Title
Design and Management of Networked Energy and Logistics Systems

Author
Qi, Wei

Publication Date
2015

Peer reviewed|Thesis/dissertation

Design and Management of Networked Energy and Logistics Systems

by

Wei Qi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Zuo-Jun Max Shen, Chair
Professor Shmuel S. Oren
Professor Philip M. Kaminsky
Assistant Professor Duncan S. Callaway

Fall 2015

Design and Management of Networked Energy and Logistics Systems

Abstract

Design and Management of Networked Energy and Logistics Systems

by

Wei Qi

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Zuo-Jun Max Shen, Chair

Key to a sustainable future is the transformative human use of energy and transportation– from a fossil-fuel dominant to a renewables-mixed portfolio of energy production, from a supply-follow-demand to a demand-responsive pattern of power consumption, and from capacity-oriented to usage-based allocation of logistics mobility. Motivated by these trends, my dissertation presents three essays to address the challenges that governments and businesses worldwide face in, respectively, 1) planning infrastructure for wind energy production, 2) designing coordination strategies for large-scale charging of plug-in electric vehicles, and 3) evaluating the economic and environmental viability of using shared mobility for retail e-commerce. Two common threads underlying my dissertation research are 1) strategic decision-making with insights into the operational level and 2) network optimization that takes into account the interdependencies both within and out of system boundaries. Solving these network optimization problems invokes the techniques of mixed-integer conic programming, decomposition algorithms, and continuous approximation modeling. Specifically, the three essays are organized as three chapters of the dissertation:

Chapter 1 studies the problem of jointly planning energy storage (ES) and transmission for wind energy generation. Regions with abundant wind resources usually have no ready access to the existing electric grid. However, building transmission lines that instantaneously deliver all geographically distributed wind energy can be costly. Energy storage systems can help reduce the cost of bridging wind farms and grids, and can mitigate the intermittency of wind outputs. We propose models of transmission network planning with colocation of ES systems. Our models determine the sizes and sites of ES systems as well as the associated topology and capacity of the transmission network under the feed-in-tariff policy instrument. We first formulate a location model as a mixed-integer second-order-conic program to solve for the ES-transmission network design with uncapacitated storage. Then we propose a method to choose ES sizes by deriving a closed-form upper bound. The major insight is that, in most cases, using even small-sized ES systems can significantly reduce the total expected cost, but their marginal values diminish faster than those of the transmission lines as their capacities expand. Despite uncertainties in climate, technologies, and construction

costs, the cost-efficient infrastructure layout is remarkably robust. We also identify the major bottleneck cost factors for different forms of ES technologies.

Chapter 2 presents a hierarchical optimal control framework to coordinate the charging of plug-in electric vehicles in multifamily dwellings. A particular scenario is considered where distributed urban residential communities access electric power supplies through a common primary distribution transformer. We first formulate a centralized finite-horizon control problem. The proposed multistage mixed integer program seeks to maximize the total utility of the charging service provider while satisfying customers' charge demands and transformer capacity constraints. By exploiting the structure of the centralized model, we decompose the centralized problem with respect to each parking deck, based on the Lagrangian relaxation method; we design an effective heuristic method to find feasible solutions to speed up convergence. Case studies on operations of five parking decks following different charging strategies are carried out. Simulation results demonstrate that the proposed distributed hierarchical charging strategy outperforms the centralized charging strategy from the perspective of computational requirements. System reliability and customer privacy protection are also discussed.

Chapter 3 studies an integrated logistics system with shared mobility for retail e-commerce. Two socioeconomic transformations, namely, the booms in sharing economy and retail e-commerce, lead to the prospect where shared mobility of passenger cars prevails throughout urban areas for home delivery services. Local governments and logistics services providers are in need of evaluating the potentially substantial impacts of this mode shift, given their societal and environmental concerns and economic objectives. We addresses this need by providing a logistics planning framework: 1) Part one presents logistics planning models. These models characterize optimal routes of short-haul trucks and passenger cars, and generate the optimal density of service zones within which passenger vehicles pick up goods and fulfill the last-mile delivery. 2) Part two, based on empirical estimates, analyzes operating costs and greenhouse gas emissions implications of this sharing logistics paradigm. The findings suggest that a transition to this paradigm has the potential for creating considerable economic and environmental benefits, although immediate savings are not as achievable as one may conjecture. If being in this paradigm, even exclusively minimizing operating costs does not significantly increase emissions relative to the minimum level of emissions. A nonlinear payment scheme can be used to efficiently induce shared mobility into passenger car home delivery services.

To my parents and grandparents

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful for having Professor Zuo-Jun Max Shen as my advisor. He introduced me to the fields of operations research and operations management. His advice and support were invaluable to me. His dedication to producing high-quality research and benefiting society has motivated me to pursue my career goals.

I would like to thank my dissertation committee members–Professor Shumel Oren, Professor Phil Kaminsky, and Professor Duncan Callaway. They gave me numerous suggestions on my research. They set high standards while having been always supportive.

I am grateful to Professor Ilan Adler, Professor Jon Burgstone, Professor Steven Wood and Professor Zuo-Jun