and long-term predictions are required for managing the operation of the power system. Short-term prediction, which ranges roughly from 30 minutes to 6 hours, is utilized in ED planning and load management decisions. Long-term forecast, e.g, one day to one week ahead, is applied to UC decisions (109). Capacity planning needs generation forecasts for years ahead. Classification of prediction tools on different time scales is provided in (23). In this work, it is shown that probabilistic wind power forecasting compared to a single value estimation can reduce the spinning reserve capacity and operating costs of power system substantially. The review study in (19) also confirms that probabilistic forecasting and scenario-based forecasting perform better than single point prediction of wind. Despite recent progress in wind power forecasting, as very recent studies show that short-term wind speed forecasting can reach more than 90%

accuracy (110), high-accuracy for long-term wind power forecasting is still a topic of interest. The forecasting information is used as input to UC and ED problems.

2.6.3 Uncertain UC Optimization Approaches

Different optimization approaches have been proposed to solve the UC and dispatch planning with the advances in recent forecasting tools. They differ in the way in which they represent future uncertainty. These include two-stage SP, multi-stage SP (111), interval-based robust optimization (112), and chance-constrained optimization (113). These optimization approaches can be classified into two main categories: stochastic and robust optimization. Next, we describe each of these methods separately.

2.6.3.1 Stochastic programming:

- *Two-stage stochastic programming:*

  In SP, the uncertain parameters are represented as scenarios in which the underlying distributions are assumed to be known, or they can be estimated. In a scenario-based SP, a number of scenarios represent the uncertainty. Each scenario describes a possible future realization of the uncertain parameter. SP often uses a two-stage formulation, where in the first stage, some actions are taken to minimize or maximize the expected value of a desirable function (e.g., minimize cost) given the probability distribution of random future events, and in the second stage after the realization of random variables, a recourse decision is made to exploit the choices of the first stage better.

  The two-stage nature of UC motivates the use of SP in which it has a similar two-stage procedure (94). In the first stage, in the day-ahead market, system operators attempt to find the commitment status of generating units in order to minimize the expected operations cost, given the scenario set of random variables (load and wind power), while in the second stage after the realization of the actual load and supply, a real-time dispatch decision will be made.

  In the scenario-based stochastic programming approach, uncertain parameters are assumed to have a specific probability description. In literature, electrical load is often assumed

to follow a normal distribution (14; 114; 115; 116). WPFE is modeled as normal (114; 115), or truncated normal distribution (117). Unit generation and transmission outages are modeled by discrete event scenarios (118), fuzzy sets (119), monte-carlo simulation (120), and/or independent Markov processes (121). Therefore, a great challenge of the scenario-based approach is to develop an appropriate probability distribution or weights on scenarios. Besides, probability distributions usually fail to represent the actual uncertain parameter. For instance, wind production is a nonlinear and non-stable process; hence, WPFE may not fit into any known probability distributions (23).

On one hand, a large number of scenarios are required to capture the uncertainty associated with the system, thus improving the quality of the solution. On the other hand, the computational time is limited to the planning horizon (few minutes for ED or few hours for UC), which can be violated with a high number of scenarios and required solution time for the corresponding problems. This limits the application of SP to a large-scale power system. To address this issue, Bender's decomposition approach is used to cut down the computational burden of SP. Since the scenarios in the second stage are not correlated, the problem can be decomposed to smaller optimization problems that can be solved faster. However, this technique is delicate to scenario generation and is usually used when the second-stage problem is a linear program with continuous variables. However, the second-stage problem includes some integer variables, e.g. commitment status of power plants. The most common algorithm in two-stage SP is dual decomposition (or Lagrangian Relaxation) approach, which decomposes the UC problem into smaller and simpler sub-problems by dualizing the coupling constraints (see (122; 102; 123)).

Besides, various scenario generation/reduction techniques have been employed by researchers to overcome the computational burden of the stochastic programming while improving the quality of the solution (see (114; 115; 124)).

- *Multi-stage SP:* In the second stage of two-stage models, future uncertainties are observed

only once. However, in power systems, information on realized uncertain parameters is updated frequently on an hourly or sub-hourly basis. In this situation, a multi-stage framework can take advantage of more disclosing uncertainties over time to adjust the commitment status as well as dispatch decisions dynamically. In general, multi-stage models perform better for longer time-horizon and capture the dependencies between different stages. In particular, when generators can start-up or shutdown frequently, the commitment decisions can be adjusted in different stages (125). However, constructing the scenario tree and assigning probabilities to different trajectories are complicated. Moreover, as the number of stages increases, the number of possible scenarios grows exponentially, resulting in a more computationally intensive problem to solve (101). That is why this approach has not gained much attention in recent optimization frameworks for UC.

- *Chance-Constrained Optimization:* Chance-constrained programming is another approach to represent the uncertainty in a power system in which constraints are forced to be held with a presumed level of probability. For instance, the reserve requirement constraint is often modeled as a chance constraint in which the load must be satisfied to a certain level, referred to as loss-of-load probability (LOLP) (117). In addition, the transmission line overload (also see (121)) and wind curtailment (see (126)) due to uncertainty is restricted by small predefined probabilities. That guarantees high utilization of wind power output. The main driving factor to this approach is to increase the robustness of the solution in SP. By properly assigning the probability that the constraints should be satisfied, system operators can provide a trade-off between cost effectiveness and reliability of the model. Moreover, this approach reduces the computing complexity of a scenario-based SP method.

  Since chance-constrained method fails to model the uncertainty by itself, a combined chance-constrained stochastic optimization is proposed to solve UC in which the two-stage stochastic optimization is used to capture the uncertainty of load/wind production (127; 128; 126). For instance, in (127), the load and WPFE are modeled by scenario-based stochastic optimization, while the chance constraints restrict the loss of load by a small probability.

In literature, a sample average approximation (SAA) model is often utilized to solve a chance-constrained stochastic program (126; 128; 129). Also, chance-constrained UC model can be transformed into an equivalent mixed integer linear deterministic problem (117; 121). However, transforming a chance constraint into equivalent deterministic constraints suffers from some challenges and only some limited forms of chance constraints can be converted to deterministic constraints (130).

2.6.3.2 Robust Optimization:

Another approach to address uncertainty is the interval-based robust optimization (RO) approach, which has gained more attention in recent years compared to SP, since it does not posses the main challenges of SP for large-scale power system problems. Unlike the SP approach, RO does not require the probability distribution for future uncertainties. When the parameter uncertainty is not stochastic or if the distributional information is hard to obtain, instead of assigning probabilities to the scenarios, the robust optimization approach attempts to find a solution that is feasible for all realization of the parameter in a given set. Also, RO guarantees the best solution under the worst-case scenario (131). In that sense, it may yield to conservative solutions in which they can cause higher operating costs to the system in the absence of the worst-case scenario. Thus, the worst-case scenario should be chosen precisely to provide a robust solution to the uncertain system as well as low scheduling cost. Computational tractability is a primary motivation behind the use of RO. A considerable number of robust optimization formulations have been proposed to handle wind power uncertainty in UC. The UC problem is formulated as a robust mixed-integer programming problem, and Benders' decomposition algorithm is often used to solve these problems (6; 132; 133). Similar to SP, RO can be formulated in two-stage (132; 134) or multi-stage structures (135).

2.6.3.3 Other Optimization Techniques:

Besides SP and RO, there have been attempts to solve unit commitment using other optimization methods such as fuzzy optimization(119), stochastic dynamic programming (136; 137), and risk-

based optimization (138). Moreover, many heuristics and meta-heuristic techniques such as particle-swarm optimization (115), priority list (139; 140), and simulated annealing (141) have been presented by researchers to find an optimal schedule in the presence of wind energy. In (16), the authors list all the methods used for UC such as mathematical and heuristic approaches. A review work in (142) provides an extensive comparison between different numerical optimization approaches to UC including SP, RO, and SDP.

Table 2.2, (adopted from (142)) provides a comparison between common UC models discussed in this section. All the references covered in Section 2.6 consider the uncertainty of wind power output, while only some of them take the load forecasting errors and system outages into account. For instance, the authors in (21) only focus on wind power uncertainty; other uncertainties (e.g., load variations, and forced generation outages) are ignored; while the study in (118) takes all sources of uncertainties into account.

2.7 Summary

As the integration of intermittent wind power grows, the effects on power system operations increase. This results in an increasingly unreliable system since both wind speed and wind power curve pose a high degree of volatility. On the other hand, forecasting information is used as input to scheduling problems such as UC and ED. Thus, robust forecasting tools are essential in the integration of wind energy into the power system. Despite the recent sign of progress in the accuracy of statistical methods such as artificial intelligence models and support vector machines, the need for more accurate forecasts in all forecasting horizons, especially long-term, is high. Given the fact that wind uncertainty and prediction error is inevitable, one general approach to accommodate wind uncertainty is to build a flexible system. This goal can be achieved by incorporating DR programs and energy storage systems into wind-integrated power grids. ESSs and DRPs seek to smooth out the variable wind power output by modifying the electricity supply and demand, respectively. ESSs have been a successful approach in recent years with the development of different storage technologies, each of which has its own applications and characteristics. However, further research studies should be carried out toward low-cost and efficient storage devices.

Table 2.2: Numerical Optimization Techniques for UC Under Wind Power Uncertainty. Adapted with permission from (142)

UC optimization models	Advantages	Disadvantages
Two-stage SP	• Minimize total expected cost; easier to understand (and compute) than minimizing the worst-case cost. • Various decomposition and sampling-based algorithms already existed with convergence and performance guarantees. • Can address robustness issues using risk measures. • Can provide expected value of perfect information (EVPI) and value of stochastic solution (VSS).	• Need to assign probabilities for scenarios. • Computationally demanding for large number of scenarios. • Difficulties in dealing with integer variables in the second stage. • Static assumption of the uncertainties.
Multi-stage SP	• Truly a decision-making model (as opposed to "what-if" analysis) over multiple time periods under uncertainty. • Ability to model the dynamic process of uncertainties and decisions. • Useful for systems with generators that can reschedule quickly.	• Curse of dimensionality, and hence computationally very expensive. • Need explicit scenario trees and random paths' probabilities. • Even more difficult with integer variables present in all stages.
Two-stage chance-constrained	• Computationally not as expensive as other SP models. • Can provide a balance between cost effectiveness and robustness of the solution.	• Only some forms of chance constraints can be transformed into deterministic constraints. • There exists no general solution approach in the case where decision variables cannot be decoupled.
SDP	• Can handle multi-stage stochastic problems with relatively low computational burden. • Can model closed-loop systems (such as real-time pricing).	• Convergence to optimal solutions may be difficult to establish. • Integer variables may present difficulties in general.
Two-level and multi-level RO	• Do not need probability distribution. • Guarantee the optimal solution for the worst-case scenario. • Computationally not as expensive as stochastic programming.	• May yield over-conservative solutions. • Need expertise and rationale on uncertainty set construction. • Need to use different algorithms for different types of uncertainty sets.

The recent developments toward smart grids and the ability for communication among system participants enable DR programs to be adopted more effectively in modifying end-use customers' electricity demand to maintain system reliability. Further developments of smart grid technologies are needed to facilitate the communication among electricity producers and consumers, yielding improved consumer participation and thus, more effective adoption of DR.

In the last decade, with increasing uncertainty in power system operations caused by the volatile load and renewable power outputs, the need for optimization techniques that take the stochastic characteristic into account has grown significantly. Stochastic optimization is widely used in literature by considering different scenarios. However, computational intensiveness of SP has limited this approach for large-scale UC problems. Also, scenario reduction techniques should be developed to decrease the computational burden of SP. Another topic of interest is the utilization

of uncertainty analysis and probabilistic prediction in UC problems. Robust optimization, on the other hand, performs superior to SP regarding computational time. However, this approach also suffers from conservative and high-cost solutions. In recent years, some different approaches have been applied to UC problem, such as SDP and fuzzy optimization. These methods are in their early stages, and limited research has been done in using these techniques to this problem domain. Therefore, there is a need for more research studies on the application of specialized optimization algorithms to large-scale power systems scheduling.

The development in the smart grid can facilitate the communication between the users and system operators to exploit the flexibility of consumers, e.g., residential electricity users and EV drivers. This requires a well-built infrastructure to communicate and control power consumption and provide suitable incentives for customers to participate in these programs. For instance, motivating EV drivers to join in V2G and connect their vehicles during peak demand by ensuring that their V2G revenue is significantly higher than the cost of battery degradation. Another example is to encourage residents to provide their load profile ahead of time and get incentives for shifting their load from high peak to off-peak periods. Moreover, demand-side management approaches will be more practical if consumers have real-time information about electricity costs to react correctly or system operators should be able to control some smart electric devices in case of emergency.

Although a relatively large number of research groups have performed research studies on the scheduling of EV charging (143; 144; 93), only a few have focused on the objective of the maximizing RES integration, especially for the bidirectional class of V2G. For instance, the authors in (93) propose a controlled charging/discharging model for a large number of EVs in an intelligent parking lot that provides financial profits for the EV owners as well as satisfying technical operation goals. However, the source of power is not considered in their paper. This is an exciting area of research, as the importance of sustainable energy and the integration rate of EVs both grow dramatically in the next few years. Furthermore, if EVs are plugged in when they are not driven, the vehicle charging optimizer can make sure the lowest possible cost for their owners as well as

spreading the charging demand more evenly throughout the day. This, however, requires a lot of public charging infrastructure options compared to the scenario that EVs are charged only when at home.

Moreover, the literature is lacking optimization algorithms for charge/discharge schedules in which real-time interactions between EV owners, aggregators, and system operators are taken into account. In most cases, studies only rely on average statistics of EV, such as the state of charge (SOC), battery capacity, the number of EVs connected to the grid during a specific time of the day, and the duration of connection (145). However, in reality, these parameters can vary by the day and by EV. This information can be collected from each EV users separately in real time by advanced sensors and smart grid control techniques.

For a framework to be successfully applied, the participation barriers should be appropriately addressed. The major obstacle to V2G participation is battery degradation since frequent cycling of the EV batteries will significantly reduce their lifespan. Almost all research papers in this domain propose scheduling algorithms in which either the charging rate is changed frequently, or the charging process is paused/stopped often. This leads to high battery degradation cost that discourages the EV owners from allowing a third-party (e.g., aggregator) control their charging. In the next two chapters, charge/discharge optimization algorithms are proposed to address the issues discussed above that are considered to be missing in the literature.

3. DETERMINISTIC MODELING *

3.1 Introduction

In recent years, several countries are committing to reduce their carbon emissions. An integral part of this effort is to reduce fossil fuel consumption. To this end, renewable energy sources (RES) are vital to replace or reduce the dependence on existing gas and coal power plants. However, volatility of RES generation affects the reliability of power systems. Since wind generation does not necessarily follow the load pattern, during periods of high generation and low demand, wind energy is curtailed to maintain the demand-supply balance (146). While the grid can absorb some RES generation, complete dependence on these intermittent sources requires massive grid-scale energy storage. High initial cost of storage units has been a barrier to their grid-level implementation. A commonly used approach to accommodate wind penetration in power system is to adjust the consumers' energy consumption through demand response programs so that demand can follow supply from wind sources (147).

As another solution to cut emission, countries are attempting to electrify the transportation sector as it accounts for a large amount of carbon emissions (20-25% share) (148). EVs look more promising than ever to replace the traditional internal combustion engine vehicles in the future since they could achieve zero emission if the electricity used is generated from RES. However, the increasing adoption of EVs may cause a potential problem for the electric grid because of the unpredictable charging schedules. With an unplanned charging process, the increase in EV penetration will lead to an increase in electricity demand and a significant change in the shape of the demand curve. This is likely to result in higher demand variability and impact electricity infrastructure, and make it difficult to accurately predict the load. Peak demand determines the system capacity requirements. Increasing the peak demand will affect the electricity infrastructure of the power system (145). Most importantly, uncoordinated charging of EVs adds a stochastic

*Reprinted with permission from "Leveraging owners' flexibility in smart charge/discharge scheduling of electric vehicles to support renewable energy integration", by Pouya Sharifi, Amarnath Banerjee, Mohammad J. Feizollahi, 2020. Computers & Industrial Engineering, 149, 106762, Copyright 2020 by Elsevier.

element to the power system, which complicates the planning and operation of power systems. Considering the volatility of RES generation and unpredictable driving habits of EV owners and their charging demand, it is expected that the system operators encounter even more complexity and uncertainty in unit commitment problem, which result in sub-optimal solutions.

To address this problem, many have proposed the vehicle-to-grid (V2G) concept to integrate future EVs into the grid successfully. The V2G concept sees the EVs as a distributed generation/storage system as well as dynamic flexible load which could be utilized to balance the electricity supply and demand (83). A large number of EVs can add a great deal of flexibility in controlling the output of wind energy by storing the excess production in their batteries, and sending power back to the grid when needed. Moreover, they can provide load-related flexibility such as load reduction, peak shaving, load shifting, and load following (149; 150). A group of EVs can act as storage units and play a role in demand response if managed properly. Therefore, in this dissertation, both charge and discharge capabilities of EVs are considered in a bidirectional V2G. By considering the discharge capability of EVs, the amount of wind curtailment can be reduced by storing the excessive energy in the batteries of the EV fleet. This chapter proposes an approach to leverage the flexibility of a large number of EVs to determine a charge/discharge schedule that can benefit the owners and the grid. The general idea is that by aggregating a large fleet of EVs that are plugged in to smart chargers for several hours (e.g., during nighttime or at the workplace) and they only require a few hours to recharge, the charging process can be scheduled to utilize more wind energy. Also during periods of low wind generation, the charging can be delayed or some EVs can provide energy by discharging their batteries.

In the literature, many authors have studied the schedule of EV charging to benefit the power system. The V2G technology can improve the power system operations along several services, such as frequency/voltage regulation, spinning reserve, peak shaving, and RES integration support (151). For instance, the work in (86) presents a game-theoretic approach in load management strategy for EV charging to reduce peak load considering dynamic behavior of EV drivers as well as electricity price. However, intermittent generation of RES is not considered in this work. The

number of research studies on EV charge/discharge scheduling with the focus on maximizing RES integration has been limited. In (90), the authors formulate the optimal charging schedule in charge-only mode and find that charging EVs overnight can absorb the excess power generated by wind power; thus, increasing the RES utilization. In (152), the authors present an optimization algorithm for the charging schedule problem to minimize energy cost from the grid based on a queuing model that does not require any information or prediction about the wind production, EV charging request, and electricity price. The study in (153) formulates the EV charging schedule as a mixed integer programming (MIP) problem maximizing the RES integration for the day-ahead problem assuming perfect generation forecast and known EV demand for the planning horizon. However, these works only consider the unidirectional V2G (G2V) and the discharge capability of EVs in a V2G setup is not considered. There are some articles in this domain that have considered the bidirectional V2G in which the EVs are also capable of providing discharge. In (154), the authors developed a MIP problem to maximize the utilization of renewable energy while satisfying EV and household demand, assuming known future trips as well as EV and household loads in a day-ahead scheduling framework. Likewise, the authors in (91) propose an energy management system to maximize the utilization of wind energy assuming known wind and load. In their work, the storage capability of EVs in the context of a distributed feeder with primary wind resource is examined. The distributions for arrival and departure times are assumed to be known and only one generic type of vehicle is considered. In (93), the authors propose an intelligent scheduling algorithm with the objective of maximizing owners' profit under the assumption of known day-ahead arrival and departure times. The work in (92) proposes to utilize the storage capability of a large fleet of EVs that can be used as distributed storage units to help keep the grid frequency within a certain limit in the presence of intermittent RES generation in a distributed grid framework. In a recent work (155), a coordinating day-ahead scheduling is proposed in the presence of uncertain wind power to minimize the curtailment of the wind power as well as the emission. However, the arrivals of EVs and their charging requirements are known day ahead of scheduling. A multi-objective stochastic optimization approach is used in (156) to minimize the operating cost as well as

the wind curtailment. In this work, the wind power uncertainty is modeled by a known distribution (i.e., Weibull), and the EV arrivals are assumed to be known.

There are some gaps in the literature, which are addressed here. First, the assumptions of known day-ahead wind generation, and/or known drivers' commuting behavior which is common in many articles, produce unrealistic results. In the literature, the day-ahead scheduling problem is usually solved assuming known EV characteristics, such as the arrival and departure times, the charging requirements, the power rates, and the battery capacities (155; 154; 156). If not deterministically known, the stochastic EV characteristics are modeled only by a few known scenarios (91; 157; 158). This is impractical in reality since not only there are many different types of EVs and each one has different charging power and battery capacity, but also the commuting behavior of users and initial SOC are highly unpredictable and varies significantly on different days depending on several factors, such as traffic conditions, and randomness in commuting behavior of drivers (159). Thus, it seems that the underlying distributions for all EV-related parameters are impossible to achieve. Moreover, assuming known departure time and desired level without users' feedback may cause great discomfort for the users. For instance, in the case that the owners need high charging level in a short time, but the optimal charging process determines to postpone the charging or impose discharging in the beginning of plug-in period since the actual departure time and users' needs are unknown. The user feedback and discomfort is taken into account in none of the works mentioned above. Motivated by the above issues, this dissertation proposes a real-time optimization framework in which the EV characteristics and charging requirements, which unlike other research studies, are only known upon vehicle arrivals, not a day ahead of scheduling. The uncertainty of wind and price forecasts as well as future EV arrivals are also considered. It is proposed that by frequently optimizing the charge schedule based on updated wind forecast, electricity price, and EV arrivals, the aggregator can make up for the sub-optimal decisions made in previous time slots; thus, reducing the effect of uncertainty significantly. Moreover, the level of user discomfort and EV-related uncertainty are reduced by designing an interactive mechanism where the users can input the time of departure and their required energy level upon arrival. A

dynamic rolling-horizon algorithm is developed for the aggregator to determine EV charging and discharging schedule. This work is similar to (144), where the authors have considered the dynamic behavior of EV owners in an optimal charge/discharge scheduling of EVs using a rolling horizon optimization method; however, the intermittent generation of RES is not considered in their work. To evaluate the proposed dynamic scheduling, we first model the optimal schedule for a day-ahead static scenario, where commuting schedule of all EVs as well as wind power output are assumed to be known day ahead of time. The results of the dynamic model is then compared to this static model. A summary of the recent articles on wind-EV integrated power system is provided in Table 3.1.

Table 3.1: Summary of literature in EV charge/discharge scheduling. Reprinted with permission from (160)

Reference	User Discomfort	Uncertainty Modeling	V2G Type	Planning Time	Objectives
(153)	No	Perfect generation forecast	Unidirectional	Day-Ahead	Wind integration
(161)	No	Perfect wind forecast, known distribution for arrivals	Bidirectional	Day-Ahead	Energy loss (curtailment)
(158)	No	Perfect wind forecast, known distribution for arrivals	Bidirectional	Day-Ahead	Operation Cost
(162)	No	Known scenarios for wind and EV	Bidirectional	Real-time	Wind Integration
(157)	No	Known scenario distribution for Wind and EV arrivals	Bidirectional	Day-Ahead	Operation Cost
(91)	SOC_{min}	Perfect wind forecast, Known scenarios for arrivals	Bidirectional	Real-time	Wind Integration
(154)	No	Perfect wind forecast, Known EV parameters	Bidirectional	Day-Ahead	Conventional generation
(155)	No	Distribution of Wind forecast error, known EV arrivals	Bidirectional	Day-Ahead	Wind Curtailment, Emission
(156)	No	Known scenario for wind, Known EV arrival	Bidirectional	Day-Ahead	Operation Cost, wind curtailment

Second, as shown in Table 3.1 most articles focus only on grid-related objectives, such as minimizing operation cost or maximizing RES integration. The EV owners' benefits in terms of revenue and low charging cost are not considered comprehensively in the literature. Rather than only focusing on maximizing wind utilization (WU), this optimization framework is based on a multi-objective problem in which both wind utilization and charging cost (and revenue) are taken into account.

Finally, user discomfort has not been considered in many of the mentioned articles. Even though EV in V2G technology is a promising approach to add demand flexibility, people participation has been limited. There are three main challenges in V2G that need to be addressed before applying

it to the real-world. 1) *Battery degradation* has been reported as the major impediment in people participation in V2G (163). An average automotive lithium ion battery lasts between 2000-3000 charging cycles (55); hence, cycling them daily will significantly reduce their lifespan. 2) *Low efficiency* and round trip power losses of all EV components make V2G a less efficient method to store electricity compared to other energy storage methods (e.g., pumped hydro storage) (164). Comparing to other methods of energy storage (e.g., pumped hydro storage 85-95%), V2G as an energy storage system might lose more in transaction. 3) *Feeling insecure for urgent needs* discourages an EV owner to allow a third party to directly control their charging process (151).

In this work, all the aforementioned V2G challenges are addressed and unlike many of the previous studies, the proposed algorithm does not attempt to start/stop charging EVs frequently, which contributes to battery degradation. The approach here is based on optimal scheduling of a large number of EVs depending on the availability of EVs and a mechanism to capture the preferences, needs and flexibility of their owners.

The main contributions of this work are summarised below:

- Design an interactive mechanism where EV owners can input their preferences in terms of charging requirement and the departure time.
- Address the main challenges for people's participation and reduce the user discomfort level.
- Propose a scheduling algorithm where aggregators can leverage the flexibility of EV drivers, and the V2G technology to support integration of intermittent wind energy into the power system.
- Propose a rolling-horizon dynamic model that frequently optimizes charge/discharge schedule based on updated wind and price forecast, reducing the effect of forecast uncertainty.
- Develop a multi-objective optimization method where both the owner and the grid can benefit from participating in V2G (i.e., low charging cost and high wind integration).

The smart charge/discharge scheduling problem where an aggregator can modify the charge/discharge rate in discrete time intervals, is studied for both static and dynamic models and the performance of

the proposed models are compared with the "Business-as-usual" (BAU) case, where the EVs start charging immediately upon arrival at full speed until reaching full charge. Moreover, the benefits of participating in bidirectional V2G is evaluated compared to unidirectional V2G, also known as grid-to-vehicle (G2V).

The remainder of this chapter is structured as follows. In Section 3.2, an overview of our system model and the proposed approach is provided. In Section 3.3, the optimal scheduling problem of the static model is formulated, while the dynamic model is discussed in Section 3.4. Section 3.5 discusses the simulation results. Finally, the concluding remarks of the chapter is presented in Section 3.6.

3.2 Approach and System Model Overview

Smart energy metering and advanced controls have enabled real-time communication in the smart grid (83). Consider a smart grid with real-time communication between its participants, and an aggregator who is responsible to offset the fluctuation of RES (e.g., wind energy) with the optimal charging/discharging schedule of EVs. A schematic of the relationship is shown in Fig. 3.1. The aggregator, which can be a utility company or a third party, gets information (forecasts) about available renewable energy, market electricity price, and data related to the status of EVs, the charging requirements, and their owners' preferences. Also, suppose that there is a set of EV owners who want to charge their vehicles with renewable energy as much as possible to address carbon footprint and sustainability concerns. They may also have an economic motivation to pay less for charging their vehicles. The EVs are connected to the smart chargers, where the real-time EV information can be obtained by metering devices and sent to the aggregator.

3.2.1 Electricity Generation and Consumption Model

This dissertation chapter and the following chapter study a Micro Grid (MG) with wind power as primary resource and local parking lots as consumers. EV charging in parking lots are the only demand for wind generation and there is no storage unit in the system other than the vehicles' batteries. Since wind energy has near zero marginal generation cost and because of the support

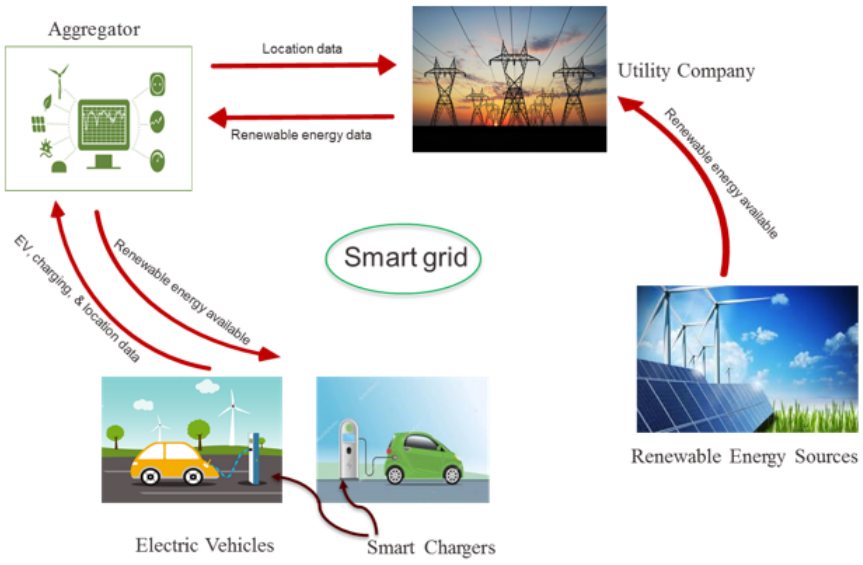


Figure 3.1: Smart grid communications. Reprinted with permission from (160)

policy from the government, it is assumed that wind energy production has no cost. The MG is connected to the external grid via transformer for back-up power, so that if charging demand exceeds the available wind energy, the remaining energy is purchased from the grid at the real-time electricity price. EV owners can participate in either unidirectional or bidirectional V2G. The former, also referred to as G2V, is when the EV can only consume energy from the electrical grid. However, in bidirectional V2G, the EVs can inject energy to the MG by providing discharge. In the context of this problem, V2G is referred to bidirectional flow of energy between power source and vehicle. It is assumed that the discharge energy is only used for charging other EVs, and is not sold to the external grid. The EV discharge energy is sold at a price slightly less than that of real-time wholesale market price. This ensures that EVs are making revenue from selling their stored energy and also other EVs are paying less compared to the market price. The maximum charging power for an EV in one hour interval is calculated as $\min(AR_i, \ CP_i)$, where AR_i is the maximum power the EV can take, and CP_i is the maximum charging power of the outlet that EV i is connected to. Thus, the maximum energy EV i can take in each Δt decision period is defined by $P_i^c = \min(AR_i, \ CP_i) \times (\Delta t / 1hour)$. Another assumption is that charging and discharging

powers are the same ($P_i^c = P_i^d$).

Local consumers in the MG are the EV owners parking their vehicles in residential places or in workplace parking lots. The electricity consumption for households is excluded from the model, since in case of known or close to known household demand, that would not add value to our optimization problem.

3.2.2 Interactive Mechanism

User discomfort seems to be the main impediment in people participation in demand response programs (i.e., V2G), as the value of electricity for the users is much higher than its price. This implies that users are willing to pay more for the electricity rather than sacrificing their comfort for a lower electricity bill (165). EV owners' willingness to participate in V2G is low due the fact that by allowing a third party (aggregator) to modify the charge/discharge rate, the vehicles might not have the desired level of charge at the time of their departure. Moreover, the aggregators might schedule the charging for frequent charge/discharge cycles that contribute towards significant battery degradation. To address these issues, the aggregator needs to know the owners' preferences and requirements, and take battery degradation into account.

Suppose, there is a simple web/mobile application where EV owners can input their needs and preferences upon arrival using their smart phones or computers. Such web/mobile applications (apps) are common today in a wide variety of application areas. Aggregators and third party vendors can easily develop a customized and easy to use app for the EV owners using a convenient front-end and tie the app and the collected data with the aggregator's back-end data repository. The specifications and interface of such an app can vary depending on the level of desired customization and the user interface. The required input information includes departure time (T_{dep}), desired level of battery at departure ($SOC_{desired}$), minimum required level of charge (SOC_{min}), and whether they want to participate in V2G or G2V. Knowing these parameters not only enhance the level of comfort for the EV owners, but also decrease the level of uncertainty in EV scheduling. The minimum required SOC_{min} is to restrict the amount of discharge at any point in time during the plug-in period, so that in case of urgent need, where the owner wants to depart earlier than the

prespecified departure time, the vehicle has sufficient charge level to reach the destination. A schema of different battery-related parameters and their methods of detection is depicted in figure 3.2. Most smart chargers and current EVs have sophisticated sensors embedded in the system and provides almost real-time status information. This information is assumed to be available to the aggregator using existing communication technology, which is ubiquitous. The sensor technology in smart chargers and EVs and communication technology are all expected to keep improving. The aggregator receives the information from the web application and smart chargers and schedule for the EVs in a way that satisfies the user requirements. It is assumed that the communication infrastructure is fast and robust, and does not contribute towards the uncertainty and cost of energy purchased from the grid.

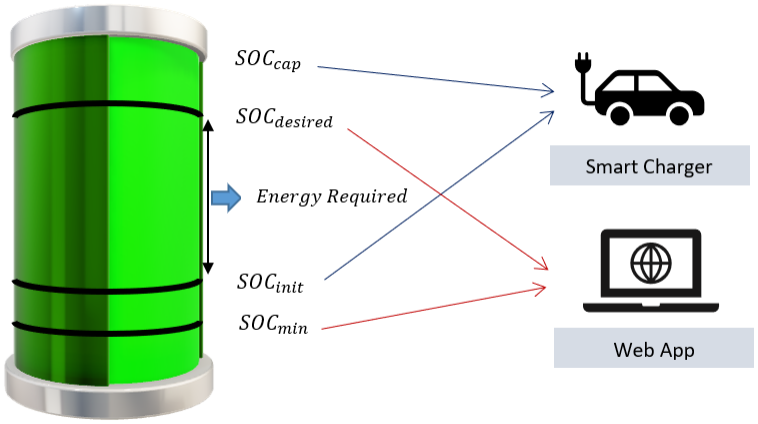


Figure 3.2: Battery-related parameters. Reprinted with permission from (160)

Since EVs are idle about 90% of the time and they only require a few hours to recharge, the goal is to use the flexibility of EVs and their drivers to absorb the fluctuations in the output of RES, while satisfying owners' requirements and concerns. The hypothesis is that the EVs charge their batteries mostly with energy from RES and discharge energy during low wind generation periods. Also, in case where there is insufficient wind to meet the demand during the whole plug-in period, the EV charging process should be shifted to low electricity price periods whenever possible.

3.2.3 Battery Degradation Model

Researchers in (163) have found that the maximum annual profit for an EV is very limited if considering the battery degradation cost, and without considering this cost, the profit is exaggerated in many studies. Moreover, battery degradation is a challenging factor in people participation, so the cost of battery degradation in the charge/discharge process is studied to reduce the total charging cost (TCC) for the owners. Finding exact battery degradation cost is out of scope for this dissertation. Here, two models for battery degradation found in the literature are used. The first model considers a quadratic function, which consists of two terms, one for charge/discharge rate and the other term captures the cost of degradation for fluctuation in energy rate (presented in equations (3.5)-(3.6)) (144). The robustness of our solution is evaluated with a second degradation model adopted from (163; 166), which models the battery wear cost as a linear function of battery replacement cost and percent of battery used. The laboratory measurements in (163) predicted a cost of 4.2 ¢$/kWh$ for a battery pack with \$5,000 replacement cost. Equations (3.1)-(3.2) define the degradation cost of EV i in the second model.

$$\Psi_i = \sum_{t \in T_i^p} 4.2 \times \frac{C_{bat,i}}{5,000} \times (percent\ of\ battery\ used)^t \tag{3.1}$$

$$(percent\ of\ battery\ used)^t = \frac{P_{c,i} X_{c,i}^t - P_{d,i} X_{d,i}^t}{SOCcap, i} \tag{3.2}$$

In the formulation, the battery degradation cost is penalized in the objective function so that the optimal solution reduces the frequency of changes in the charge rate.

3.3 Static Scheduling Optimization

A scheduling algorithm is proposed for the aggregator that determines the day ahead EV (dis)charge schedule. In the static model, the following assumptions are made:

- The EV owners are obligated to provide their arrival time, departure time, the desired level of charge, and minimum level of charge, a day ahead of scheduling.
- The initial state of charge and the EV characteristics, such as battery capacity, and acceptance

rate are known.

- The wind production and grid energy price for the next day are predicted with perfect accuracy.

Given the information input by the owners, the aggregator determines the charging schedule for the next planning horizon (e.g., next day). The planning horizon is evenly divided into discrete time intervals (Δt). This smart scheduling is determined to maximize WU, minimize consumption from the grid supply, and minimize TCC for EV owners. Although the assumptions made in this deterministic case are unrealistic, this model provides the global optimal solution for the case where all model parameters are known. The results of the static model will be used later to evaluate the performance of the dynamic model.

3.3.1 Modeling & Mathematical Formulation

The EV charging and discharging schedule is studied for the planning horizon that is evenly divided into time intervals of Δt minutes. The aggregator finds the optimal charge/discharge rate for all EVs in each period t. Throughout this chapter, the time slot $[t, t+1]$ is referred to as period t (Figure 3.3). It is assumed that grid electricity price and charge/discharge rate remain constant for the entire interval of Δt. Wind power availability in Δt time interval is known and is utilized to charge the EV fleet during that interval.

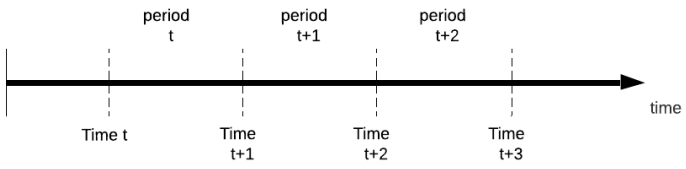


Figure 3.3: Illustration of difference between time and period. Reprinted with permission from (160)

The decision variables used in the model are described as follows. $X_{c,i}^{t}$ and $X_{d,i}^{t}$ are continuous

variables between 0 and 1 that determine the charge and discharge rate of EV i in period t with a value of 1 meaning full-speed charging/discharging, and 0 meaning remaining untouched. The state of charge for EV i at time t is captured by SOC_i^t. G^t and Ω^t are grid supply and wind curtailment in period t, respectively. Finally, Z_i, $Y_{c,i}^t$, and $Y_{d,i}^t$ are auxiliary binary variables.

The objective function is to minimize a linear combination of the charging cost of energy purchased from the electric grid, the battery degradation cost, and the wind curtailment penalty cost.

$$\min_{X_c, X_d} \quad \sum_{t \in T} pr^t G^t + \sum_{i \in I} \lambda_i \Psi_i + \sum_{t \in T} \delta \Omega^t \tag{3.3}$$

The variable G^t is the energy supplied from the external grid (conventional generators). When the charging demand exceeds the sum of available wind energy and discharged energy, the remaining energy is purchased from the external grid. Thus, it can be defined by inequality (3.4) and non-negativity constraint (3.27).

$$G^t \geq \sum_{i \in EV_{all}^t} (X_{c,i}^t P_i^c) - \sum_{i \in EV_{v2g}^t} (X_{d,i}^t P_i^d) - W^t, \quad \forall t \in T \tag{3.4}$$

It should be noted that G^t is the maximum of zero and the total net charging demand (the difference between charged and discharged energy) minus the wind energy in any period. Inherited in the first term of the objective function, the problem attempts to minimize the charging cost for EV owners that is purchased from the grid. Also, imposing charging cost for the grid energy causes the demand to seek cheap wind energy and increases WU.

The next term in the objective function is the sum of the battery degradation costs multiplied by λ_i, which is input by the owner indicating the level of tolerance for their battery degradation. Value of $\lambda_i = 1$ means no tolerance for the owner while $\lambda_i = 0$ sets the degradation cost to zero. Battery degradation is calculated using a quadratic function as defined in (3.5)-(3.6) (144).

$$\begin{aligned} \Psi_i = \sum_{t \in T_i^p} & \alpha(\eta_i^c P_i^c (X_{c,i}^t - X_{c,i}^{t-1}))^2 \\ & + \beta(\eta_i^c P_i^c X_i^t)^2, \ \forall i \in I_{g2v} \end{aligned} \tag{3.5}$$

$$\Psi_i = \sum_{t \in T_i^p} \alpha[\eta_i^c P_i^c (X_{c,i}^t - X_{c,i}^{t-1})]^2 + \beta[\eta_i^c P_i^c X_{c,i}^t]^2$$
$$+\alpha[P_i^d/\eta_i^d (X_{d,i}^t - X_{d,i}^{t-1})]^2 + \beta[(P_i^d/\eta_i^d) X_{d,i}^t]^2, \ \forall i \in I_{v2g} \tag{3.6}$$

The equation (3.5) measures the degradation cost for the vehicles in G2V. The first term measures the degradation cost caused by fluctuations in the charging rate throughout the plug-in period, while the second part measures the cost of degradation associated with the charge rate. Constraint (3.6) is for vehicles in V2G mode and is quite similar to (3.5) except for the fact that it also takes the discharge rate into account.

In the third term in (3.3), a penalty δ is considered for wind curtailment. The wind curtailment denoted by Ω^t is the maximum of zero and wind production minus the charging demand at each period. $\Omega^t = \max\{0, W^t - D^t\}$ (see constraint (3.7)).

$$\Omega^t \geq W^t + \sum_{i \in EV_{v2g}^t} (X_{d,i}^t P_i^d) - \sum_{i \in EV_{all}^t} (X_{c,i}^t P_i^c), \ \ \forall t \in T \tag{3.7}$$

Note that the demand for wind energy in period t can be easily calculated by $W^t - \Omega^t$. It is worth mentioning that in each time interval t, at most one of Ω^t and G^t can be positive. Since the EV charge load is the only demand for wind energy, minimizing wind curtailment implies maximizing WU. Similar to λ, δ is a hyper-parameter that determines the weight for wind curtailment penalty. Considering a small value for δ, this term comes into play only when there is enough wind production, and it ensures to reduce the wind curtailment by charging the vehicle to their full battery capacity instead of the user specified desired level.

Assuming a linear charging behavior for the batteries, the state of charge is initialized and

updated by constraints (3.8)-(3.10).

$$SOC_i^{t_i^{arr}} = SOC_{init,i}, \quad \forall i \in I \tag{3.8}$$

$$SOC_i^t = SOC_i^{t-1} + \eta_i^c P_i^c X_{c,i}^{t-1}, \ \ \forall i \in I_{g2v}, \ \ \forall t \in T_i^p \tag{3.9}$$

$$\begin{aligned} SOC_i^t = SOC_i^{t-1} + \eta_i^c P_i^c X_{c,i}^{t-1} \\ -P_i^d X_{d,i}^{t-1}/\eta_i^d, \quad \forall i \in I_{v2g}, \ \ \forall t \in T_i^p \end{aligned} \tag{3.10}$$

Constraint (3.8) sets the initial state of battery charge upon arrival to the state of charge at arrival time ($SOC_i^{t_i^{arr}}$). SOC at each time is updated in constraints (3.9)-(3.10) by adding the charging energy for the vehicle in the current period to the charge level of the previous time. Constraint (3.11) limits the total grid supply by the transformer's capacity denoted by P_{max}^G.

$$G^t \leq P_{max}^G \quad \forall t \in T \tag{3.11}$$

The next two constraints (3.12)-(3.13) guarantee that if the vehicle cannot reach the desired level in its designated charging period, the vehicle is charged with full speed for the entire plug-in period. The auxiliary variable Z_i takes a value of 1 if the vehicle cannot reach the desired level, hence, constraint (3.13) imposes charge with full speed ($X_{c,i}^t = 1$).

$$SOC_{init,i} + P_i^c \eta_i^c (t_i^{dep} - t_i^{arr}) \geq SOC_{desired,i} - MZ_i \tag{3.12}$$

$$X_{c,i}^t \geq Z_i, \quad \forall t \in T_i^p \tag{3.13}$$

On the contrary, if the vehicle can reach the desired level, Z_i becomes zero, and constraint (3.14) ensures that the final state of charge (SOC at departure time) is at least as much as the desired level of battery charge requested by the owner.

$$SOC_i^{t_i^{dep}} \geq SOC_{desired,i} - MZ_i \tag{3.14}$$

Constraint (3.15) limits the state of charge at any time to the battery capacity (SOC_{cap}) of an EV.

$$SOC_i^t \leq SOC_{cap,i}, \quad \forall t \in \{t_i^{arr}, ..., t_i^{dep}\} \tag{3.15}$$

The next group of constraints sets a minimum level of charge for all vehicles. Constraint 3.16 restricts the state of charge by the minimum required value if the vehicle starts with an initial SOC higher than the SOC_{min}. If a vehicle arrives with a charge level less than the minimum level ($\forall i \in B$), then it has to charge with full speed to get to the minimum level for the first T_{min} periods (equation (3.17)). After reaching the minimum state of charge, the SOC should never drop below SOC_{min} (constraint (3.18)).

$$SOC_i^t \geq SOC_{min,i} \quad \forall i \in I \setminus B, \quad \forall t \in \{t_i^{arr}, ..., t_i^{dep}\} \tag{3.16}$$

$$X_{c,i}^t = 1, \quad \forall i \in B, \quad \forall t \in \{t_i^{arr}, ..., t_i^{arr} + T_{min,i}\} \tag{3.17}$$

$$SOC_i^t \geq SOC_{min,i} \; \forall i \in B, \; \forall t \in \{t_i^{arr} + T_{min,i}, ..., t_i^{dep}\} \tag{3.18}$$

$T_{min,i}$ denotes the minimum number of periods that EV i has to charge with full speed to reach $SOC_{min,i}$, and is calculated by (3.19).

$$T_{min,i} = \left\lceil \frac{SOC_{min,i} - SOC_{init,i}}{\eta_i^c P_i^c} \right\rceil \tag{3.19}$$

Constraints (3.20)-(3.21) set the charge/discharge rates to zero for the period prior to arrival.

$$X_{c,i}^{t_i^{arr}-1} = 0 \qquad \forall i \in I_{g2v} \tag{3.20}$$

$$X_{c,i}^{t_i^{arr}-1} + X_{d,i}^{t_i^{arr}-1} = 0 \quad \forall i \in I_{v2g} \tag{3.21}$$

Constraints (3.22)-(3.24) ensure that in any period during plug-in time, the vehicles in V2G mode can either charge, discharge, or do nothing. The variables $Y_{c,i}$ and $Y_{d,i}$ take value of 0 if the vehicle is charging and discharging, respectively. The constraint 3.24 dictates that at least one of the

charging and discharging rate has to be zero.

$$X_{c,i}^t \leq 1 - Y_{c,i}^t \quad \forall i \in I_{v2g}, \ \ \forall t \in T_i^p \tag{3.22}$$

$$X_{d,i}^t \leq 1 - Y_{d,i}^t \quad \forall i \in I_{v2g}, \ \ \forall t \in T_i^p \tag{3.23}$$

$$Y_{c,i}^t + Y_{d,i}^t = 1 \quad \forall i \in I_{v2g}, \ \ \forall t \in T_i^p \tag{3.24}$$

Finally, constraints (3.25)-(3.27) specify the binary and non-negativity constraints for the decision variables.

$$0 \leq X_{c,i}^t \leq 1 \quad \forall i \in I, \ \ \forall t \in T_i^p \tag{3.25}$$

$$0 \leq X_{d,i}^t \leq 1 \quad \forall i \in I_{v2g}, \ \ \forall t \in T_i^p \tag{3.26}$$

$$G^t, \Omega^t \geq 0 \quad \forall t \in T \tag{3.27}$$

Although, the assumptions made in this deterministic case were somewhat impractical, this model provides the global optimal solution for the case where all model parameters are known. It also provides a baseline for the dynamic model which is described in the next section. In the dynamic model case, some of the impractical aspects of the static model are relaxed.

3.4 Dynamic Scheduling Optimization

In a more realistic scenario, the assumption for the obligation of providing perfect information a day ahead by the EV owners is relaxed. In this dynamic model, EV owners input their needs and preferences (departure time, desired level of battery charge, minimum required level of charge) upon arrival. The smart chargers automatically detect the necessary EV characteristics, such as battery capacity (SOC_{cap}), acceptance rate, and state of charge (SOC). At any planning time j, the aggregator batches all the vehicles that have arrived during the $[j-1, j]$ time slot. The aggregator also considers those vehicles that have not departed from the previous periods and are still in charging. With updated information regarding the number of EVs and their requirements, the renewable energy generated, and the price of electricity, the scheduling algorithm (discussed

in 3.4.1) optimizes the charging schedule for the current planning window (rolling window). The planning window is defined by the period from time j till the time that all vehicles in set E^j departs. E^j is the set of vehicles considered in planning at time j. This schedule is called dynamic because it can be updated as wind production forecast, the electric price forecast, and EV availabilities are updated. Also, note that the charging schedule of all vehicles gets updated frequently at each planning time until they depart, thus minimizing the effect of uncertainty in wind generation and electricity price. The uncertainty of wind generation forecast and electricity price forecast is considered in section 3.5.2.3, but for now, the assumption is that the forecasts are perfectly accurate, which provides the baseline model that will be used for performance evaluation.

3.4.1 Algorithm

Considering 1-hour planning intervals Δj, a day is divided equally into 24 intervals so that planning times are at each exact hour of the day. For ease of notation, $j+1$ denotes the next planning time, which is one hour after current planning time j. However, to be consistent with the notation of our decision periods t, which increment by one every $\Delta t = 15$ minutes, another variable is defined $\phi_j = 4j$. For instance, $j = 0 \implies \phi_j = 0$ denotes planning at time 12:00 AM, and $j = 1 \implies \phi_j = 4$ denotes the next planning time at 1:00 AM.

At any planning time j, there is a set of EVs that arrived during time slot $[j-1, j]$, called A^j. Vehicles in A^j are added to the set E^j, which accounts for the vehicles that need to be planned at time j. The aggregator also batches those vehicles that have already arrived and planned in the previous period $j-1$ (vehicles in set E^{j-1}) if the vehicle remains plugged-in after time j. Based on the owners' input, the aggregator determines the planning window, which is from time j to the time the last vehicle departs, denoted by τ^j_{max}. It is defined as $\tau^j_{max} = \max\{t_i^{dep} |\ \ i \in E^j\}$. As an example, consider planning for EVs at $j = 8$ (8:00 AM). In Fig. 3.4, arrival and departure times for vehicles {1,2,3} are shown by down arrows and up arrows, respectively. Set $A_8 = \{1, 2, 3\}$ is the set of EVs arriving in period $j =$ [7:00 AM, 8:00 AM]. Assuming that there are no arrivals before 7:00 AM, set E_8 is equal to A_8. The end of planning window, τ^8_{max}, is defined by $\tau^8_{max} = \max\{t_i^{dep} |\ \ i \in E_8\}$. The last departure is when vehicle 2 departs at 2:00 PM ($t_2^{dep} = 56$). Thus,

the planning window is from $\phi_8 = 32$ to $\tau^8_{max} = 56$.

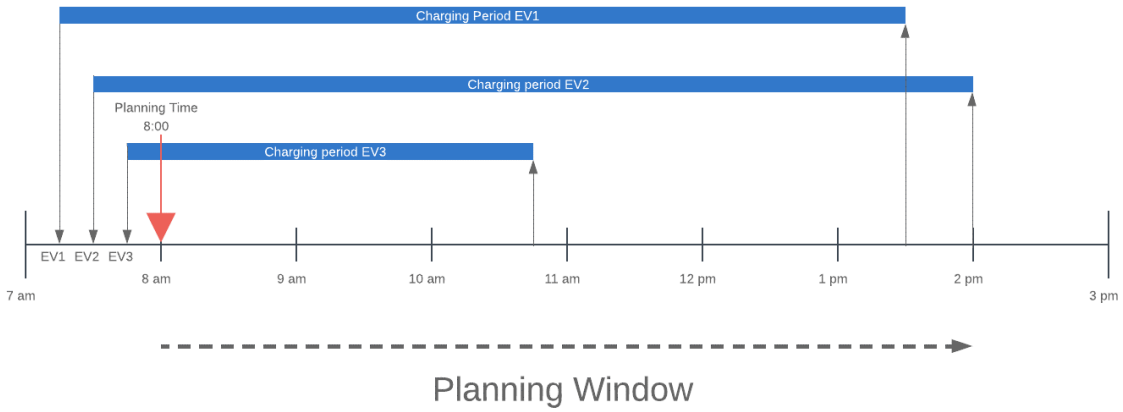


Figure 3.4: Illustration of planning (rolling) window in system with three arrivals. Reprinted with permission from (160)

Since the arrivals after j are unknown, the average charge demand of future arrivals for the current planning window should be estimated. To estimate the amount of charge, one needs to calculate the average number of EVs in each Δt period, the average charge required, and the average plug-in period by analyzing the historical data.

$$E[D^t_f] = \frac{\hat{ER} \cdot \hat{N^t_j}}{\hat{PT}} \tag{3.28}$$

where $\hat{ER}$ is the estimated charge required by an EV, and it can be calculated by $\hat{ER} = \hat{SOC}_{des} - \hat{SOC}_{init}$. After finding the set of vehicles to plan at time j, their state of charge, the charge/discharge rate for the period prior to j, and the estimated future charge demand, the aggregator runs the optimization algorithm that solves the mixed-integer quadratic programming (MIQP) problem (described in 3.4.2) to find the optimal charging procedure for each EV. The aggregator repeats this process for the next planning time $j + 1$. This repetitive algorithm will run for the planning horizon J, which can be from a few hours to a couple of years. The steps of the rolling horizon approach are provided in Algorithm 1.

Algorithm 1 Rolling Horizon Algorithm

Initialize $E^j = \emptyset, \ \ \forall j \in J + \{0\}$, and set $j = 1$.
for $j \in J$ **do**
  $E^j \leftarrow A^j$
  for $i \in E_{j-1}$ **do**
    if $t_i^{dep} > \phi_j$ **then**
      Add i to the E^j. $E^j \leftarrow E^j + \{i\}$
    end if
  end for
  for $i \in E^j$ **do**
    Set $LC_i^{\{j\}}, LD_i^{\{j\}}$ as the charge/discharge rate in the previous period.
    Update the $SOC_{init,i}$ as the SOC of EV i at time j.
  end for
  Update the uncertain parameters (W^t and pr^t) with their recent forecast (W_f^t and pr_f^t).
  Set $B = \{i| \ \ i \in E^j, \ \ SOC_{min,i} > SOC_{init,i}\}$
  Find the end of current planning window by $\tau_{max}^j = \max\{t_i^{dep}| \ \ i \in E^j\}$
  Calculate the expected charge demand of future arrivals from (3.28).
  Run the optimization algorithm described in 3.4.2.
  Charge and discharge the EVs according to optimal X_c, X_d values.
end for

3.4.2 Modeling & Mathematical Formulation

At each planning time j, the algorithm needs to solve a MIP or MIQP problem (depending on the battery degradation model) similar to the static case. The optimization problem at planning time j considers the following objective function:

$$\min \quad \sum_{t \in T_j} pr_f^t G^t + \sum_{i \in E^j} \lambda \Psi_i + \sum_{t \in T_j} \delta \Omega^t \tag{3.29}$$

The objective function (3.29) is similar to the static case (3.3), except for the fact that it minimizes the cost for the planning window, which is $T_j = \{\phi_j, ..., \tau_{max}^j - 1\}$ instead of the entire planning horizon T. Here, τ_{max}^j is the end of planning window and is defined by $\tau_{max}^j = \max\{t_i^{dep}| \ \ i \in E^j\}$. Also, note that the total plug-in period (T_i^p) is no longer from arrival time to

departure time; it is from planning time j till departure. Thus, T_i^p is updated by $\{\phi_j, ..., t_i^{dep} - 1\}$.

$$\Psi_i = \sum_{t=\phi_j}^{t_i^{dep}-1} \alpha(\eta_i^c P_i^c (X_{c,i}^t - X_{c,i}^{t-1}))^2 + \beta(\eta_i^c P_i^c X_i^t)^2, \forall i \in E_{g2v}^j \tag{3.30}$$

$$\Psi_i = \sum_{t=\phi_j}^{t_i^{dep}-1} \alpha[\eta_i^c P_i^c (X_{c,i}^t - X_{c,i}^{t-1})]^2 + \beta[\eta_i^c P_i^c X_{c,i}^t]^2 + \alpha[P_i^d/\eta_i^d (X_{d,i}^t - X_{d,i}^{t-1})]^2 + \beta[(P_i^d/\eta_i^d) X_{d,i}^t]^2, \quad \forall i \in E_{v2g}^j \tag{3.31}$$

$$G^t \geq \sum_{i \in EV_{all}^t} (P_i^c X_{c,i}^t) + E[D_f^t] - \sum_{i \in EV_{v2g}^t} (P_i^d X_{d,i}^t) - W_f^t \tag{3.32}$$

$$\Omega^t \geq W_f^t + \sum_{i \in EV_{v2g}^t} (P_i^d X_{d,i}^t) - \sum_{i \in EV_{all}^t} (P_i^c X_{c,i}^t) - E[D_f^t] \tag{3.33}$$

The problem is subject to a set of constraints, most of which are similar to the static case with a few modifications. Since planning occurs at time j, only vehicles in the set E^j from time ϕ_j to τ_{max}^j are included in the model. Constraints (3.34)-(3.35) restore the charging and discharging rates for the last period prior to the planning time j.

$$X_{c,i}^{\phi_j - 1} = LC_i^{\{j\}}, \qquad \forall i \in E^j \tag{3.34}$$

$$X_{d,i}^{\phi_j - 1} = LD_i^{\{j\}}, \qquad \forall i \in E_{v2g}^j \tag{3.35}$$

The rest of the constraints are similar to the static case. One needs to consider the modifications mentioned above and replace I with E^j.

$$SOC_i^{\phi_j} = SOC_{init,i}, \quad \forall i \in E^j \tag{3.36}$$

$$SOC_i^t = SOC_i^{t-1} + \eta_i^c P_i^c X_{c,i}^{t-1}, \quad \forall i \in E_{g2v}^j, \quad \forall t \in \{\phi_j + 1, ..., t_i^{dep}\} \tag{3.37}$$

$$SOC_i^t = SOC_i^{t-1} + \eta_i^c P_i^c X_{c,i}^{t-1} - P_i^d X_{d,i}^{t-1}/\eta_i^d, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in \{\phi_j + 1, ..., t_i^{dep}\} \tag{3.38}$$

$$SOC_{init,i} + P_i^c \eta_i^c (t_i^{dep} - \phi_j) \geq SOC_{desired,i} - MZ_i, \quad \forall i \in E^j \tag{3.39}$$

$$X_{c,i}^t \geq Z_i, \quad \forall i \in E^j, \ \ \forall t \in T_i^p \tag{3.40}$$

$$SOC_i^{t_i^{dep}} \geq SOC_{desired,i} - MZ_i, \quad \forall i \in E^j \tag{3.41}$$

$$X_{c,i}^t \leq 1 - Y_{c,i}^t, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in T_i^p \tag{3.42}$$

$$X_{d,i}^t \leq 1 - Y_{d,i}^t, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in T_i^p \tag{3.43}$$

$$Y_{c,i}^t + Y_{d,i}^t = 1, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in T_i^p \tag{3.44}$$

$$SOC_i^t \geq SOC_{min,i}, \quad \forall i \in E^j \setminus B, \ \ \forall t \in T_i^p \tag{3.45}$$

$$X_{c,i}^t = 1, \quad \forall i \in B, \ \ \forall t \in \{t_i^{arr}, ..., t_i^{arr} + T_{min,i}\} \tag{3.46}$$

$$SOC_i^t \geq SOC_{min,i}, \quad \forall i \in B, \ \ \forall t \in \{t_i^{arr} + T_{min,i} + 1, ..., t_i^{dep}\} \tag{3.47}$$

$$SOC_i^t \leq SOC_{cap,i}, \quad \forall i \in E^j, \ \ \forall t \in \{\phi_j, ..., t_i^{dep}\} \tag{3.48}$$

$$0 \leq X_{c,i}^t \leq 1, \quad \forall i \in E^j, \ \ \forall t \in T_i^p \tag{3.49}$$

$$0 \leq X_{d,i}^t \leq 1, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in T_i^p \tag{3.50}$$

$$Y_{c,i}^1, \ Z_i \ \ Binary, \quad \forall i \in E^j, \ \ \forall t \in T_i^p \tag{3.51}$$

$$Y_{d,i}^t \ \ Binary, \quad \forall i \in E_{v2g}^j, \ \ \forall t \in T_i^p \tag{3.52}$$

$$\Omega^t, \ G^t \geq 0, \quad \forall t \in \{\phi_j, ..., \tau_{max}^j - 1\} \tag{3.53}$$

Considering perfect forecasts for wind and price, the problem is a deterministic mean-value optimization in which the only uncertain element is the future EV load. This problem can also be solved using multi-stage SP approaches. However, multi-stage SP approaches require generating a large number of scenarios (scenario tree) and often suffer from curse of dimensionality. A two-

stage approximation method can mitigate the curse of dimensionality by only considering the stochastic scenarios for the next stage $(t+1)$ while fixing the value of uncertain parameter for the stages $t+2, ..., T$.

3.5 Results

In this section, a set of comprehensive simulation experiments are performed to examine the performance of the proposed controlled charge/discharge scheduling algorithm.

3.5.1 Simulation Settings

In the simulation, a day is evenly divided into 96 time intervals of $\Delta t = 15$ minutes. The algorithm's goal is to decide the battery charge/discharge rate of EV i during interval t ($X^t_{c,i}$, $X^t_{d,i}$). *Generation & Consumption data:* The variations for electricity price and wind power generation are captured in 10 different scenarios (10 months of the year). Hourly wind power production is simulated using the Grid Lab System Advisor Model (SAM) in northern California for a single wind turbine based on specifications of Endurance X33 turbine with 230 kW power capacity for 10 consecutive days for all 10 scenarios (167). Hourly electricity price is collected using historical Locational Marginal Price (LMP) data for day-ahead market at node "PLAINFLD_6_N001" from the California Independent System Operator (CAISO) for the same days (1). The 10 scenarios include five months in Spring and Summer seasons, and five months in Fall and Winter seasons. Thus, the simulation consists of a total of 100 days, with each day having random arrival and departure scenarios as well as different wind and electricity price profiles. The discharged energy of an EV is sold with a price of 95% of the real-time electricity price. The proposed selling price is high enough to motivate the owners to participate in V2G as they can make profit by charging their EVs during low-price periods, and discharging when the prices are higher. Since the revenue earned by the EV owners in discharge mode is equal to the charge cost of discharged energy for the EVs in charge mode, the revenue term is not included in the mathematical formulation.
EV-related data: Considering a total of 100 EV trips per day, the battery and charging characteristics of EVs are based on the specifications of different EVs available in the market in 2018 (168). The

EV battery info is summarized in Table 3.2. The EV arrival times are captured using data from the National Household Travel Survey (NHTS) (169). It is assumed that 50 of the EV charging occur at workplace and the other 50 are vehicles parked at home. Vehicles are assumed to be plugged-in as soon as they arrive. To simulate the EV fleet's arrival times, the travel time data related to trips with home or work purposes from NHTS is used. Figure 3.5 shows the probability of trip to home for each hour of the day. In figure 3.6, the x-axis represents hour of the day, and the probability of trip to workplace is on the y-axis. In each scenario of the simulation, the optimal schedule algorithm is run for 10 days. The EV plug-in period is modeled as discrete uniform distribution between 4 and 12 hours with a resolution of 15 minutes. The desired level of battery at departure is modeled using uniform distribution in the range of 0.75 to 0.95 of battery capacity. A charging and discharging efficiency of 90% is assumed. The initial state of charge is also assumed to follow a uniform distribution between 0 and 0.65 of the battery capacity, while the minimum required state of charge (SOC_{min}) for all vehicles is assumed to be 5 kWh, which in case of emergency seems sufficient to reach a distance of approximately 25 miles. For battery degradation cost, as discussed in (144), values of 0.05 and 0.1 ¢ are used for α and β, respectively.

Table 3.2: EV characteristics. Reprinted with permission from (160)

Vehicle	Acceptance Rate AR	Battery Size SOC_{cap}	Charger Capacity CP	Battery Cost($)
BMW i3 2017	7.4	32	7.7	4,640
Ford Focus EV 2017	6.6	33.5	7.7	4,850
Nissan Leaf S 2016	6.6	24	7.7	3,500
Tesla Model S 70 Single	9.6	70	11.5	10,150
Tesla Model S 90 Dual	19.2	90	15.4	13,000

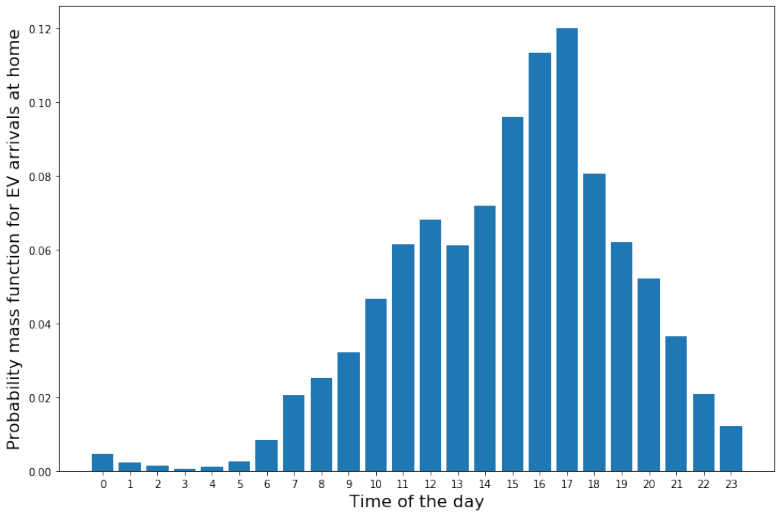


Figure 3.5: Distribution of arrivals at home. Reprinted with permission from (160)

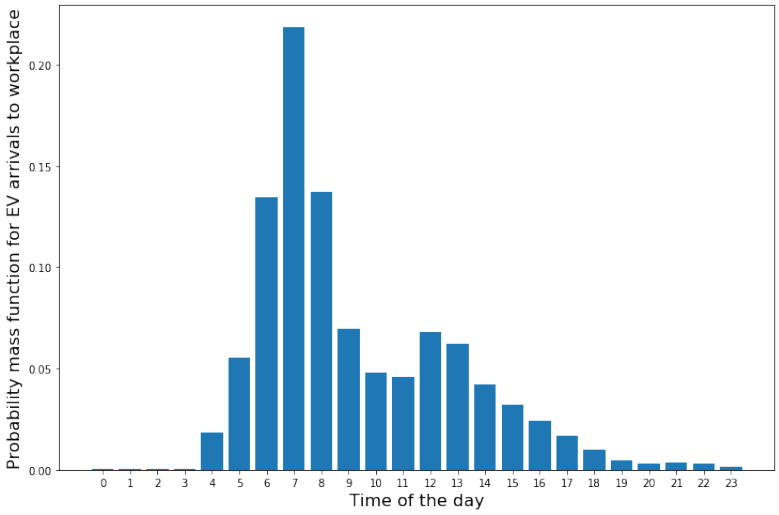
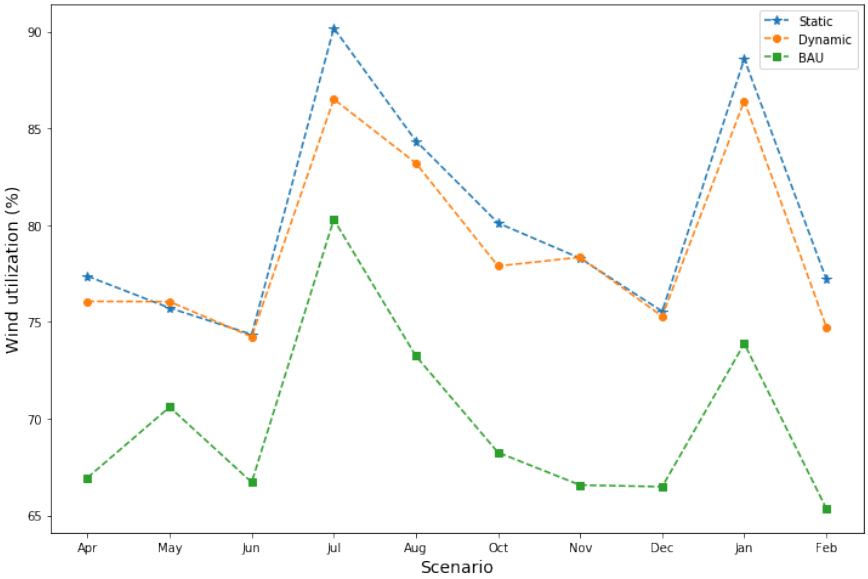


Figure 3.6: Distribution of arrivals at workplace. Reprinted with permission from (160)

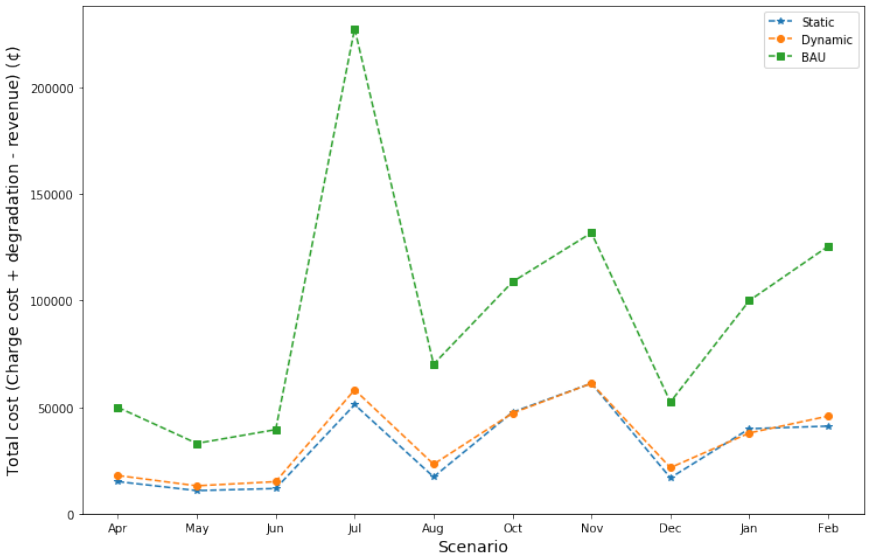
3.5.2 Simulation Results and Performance Analysis

To evaluate the performance of the proposed approach, the results of the dynamic charging algorithm are compared with the static and BAU charging scenario. All models are run for the 10 scenarios mentioned above. The simulation is coded in Python 3.7 using Gurobi optimizer. The results for optimal solutions are shown in Fig. 3.7. Comparing the results, the proposed charging

algorithm shows significant improvement in all the objective measures including WU and TCC for EV owners. TCC can be calculated as sum of charging cost and degradation cost minus revenue earned from selling discharging energy. Note that charging cost consists of the cost of energy from external grid as well as cost of energy purchased from EVs in discharge mode.



(a) Wind utilization (%)

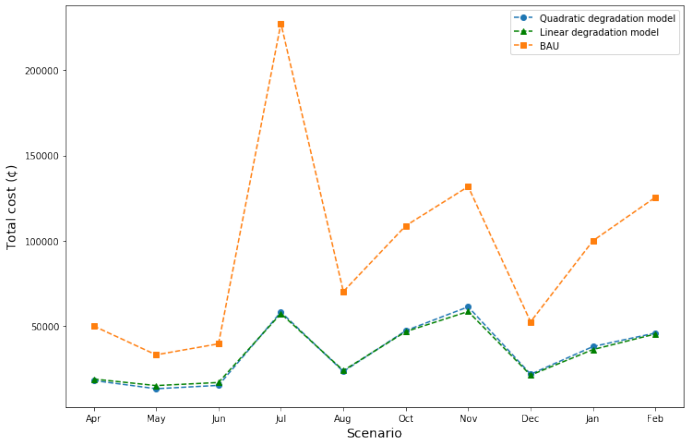


(b) Total cost (¢)

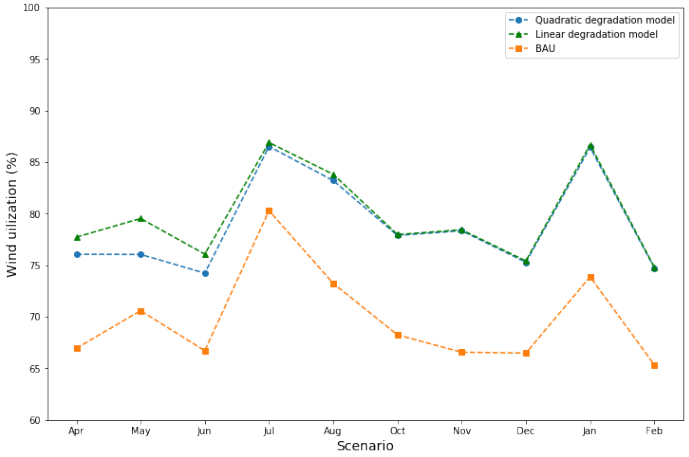
Figure 3.7: Performance evaluation for all three charging cases. a) wind utilization, and b) total cost. Reprinted with permission from (160)

The results demonstrate that the dynamic model achieves a similar performance compared to the static scheduling model. It is worthwhile to mention that the performance of the static case for some objective metrics is worse than the dynamic case for a few scenarios due to the fact that the problem is formulated as a multi-objective optimization. However, the total objective value for the static case is always better than the dynamic case.

To validate our degradation model, figure 3.8 shows that considering the linear degradation model described in section 3.2.3 leads to similar results compared to the quadratic model, which indicates that our model is robust with respect to different degradation cost functions.



(a) Total cost



(b) Wind utilization

Figure 3.8: Comparison of the models with quadratic vs linear degradation function. Reprinted with permission from (160)

3.5.2.1 Discussion of Objective Function Hyper-Parameters

Considering that the scheduling problem is a weighted multi-objective optimization problem, it is required to evaluate the performance of the model for different values of hyper-parameters. The values of δ and λ determine the importance of the corresponding term in the objective function.

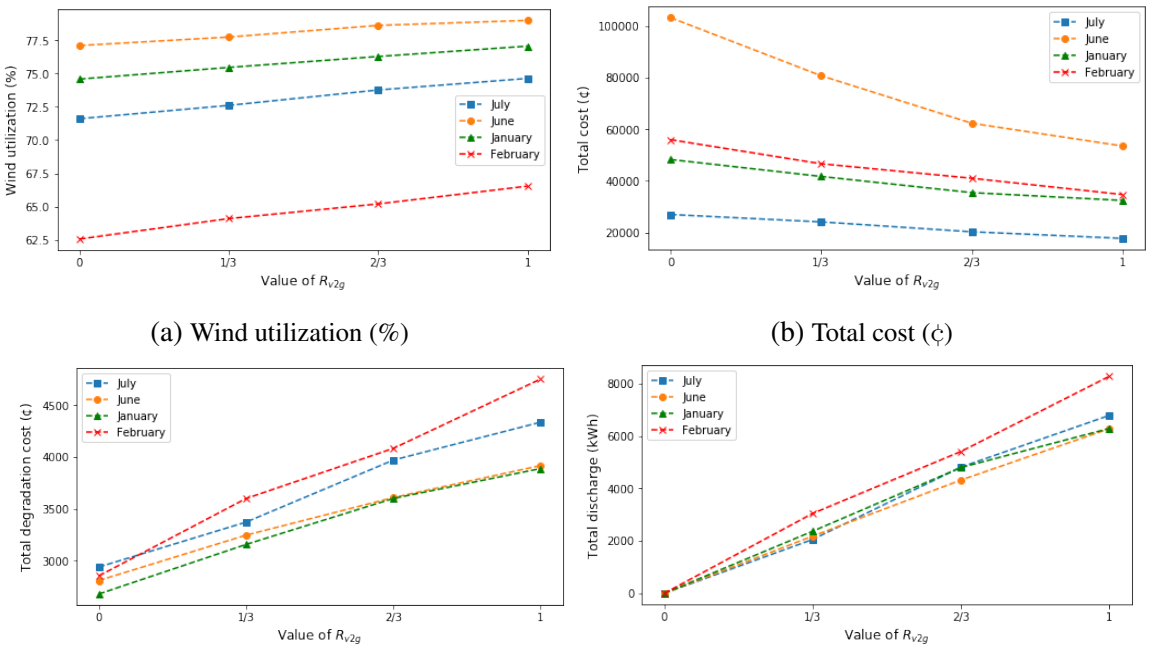
Performance evaluation under different values of δ: A constant weight for wind curtailment does not capture the dynamics of our model well. Thus, the penalty should be relative to the electricity price such that wind curtailment penalty and grid supply have the same ratio through all periods. Here, δ is considered as multiplier of pr^t in the third term of objective function. A small value of δ provides a small weight to wind curtailment minimization, while making sure that grid supply is as low as possible. If a large value is assigned to δ, wind curtailment is penalized more causing to increase WU at the cost of higher discharged energy and higher degradation cost, which results in higher total charging cost. If WU is of top priority, then a high value should be used for δ, and vice versa. Experimentally, a value of 0.25 seems to have the best performance among all options.

Performance evaluation under different values of λ: The value of λ indicates the tolerance of EV owners for their battery degradation. A value of 0 means high tolerance, while value of 1 means battery degradation cost is as important as charging cost for the owner. A low value leads to higher WU at the cost of more degradation and TCC.

3.5.2.2 V2G Benefits

To assess the benefits of V2G in this MG, different scenarios are compared in which the percentage of vehicles participating in V2G versus G2V varies. Different values of R_{v2g} are used to evaluate the benefits of V2G to both the owner and the grid. The performance evaluation under different values of R_{v2g} is shown in Fig. 3.9. In order to maintain clarity, the plots are shown for four months (scenarios) instead of all ten scenarios. In case where all vehicles participate in V2G, the grid benefits from increasing the WU and reducing the total energy curtailment. From the owners perspective, the degradation cost increases as the vehicles discharge their energy; however, TCC decreases mainly due to the fact that the energy purchased from discharged energy reduces

the need for expensive supply from the external grid. In V2G case, the EVs store energy from wind source and in low wind periods, they inject energy back to the system to charge other EVs.



(a) Wind utilization (%)

(b) Total cost (¢)

(c) Total degradation cost (¢)

(d) Energy discharged (kWh)

Figure 3.9: Performance evaluation under different values of R_{V2G}. a) wind utilization, b) total cost, c) degradation cost, d) energy discharged. Reprinted with permission from (160)

3.5.2.3 Wind & Price Forecast Uncertainty

In reality, the wind and LMP forecasts for the following couple of hours are not perfect. Thus, this section considers a model where the updated forecast is only accurate for the current planning interval. If planning at time t, wind and LMP are known only till time $t+1$ and are unknown for periods beyond. The stochastic data $\xi_t = \{\tilde{w}^t, \tilde{pr}^t\}$ is observed only at time t (Figure 3.10). Hence, to use the stochastic programming methods when planning at time j for the rolling horizon $\{j, ..., \tau_{max}\}$, one needs to generate scenarios for all planning times $\{j+1, ..., \tau_{max}\}$. This leads to an explosion in number of scenarios even if the underlying distributions of wind and price are known. Given that wind power is a non-stable non-linear random process, it does not follow any

known probability distribution. Thus, the mean-value optimization approach is used to handle the uncertainty and it is believed that in the dynamic model, since the scheduling of EVs are updated very often (every one hour), the effects of intermittent wind generation and price volatility are significantly reduced and the model can accommodate the forecast uncertainty to a higher degree. In other words, the aggregator can make up for the sub-optimal decisions made in previous planning times and obtain a close to optimal solution. To prove this point, let us consider a model with perfect forecast for the next one hour (Δj), beyond that, a discrete time Markov Decision Process (MDP) model is applied to estimate the wind generation (170). In a simple case, the state space for wind production is discretized into 20 states and the transition probability is estimated based on historical (training) data. The forecasts for the next k hours are estimated by

$$W_f^{t+k} = \sum_{w \in S_w} P^k_{W^t,w} \cdot w, \quad k = \{1, 2, 3, ...\} \tag{3.54}$$

where, the state space for wind energy is denoted by S_w, and consists of 20 scenario representatives. The 30 days prior to the simulation start date are used as training to calculate the transition probability matrix. The problem is optimized for the next 10 days with MDP wind forecast (W_f). In addition, for predicting the real-time LMP, a similar-day approach is used, which the price in each interval of the rolling window is predicted as the average of the LMPs for the same interval of 10 days with similar weather conditions. The results are shown in Fig. 3.11, where it is observed that the dynamic model with imperfect forecast performs similar to the perfect forecast model. Considering the total of 10 scenarios (100 days), the paired t-test is used to compare the performance of models for WU and TCC metrics. It indicates that the performance of the model with imperfect forecast is significantly better than the BAU scenario (P-value < 0.00001 for WU, and 0.008 for TCC). For proper evaluation of the mean-value dynamic model, it is required to measure the expected value of perfect information (EVPI). The estimate of the EVPI calculated by the results is 0.9% for WU, meaning that in case of perfect forecast, the WU is only 0.9% higher than the mean-value model. The EVPI estimate for TCC is less than 1000¢for 100 EVs over the

course of 5 days. Even though solving this problem using multi-stage SP approaches might lead to slightly better results, it requires more computational time. The value of stochastic solution (VSS) is definitely less than EVPI value; thus, the low value of EVPI suggests that there might be very little gain in solving the problem using the SP methods and it does not justify the need to sacrifice the computational time for less than 0.9 WU increase. Also, note that a very simple MDP process is used as wind forecasting tool that does not result in highly accurate forecasts. A better forecast for wind generation (e.g., using recurrent neural network) is likely to further improve the quality of solutions. Moreover, one can easily solve a two-stage approximation of the scheduling problem using sample average approximation (SAA) and formulating the equivalent deterministic problem if a sample of random vector ξ can be generated. The only concern in using the SAA approach is the possibility that the second-stage problem is infeasible for some scenarios. However, one can always make the second-stage problem feasible by including the infeasible constraint in the objective function with a penalty parameter.

Another major reason behind using the mean-value approach instead of SP methods is the computational time. Since the problem is a real-time scheduling problem, computational time is of importance and in case of a large number of scenarios, the problem becomes computationally very expensive and thus, not suitable for real-time scheduling. The average solving time for all the simulation scenarios is around 5 seconds running the algorithm using Gurobi Optimizer in Python 3.7 on a 64-bit operating system with a 2.80GHz processor. Tested on a problem with 200 EV trips per day and two wind turbines with the same characteristics, it takes about 11 seconds to solve the problem, which is acceptable.

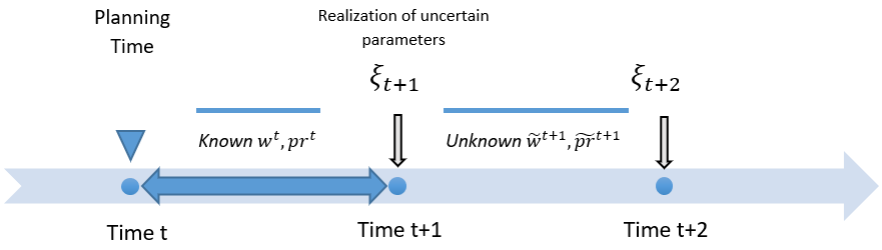
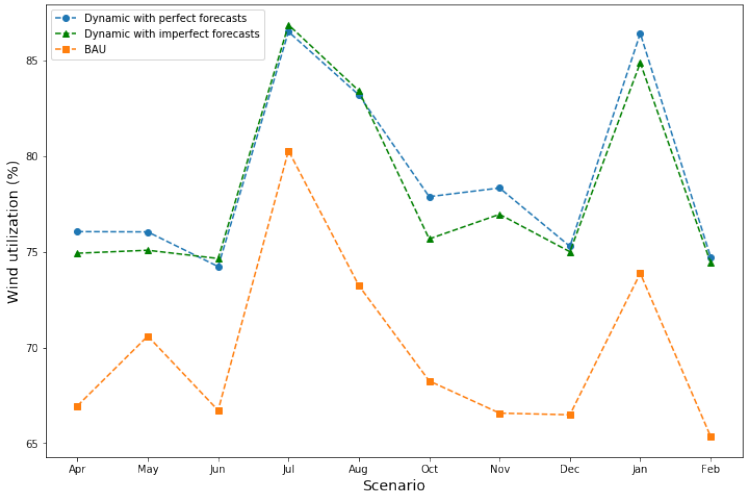


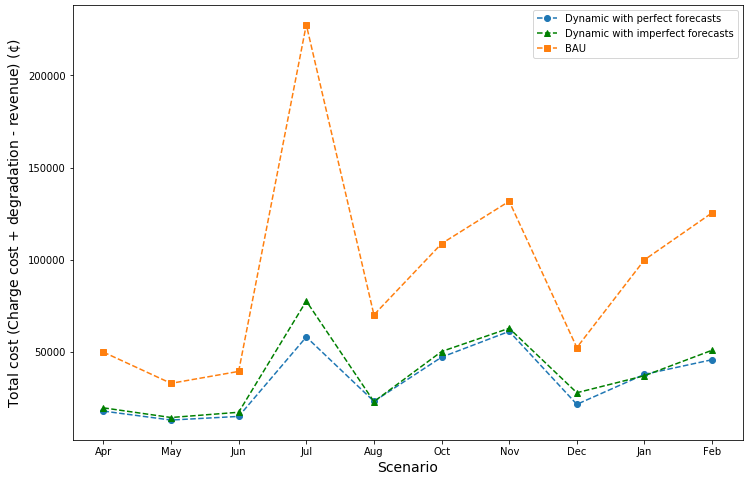
Figure 3.10: Uncertainty realization when planning at time t. Reprinted with permission from (160)

3.6 Conclusion

In this chapter, an optimization algorithm for EV charge/discharge scheduling is proposed to support wind energy integration using a rolling horizon optimization method. The problem is formulated as a MIQP and the results show significant improvement compared to the BAU case in terms of TCC, WU, and demand from conventional generators. The main contribution here is to design the scheduling algorithm that aggregators can leverage the presence of advanced communication technology in smart grid, flexibility of EV drivers, and the V2G technology to support high integration of intermittent wind energy into the power system. To mitigate the barriers in people participation in V2G, the battery degradation, minimum required level of charge, and/or financial incentives for the EV drivers are considered. The financial incentive for the owners is that they can make charge their EVs with cheap wind power and make revenue by selling the excessive energy. Furthermore, in this work, a multi-objective optimization is considered to maximize WU, minimize the demand from conventional generators, and minimize TC while satisfying the drivers' requirements. A simulation of the proposed algorithm for different scenarios of EV characteristics, arrivals, departures, and charging requirements are performed to check the quality of solutions and schedules. The results show that the proposed model leads to significant improvement in all metrics, and benefits both the owner and the grid. Moreover, the results indicate that frequent updates of wind power forecast (and LMP) in the deterministic problem significantly



(a) Wind utilization (%)



(b) Total cost (¢)

Figure 3.11: Comparison of dynamic models with perfect forecast vs imperfect forecasts. Reprinted with permission from (160)

reduce the effects of forecast uncertainty. Future research is needed to consider a large-scale grid with multiple generators and consumers with the presence of battery storage and demand side management for residential load to evaluate the users flexibility in the power system. A financial incentive framework should also be developed to encourage a larger participation in V2G.

4. STOCHASTIC MODELING

4.1 Introduction

As discussed in the previous chapter, obtaining an optimal schedule for real-time charge/discharge of EVs is challenging due to the high degree of uncertainty in the power system, such as wind forecast, electricity price forecast, commuting behavior of EV drivers, and different EV characteristics and owners' requirements.

Let us take a look at the scheduling problem from a stochastic optimization perspective. At each planning time, a decision should be made followed by a set of observations of uncertain parameters, which are revealed over time. If planning at time j, wind and LMP are known till time $j+1$ and they are unknown for periods beyond. Moreover, the arrivals of future EVs, their charging demand, and the characteristics of EVs (i.e, power rate, departure time, desired level) are also unknown at time j. The stochastic data $\xi_j = \{\tilde{w}^j, \tilde{pr}^j, E^j, \{E_i^{req}, P_i^c, SOC_{cap,i}, t_i^{dep} \quad \forall i \in E^j\}\}$ is observed only at time j (Figure 3.10). E^{req} is the the remaining energy required to reach the desired level and is calculated by $E^{req} = SOC_{desired} - SOC$. Other notations are already defined in Chapter 3.

This scheduling problem is a classic example of a multi-stage stochastic problem, where at each planning time (stage) a decision needs to be made given the uncertainty in the future time steps or stages. To use conventional stochastic programming (SP) approaches for obtaining the optimal scheduling of EVs at time j, one needs to generate scenarios of all stochastic elements for all decision periods $\{j+1, ..., \tau_{max}^j\}$. In the scenario-based SP approach, uncertain parameters are assumed to have a specific probability description. For instance, wind power forecast error (WPFE) is usually modeled as normal (114; 115), or truncated normal distribution (117). Note that other than intermittent uncertain wind power, the greatest uncertainty in EV scheduling is the commuting behavior of the drivers as well as the vehicles characteristics. Therefore, a significant challenge of the scenario-based approach is to develop an appropriate probability distribution or weights on scenarios. Besides, probability distributions usually fail to represent the actual

uncertain parameter (random variables) as is the case for wind production because wind power is a nonlinear and non-stable process that does not fit into any known probability distribution (102). Moreover, considering the heterogeneity and different patterns of commuting behavior of EV owners (102; 159), scenario generation for EV arrivals and EV-related characteristics is also really challenging if not impossible.

Even if the distribution of uncertain parameters are known, the number of scenarios increases significantly with the number of stages considering the multi-stage nature of the problem. Assuming the rolling horizon containing K stages, N possible scenarios for each of the uncertain parameters at any stage, and M uncertain parameter, a total of $K^{N\times M}$ scenarios needs to be generated. Thus, generating a scenario tree for a K-stage stochastic problem with multiple uncertain parameters suffers greatly from the curse of dimensionality. In addition, SP approaches suffer from high computational time and in the case of real-time scheduling, where decisions need to made every few minutes, SP might be computationally too expensive to be practically acceptable (e.g., decision periods are every 15 minutes).

In summary, the model-based SP approaches are not suitable for such real-time scheduling problem due to the following reasons:

- Setting up a model-based method includes selecting accurate models and estimating the model parameters. Given the heterogeneity of the EV driving patterns and the EV-related features, each EV might need a different model and model parameters to capture the uncertainty.
- Determining the underlying distribution of wind and electricity price is very complex due to the high volatility of real-time price and nonlinear, non-stable nature of wind speed.
- Even if the underlying distributions are known, since there are so many uncertain elements, constructing the scenario tree for the multi-stage problem and assigning probabilities to different trajectories are complicated.
- SP methods suffer from curse of dimensionality and hence are computationally very expensive, which makes them impractical for real-time scheduling.

There are some approaches to solve multi-stage SP that makes it less computationally expensive and tractable. As a simple solution to solve a multi-stage stochastic problem, deterministic optimization approaches that are based on point forecasts of uncertain elements (e.g., mean-value or expected-value problem) can be used. However, deterministic mean-value (MV) optimization do not result in global optimal charging strategy. Another approach is two-stage approximation (TSA) with rolling horizon, where only the second-stage uncertainty is modeled using generated scenarios, and for stages after that the uncertain elements are assumed to be known (e.g., mean value). Generally, if the scenarios are properly constructed, then TSA reaches a better solution than MV problem at the cost of higher computational time. However, neither TSA nor MV result in a global optimal charging strategy.

The deterministic approach implemented in Section 3.5.2.3 was developed based on the expected value or mean value of the uncertain parameters. As stated before, the deterministic approach is easy and quick to solve, but might be far from the optimal solution of the true stochastic problem. In this chapter, the goal is to develop an approach that gives better results than a MV problem and can be applied in real-time settings without having to generate a large number of scenarios.

In contrast to model-based approaches, model-free Reinforcement Learning (RL) methods are based on pure trial-and-error without the need for a model of the system. Recent advancements in the field of Artificial Intelligence (AI) have made model-free RL a great approach to solve complex problems. The number of research articles on applying RL to power system scheduling, specially demand response problems, has increased significantly in recent years (165). However, standard RL methods (i.e., Q-learning) suffer from slow convergence and curse of dimensionality when applied to a large-scale problem with large state/action space, hence require a large number of trials to converge. Moreover, there is a chance that Q-learning converges and gets trapped into a local optimal solution, rather than the global optimal. Inspired by these issues, this chapter proposes an approach that uses the forecasts of exogenous data (wind and price) as well as the results of the MV optimization problem, to improve the performance of RL. The proposed approach also mitigates some of the challenges of the RL approach, such as large state space, large action space,

feature engineering, and high-variance reward (cost) in dynamic environments by smoothing the learning process for the RL agent using the results of deterministic MV optimization to estimate the action-value function. The proposed approach, which is referred to as combined RL-optimization method, can be applied in many other multi-stage stochastic optimization problems, where building an accurate model and/or a scenario tree to capture all the uncertainties in the system is challenging and conventional SP methods are computationally unacceptable.

The structure of this chapter is as follows. First, the background in RL and literature review of RL in EV charge/discharge scheduling are provided in Section 4.2. The problem formulation and the proposed approach are presented in Section 4.3. The combined RL-optimization algorithm and the experimental results are presented in Section 4.4. Section 4.5 provides a heuristic approach in conjunction with the proposed RL-optimization method to solve a large-scale EV scheduling problem. Section 4.6 discusses the application of the proposed approach in online learning and further ideas to improve the approach. Finally, a conclusion of this chapter and future research are discussed in Section 4.7.

4.2 Background & Literature Review

This section presents the background in RL and briefly reviews the literature on application of RL in EV charge/discharge scheduling.

4.2.1 Reinforcement Learning

Reinforcement Learning is an AI algorithm in which an agent learns the optimal policy that can maximize long-term additive reward through interacting with the environment (171). The interaction between the agent and the environment in RL is usually formalized by a Markov Decision Process (MDP) containing a tuple of elements $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$, where $\mathcal{S}$ is the set of environment states, $\mathcal{A}$ is the set of actions, $\mathcal{P}(s_{t+1}|s_t,\ \ a_t) : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R} \in [0, 1]$ is the transition probability between the states, and $\mathcal{R}(s, a) : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the immediate reward of taking action a at state s. A policy π determines the way the learning agent is behaving at a given state. In other words, π is a mapping from states to actions ($\pi : \mathcal{S} \rightarrow \mathcal{A}$) and the goal of a RL

agent is to learn the optimal policy that maximizes the discounted accumulated rewards over long run $\pi^* = \arg\max_\pi \mathbb{E}_\pi\{\sum_{t=0}^{\infty} \gamma^t r_t\}$. γ is the discount factor determining the trade-off between the current and future rewards. While r (i.e., $\mathcal{R}$) is the immediate reward the agent receives by executing an action in the current state, the action-value function (Q value function), on the other hand, determines how good an action is in the long run when in state s. The Q value can be expressed as $Q^\pi(s,a) = \mathbb{E}_\pi[\sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_\tau | s_t = s, a_t = a]$.

When the transition probabilities $\mathcal{P}$ are known, iterative approaches can be used to find the optimal solution. According to the Markov property, Q value can be recursively calculated with the Bellman equation: $Q^\pi(s,a) = \mathbb{E}_{s'}[r + \gamma \mathbb{E}_{a' \sim \pi(s')}[Q^\pi(s', a')]]$, where s' and a' are state and action in the next time step. Then, the optimal policy can be derived from $\pi(s) = \arg\max_a Q^\pi(s,a)$.

RL can also be applied to problems where the system dynamics and transition probability between states are unknown. Model-free RL methods are learning explicitly from trial and error and they do not require any knowledge of the environment (system). Thus, RL is advantageous when applied to complex decision-making problems where the real challenge is to construct an adequately precise system model. Most model-free approaches attempt to estimate the Q value function and among those, Q-learning (172) has gained a lot of interest due to its simplicity and effectiveness. Q-learning is an off-policy algorithm, meaning that to update the Q values, there is no need to follow a policy as they are updated using the maximum Q-value in the new state s'. That is an advantage of off-policy Q-learning over on-policy methods (e.g., SARSA), which enables learning from historical data collected where the actions are not selected based on any specific policy. In simple problems with discrete state-action spaces, the Q values can be stored in a Q table with the values being updated by:

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r(s,a) + \gamma \max_a Q(s',a) - Q(s,a)] \tag{4.1}$$

where α is the learning rate.

However, for problems where the number of state-action pairs is too large or the state/action

space is continuous, a function estimator can replace the Q table to address the curse of dimensionality. Simple linear regression or complex neural network can be used for function approximation. Neural networks have been widely used in RL due to its ability to capture non-linearity and complex interactions between the state features. The Q-learning algorithm combined with deep neural network resulted in its excellence on the growing applications in the Atari Games (173). This combined algorithm was further developed by the authors in (174) and is termed as deep Q-network (DQN). The DQN algorithm addressed the instability issue of traditional RL methods in which both inputs and targets change constantly. In DQN, the target value $\hat{q} = r + \gamma \max_a Q(s', a'; \theta')$ is calculated based on a *target network* with parameters θ'. The *target network* is different from the online network $Q(s', a'; \theta)$. The parameters in *target network* (θ') are copied from the online network every fixed number of steps. This makes the learning process similar to supervised learning in a sense that the targets are kept fixed for a number of steps. The algorithm then attempts to minimize the loss function, which is usually defined as the mean squared error (MSE) on ($\hat{q}$, $Q(s, a; \theta)$).

The authors in (174) propose the idea of experience replay in DQN, which was first presented in (175). Identified as another import factor of DQN, experience replay has been demonstrated to lessen the correlation among consecutive states. In experience replay, the set of transitions $(s^l, a^l, r^l s^{l+1})$ of past episodes are stored in a replay memory $\mathcal{D}$, and the network at any iteration is trained on a mini-batch of transitions $\mathcal{F}$ sampled uniformly from the memory $\mathcal{D}$. Another benefit of using experience replay is the data efficiency as each transition is possibly used multiple times to update the Q network parameters, instead of being used only once and then discarded. Therefore, the MSE loss function in DQN with experience replay at iteration i is defined by:

$$L_i(\theta_i) = \mathbb{E}_{s,a,r,s' \sim \mathcal{U}(\mathcal{D})}[(\hat{q}_i - Q(s, a; \theta_i))^2] \tag{4.2}$$

Double deep Q network (DDQN) is another variant of DQN proposed in (176) and further improves the DQN algorithm by mitigating the issue of Q value overestimation. In DDQN, the

target is redefined by: $\hat{q} = r + \gamma Q(s', \arg\max_{a'} Q(s', a'; \theta); \theta')$. The effectiveness of DDQN has attracted a lot of attention in power system applications (177; 178; 179; 180). This chapter builds upon the existing DDQN method and contributes to the application of RL in the charge/discharge scheduling problem. The possible challenges of using standard DDQN, which makes that approach inefficient for this scheduling problem and possibly other multi-stage decision problems, are investigated and the solution to alleviate each issue is provided. Since the charge/discharge scheduling is a minimization problem, hereafter, the term "cost" is used instead of "reward" and all the "max" operators are replaced by "min".

4.2.2 Literature Review

In recent years, model-free RL has gained much attention due to its ability for self learning from historical data, and the number of research articles in applying RL to power system scheduling (e.g., demand response) has increased significantly (181; 182; 183; 184; 185; 186). Many researchers formulated the EV scheduling problem using MDP with known transition probability (i.e., model-based). The model-based approaches achieve good results when the uncertain parameters and the environment have known underlying distributions (187; 188). The more the number of random variables, the more challenging it is to obtain an accurate model of the system. Very few articles have used model-free RL methods in EV charge/discharge scheduling problems. Some of the research studies done in this domain consider finding the optimal action for individual EVs that can take actions independent of other EVs (189; 190; 159). For instance, the authors in (189) develop a RL-based charging method that minimizes the charging cost for an individual EV based on forecasted price. In a similar study, the authors in (159) propose a model-free RL approach using DDQN to find the optimal schedule that minimizes the charging cost for individual EVs in the presence of uncertain price and EV commuting behavior. They use LSTM on past price data to extract features that are useful in capturing the price dynamics. The number of possible actions in this study is seven different charge rates for an EV. The only source of uncertainty in these works is the price signal.

In the absence of coupling constraints, the action space is small. The decoupling of EVs is

possible in case where the goal is to minimize the charging cost and the price is independent of the EV demands. However, by considering wind energy curtailment, the decoupling is not possible since the actions of EVs should be dependent of each other to minimize the curtailment of wind. When considering a centralized charging schedule for EV fleet, the curse of dimensionality becomes a significant factor as the action space gets large. The authors in (191) considered a centralized scenario where an aggregator is responsible for purchasing the electricity in a day-ahead market and dispatching the power in real-time to the EVs, with the goal of minimizing the charging cost. To deal with the large action space, the authors proposed a heuristic scheme for EV fleet charging. However in their work, the EVs are only capable of charging, and the wind uncertainty is not taken into consideration.

Most research studies focus on minimizing the charging costs of the EVs under stochastic prices. However, in this work, both charging cost and wind integration is considered under uncertain price and wind power. Moreover, in the literature, a double deep Q-network (DDQN) application in real-time EV scheduling has not been reported. Also, the comparison between simple mean-value optimization and RL is not provided elsewhere. The summary of the contributions of this chapter is as follows:

1. The real-time EV scheduling problem is formulated as an MDP with unknown transition probability under stochastic wind energy and price.

2. A DDQN RL approach combined with optimization is proposed to learn the optimal charge/discharge schedule.

3. The challenges of RL such as large action-space, feature engineering, and slow convergence are addressed.

4. A heuristic EV fleet charging/discharging scheme is developed in conjunction with a double-agent RL to address the curse of dimensionality issue when considering large number of EVs.

4.3 Proposed Approach

The charge/discharge EV scheduling problem studied here is the same as in Chapter 3. A deterministic optimization approach based on the mean value or expected value of the uncertain parameters was developed in Section 3.5.2.3. Alternatively, one could use point forecasts of exogenous variables, such as wind production and real-time electricity price. In this chapter, the deterministic problem is referred to as the mean-value (MV) problem.

The approach to this dynamic stochastic programming problem is based on a combined RL-optimization method. The optimal solution from the MV is incorporated into the RL algorithm in order to further facilitate the agent task in finding the optimal policy and achieve faster convergence. The RL agent learns the dynamics of the system through interaction with the environment and achieves a solution close to the global optima. The scheduling problem is formulated as a Markov Decision Process (MDP) with unknown transition probabilities, and a DDQN method is used to estimate the optimal action-value function. To evaluate the proposed RL-optimization approach, the combined RL-optimization method is compared with the standard DDQN approach.

4.3.1 Problem Formulation

An MDP framework is used to formulate the scheduling problem. The objective of the RL agent is the same as the optimization objective 3.3. At each time step, the agent, which is the decision maker aggregator, needs to decide whether to charge or discharge the EVs for the future action periods based on the available information. The MDP is defined by the 4-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}(\mathcal{S}, \mathcal{A}), \mathcal{C}(\mathcal{S}, \mathcal{A}))$ with state space ($\mathcal{S}$), action space ($\mathcal{A}$), the unknown state transition probability ($\mathcal{P}$), and cost function $\mathcal{C}$. The details of the MDP are provided below.

4.3.1.1 State:

The agent needs to learn the state transitions by interacting with the environment, executing action a in state s and observing the cost c. The states can be any information available to the agent that might help to make a decision. The system state space contains time-dependent information X_{time}, EV requirements and availability X_{ev}, and wind and price information X_{wp}.

$$\mathcal{S} = X_{time} \times X_{ev} \times X_{wp} \tag{4.3}$$

The time-dependent component, X_{time}, consists of information regarding the exact time step or the quarter of the day X_q and the day of week X_{dow}. The time-related information helps the agent capture the system dynamics since the EV arrivals, wind production, and price usually follow a daily pattern or a weekly trend.

$$X_{time} = X_q \times X_{dow} \tag{4.4}$$

X_{ev} contains information about whether the EV is plugged-in or not (U), the remaining energy requirement of each EV (E^{req}), and the remaining parking time of each EV (T^{rem}). U, E^{req} and T^{rem} are all vectors of dimension N_{max}, where N_{max} is the maximum number of EVs that arrive in the parking lot in a day or one can think of it as the parking lot capacity at any time. Each element of the vector corresponds to an EV connected to a smart charger.

Moreover, the vector of optimal actions of the MV optimization problem (a^*_{opt}) is also included in the state space. The hypothesis is that by knowing the optimal solution to the forecast model, the agent can better learn how to make decisions given the optimal solution, and other available information.

$$X_{ev} = U \times E^{req} \times T^{rem} \times a^*_{opt} \tag{4.5}$$

Since the outcome of the next step for real-time price and wind energy depends on their current values to some degree, the current values (W, pr) need to be included in the state space. The state description for uncontrollable wind and price state is defined by X_{wp}. Moreover, the optimal solution depends on future realization of uncertain parameters, thus good forecast of uncertain parameters further aid the agent to make a decision in the current state. Different forecasting tools can be used, but in this chapter, the same forecasting techniques discussed in chapter 3 (MDP for wind and similar day approach for wind) are used. Therefore, X_{wp} is split into four components as

follows:

$$X_{wp} = W \times pr \times W_{forecast} \times pr_{forecast} \tag{4.6}$$

4.3.1.2 Action:

For each vehicle, the set of actions contain full charge ($a = 1$), full discharge ($a = -1$), or remaining untouched ($a = 0$). Thus, the action space for the aggregated EVs is defined by $\mathcal{A} = \{a \in \mathbb{R}^{N_{max}} |\ a_i \in \{-1, 0, 1\},\ \ i = 1,\ ..., N_{max}\}$. Given N_{max} vehicles in the parking lot, the total number of actions is $3^{N_{max}}$. The action space becomes quite large with a large number of vehicles and makes it challenging for the RL algorithm to estimate action-value function for all state-action pairs. Moreover, our optimization problem has some constraints that need to be satisfied. One way of addressing the large state-action pairs, is to limit the set of possible actions denoted by A(s) based on the constraints of the optimization model as well as the feasibility of the action. For instance, if $U_i = 0$, the only possible action for EV i is $a_i = 0$. If the state of charge for an EV is going to exceed the battery capacity by charging, then the feasible actions are either to discharge or remain untouched. Moreover, if the only way to get to the desired level of charge is to charge the EV in all remaining steps, then the action should always be 1 (full charge) so as not to violate the charging requirements. Also, if the EV has SOC less than the SOC_{min}, it should charge with full speed until it reaches that level. A complete set of states and the corresponding possible actions are provided below. With this approach, we reduce the size of action-state pairs significantly for each state, where the agent only explores the possible actions instead of the whole

action space, thus making it easier and faster to approximate the action-value function Q.

$$U_i = 0 \implies \mathcal{A}_1(s) = \{a \in \mathcal{A} | a_i = 0\} \tag{4.7}$$

$$SOC_i - P_i^d < SOC_{min,i} \implies \mathcal{A}_2(s) = \{a \in \mathcal{A} | a_i \in \{0, 1\}\} \tag{4.8}$$

$$SOC_i < SOC_{min,i} \implies A_3(s) = \{a \in \mathcal{A} | a_i = 1\} \tag{4.9}$$

$$SOC_i + P_i^c \, T_i^{rem} < SOC_{desired,i} \implies \mathcal{A}_4(s) = \{a \in \mathcal{A} | a_i = 1\} \tag{4.10}$$

$$(SOC_i - P_i^d) \times (T_i^{rem} - 1) < SOC_{desired,i} \implies \mathcal{A}_5(s) = \{a \in \mathcal{A} | a_i = 1\} \tag{4.11}$$

$$\sum_{i \in E^t} P_i^c - W_t > P_{max}^G \implies \mathcal{A}_6(s) = \{a \in \mathcal{A} | \sum_{i \in E^t} a_i \times P_i^c <= P_{max}^G\} \tag{4.12}$$

Equation 4.8 ensures that the SOC level of the EVs never drops below the SOC_{min}, while Equation 4.9 enforces full-speed charging when the vehicle arrives with SOC level less than SOC_{min}. Equation 4.10 checks if the vehicle cannot reach the desired level even with full-speed charge during the whole parking period. In that case, the EV has to be charged ($a = 1$). Moreover, if discharging an EV will cause violating the charging requirement, discharge is not a possible action (equation 4.11). Equation 4.12 makes sure that the total charging from the external grid does not exceed the transmission capacity P_{max}^G.

For an action to be feasible, all the constraints have to be satisfied, thus, the set of possible actions $\mathcal{A}(s)$ is the intersection of $\mathcal{A}_1$, ..., $\mathcal{A}_6$. However, there is a chance that no feasible action exists in a given state. In that case, a random action can be taken but the penalty for violation of the constraint should be included in the cost function. For instance, the penalty of not reaching the desired level can be added.

4.3.1.3 Cost Function:

In the model-free system, the agent interacts with the environment continually and observes the immediate cost associated with its actions. The immediate cost signal the agent receives is a function of current state and current action as $C_t = f_c(a_t, s_t)$. The goal is to minimize the charging cost for EVs as well as the wind curtailment while taking the battery degradation into account.

Thus, the cost in RL is the same as the objective function of the optimization model but only for the current time step, which is the weighted sum of wind curtailment cost, the charging cost, and the battery degradation. The penalty for not reaching the desired level is also added. The difference between the desired charge level and the SOC level at departure determines the energy deficiency, and a penalty factor ν determines the per kWh cost of energy deficiency. Thus, the cost function can be expressed as follows:

$$c_t = \sum_i (a_{i,t} P_i^c - W^t) pr^t + \delta(W^t - \sum_i a_{i,t} P_i^c) + \sum_i \Psi_i + \sum_{i \in \{i | t_i^{dep} = t\}} \nu(SOC_{desired,i} - SOC_i^t) \tag{4.13}$$

where Ψ_i is cost of battery degradation and can be calculated using equation 3.6. In the above equation, the charging and discharging efficiencies are removed for simplicity.

4.3.1.4 System Dynamics:

The state transition from state s_t to s_{t+1} depends not only on the current action a_t, but also on the randomness of uncontrollable states (wind, price, and arrivals). The state transition for X_{ev} state depends only on the current action, and the system dynamics for EV-related states are described by:

$$E_{i,t+1}^{req} = E_{i,t}^{req} - \eta_i^c \; a_{i,t} \; P_i^c \tag{4.14}$$

$$T_{i,t+1}^{rem} = T_{i,t}^{rem} - 1 \;\; if \;\; U_{i,t} = 1 \tag{4.15}$$

where P^c is the maximum charging power of the EV in each action period (Δt) and η^c is charging efficiency. The state U_i is partially controllable, meaning that when $U_{i,t}$ equals to 1 and the remaining parking time is more than one time period, then $U_{i,t+1}$ is also 1. However, when $U_{i,t}$ = 0, the arrival of EV is not known and the state transition is influenced by randomness ω^t. For uncontrollable states X_{wp}, the state transition depends on current state and the random disturbance ω. Using the notations, we have $X_{wp,t+1} = f(X_{wp,t}, \omega^t)$. Defining an accurate model for the

random disturbance parameter is difficult since it depends on many external factors, such as, volatile electricity load, the bidding market, wind speed, commuting behavior, etc. This is the main reason behind using model-free RL.

4.3.1.5 Action-Value Function:

The quality of charge/discharge action in any state is evaluated by the action-value function, which is the expected total cost starting from state s, taking action a and following policy π afterwards (Equation 4.16).

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}[\sum_{k=t}^{T} \gamma^{k-t} c_k | s_t = s, a_t = a] \tag{4.16}$$

where π is the scheduling policy mapping each state to an action. A policy determines how the agent should act at a given state. A discount factor $\gamma \in [0,1]$, is used to weight the immediate costs and future step costs. The RL agent attempts to find the optimal policy that results in minimum action-value function.

4.4 RL-Optimization Algorithm

The proposed algorithm is an extension to the DDQN algorithm with experience replay (176) using the forecast of the wind and price and the optimal results from the MV problem.

This scheduling problem was first solved using standard DDQN with ε-greedy action selection, and it was observed that the Q-values do not converge to a good solution and are probably trapped in a local optimal solution. The results will be shown later in section 4.4.2.1. The two main reasons for the inferior quality solution are perceived to be high variance cost and inefficient exploration.
High Variance Cost: The DDQN and other RL approaches perform well on problems with static environments. In static environment problems, the agent seeks to maximize the total reward where the total reward of the optimal policy does not vary between episodes. However, in the EV scheduling problem, the environment is highly dynamic and the cost significantly changes for different scenarios of wind, price, and EV arrivals. The wind and price profiles vary significantly daily and seasonally, potentially causing a change in the underlying distribution. Changes in the

underlying distributions result in high variance in cost estimation.

To better understand this, let us consider two different scenarios. In the first scenario, wind energy is low, a high number of EVs arrive at the parking lot at the same time, and they remain plugged in for short periods. To satisfy the charging demand, a lot of energy needs to be purchased from the external grid, thus the total cumulative cost will be high. In the second scenario, however, there is enough wind energy generated and the LMP is low. Even if the actions in the first scenario are optimal, the cost might be much higher compared to the second scenario. This is true even in the case where sub-optimal actions were taken by the agent in the second scenario. This implies that the cost signal is not significantly affected by the scheduling algorithm. As a result, the agent has a difficult time determining if the cost is associated with the change in the environment or the actions it has taken. When the agent observes a low cost, it cannot determine whether it was because of a better action or because of a change in environment. In the literature, some authors have proposed the idea of regretted reward (cost) to alleviate high variance reward estimation (192). The regretted cost is calculated as the expected difference between the sum of estimated cost for the optimal policy and the cost collected by the agent. The regret can somehow measure the gap between the current policy and the optimal policy. It should be noted that the optimal policy is not known, and it is difficult to obtain an estimate for the optimal policy. However, by solving the MV optimization problem, a close-to-optimal cost can be achieved. Even though it is not the optimal policy, it might be close enough to measure the changes in the dynamic environment and reduce the variance in cost. The regretted cost at time t can be approximated by:

$$\tilde{c}_t = c_t - c_t^* \tag{4.17}$$

where c_t^* is the optimal cost obtained from solving the MV problem at time t. The regretted cost replaces the original cost signal.

Inefficient Exploration:

Exploration and exploitation trade-off is an important topic in RL as the agent interacts with the

environment and improves its actions in a trial and error fashion. The exploitation phase is about taking the actions that result in minimum total expected cost in the future (minimum Q-value). However, since there is always uncertainty about the estimates for action-value function, the agent needs to explore other possible actions. This implies non-greedy actions should be selected using ε-greedy action selection. In a standard RL approach, the non-greedy actions are taken randomly.

Generally speaking, in sequential decision-making problems with uncertainty in the future, when the agent is exploring at first, the random actions can lead to costs that is far from the minimum cost obtained by the optimal policy. To give an example, an EV during its parking time requires charging more often than discharging to reach the desired level. However, by random exploration, the agent charges and discharges with the same probability, leading to violating the charging requirement or the minimum SOC constraints. A rare combination of actions would result in a cost better than the MV problem. Thus, when actions are chosen randomly, there is a very small chance that the agent takes that rare combination of actions. As a result, by random exploration, the agent may never try actions that lead to a lower total cost than the MV solution.

It would be preferable to select random actions based on their potential to be actually an optimal action (171). Thus, in the proposed approach, the actions are taken based on how close they are to the optimal actions of the MV problem. This implies that in the exploration phase, the actions that are optimal to the MV problem are chosen more frequently than other possible actions. This is another way to incorporate the results of the optimization model in the RL approach. More precisely, non-greedy actions are taken proportional to the inverse of the absolute difference from the optimal action. For instance, if the optimal action to the forecast model for an EV at time t is 1, then in the exploration phase, action 1 should be picked with a higher probability compared to action 0, or -1. In the exploration phase, non-greedy actions are taken based on the stochastic policy stated below.

$$\pi(a_i|s) = \frac{\left(\frac{1}{|a_i - a_i^*| + \epsilon}\right)}{\sum_{a \in \{-1,0,1\}} \left(\frac{1}{|a - a_i^*| + \epsilon}\right)} \qquad \forall a_i \in \{-1, 0, 1\} \tag{4.18}$$

where a_i^* is the optimal solution to the MV problem and a small value of ϵ is added to avoid 0

denominator values. Like the regretted cost, this action selection method reduces the variance in Q-value estimation. It also ensures that the cost of exploration is not very high, which makes this approach more desirable for online learning settings. In the exploitation phase, however, the actions are taken using a greedy approach based on the action-value estimate $\hat{Q}(s, a)$.

$$a^* = \underset{a \in \mathcal{A}(s)}{\arg\min}\, \hat{Q}(s, a) \tag{4.19}$$

4.4.1 Training Algorithm

The training of the neural network (NN) used to estimate the Q-value function is provided in Algorithm 2. The parameters of the online and target NN are denoted by θ and θ', respectively. The inputs of the NN are the state S in equation 4.3. The algorithm outputs the parameters of the online Q network (θ).

First, the parameters θ and θ' are initialized with random values (e.g., 0) everywhere on $\mathcal{S} \times \mathcal{A}$ (lines 1,2). Assuming starting with no vehicles in the parking lot, each episode starts when the first EV arrives, and ends when there is no vehicle in the parking lot. For each episode, the following process repeats. At any time step t, the MV problem is solved and the optimal action and cost for the current action period are stored in $a^*_{opt,t}$ and c^*_t. Then, forecasts of future wind and price are updated and added to the state vector. If forecasts are not provided, one can replace the forecast with the estimate or expected value of wind and price. Then, a_t is selected based on the ε-greedy approach, where the actions are selected based on equations 4.7-4.12 with probability ε, and from the Q network with probability $1 - \varepsilon$ (equation 4.18). After executing the action a_t, the immediate cost c_t and the new state s_{t+1} are observed. After that, the regretted reward is calculated using equation 4.17 and the transition tuple $(s_t, a_t, \tilde{c}_t, s_{t+1})$ is stored in replay memory $\mathcal{D}$. A minibatch of stored transitions are randomly chosen to update the parameter θ. First, target values q^l are measured as in line 14 of the algorithm. Then, the network parameters θ are updated by performing gradient descent (or any other optimization method) on MSE, which is the error between the target value q^l and the Q value estimated from NN ($\hat{Q}(s, a; \theta)$). The MSE loss is defined by:

$$L(\theta) = \sum_{l=1}^{\#\mathcal{F}} [q^l - \hat{Q}(s^l, a^l; \theta)]^2 \tag{4.20}$$

The gradient descent parameter update is according to:

$$\theta \leftarrow \theta - \alpha \Delta_\theta L(\theta) \tag{4.21}$$

where $\Delta_\theta L(\theta)$ is the gradient of the MSE loss with respect to the parameter θ, and α is the learning rate. After a specific number of time steps ($\mathcal{B}$), the parameters θ' are copied from θ. The process keeps repeating until the convergence of Q-values is reached.

Algorithm 2 Combined DDQN-Optimization Algorithm

Randomly initialize $\hat{Q}$ weights (θ) everywhere on $\mathcal{S} \times \mathcal{A}$. $\forall s \in \mathcal{S}, \forall a \in \mathcal{A}(s)$.
Set $\theta' = \theta$.
for each episode **do**
    for $t = 1, ..., T$ **do**
        Update the wind forecast and price forecast for the next few hours.
        Run the MV optimization model and find the optimal solution for the planning horizon and find the current optimal action $a^*_{opt,t}$ and optimal cost c^*_t.
        Update s_t with the forecasts and optimal action a^*_t (Equations 4.5 , 4.6).
        Limit the set of possible actions based on model constraints (Section 4.3.1.2)
        With probability ε select an action based on optimality potentials (Equation 4.18). with $1 - \varepsilon$ probability select greedy action (Equation 4.19).
        Observe immediate cost c_t and the next state s_{t+1}.
        Calculate the regretted reward. $\tilde{c}_t = c_t - c^*_t$.
        Store transition $(s_t, a_t, \tilde{c}_t, s_{t+1})$ in $\mathcal{D}$.
        Sample random minibatch of transitions $\mathcal{F}$ from $\mathcal{D}$. $\mathcal{F} = \{(s^l, a^l, \tilde{c}^l, s'^l\}_{l=1}^{\#\mathcal{F}}$.
        $q^l \rightarrow \tilde{c}^l + \gamma \hat{Q}(s', \arg\min_{a' \in A(s'^l)} \hat{Q}(s'^l, a'; \theta); \theta') \quad l = 1, ..., \#\mathcal{F}$
        Use MSE loss on $(q^l, \hat{Q})$ on the minibatch $\mathcal{F}$ and perform gradient descent to update parameters θ.
        Copy θ' from θ every $\mathcal{B}$ steps.
    end for
end for

It is worth stating that one can replace the results of the MV problem with TSA. This would

result in faster convergence if scenarios are generated properly.

Once the NN of Q values is trained using Algorithm 2, the Q network is used for real-time scheduling. The implementation of real-time EV charge/discharge scheduling is straightforward. The aggregator, at each time step, gathers the state information of the environment (s) which includes all the forecast information, time-related features, and EV-related data. Then, a feed-forward in the Q network measures the action-value $Q(s, a)$ for all possible actions in $\mathcal{A}(s)$. After that, the action with minimum Q value is taken: $a_t = \arg\min_{a \in \mathcal{A}(s)} Q(s_t, a; \theta)$.

4.4.2 Experiments

This section evaluates the performance of the proposed approach and compares it to the standard DDQN method. The results of the proposed approach is also compared with the solutions of the benchmark problems, such as MV and static day-ahead problems.

4.4.2.1 Simulation Setting:

For the simulation, the problem in Chapter 3 with a few modification is studied here. A total of 4 EV arrivals (instead of 100) for a day is considered. The arrival data is generated on a daily basis using the National Household Travel Survey (NHTS) for work trips. The energy required for an EV is modeled using uniform distribution between 75% to 95% of the battery capacity, and the parking time is randomly generated between 20 to 44 action periods (time steps). Hourly price data is collected using real-time LMP data from the California ISO (CAISO). Wind data is simulated using the Grid Lab System Advisor Model (SAM) with the same specifications as in Chapter 3. However, wind data is scaled down by a factor of 25 since the number of EVs in this experiment is 4 compared to 100 EVs in the simulation settings of Chapter 3. All EVs are assumed to have a battery capacity of 40 kWh and the maximum charge power is 9.2 kW. One year data for wind and price (365 days) is split into training (300 days) and testing (65 days). In other words, the training environment is simulated using the data from the first 300 days of the year and the testing environment is built upon the data from the next 65 days. Using the training environment, the proposed combined DDQN-optimization algorithm is used to train and learn the optimal schedule,

then the trained Q network is applied on test days, which was previously unseen by the agent. It is worth noting that if the distribution of wind and price was available, one could randomly draw scenarios from the distribution to simulate the environment. However, the aforementioned assumption is that the underlying distributions are unknown; thus, the historical data is used to simulate the environment. The training is simulated in Python using PyTorch 9.2.

4.4.2.2 Q Network Architecture & Hyper-parameters:

A five-layer fully-connected NN with three hidden layers of size 64, 256, 128 is used as the Q network (Figure 4.1). The input layer is a 47-dimensional state vector and the output layer is of size 3^4, which is the size of the action space. A value of 0.98 is set for γ factor so that the agent takes the future costs into consideration. The parameters of the *target network* is copied from the Q-network every $\mathcal{B} = 2000$ number of steps. A minibatch of size 32 is randomly sampled from the memory $\mathcal{D}$ at each time step to update the parameters. Each episode starts when the first EV arrives, and ends when the last EV leaves the parking lot. When starting the training process, exploration is needed to estimate the Q values, then exploitation is used to take the greedy action with minimum Q value. A value of $\varepsilon = 1$ ensures pure exploration and $\varepsilon = 0$ is pure exploitation. In this simulation setting, the ε reduces from 1 to 0.1 using equation 4.22 moving from pure exploration to exploitation 90% of the time.

$$\varepsilon(x) = 0.1 + 0.90 * e^{-\frac{x}{200}} \tag{4.22}$$

where x is the current episode in training.

4.4.2.3 Performance Evaluation on Training Data:

To evaluate the performance of the proposed approach, the total cost (battery degradation plus the charging cost and curtailment cost) is plotted for each epoch in training. Each epoch in Figure 4.2 is 300 days (episodes) of training. The cost from the proposed approach is benchmarked against the solutions from the MV problem, and the static day-ahead problem. The static problem is the day-ahead optimization problem with a perfect forecast of wind and data and known future

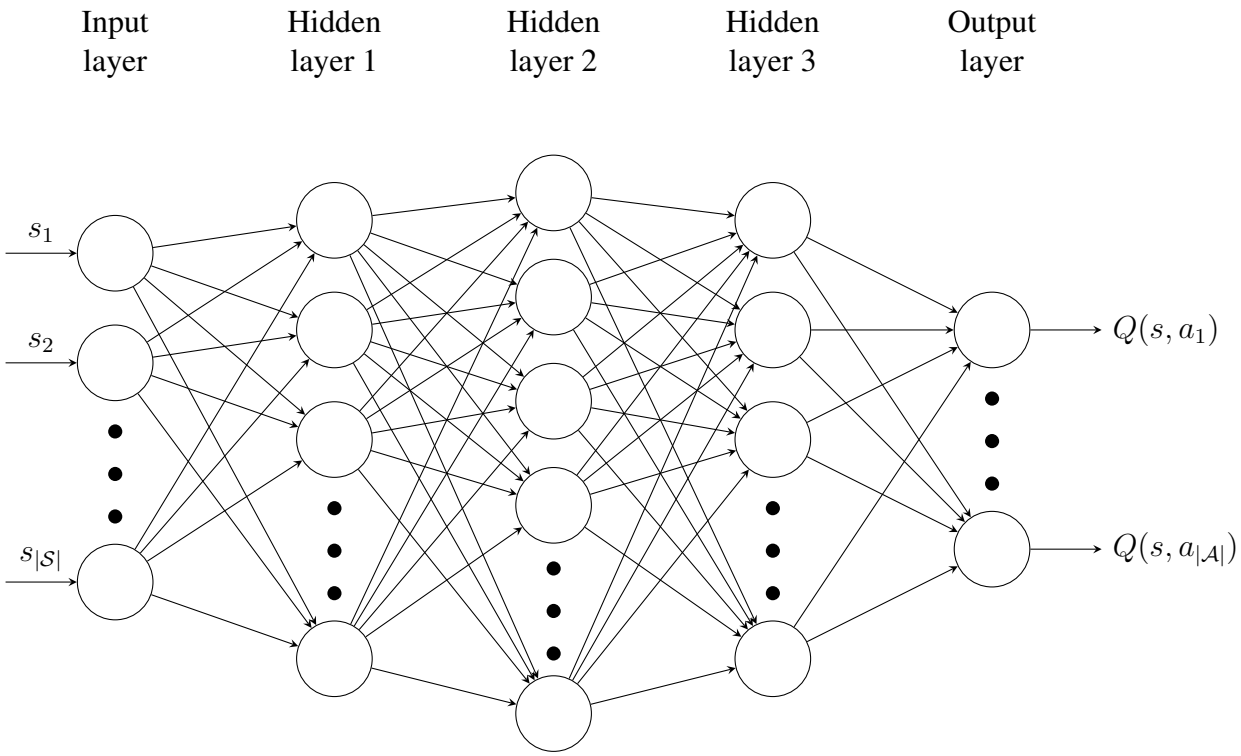


Figure 4.1: Q-Network

arrivals, which was discussed in Section 3.3. Figure 4.3 shows how the ε changes during the training process.

To assess the performance of the proposed algorithm, a standard DDQN algorithm with the same NN architecture and same hyper-parameters as the DDQN-Optimization is used. Two versions of the standard DDQN are considered. In the first version, referred to as DDQN-1, the forecasts of exogenous data is not added into the state space, and the penalty of violating the constraints is added to the cost function. Figure 4.4 shows the changes in total cost for the DDQN-1 during the training process. As seen here, this method results in a poor and unstable performance mainly due to inefficient exploration and high-variance cost discussed in the previous section. Since the penalties of violating the constraints are added without restricting the set of possible actions $A(s)$, when the agent explores with random actions, there is a high chance of violating the constraints leading to very high cost compared to the proposed approach. Even though the choice of penalty

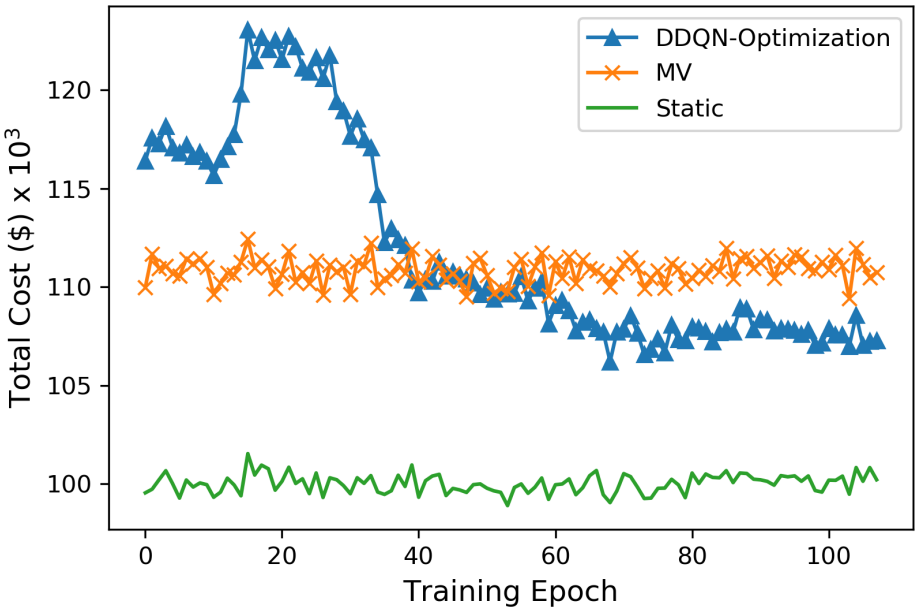


Figure 4.2: Changes in the total cost of proposed approach during the training

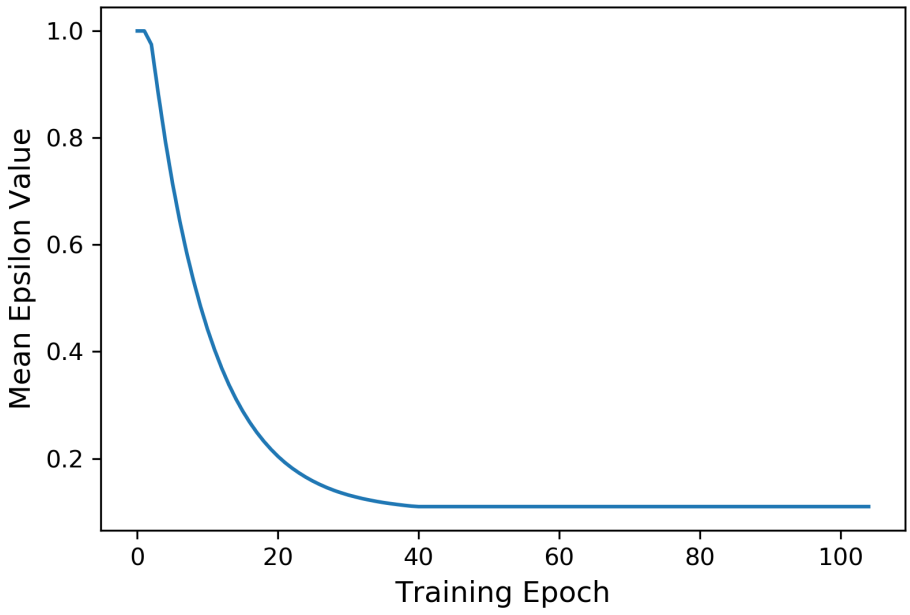


Figure 4.3: Average ε used during the training

factor influences the performance, similar unstable performance was observed with different penalty factors. Compared to the proposed approach, the DDQN-1 does not converge to the optimal solution even with a lot more training epochs (600 vs. 105). This makes the approach unsuitable specially in on-line learning, where the agents needs to learn quickly.

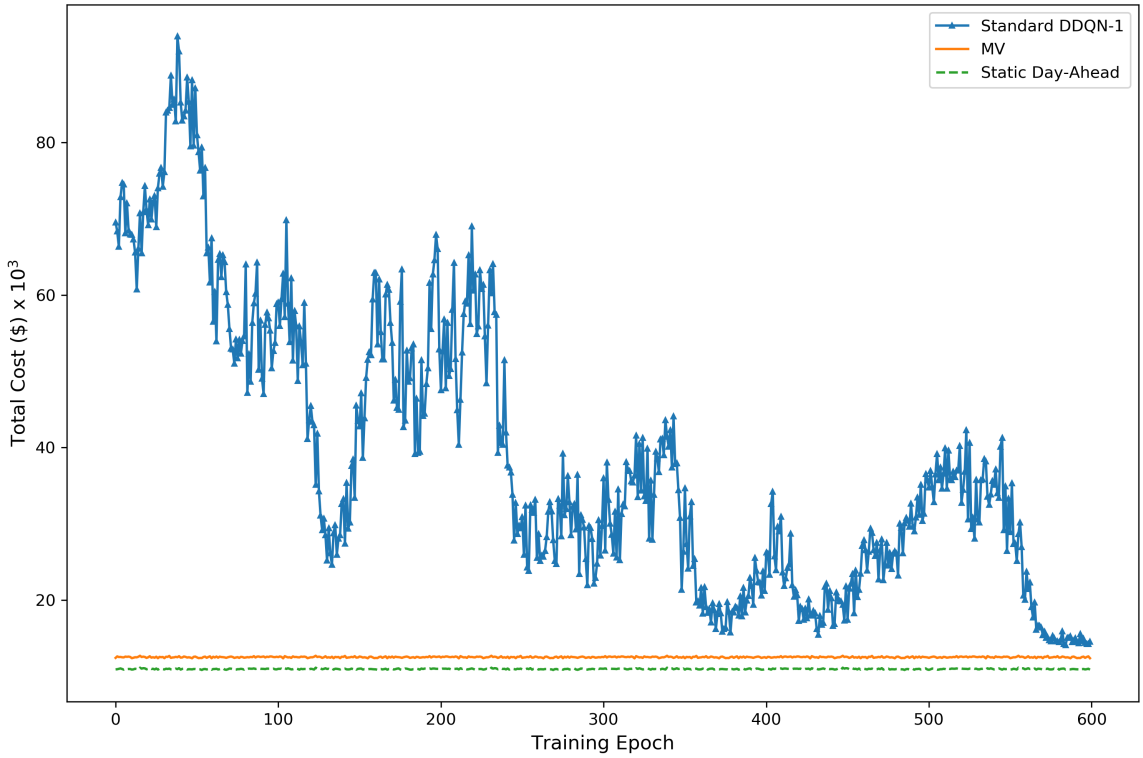


Figure 4.4: Changes in the total cost of DDQN-1 during the training

Another version of the standard DDQN (DDQN-2) is also examined in which the forecasts are added to the state space and the constraints are dealt by restricting the set of possible actions as in Section 4.3.1.2. The difference between DDQN-2 and the proposed DDQN-Optimization is that the results of the MV problem is not used for action selection and for measuring regretted cost. As shown in the Figure 4.5, it is observed that the proposed approach converges faster to the optimal

solution compared to the standard method. Also, it is observed that the cost in DDQN-2 method oscillates more than the proposed method, which is caused due to the high-variance cost. In other words, integrating the MV problem's solution in DDQN resulted in more stable, lower variance Q value approximation.

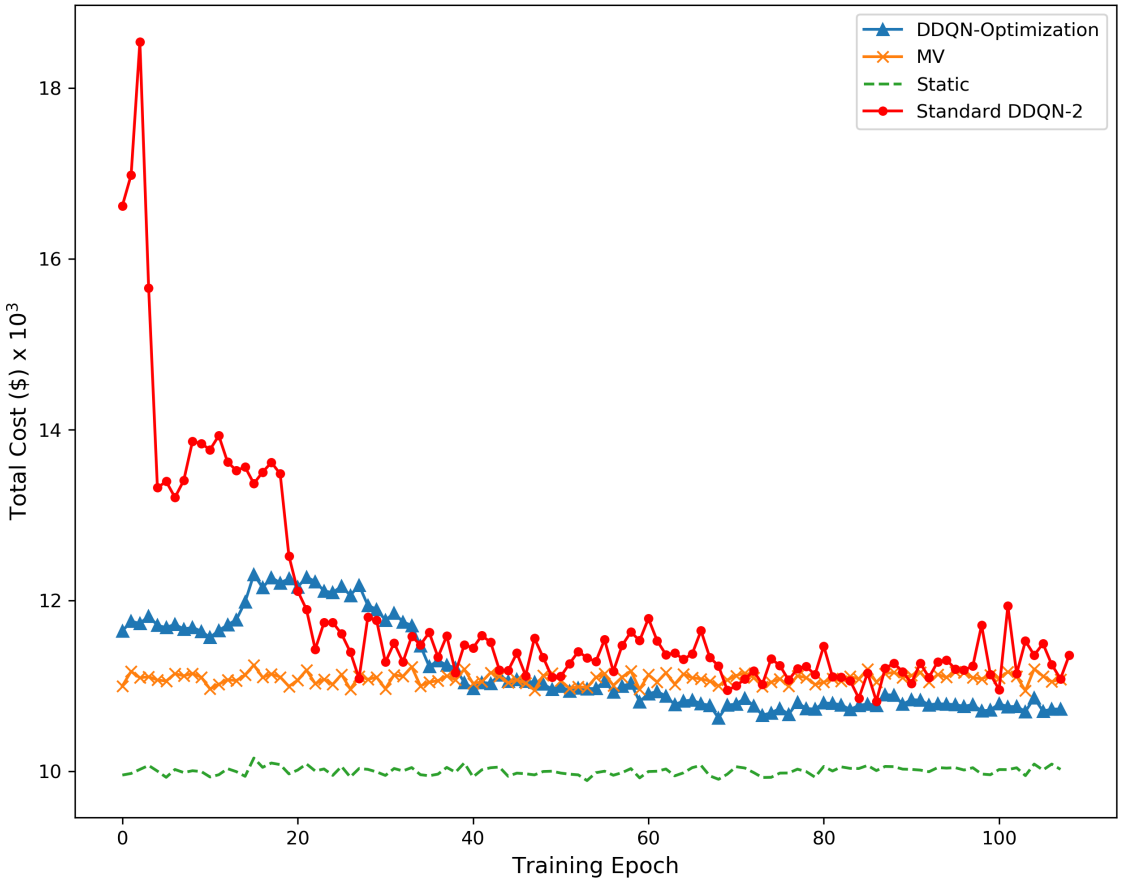


Figure 4.5: Changes in the total cost for proposed approach and DDQN-2 during the training.

Comparing the performance of DDQN-2 to DDQN-1, the results demonstrate that by restricting the set of possible actions at each state and by adding the forecast of exogenous variables into the

state space, the agent is able to learn the optimal policy in a more efficient and stable manner.

4.4.2.4 Performance Evaluation on Test Data:

Even though the proposed method performs well on the training data, there is a need to further investigate the performance on the test data to see how well the approach can generalize on unseen data. To do so, the learned Q network from training is used to generate charge/discharge schedule solutions over the test days. The cumulative cost of the proposed approach and the benchmark solutions is plotted in figure 4.6. It is observed that the proposed approach leads to lower total cost compared to the deterministic MV problem. Another point to note is that quality MV solutions rely heavily on the forecast accuracy of the uncertain parameters. However, the proposed approach does not require such accurate forecasts.

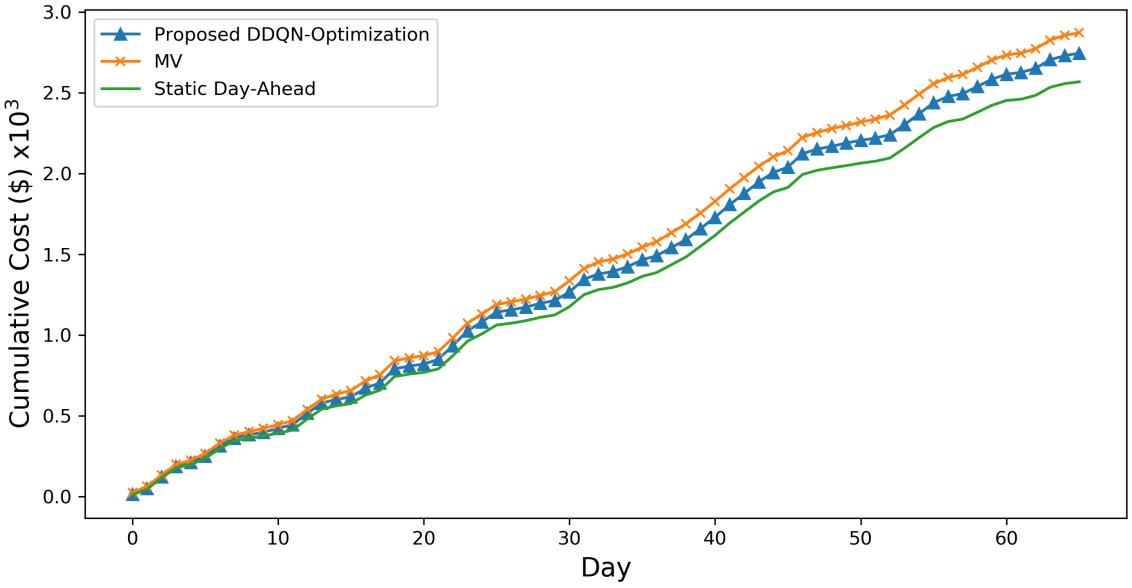


Figure 4.6: Cumulative cost of the proposed solution over test days benchmarked against MV and static solutions.

4.4.2.5 *Discussion on Effect of Large Action/State Space:*

The action that the agent takes at each time step consists of actions for each of the EVs plugged in to the smart chargers. Thus, considering only 3 possible actions for each EV (charge, discharge, and do nothing), the action space is of size $3^{N_{max}}$. As the number of EVs, N_{max} increases, the action space grows exponentially making it more challenging for the agent to learn the optimal policy. To investigate the effect of action space, another training of the proposed approach in a setting with $N_{max} = 6$ is performed. Proportional to the increase in N_{max}, the wind data is scaled up with 1.5 factor and the size of hidden layers in Q network were also increased proportionally. The results of the training for 200 epochs is plotted in Figure 4.7. Comparing the solutions with the previously studied problem ($N_{max} = 4$), as one would expect, the agent is having a more difficult time to converge to the optimal solution. The cost of the training epoch even after 200 epochs is just slightly better than the MV solution. Even though with more training epochs the agent can reach better solutions, the number of required trials to converge increases significantly with large number of vehicles.

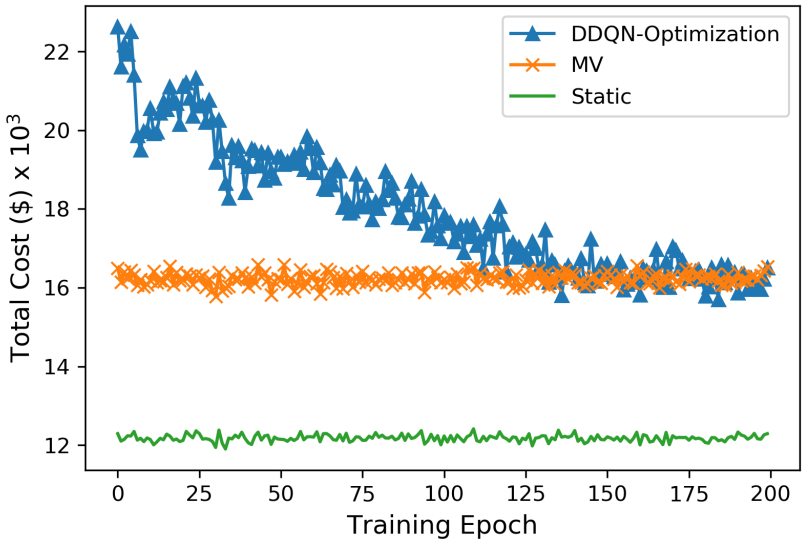


Figure 4.7: Changes in the total cost of proposed approach during the training for the problem with $N_{max} = 6$.

The curse of dimensionality issue for large action space leads us to the next section of this chapter, where this issue is addressed by using a heuristic approach along with the proposed RL-optimization.

4.5 Heuristic with RL-Optimization Approach

The proposed approach in the previous section demonstrated a great performance compared to the MV problem and some of the challenges of the standard DDQN were addressed. However, there is still the issue of curse of dimensionality when considering a large number of EVs (e.g., $N_{max} = 50 \rightarrow |\mathcal{A}| = 7.17 \times 10^{23}$). This section introduces a compact action-space representation as well as a heuristic scheme that is applied in conjunction with the RL-Optimization method to deal with large action space.

4.5.1 Proposed Approach

The general concept behind the proposed approach is to reduce the action space to a lower dimensional action space from which the agent can learn more quickly. Then, by using a heuristic method, the RL actions are mapped into the original action space, referred to as control actions.

Let us consider the scheduling problem from a slightly different perspective. At any decision period, there are three possible actions for each vehicle. Given N_{ev}^t EVs parked at time t, the aggregator needs to determine which vehicles to charge, discharge, and remain untouched. The number of possible combinations to do so is $3^{N_{ev}^t}$. The already proposed approach attempts to determine the best action among the $3^{N_{ev}^t}$ possible actions that results in minimum cumulative cost. However, if the agent was supposed to determine the number of vehicles to charge (N_{ch}) and number of EVs to discharge (N_{dc}), then the number of possible actions would be less than ${N_{max}}^2$ since both N_{ch} and N_{dc} are bounded by N_{max}. This would result in much lower action space. For instance, considering 50 EVs, the size of the action space reduces from 7.17×10^{23} to less than 2500. Once the aggregator knows how many to charge and discharge, a heuristic function can be applied to assign each EV into one of the three groups of "charging", "discharging", and "untouched", based on their charging priorities. The charging priority of an EV can be measured

based on how much charge it requires (E^{req}), and how many more periods it remains plugged-in (T^{rem}). Then, the N_{ch} EVs with the highest charging priority can be charged, and N_{dc} EVs with the lowest charging priorities or highest discharge potentials are assigned to the discharge group.

To further reduce the action space, a double-agent RL framework is proposed in which two separate Q networks are used to learn the optimal policy. The first Q network, referred to as Q_{ch}, attempts to learn the Q values associated with the number of EVs in charge mode (N_{ch}), and another network, referred to as Q_{dc}, is used to learn the optimal number of EVs in discharge mode (N_{dc}). The possible values for N_{ch} is $\{0, ..., N_{max}\}$. The same applies to N_{dc}. Thus, the output layers of the Q_{ch} and Q_{dc} networks are of size $N_{max} + 1$.

4.5.2 Heuristic Scheme

At any time step t, the agents of the two networks take actions using action-selection (i.e., ε-greedy). The actions here are not the control actions that define the charge/discharge schedule for the EV fleet. The actions only tell the aggregator the number of vehicles to charge and discharge. To determine the control actions, a heuristic approach is used. To use the heuristic scheme, the charging priorities and discharging capability of EVs need to be determined. The charging priority of EV i can be measured by variable τ_i^{ch} defined as:

$$\tau_i^{ch} = \frac{E_i^{req}}{T_i^{rem} * P_i^c} \tag{4.23}$$

The variable τ^{ch} measures the level of flexibility in charging. When $\tau^{ch} = 1$, it indicates that the vehicle should be charged with full-speed for all the remaining parking periods to reach the required level. $\tau^{ch} > 1$ implies that the vehicle will not have the required level of charge upon departure. When the battery is charged more than the desired level, E^{req} is negative, thus $\tau^{ch} < 0$. For negative values of τ^{ch}, the vehicle has enough energy, and there is no urgency to charge. As τ^{ch} increases from 0 to 1, the need for charging increases. The discharging capability of the EVs can also be measured by τ^{dc}.

$$\tau_i^{dc} = \frac{E_i^{req} - P_i^d}{(T_i^{rem} - 1) * P_i^c} \tag{4.24}$$

Highly correlated with τ^{ch}, the variable τ^{dc} measures the τ^{ch} in the next time step if the vehicle discharges in the current time step. The average of the two variables can be a good indicator of the charge/discharge flexibility. A low value of that variable indicates high discharging potential and a high value implies high charging need for the EV.

Algorithm 3 presents the heuristic scheme of assigning the EVs in each of the charging vs discharging groups. The algorithm takes N_{ch}, N_{dc}, E^{req}, T^{rem}, and P^c as input and outputs the control actions for all vehicle.

Algorithm 3 Heuristic Scheme

Input: N_{ch}, N_{dc}, N_{ev}, E^{req}, T^{rem}, and P^c.
Measure the τ^{ch} and τ^{dc} for all EVs plugged in.
Set $\tau = (\tau^{ch} + \tau^{dc})/2$
Set $a_i = 0 \quad \forall i \in \{1, ..., N_{ev}\}$.
$\mathcal{E}^{sort} \leftarrow$ Sorted EVs based on τ values in descending order.
for $i = \mathcal{E}_1^{sort}, ..., \mathcal{E}_{N_{ch}}^{sort}$ **do**
    Set a_i = 1
end for
for $i = \mathcal{E}_{N_{max}-Ndc}^{sort}, ..., \mathcal{E}_{N_{max}}^{sort}$ **do**
    Set a_i = -1
end for
Output: $\{a_i | \ i \in \{1, ..., N_{ev}\}\}$

Two similar MDP frameworks are used to formulate the double-agent RL. The state space S as defined previously in section 4.3.1.1 can be used; however, the state feature dimension becomes too large if the EV-related feature contains data for individual EVs. The previously defined EV-related

features are E^{req}, T^{rem}, and U.

$$E^{req} = \{E_i^{req} \mid \forall i \in \{1, ..., N_{max}\} \tag{4.25}$$

$$T^{rem} = \{T_i^{rem} \mid \forall i \in \{1, ..., N_{max}\} \tag{4.26}$$

$$U = \{U_i \mid \forall i \in \{1, ..., N_{max}\} \tag{4.27}$$

4.5.3 Feature Engineering

With a large number of EVs, including the above features into the state space causes scalability issues. Thus, there is the need to reduce the state space into a compact aggregated feature space by applying feature engineering (FE). FE is usually done for two purposes. 1) dimensionality reduction and 2) extracting meaningful features. The latter needs a domain knowledge to know what features might be discriminative. The dimensionality reduction can be done by handcrafting the features or by applying principle component analysis (PCA). For instance, the work in (191) reduces the dimension of state space by considering the total charged energy of EV fleet instead of including the charged energy of all EVs in the state space. By converting a N_{max}-sized vector into a single scalar, a large amount of information will be lost. Moreover, two completely different states might have the same scalar value, making the agent unable to distinguish between the two states. To give an example, if one uses the total energy required of the EV fleet instead of the energy required for each of the EVs, then the agent is not able to distinguish between these two completely different states: $E_1^{req} = 10, E_2^{req} = 0$ vs $E_1^{req} = 5, E_2^{req} = 5$. Assuming both EVs have only one decision period left before they depart, and the charging power of both vehicles is 5kW, then the optimal schedule to the latter case must be to charge both, while charging only EV 1 in the first case is perhaps the optimal solution. Therefore, the imperfect state information caused by feature extraction can lead to poor performance. Inspired by this issue, careful feature extraction is required to reduce the dimensions without losing too much information.

Instead of converting the N_{max}-sized vectors to scalar values (i.e., summation over the entire vector), we propose to reduce the vector to a M-sized array, where $M << N_{max}$. To do so,

the parameter (i.e., E^{req}) is divided into M intervals. Then, the count of EVs in each of the M intervals simply generates the feature vectors. For instance, the energy required E^{req} for any EV can range from 0 (EV fully charged) to $\max(SOC_{cap})$ (drained battery), where $\max(SOC_{cap})$ is the maximum of battery capacity for the EV fleet. The interval $[0, \max(SOC_{cap})]$ is then divided into M slots. The generated M dimensional vector is filled with the count of EVs with E^{req} values in each slot. For instance, consider the case where $N_{max} = 3$ and the charging requirement vector $E^{req} = \{10, 0, 3\}$. In a very simple case, let us assume $M = 2$ and the two intervals are $[0, 5)$ and $[5, \infty)$. Two vehicles fall into the first interval and one vehicle falls into the second interval. Thus, the reduced feature vector becomes $\{2, 1\}$. By normalizing the vector, it becomes $\{0.66, 0.33\}$. Note that this example considered only 3 vehicles, this method will be more advantageous when applied to the problem with a large number of EVs (e.g., >100). Also note that regardless of number of EVs, the new generated feature has M dimensions, but can still capture the most essential information.

4.5.4 Algorithm

Applying the above approach on EV-related features (E^{req} and T^{rem}), the state space $\mathcal{S}$ used in MDP frameworks is expressed as:

$$\mathcal{S} = X_{wp} \times X_{time} \times N_{ev} \times E^{req} \times T^{rem} \tag{4.28}$$

where N_{ev} is the number of EVs plugged-in, and X_{wp} and X_{time} are defined as before.

To learn the optimal real-time schedule (i.e., optimal Q values), the combined DDQN-Optimization can be applied with a few modifications. First, there are two Q networks here that should be parameterized separately with each one having a *target network*. Second, since the actions of RL agents are numbers to charge and discharge, the set of possible actions should change accordingly. At each time step t, the number of EVs to charge must definitely be less than the number of EVs plugged-in. Thus, the possible set of actions for the Q_{ch} is $\{0, 1, ..., N^t_{ev}\}$. Once the action (N_{ch}) is selected from the Q_{ch} network, the total number of EVs that can be discharged is bounded by

$N^t_{ev} - N_{ch}$. Thus the set of possible actions for the Q_{dc} network is $\{0, 1, ..., N^t_{ev} - N_{ch}\}$. A complete description of the proposed approach is presented in Algorithm 4.

Algorithm 4 Combined DDQN-Optimization with heuristic

Randomly initialize $\hat{Q}_{ch}$ and $\hat{Q}_{dc}$ weights (θ_1, θ_2) everywhere on $\mathcal{S} \times \mathcal{A}$. $\forall s \in \mathcal{S}, \forall a \in \mathcal{A}(s)$.
Set $\theta'_1 = \theta_1$ and $\theta'_2 = \theta_2$.
for each episode **do**
    for $t = 1, ..., T$ **do**
        Run the first stage optimization model and find the optimal solution for the planning horizon and store the current optimal action $a^*_{opt,t}$ and optimal cost c^*_t.
        Update the wind forecast and price forecast for the next few hours (using forecasting tools).
        Update s_t with the forecasts and optimal action $a^*_{opt,t}$.
        Limit the set of possible actions for Q_{ch} network. $a_{ch,t} \in \{0, 1, ..., N^t_{ev}\}$.
        With probability ε select an action based on optimality potentials (Equation 4.18). with $1 - \varepsilon$ probability select greedy action.
        Limit the set of possible actions for Q_{dc} network based on the action selected (a_{ch}). $a_{dc,t} \in \{0, 1, ..., N^t_{ev} - a_{ch,t}\}$.
        Find and take the control actions using heuristic scheme 3.
        Observe immediate cost c_t and the next state s_{t+1}.
        Calculate the approximate regretted reward. $\tilde{c}_t = c_t - c^*_t$.
        Store transition $(s_t, a_{ch,t}, \tilde{c}_t, s_{t+1})$ in $\mathcal{D}_{ch}$.
        Store transition $(s_t, a_{dc,t}, \tilde{c}_t, s_{t+1})$ in $\mathcal{D}_{dc}$.
        Sample two random minibatches of transitions ($\mathcal{F}_{ch}$ and $\mathcal{F}_{dc}$) from $\mathcal{D}_{ch}$ and $\mathcal{D}_{dc}$, respectively. $\mathcal{F}_{ch} = \{(s^l, a^l_{ch}, \tilde{c}^l, s'^l)\}_{l=1}^{\#\mathcal{F}_{ch}}$. $\mathcal{F}_{dc} = \{(s^l, a^l_{dc}, \tilde{c}^l, s'^l)\}_{l=1}^{\#\mathcal{F}_{dc}}$
        $q^l_{ch} \rightarrow \tilde{c}^l + \gamma \hat{Q}_{ch}(s', \arg\min_{a' \in A(s'^l)} \hat{Q}_{ch}(s'^l, a'; \theta); \theta') \quad l = 1, ..., \#\mathcal{F}_{ch}$
        $q^l_{dc} \rightarrow \tilde{c}^l + \gamma \hat{Q}_{dc}(s', \arg\min_{a' \in A(s'^l)} \hat{Q}_{dc}(s'^l, a'; \theta); \theta') \quad l = 1, ..., \#\mathcal{F}_{dc}$
        Measure MSE losses on $(q^l_{ch}, \hat{Q}_{ch})$ and $(q^l_{dc}, \hat{Q}_{dc})$ for minibatches $\mathcal{F}_{ch}$ and $\mathcal{F}_{dc}$
        Perform gradient descent to update parameters θ_1 and θ_2.
        Copy θ'_1 and θ'_2 from θ_1 and θ_2 every B steps.
    end for
end for

4.5.5 Experiments

To evaluate the performance of the proposed approach, the experiments with the same settings as in section 4.4.2 is implemented. In this experiment, the total number of EV arrivals during the day is set to 50 (instead of 4 in the previous experiment) and the wind production data is scaled up

proportionally. The M value for the FE (Section 4.5.3) is set to 10. The same 300 days of training data are used to simulate the training environment, and the results is tested for the next 65 days. Another difference from the previous setup is that the degradation cost is removed from the cost function. The reason is that including the battery degradation cost would add more complexity in terms of defining the proper state features. The solution from the proposed approach is again benchmarked against the solutions from MV and day-ahead static problems. Even though using the heuristic scheme does not guarantee optimality, the results show that our heuristic method combined with DDQN-Optimization approach leads to close to the optimal solutions. As seen in the Figure 4.8, the heuristic method with RL-Optimization (the proposed approach in the figure) converges more quickly than the standard DDQN. In the standard DDQN, the features are extracted in an aggregated sense. For instance, the sum of all the energy required for the EV fleet is used instead of the individual EVs. However, in the proposed approach, the features are extracted based on method discussed in Section 4.5.3. The results proves that the proposed FE method lessens the amount of information loss and leads to better solutions. Note that the ε reduces from 1 to 0.1 and is set to 0.1 after that (after about 80 training epochs in Figure 4.9), meaning that 10% of the time, the non-greedy actions are taken.

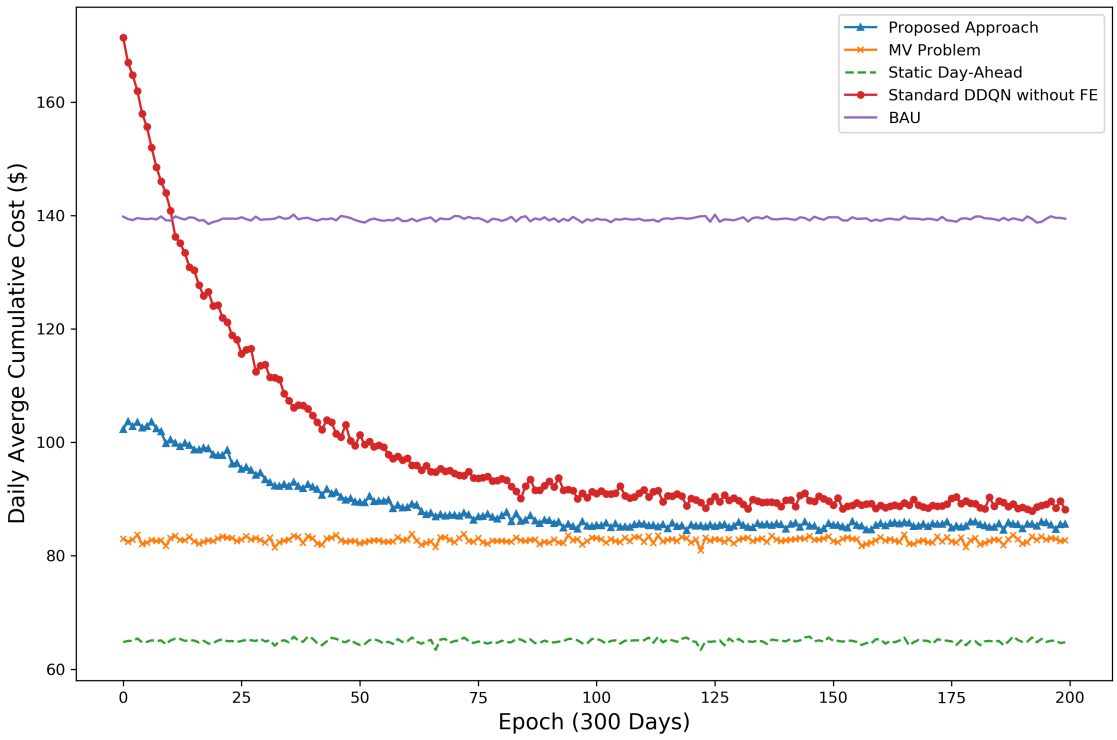


Figure 4.8: Changes in the total cost of proposed approach during the training process a benchmark solutions.

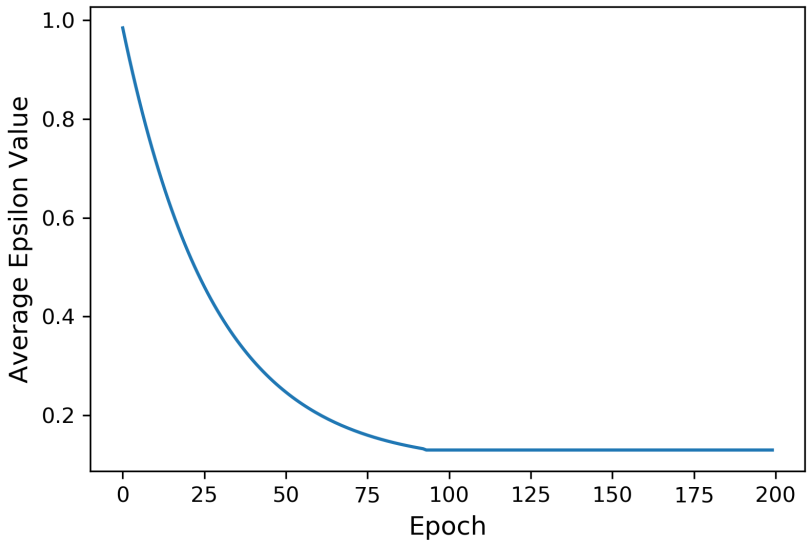


Figure 4.9: Average ε used during the training.

Setting $\varepsilon = 0$ and taking the optimal actions from the trained Q networks, Figure 4.10 shows the performance of the proposed approach on the test days benchmarked against MV and static day-ahead solutions. As seen, the trained Q network using the proposed algorithm generalizes well on the test days. The RL solutions and the MV optimization solutions are very close to each other. In details, the total cumulative costs for the proposed approach and the MV optimization are $\$4,758$ and $\$4,817$, respectively. Since using the heuristic approach does not guarantee optimality and also because some important information is lost by extracting features, the optimal Q values do not result in much better solutions than the MV problem.

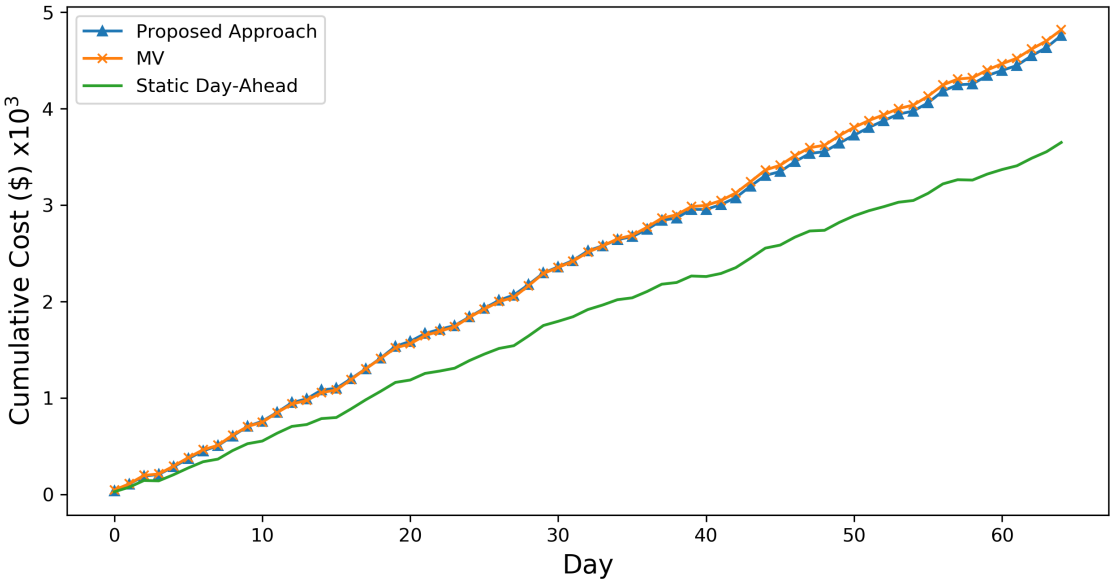


Figure 4.10: Cumulative cost of the proposed approach over test days benchmarked against MV and static solutions.

4.6 Discussion

Combining RL with optimization, the proposed approach improved the performance of RL through incorporating forecasts into the state space, modifying action-selection based upon optimality

potential, and reducing the variance in Q value approximations by calculating the regretted reward. Moreover, the curse-of-dimensionality issue for large-scale scheduling problem was addressed using a heuristic scheme. This section discusses the potential of the approach in online-learning applications, the scalability of the proposed approach, and a few ideas for further improving the approach.

Online Learning: The experimental results presented in Section 4.4.2 indicate that the proposed DDQN-optimization approach can converge quickly and in a stable way to the optimal solution, while the standard DDQN shows unstable convergence with a higher chance of getting trapped into a local optimal. The advantages of this combined approach over standard DDQN will be even more tangible in online learning, where the agent needs to learn the optimal policy through trying different actions in real-time. The combined approach not only converges faster but also ensures that when the agent is exploring the environment, the real-time cost is close to the optimal cost in MV, whereas the standard DDQN explores with random actions which results in higher cost. However, the fact that the RL approach requires so many trials to converge limits the application of RL in online-learning. The two main reasons RL cannot converge quickly are lack of key discriminative features and large action-state space. The previous section attempted to address these issues to some extent but further improvement can be achieved by better feature engineering and further limiting the set of actions.

Feature Engineering: Generally, feature engineering has two purposes, feature extraction and feature reduction. In RL literature, the former is usually done in a subjective way; for instance, based on the domain knowledge, and the latter is usually handcrafted (193; 191) or is performed using unsupervised learning methods, such as auto-encoders and principle component analysis (PCA). Handcrafted features might cause to lose a lot of important information about the system model. Auto-encoder and PCA, on the other hand, are unsupervised methods in which the relation between the features and the optimal actions is not taken into consideration. Hence, there exists the need for feature extraction in a more supervised learning way. The need for further studies on automated feature extraction and feature reduction techniques to overcome the dimensionality

issue has been highlighted in papers (194; 165). Since only a few articles in the demand response domain considered applying feature reduction to reduce the dimensionality of state space, here a method for feature engineering in a more supervised way is suggested. In order to determine the sets of features that would help the agent learn the optimal policy, one needs to identify the specific features that are predictive of the optimal actions. Theoretically, this means that if we had a labeled data set with optimal actions as the response and the state features as predictors, machine learning tools should be able to accurately predict the response. Thus, to treat the feature extraction in a supervised way, optimization methods using historical data can provide us with a data set of optimal actions in each state. Including all possible features as predictors, supervised feature reduction techniques can be used to extract or reduce the set of features that are discriminative of the optimal action.

Moreover, in the EV charge/discharge scheduling problem, since the agent needs to learn the dynamics of the environment, important predictive features need to be extracted that can capture the system dynamics (wind, price, future arrivals). Price signals are usually demand/time-dependent. Assuming that no information on external grid demand is available, only time-relevant features can be included in the state space. Price signals usually have a daily pattern, thus time-independent information such as "time of day" and "day of week" are examples of good features. In addition, to better capture time dependencies of wind and price, sequential recurrent neural networks (195) (e.g., LSTM) can be trained on past data to extract more predictive features; thereby improving the performance of RL.

Discussion on Scalability: The feature engineering technique presented in this chapter is independent of the number of EVs, so it can be applied on larger problems as well. However, the heuristic EV fleet charging/discharging scheme is dependent on the capacity of the parking lot. In other words, the size of the action space for the two Q networks is $N_{max} + 1$. Thus, considering 50 EVs there are 51 possible values for actions N_{ch} and N_{dc}. Although with enough training the agent will be able to learn a good policy, the training process would be really slow for large number of EVs (i.e., 1000 EVs). Since the training process needs to be done only once in a simulated environment,

the scalability issue is somehow alleviated. Moreover, with a small twist, the action space can be defined independent of N_{max}. Instead of deciding how many vehicles to charge or discharge, the agent decides what percentage of EVs to charge and discharge. Thus, the interval [0, 1] can be discretized into N_p percentages, which is independent of the number of EVs. For instance, considering a system with 10000 EVs, the percentage of EVs to charge can be chosen from the set $\{0, 0.5, 1, 1.5, 2, ..., 100\}$ with 200 possible actions. Then, simply the percentage is translated to the number of EVs to charge by multiplying the percentage by $N_{ev,t}$ at any time step t. Thus, the size of the action space is reduced significantly. The true optimal $N_{ch,t}$ may not be among the possible values, but the agent can reach a very close-to-optimal solution. Increasing the size of N_p, more optimal solution can be reached with the cost of higher training time. Thus, one must find the right compromise for the percentage discretization and the training time.

Limiting the set of actions: In this chapter, a state-dependent action space based on the feasibility of the action was defined. However, we can further limit the set of actions by removing not only those actions that are infeasible (e.g., violating the charging requirement), but also the actions that are very unlikely to be optimal actions. For instance, if there is only one EV in the system with high required energy and the available wind is high, then any action other than charging the EV is actually a bad decision, even if all actions are feasible. The limitation of this method is that it requires a good amount of prior knowledge about the possible optimal actions, and there is no guarantee that a perceived bad decision is actually non-optimal. Another approach to limit the set of actions is to use the labeled data set of optimal actions and states of past data. Then, machine learning tools can train on this data set to determine the probability of being an optimal action in a given state. Using the trained model, if at any state, the probability of an action to be an optimal action is very low, that action can be removed from the set of possible actions. However, the threshold probability to determine low-probability actions is a hyper-parameter that needs to be tuned.

4.7 Summary & Conclusion

In this chapter, the charge/discharge scheduling is formulated as a MDP with unknown state transition probabilities. A combined model-free RL with mean-value optimization is proposed to further improve the quality of solutions from the deterministic scheduling problem with the use of information from optimal solutions in the MV optimization model. The issue of high-variance cost in EV-wind environment is addressed by introducing the regretted cost. The action selection method is modified based on the potential to be the optimal action. Furthermore, a state-dependent action space is used to limit the set of actions at each state based on the model constraints and feasibility of the action, which addresses the challenges of the RL approach with large action space. Also, the point forecast of wind and the price is included in the state space. A combined DDQN-optimization is applied on simulated environment using historical data to approximate the optimal action-value function. Then, the optimal actions from the optimal action-value function is selected for real time scheduling. The simulation results demonstrate that the proposed algorithm outperforms the standard DDQN method in terms of convergence speed and solution quality, and achieves a better solution than the deterministic MV optimization. The proposed combined RL-optimization approach can be applied in other multi-stage stochastic problems, where constructing the scenarios to model uncertainty is the real bottle-neck in solving the problem or it is computationally too expensive to be practical for real-time implementation. Finally, a heuristic EV fleet charging/discharging scheme is integrated with the proposed RL-optimization approach to reduce the size of the action space that can be applied to large-scale problems. The proposed approach is tested on a scenario with 50 EVs and demonstrated a great performance compared to benchmark solutions. Further studies are required to evaluate the performance of this approach compared to model-based SP approach with scenario reduction techniques in terms of quality of solutions and computational time.

5. SUMMARY & CONCLUSION

5.1 Summary

This dissertation focuses on the wind integration support by leveraging the flexibility of large number of electric vehicles in the charging process. First, a review of recent approaches to accommodate high integration of wind energy is provided in Chapter 2. In Chapter 3, optimization algorithms for EV charge/discharge scheduling in a deterministic case are proposed to support wind energy integration for both unidirectional and bidirectional V2G scenarios. The main contributions of Chapter 3 are to design the scheduling algorithm that aggregators can exploit the presence of advanced communication technology in smart grid, flexibility of EV drivers, and the V2G technology to support high integration of wind energy into the power system. The approach here is based on optimal scheduling of a large number of EVs depending on the availability of EVs and preferences, needs and flexibility of their owners. To mitigate the barriers in people participation in V2G, the battery degradation, minimum charging requirement for urgent needs, and/or financial incentives for the EV drivers are considered. Furthermore, a multi-objective optimization is considered to maximize wind utilization, minimize the demand from conventional generators, and minimize charging cost while satisfying the driver needs and preferences. Simulation of the proposed algorithms for different scenarios of EV characteristics, arrivals, departures, and charging requirements are performed to check the quality of solutions and schedules. The results show significant reductions in charging cost as well as wind curtailment compared to the uncontrolled charging scenario.

The EV charge/discharge optimization problem is further extended in Chapter 4 by considering the uncertainty in wind and price forecasts. A model-free reinforcement learning integrated with the rolling-horizon mean-value optimization is proposed where any prior system information is not required. Integrating the results of the mean-value problem into the RL algorithm (DDQN) enables the agent to learn more quickly and is more stable compared to the standard RL methods. The

simulation results in a parking lot with 4 EVs showed the effectiveness of the proposed approach as it achieved better performance than the mean-value optimization model. However, RL suffers from the curse-of-dimensionality as the state-action space becomes too large when there is a large number of vehicles in the system. To alleviate this issue, a reduced action-space MDP framework is proposed in which the reduced action space is mapped into the original action space by a heuristic algorithm. Even though using heuristic approach does not guarantee optimality, the experimental results show that the proposed algorithm can reach close-to-optimal solutions.

5.2 Limitations & Future Research

In this dissertation, only the flexibility of EV owners in the charging process in considered for wind energy support. However, in the power system, the system operators have many other flexible resources at their disposals, such as storage devices, flexible conventional generators, demand response for multiple demand resources (e.g., HVAC systems and smart appliances), and other options. Thus, further studies are required to quantify the potential of each of the flexible resources (individually or integrated) for wind energy support and the cost of implementing each resource. Moreover, another suitable and widely used renewable energy source, solar energy, is not considered in this work. The wind and solar options have completely different generation patterns, and further studies are required to expand the models to support integration of solar energy and to investigate the effect of V2G on solar energy. Furthermore, different pricing mechanisms and incentive frameworks need to be compared.

One limitation of this work is the actual EV commuting data. The arrivals of EVs are simulated using the commuting behavior of all vehicles including ICEs. The commuting behavior and the driving range of EVs differ from other vehicles. Moreover, the actual data on EVs parking time, their initial state of charge, and their charging requirement was unattainable at the time of this research. Thus, the real-life charging requirements and commuting behavior might be inconsistent with the simulations. This can be interpreted as another advantage of the model-free RL approach since it can adapt to different modeling mechanisms and do not rely on accurate model for the distribution of the random variables. Moreover, this work studied a micro grid connected to an

external grid where the flow of energy was unidirectional coming only from the external grid. However, the micro grid can sometimes be used as reserve for the external grid and the excessive wind energy and/or the energy discharged from the EVs can be sold to the external grid. So future works can consider the bidirectional flow of energy between the two sources, which could result in further reduction in wind curtailment and more revenue for the owners.

Since RL is a quickly progressing field and new RL methods are being developed every year, it is important to investigate the application of state of the art RL algorithm in this domain. Even though the heuristic scheme proposed in Chapter 4 is able to lower the size of the action space significantly, the reduced action-space is still dependent on the number of EVs in the parking lot. The performance of the proposed approach should be tested for large-scale problems and future work should focus on developing a heuristic scheme independent of the number of vehicles. Another approach to deal with large number of EVs could be to cluster the EV fleets into a small set of groups based on their charging priorities. Then, the same action is taken for the EVs in each group. The issue with that approach is that clustering is an unsupervised learning method, hence, the number of EVs in each group is determined purely based on their charging needs, which can result in high cost solutions. Therefore, future work to investigate the clustering approach in a more supervised way is required.

REFERENCES

[1] "California iso." http://www.caiso.com/informed/Pages/CleanGrid/default.aspx, 2019. Accessed: 2019-04-27.

[2] N. Panwar, S. Kaushik, and S. Kothari, "Role of renewable energy sources in environmental protection: a review," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 3, pp. 1513–1524, 2011.

[3] S. Sen and S. Ganguly, "Opportunities, barriers and issues with renewable energy development–a discussion," *Renewable and Sustainable Energy Reviews*, vol. 69, pp. 1170–1181, 2017.

[4] J. M. Carrasco, L. G. Franquelo, J. T. Bialasiewicz, E. Galván, R. C. PortilloGuisado, M. M. Prats, J. I. León, and N. Moreno-Alfonso, "Power-electronic systems for the grid integration of renewable energy sources: A survey," *IEEE Transactions on industrial electronics*, vol. 53, no. 4, pp. 1002–1016, 2006.

[5] R. O'Connell, R. Pletka, S. Block, R. Jacobson, P. Smith, S. Tilley, and A. York, "20 percent wind energy penetration in the united states: A technical analysis of the energy resource." https://www1.eere.energy.gov/wind/pdfs/42864.pdf, 2008.

[6] C. Zhao, J. Wang, J.-P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2708–2717, 2013.

[7] P. A. Ruiz, C. R. Philbrick, E. Zak, K. W. Cheung, and P. W. Sauer, "Uncertainty management in the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 642–651, 2009.

[8] G. Osório, J. Lujano-Rojas, J. Matias, and J. Catalão, "A fast method for the unit scheduling problem with significant renewable power generation," *Energy Conversion and Management*, vol. 94, pp. 178–189, 2015.

[9] F. Díaz-González, A. Sumper, O. Gomis-Bellmunt, and R. Villafáfila-Robles, "A review

of energy storage technologies for wind power applications," *Renewable and sustainable energy reviews*, vol. 16, no. 4, pp. 2154–2171, 2012.

[10] W. Qi, Y. Liang, and Z.-J. M. Shen, "Joint planning of energy storage and transmission for wind energy generation," *Operations Research*, vol. 63, no. 6, pp. 1280–1293, 2015.

[11] M. Y. Suberu, M. W. Mustafa, and N. Bashir, "Energy storage systems for renewable energy power sector integration and mitigation of intermittency," *Renewable and Sustainable Energy Reviews*, vol. 35, pp. 499–514, 2014.

[12] X. Luo, J. Wang, M. Dooner, and J. Clarke, "Overview of current development in electrical energy storage technologies and the application potential in power system operation," *Applied Energy*, vol. 137, pp. 511–536, 2015.

[13] Q. Qdr, "Benefits of demand response in electricity markets and recommendations for achieving them," *US Dept. Energy, Washington, DC, USA, Tech. Rep. 2006*, 2006.

[14] Y. Ikeda, T. Ikegami, K. Kataoka, and K. Ogimoto, "A unit commitment model with demand response for the integration of renewable energies," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–7, IEEE, 2012.

[15] D. Pozo, J. Contreras, and E. E. Sauma, "Unit commitment with ideal and generic energy storage units," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2974–2984, 2014.

[16] S. Y. Abujarad, M. Mustafa, and J. Jamian, "Recent approaches of unit commitment in the presence of intermittent renewable energy resources: A review," *Renewable and Sustainable Energy Reviews*, vol. 70, pp. 215–223, 2017.

[17] T. Ackermann, *Wind power in power systems*. John Wiley & Sons, 2005.

[18] M. Albadi and E. El-Saadany, "Overview of wind power intermittency impacts on power systems," *Electric Power Systems Research*, vol. 80, no. 6, pp. 627–632, 2010.

[19] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[20] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumaili, and V. Miranda, "Wind power forecasting uncertainty and unit commitment," *Applied Energy*,

vol. 88, no. 11, pp. 4014–4023, 2011.

[21] C. Lowery and M. O'Malley, "Impact of wind forecast error statistics upon unit commitment," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 760–768, 2012.

[22] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1166–1174, 2014.

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

[24] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, "Economic valuation of reserves in power systems with high penetration of wind power," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 900–910, 2009.

[25] R. Doherty and M. O'malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 587–595, 2005.

[26] E. Denny and M. O'Malley, "Wind generation, power system operation, and emissions reduction," *IEEE Transactions on power systems*, vol. 21, no. 1, pp. 341–347, 2006.

[27] H. Banakar, C. Luo, and B. T. Ooi, "Impacts of wind power minute-to-minute variations on power system operation," *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 150–160, 2008.

[28] M. S. Rawat and S. Vadhera, "Analysis of wind power penetration on power system voltage stability," in *Power Systems (ICPS), 2016 IEEE 6th International Conference on*, pp. 1–6, IEEE, 2016.

[29] Y. Chi, Y. Liu, W. Wang, and H. Dai, "Voltage stability analysis of wind farm integration into transmission network," in *Power System Technology, 2006. PowerCon 2006. International Conference on*, pp. 1–7, IEEE, 2006.

[30] Y.-z. Sun, Z.-s. Zhang, G.-j. Li, and J. Lin, "Review on frequency control of power

systems with wind power penetration," in *Power System Technology (POWERCON), 2010 International Conference on*, pp. 1–8, IEEE, 2010.

[31] A. Molina-Garcia, F. Bouffard, and D. S. Kirschen, "Decentralized demand-side contribution to primary frequency control," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 411–419, 2011.

[32] H. Holttinen, P. Meibom, A. Orths, and et al, "Impacts of large amounts of wind power on design and operation of power systems, results of iea collaboration," *Wind Energy*, vol. 14, no. 2, pp. 179–192, 2011.

[33] F. Santos-Alamillos, D. Pozo-Vázquez, J. Ruiz-Arias, V. Lara-Fanego, and J. Tovar-Pescador, "A methodology for evaluating the spatial variability of wind energy resources: Application to assess the potential contribution of wind energy to baseload power," *Renewable Energy*, vol. 69, pp. 147–156, 2014.

[34] L. Wu, M. Shahidehpour, and T. Li, "Stochastic security-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 22, no. 2, pp. 800–811, 2007.

[35] S. Takriti, J. R. Birge, and E. Long, "A stochastic model for the unit commitment problem," *IEEE Transactions on Power Systems*, vol. 11, no. 3, pp. 1497–1508, 1996.

[36] L. Xie, P. M. Carvalho, L. A. Ferreira, J. Liu, B. H. Krogh, N. Popli, and M. D. Ilic, "Wind integration in power systems: Operational challenges and possible solutions," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 214–232, 2011.

[37] D. K. Critz, S. Busche, and S. Connors, "Power systems balancing with high penetration renewables: The potential of demand response in hawaii," *Energy conversion and management*, vol. 76, pp. 609–619, 2013.

[38] P. Denholm, E. Ela, B. Kirby, and M. Milligan, "Role of energy storage with renewable electricity generation," tech. rep., National Renewable Energy Laboratory (NREL), Golden, CO., 2010.

[39] M. Beaudin, H. Zareipour, A. Schellenberglabe, and W. Rosehart, "Energy storage for mitigating the variability of renewable electricity sources: An updated review," *Energy for*

sustainable development, vol. 14, no. 4, pp. 302–314, 2010.

[40] S. O. Amrouche, D. Rekioua, T. Rekioua, and S. Bacha, "Overview of energy storage in renewable energy systems," *International Journal of Hydrogen Energy*, vol. 41, no. 45, pp. 20914–20927, 2016.

[41] "IRENA (2017). Electricity Storage and Renewables: Costs and Markets to 2030," *International Renewable Energy Agency, Abu Dhabi*.

[42] D. Bhatnagar, A. Currier, J. Hernandez, O. Ma, and B. Kirby, "Market and policy barriers to energy storage deployment," tech. rep., Sandia National Laboratories/Office of Energy Efficiency and Renewable Energy, 2013.

[43] S. Rehman, L. M. Al-Hadhrami, and M. M. Alam, "Pumped hydro energy storage system: A technological review," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 586–598, 2015.

[44] H. Chen, X. Zhang, J. Liu, and C. Tan, "Compressed air energy storage," in *Energy Storage* (A. F. Zobaa, ed.), ch. 4, Rijeka: IntechOpen, 2013.

[45] K. Divya and J. Østergaard, "Battery energy storage technology for power systems—an overview," *Electric Power Systems Research*, vol. 79, no. 4, pp. 511–520, 2009.

[46] K.-L. Huang, X.-g. Li, S.-q. Liu, N. Tan, and L.-q. Chen, "Research progress of vanadium redox flow battery for energy storage in china," *Renewable energy*, vol. 33, no. 2, pp. 186–192, 2008.

[47] L. Li, S. Kim, W. Wang, M. Vijayakumar, Z. Nie, B. Chen, J. Zhang, G. Xia, J. Hu, G. Graff, *et al.*, "A stable vanadium redox-flow battery with high energy density for large-scale energy storage," *Advanced Energy Materials*, vol. 1, no. 3, pp. 394–400, 2011.

[48] G. O. Cimuca, C. Saudemont, B. Robyns, and M. M. Radulescu, "Control and performance evaluation of a flywheel energy-storage system associated to a variable-speed wind generator," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, pp. 1074–1085, 2006.

[49] H. Liu and J. Jiang, "Flywheel energy storage—an upswing technology for energy

sustainability," *Energy and buildings*, vol. 39, no. 5, pp. 599–604, 2007.

[50] L. Qu and W. Qiao, "Constant power control of dfig wind turbines with supercapacitor energy storage," *IEEE Transactions on Industry Applications*, vol. 47, no. 1, pp. 359–367, 2011.

[51] C. Abbey and G. Joos, "Supercapacitor energy storage for wind energy applications," *IEEE Transactions on Industry Applications*, vol. 43, no. 3, pp. 769–776, 2007.

[52] H. Zhao, Q. Wu, S. Hu, H. Xu, and C. N. Rasmussen, "Review of energy storage system for wind power integration support," *Applied Energy*, vol. 137, pp. 545–553, 2015.

[53] S. Sundararagavan and E. Baker, "Evaluating energy storage technologies for wind power integration," *Solar Energy*, vol. 86, no. 9, pp. 2707–2717, 2012.

[54] N. S. Hasan, M. Y. Hassan, M. S. Majid, and H. A. Rahman, "Review of storage schemes for wind energy systems," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 237–247, 2013.

[55] A. Castillo and D. F. Gayme, "Grid-scale energy storage applications in renewable energy integration: A survey," *Energy Conversion and Management*, vol. 87, pp. 885–894, 2014.

[56] P. Meneses de Quevedo and J. Contreras, "Optimal placement of energy storage and wind power under uncertainty," *Energies*, vol. 9, no. 7, p. 528, 2016.

[57] H. Zhao, Q. Wu, S. Huang, Q. Guo, H. Sun, and Y. Xue, "Optimal siting and sizing of energy storage system for power systems with large-scale wind power integration," in *PowerTech, 2015 IEEE Eindhoven*, pp. 1–6, IEEE, 2015.

[58] M. Nick, M. Hohmann, R. Cherkaoui, and M. Paolone, "Optimal location and sizing of distributed storage systems in active distribution networks," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, pp. 1–6, IEEE, 2013.

[59] E. D. Castronuovo and J. P. Lopes, "On the optimization of the daily operation of a wind-hydro power plant," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1599–1606, 2004.

[60] E. D. Castronuovo and J. A. P. Lopes, "Optimal operation and hydro storage sizing of a

wind–hydro power plant," *International Journal of Electrical Power & Energy Systems*, vol. 26, no. 10, pp. 771–778, 2004.

[61] N. Shi and Y. Luo, "Energy storage system sizing based on a reliability assessment of power systems integrated with wind power," *Sustainability*, vol. 9, no. 3, p. 395, 2017.

[62] M. Korpaas, A. T. Holen, and R. Hildrum, "Operation and sizing of energy storage for wind power plants in a market system," *International Journal of Electrical Power & Energy Systems*, vol. 25, no. 8, pp. 599–606, 2003.

[63] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Optimal operation of independent storage systems in energy and reserve markets with high wind penetration," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1088–1097, 2014.

[64] H. Akhavan-Hejazi and H. Mohsenian-Rad, "A stochastic programming framework for optimal storage bidding in energy and reserve markets," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, pp. 1–6, IEEE, 2013.

[65] V. M. Balijepalli, V. Pradhan, S. Khaparde, and R. Shereef, "Review of demand response under smart grid paradigm," in *Innovative Smart Grid Technologies-India (ISGT India), 2011 IEEE PES*, pp. 236–243, IEEE, 2011.

[66] R. Sioshansi and W. Short, "Evaluating the impacts of real-time pricing on the usage of wind generation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 516–524, 2009.

[67] K. Herter, P. McAuliffe, and A. Rosenfeld, "An exploratory analysis of california residential customer response to critical peak pricing of electricity," *Energy*, vol. 32, no. 1, pp. 25–34, 2007.

[68] M. H. Albadi and E. F. El-Saadany, "Demand response in electricity markets: An overview," in *Power Engineering Society General Meeting, 2007. IEEE*, pp. 1–5, IEEE, 2007.

[69] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with renewable energy sources: A review," *Renewable and Sustainable Energy Reviews*, vol. 18, pp. 64–72, 2013.

[70] A. J. Conejo, J. M. Morales, and L. Baringo, "Real-time demand response model," *IEEE*

Transactions on Smart Grid, vol. 1, no. 3, pp. 236–242, 2010.

[71] J. Ning, Y. Tang, and B. Gao, "A time-varying potential-based demand response method for mitigating the impacts of wind power forecasting errors," *Applied Sciences*, vol. 7, no. 11, p. 1132, 2017.

[72] G. R. Newsham and B. G. Bowker, "The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: a review," *Energy policy*, vol. 38, no. 7, pp. 3289–3296, 2010.

[73] P. Pinson, H. Madsen, *et al.*, "Benefits and challenges of electrical demand response: A critical review," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 686–699, 2014.

[74] K. Dietrich, J. M. Latorre, L. Olmos, and A. Ramos, "Demand response in an isolated system with high wind integration," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 20–29, 2012.

[75] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with renewable energy sources: A review," *Renewable and Sustainable Energy Reviews*, vol. 18, pp. 64–72, 2013.

[76] N. Mahmoudi, M. Shafie-khah, T. K. Saha, and J. P. Catalão, "Customer-driven demand response model for facilitating roof-top pv and wind power integration," *IET Renewable Power Generation*, vol. 11, no. 9, pp. 1200–1210, 2017.

[77] H. Falsafi, A. Zakariazadeh, and S. Jadid, "The role of demand response in single and multi-objective wind-thermal generation scheduling: A stochastic programming," *Energy*, vol. 64, pp. 853–867, 2014.

[78] A. Zakariazadeh, O. Homaee, S. Jadid, and P. Siano, "A new approach for real time voltage control using demand response in an automated distribution system," *Applied Energy*, vol. 117, pp. 157–166, 2014.

[79] A. Rabiee, A. Soroudi, B. Mohammadi-Ivatloo, and M. Parniani, "Corrective voltage control scheme considering demand response and stochastic wind power," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2965–2973, 2014.

[80] S. A. Pourmousavi, S. N. Patrick, and M. H. Nehrir, "Real-time demand response through aggregate electric water heaters for load shifting and balancing wind generation," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 769–778, 2014.

[81] J. Torriti, M. G. Hassan, and M. Leach, "Demand response experience in europe: Policies, programmes and implementation," *Energy*, vol. 35, no. 4, pp. 1575–1583, 2010.

[82] T. Broeer, F. K. Tuffner, A. Franca, and N. Djilali, "A demand response system for wind power integration: greenhouse gas mitigation and reduction of generator cycling," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, pp. 121–129, 2018.

[83] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 501–516, 2014.

[84] K. M. Tan, V. K. Ramachandaramurthy, and J. Y. Yong, "Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 720–732, 2016.

[85] K. Mets, T. Verschueren, F. De Turck, and C. Develder, "Exploiting v2g to optimize residential energy consumption with electrical vehicle (dis) charging," in *Smart Grid Modeling and Simulation (SGMS), 2011 IEEE First International Workshop on*, pp. 7–12, IEEE, 2011.

[86] A. Sheikhi, S. Bahrami, A. Ranjbar, and H. Oraee, "Strategic charging method for plugged in hybrid electric vehicles in smart grids; a game theoretic approach," *International Journal of Electrical Power & Energy Systems*, vol. 53, pp. 499–506, 2013.

[87] P. Finn, C. Fitzpatrick, and D. Connolly, "Demand side management of electric car charging: Benefits for consumer and grid," *Energy*, vol. 42, no. 1, pp. 358–363, 2012.

[88] H. Lund and W. Kempton, "Integration of renewable energy into the transport and electricity sectors through v2g," *Energy policy*, vol. 36, no. 9, pp. 3578–3587, 2008.

[89] C. Jin, X. Sheng, and P. Ghosh, "Energy efficient algorithms for electric vehicle charging with intermittent renewable energy sources," in *Power and Energy Society General Meeting*

(PES), 2013 IEEE, pp. 1–5, IEEE, 2013.

[90] B. S. M. Borba, A. Szklo, and R. Schaeffer, "Plug-in hybrid electric vehicles as a way to maximize the integration of variable renewable energy in power systems: The case of wind generation in northeastern brazil," *Energy*, vol. 37, no. 1, pp. 469–481, 2012.

[91] M. Ghofrani, A. Arabali, and M. Etezadi-Amoli, "Electric drive vehicle to grid synergies with large scale wind resources," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–6, IEEE, 2012.

[92] J. P. Lopes, P. R. Almeida, and F. J. Soares, "Using vehicle-to-grid to maximize the integration of intermittent renewable energy resources in islanded electric grids," in *Clean Electrical Power, 2009 International Conference on*, pp. 290–295, IEEE, 2009.

[93] M. Honarmand, A. Zakariazadeh, and S. Jadid, "Optimal scheduling of electric vehicles in an intelligent parking lot considering vehicle-to-grid concept and battery condition," *Energy*, vol. 65, pp. 572–579, 2014.

[94] A. Tuohy, P. Meibom, E. Denny, and M. O'Malley, "Unit commitment for systems with significant wind penetration," *IEEE Transactions on power systems*, vol. 24, no. 2, pp. 592–601, 2009.

[95] N. P. Padhy, "Unit commitment-a bibliographical survey," *IEEE Transactions on power systems*, vol. 19, no. 2, pp. 1196–1205, 2004.

[96] S. Sen and D. Kothari, "Optimal thermal generating unit commitment: a review," *International Journal of Electrical Power & Energy Systems*, vol. 20, no. 7, pp. 443–451, 1998.

[97] R. Baldick, "The generalized unit commitment problem," *IEEE Transactions on Power Systems*, vol. 10, no. 1, pp. 465–475, 1995.

[98] M. Ahlstrom, D. Bartlett, C. Collier, J. Duchesne, D. Edelson, A. Gesino, M. Keyser, D. Maggio, M. Milligan, C. Möhrlen, *et al.*, "Knowledge is power: Efficiently integrating wind energy and wind forecasts," *IEEE Power and Energy Magazine*, vol. 11, no. 6, pp. 45–52, 2013.

[99] G. B. Sheble and G. N. Fahd, "Unit commitment literature synopsis," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 128–135, 1994.

[100] M. A. Ortega-Vazquez and D. S. Kirschen, "Optimizing the spinning reserve requirements using a cost/benefit analysis," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 24–33, 2007.

[101] C. Uckun, A. Botterud, and J. R. Birge, "An improved stochastic unit commitment formulation to accommodate wind uncertainty," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2507–2517, 2016.

[102] A. Papavasiliou, S. S. Oren, and B. Rountree, "Applying high performance computing to transmission-constrained stochastic unit commitment for renewable energy integration," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1109–1120, 2015.

[103] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 2009.

[104] K. Moustris, D. Zafirakis, D. Alamo, R. N. Medina, and J. Kaldellis, "24-h ahead wind speed prediction for the optimum operation of hybrid power stations with the use of artificial neural networks," in *Perspectives on Atmospheric Sciences*, pp. 409–414, Springer, 2017.

[105] A. Zameer, J. Arshad, A. Khan, and M. A. Z. Raja, "Intelligent and robust prediction of short term wind power using genetic programming based ensemble of neural networks," *Energy conversion and management*, vol. 134, pp. 361–372, 2017.

[106] P. Jiang, S. Qin, J. Wu, and B. Sun, "Time series analysis and forecasting for wind speeds using support vector regression coupled with artificial intelligent algorithms," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[107] Q. Han, F. Meng, T. Hu, and F. Chu, "Non-parametric hybrid models for wind speed forecasting," *Energy Conversion and Management*, vol. 148, pp. 554–568, 2017.

[108] M. Afshari-Igder, T. Niknam, and M.-H. Khooban, "Probabilistic wind power forecasting using a novel hybrid intelligent method," *Neural Computing and Applications*, vol. 30, no. 2,

pp. 473–485, 2018.

[109] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American power symposium (NAPS), 2010*, pp. 1–8, IEEE, 2010.

[110] A. Khosravi, R. Koury, L. Machado, and J. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," *Sustainable Energy Technologies and Assessments*, vol. 25, pp. 146–160, 2018.

[111] Q. P. Zheng, J. Wang, P. M. Pardalos, and Y. Guan, "A decomposition approach to the two-stage stochastic unit commitment problem," *Annals of Operations Research*, vol. 210, no. 1, pp. 387–410, 2013.

[112] L. Zhao and B. Zeng, "Robust unit commitment problem with demand response and wind energy," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–8, IEEE, 2012.

[113] C. Zhao, Q. Wang, J. Wang, and Y. Guan, "Expected value and chance constrained stochastic unit commitment ensuring wind power utilization," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2696–2705, 2014.

[114] H. Siahkali and M. Vakilian, "Stochastic unit commitment of wind farms integrated in power system," *Electric Power Systems Research*, vol. 80, no. 9, pp. 1006–1017, 2010.

[115] V. S. Pappala, I. Erlich, K. Rohrig, and J. Dobschinski, "A stochastic model for the optimal operation of a wind-thermal power system," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 940–950, 2009.

[116] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang, and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 433–442, 2011.

[117] Z. Wu, P. Zeng, X.-P. Zhang, and Q. Zhou, "A solution to the chance-constrained two-stage stochastic program for unit commitment with wind energy integration," *IEEE Transactions*

on Power Systems, vol. 31, no. 6, pp. 4185–4196, 2016.

[118] H. Quan, D. Srinivasan, A. M. Khambadkone, and A. Khosravi, "A computational framework for uncertainty integration in stochastic unit commitment with intermittent renewable energy sources," *Applied energy*, vol. 152, pp. 71–82, 2015.

[119] R.-H. Liang and J.-H. Liao, "A fuzzy-optimization approach for generation scheduling with wind and solar energy systems," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1665–1674, 2007.

[120] M. Parvania and M. Fotuhi-Firuzabad, "Demand response scheduling by stochastic scuc," *IEEE Transactions on smart grid*, vol. 1, no. 1, pp. 89–98, 2010.

[121] H. Wu, M. Shahidehpour, Z. Li, and W. Tian, "Chance-constrained day-ahead scheduling in stochastic power system operation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1583–1591, 2014.

[122] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197–2206, 2011.

[123] A. Papavasiliou and S. S. Oren, "Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network," *Operations Research*, vol. 61, no. 3, pp. 578–592, 2013.

[124] G. Osório, J. Lujano-Rojas, J. Matias, and J. Catalão, "A new scenario generation-based method to solve the unit commitment problem with high penetration of renewable energies," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 1063–1072, 2015.

[125] J. Ramakrishnan and J. Luedtke, "A dual approximate dynamic programming approach to multi-stage stochastic unit commitment," *arXiv preprint arXiv:1801.02259*, 2018.

[126] Q. Wang, Y. Guan, and J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 206–215, 2012.

[127] Y. Zhang, J. Wang, B. Zeng, and Z. Hu, "Chance-constrained two-stage unit commitment

under uncertain load and wind power output using bilinear benders decomposition," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3637–3647, 2017.

[128] C. Zhao, Q. Wang, J. Wang, and Y. Guan, "Expected value and chance constrained stochastic unit commitment ensuring wind power utilization," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2696–2705, 2014.

[129] D. Pozo and J. Contreras, "A chance-constrained unit commitment with an nk security criterion and significant wind generation," *IEEE Transactions on Power systems*, vol. 28, no. 3, pp. 2842–2851, 2013.

[130] D. He, Z. Tan, and R. G. Harley, "Chance constrained unit commitment with wind generation and superconducting magnetic energy storages," in *Power and Energy Society General Meeting, 2012 IEEE*, pp. 1–6, IEEE, 2012.

[131] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *SIAM review*, vol. 53, no. 3, pp. 464–501, 2011.

[132] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 800–810, 2012.

[133] B. Hu, L. Wu, and M. Marwali, "On the robust solution to scuc with load and wind uncertainty correlations," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2952–2964, 2014.

[134] G. Morales-Espana, Á. Lorca, and M. M. de Weerdt, "Robust unit commitment with dispatchable wind power," *Electric Power Systems Research*, vol. 155, pp. 58–66, 2018.

[135] A. Lorca, X. A. Sun, E. Litvinov, and T. Zheng, "Multistage adaptive robust optimization for the unit commitment problem," *Operations Research*, vol. 64, no. 1, pp. 32–51, 2016.

[136] A. Papavasiliou, Y. Mou, L. Cambier, and D. Scieur, "Application of stochastic dual dynamic programming to the real-time dispatch of storage under renewable supply uncertainty," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 547–558, 2018.

[137] L. Zéphyr and C. L. Anderson, "Stochastic dynamic programming approach to managing power system uncertainty with distributed storage," *Computational Management Science*,

vol. 15, no. 1, pp. 87–110, 2018.

[138] N. Zhang, C. Kang, Q. Xia, Y. Ding, Y. Huang, R. Sun, J. Huang, and J. Bai, "A convex model of risk-based unit commitment for day-ahead market clearing considering wind power uncertainty," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1582–1592, 2015.

[139] S. Abujarad, M. Mustafa, and J. Jamian, "Unit commitment problem solution in the presence of solar and wind power integration by an improved priority list method," in *Intelligent and Advanced Systems (ICIAS), 2016 6th International Conference on*, pp. 1–6, IEEE, 2016.

[140] E. Delarue, D. Cattrysse, and W. D'haeseleer, "Enhanced priority list unit commitment method for power systems with a high share of renewables," *Electric Power Systems Research*, vol. 105, pp. 115–123, 2013.

[141] B. Garlík and M. Křivan, "Renewable energy unit commitment, with different acceptance of balanced power, solved by simulated annealing," *Energy and Buildings*, vol. 67, pp. 392–402, 2013.

[142] Q. P. Zheng, J. Wang, and A. L. Liu, "Stochastic optimization for unit commitment—a review," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913–1924, 2015.

[143] M. Hou, Y. Zhao, and X. Ge, "Optimal scheduling of the plug-in electric vehicles aggregator energy and regulation services based on grid to vehicle," *International Transactions on Electrical Energy Systems*, vol. 27, no. 6, p. e2364, 2017.

[144] Y. He, B. Venkatesh, and L. Guan, "Optimal scheduling for charging and discharging of electric vehicles," *IEEE transactions on smart grid*, vol. 3, no. 3, pp. 1095–1105, 2012.

[145] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy*, vol. 3, no. 3, pp. 193–201, 2018.

[146] C. Zhao, J. Wang, J.-P. Watson, and Y. Guan, "Multi-stage robust unit commitment considering wind and demand response uncertainties," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2708–2717, 2013.

[147] J. Aghaei and M.-I. Alizadeh, "Demand response in smart electricity grids equipped with

renewable energy sources: A review," *Renewable and Sustainable Energy Reviews*, vol. 18, pp. 64–72, 2013.

[148] B. Avci, K. Girotra, and S. Netessine, "Electric vehicles with a battery switching station: Adoption and environmental impact," *Management Science*, vol. 61, no. 4, pp. 772–794, 2014.

[149] A. Sheikhi, S. Bahrami, A. Ranjbar, and H. Oraee, "Strategic charging method for plugged in hybrid electric vehicles in smart grids; a game theoretic approach," *International Journal of Electrical Power & Energy Systems*, vol. 53, pp. 499–506, 2013.

[150] K. Mets, T. Verschueren, F. De Turck, and C. Develder, "Exploiting v2g to optimize residential energy consumption with electrical vehicle (dis) charging," in *2011 IEEE first international workshop on smart grid modeling and simulation (SGMS)*, pp. 7–12, IEEE, 2011.

[151] K. M. Tan, V. K. Ramachandaramurthy, and J. Y. Yong, "Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 720–732, 2016.

[152] C. Jin, X. Sheng, and P. Ghosh, "Energy efficient algorithms for electric vehicle charging with intermittent renewable energy sources," in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

[153] A. Schuller, C. M. Flath, and S. Gottwalt, "Quantifying load flexibility of electric vehicles for renewable energy integration," *Applied Energy*, vol. 151, pp. 335–344, 2015.

[154] S. Gottwalt, A. Schuller, C. Flath, H. Schmeck, and C. Weinhardt, "Assessing load flexibility in smart grids: Electric vehicles for renewable energy integration," in *2013 IEEE Power & Energy Society General Meeting*, pp. 1–5, IEEE, 2013.

[155] Y. Li, Z. Ni, T. Zhao, M. Yu, Y. Liu, L. Wu, and Y. Zhao, "Coordinated scheduling for improving uncertain wind power adsorption in electric vehicles—wind integrated power systems by multiobjective optimization approach," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2238–2250, 2020.

[156] L. Ju, H. Li, J. Zhao, K. Chen, Q. Tan, and Z. Tan, "Multi-objective stochastic scheduling optimization model for connecting a virtual power plant to wind-photovoltaic-electric vehicles considering uncertainties and demand response," *Energy Conversion and Management*, vol. 128, pp. 160–177, 2016.

[157] E. Heydarian-Forushani, M. E. Golshan, and M. Shafie-khah, "Flexible interaction of plug-in electric vehicle parking lots for efficient wind integration," *Applied energy*, vol. 179, pp. 338–349, 2016.

[158] M. Aunedi and G. Strbac, "Efficient system integration of wind generation through smart charging of electric vehicles," in *2013 Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies (EVER)*, pp. 1–12, IEEE, 2013.

[159] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time ev charging scheduling based on deep reinforcement learning," *IEEE Transactions on Smart Grid*, pp. 1–12, 2018.

[160] P. Sharifi, A. Banerjee, and M. J. Feizollahi, "Leveraging owners' flexibility in smart charge/discharge scheduling of electric vehicles to support renewable energy integration," *Computers & Industrial Engineering*, vol. 149, p. 106762, 2020.

[161] H. Cai, W. Du, X. Yu, S. Gao, T. Littler, and H. Wang, "Day-ahead optimal charging/discharging scheduling for electric vehicles in micro-grids," 2013.

[162] M. Singh, P. Kumar, I. Kar, and N. Kumar, "A real-time smart charging station for evs designed for v2g scenario and its coordination with renewable energy sources," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2016.

[163] S. B. Peterson, J. Whitacre, and J. Apt, "The economics of using plug-in hybrid electric vehicle battery packs for grid storage," *Journal of Power Sources*, vol. 195, no. 8, pp. 2377–2384, 2010.

[164] E. Apostolaki-Iosifidou, P. Codani, and W. Kempton, "Measurement of power loss during electric vehicle charging and discharging," *Energy*, vol. 127, pp. 730–742, 2017.

[165] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Applied energy*, vol. 235, pp. 1072–1089, 2019.

[166] E. Sortomme and M. A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 131–138, 2011.

[167] "System Advisor Model Version 2017.9.5 (SAM 2017.9.5), National Renewable Energy Laboratory. Golden, CO. Accessed May 20, 2019." `https://sam.nrel.gov/download`, 2017.

[168] "Estimated Charge Times Chart for Electric Vehicles, ." `https://www.clippercreek.com/wp-content/uploads/2018/06/Time-to-Charge-Chart-20180615_FINAL_LOW-RES.jpg`, 2018.

[169] "U.S. Department of Transportation, Federal Highway Administration, 2017 National Household Travel Survey." `https://nhts.ornl.gov/`, 2017.

[170] S. Kwon, Y. Xu, and N. Gautam, "Meeting inelastic demand in systems with storage and renewable sources," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1619–1629, 2015.

[171] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[172] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[173] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[174] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[175] J. O'Neill, B. Pleydell-Bouverie, D. Dupret, and J. Csicsvari, "Play it again: reactivation of waking experience and memory," *Trends in neurosciences*, vol. 33, no. 5, pp. 220–229, 2010.

[176] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Thirtieth AAAI conference on artificial intelligence*, 2016.

[177] B. Wang, Y. Li, W. Ming, and S. Wang, "Deep reinforcement learning method for demand

response management of interruptible load," *IEEE Transactions on Smart Grid*, 2020.

[178] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power and Energy Systems*, 2020.

[179] W. Liu, D. Zhang, X. Wang, J. Hou, and L. Liu, "A decision making strategy for generating unit tripping under emergency circumstances based on deep reinforcement learning," *Proceedings of the CSEE*, vol. 38, no. 1, pp. 109–119, 2018.

[180] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system: An overview," *CSEE Journal of Power and Energy Systems*, 2019.

[181] F. Ruelens, B. Claessens, S. Vandael, B. De Schutter, R. Babuska, and R. Belmans, "Residential demand response applications using batch reinforcement learning," *arXiv preprint arXiv:1504.02125*, 2015.

[182] F. Ruelens, B. J. Claessens, S. Vandael, B. De Schutter, R. Babuška, and R. Belmans, "Residential demand response of thermostatically controlled loads using batch reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2149–2159, 2016.

[183] S. Bahrami, V. W. Wong, and J. Huang, "An online learning algorithm for demand response in smart grid," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4712–4725, 2017.

[184] H. Kazmi, F. Mehmood, S. Lodeweyckx, and J. Driesen, "Gigawatt-hour scale savings on a budget of zero: Deep reinforcement learning based optimal control of hot water systems," *Energy*, vol. 144, pp. 159–168, 2018.

[185] B. V. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, "Battery energy management in a microgrid using batch reinforcement learning," *Energies*, vol. 10, no. 11, p. 1846, 2017.

[186] H. Ko, S. Pack, and V. C. Leung, "Mobility-aware vehicle-to-grid control algorithm in microgrids," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 7, pp. 2165–2174, 2018.

[187] Q. Huang, Q.-S. Jia, Z. Qiu, X. Guan, and G. Deconinck, "Matching ev charging load with uncertain wind: A simulation-based policy improvement approach," *IEEE Transactions on*

Smart Grid, vol. 6, no. 3, pp. 1425–1433, 2015.

[188] E. Walraven and M. T. Spaan, "Planning under uncertainty for aggregated electric vehicle charging with renewable energy supply," in *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pp. 904–912, IOS Press, 2016.

[189] A. Chiş, J. Lundén, and V. Koivunen, "Reinforcement learning-based plug-in electric vehicle charging with forecasted price," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3674–3684, 2016.

[190] W. Shi and V. W. Wong, "Real-time vehicle-to-grid control algorithm under price uncertainty," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 261–266, IEEE, 2011.

[191] S. Vandael, B. Claessens, D. Ernst, T. Holvoet, and G. Deconinck, "Reinforcement learning of heuristic ev fleet charging in a day-ahead electricity market," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1795–1805, 2015.

[192] S.-Y. Chen, Y. Yu, Q. Da, J. Tan, H.-K. Huang, and H.-H. Tang, "Stabilizing reinforcement learning in dynamic environment with application to online recommendation," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1187–1196, 2018.

[193] B. J. Claessens, D. Vanhoudt, J. Desmedt, and F. Ruelens, "Model-free control of thermostatically controlled loads connected to a district heating network," *Energy and Buildings*, vol. 159, pp. 1–10, 2018.

[194] G. T. Costanzo, S. Iacovella, F. Ruelens, T. Leurs, and B. J. Claessens, "Experimental analysis of data-driven control for a building heating system," *Sustainable Energy, Grids and Networks*, vol. 6, pp. 81–90, 2016.

[195] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv preprint arXiv:1506.00019*, 2015.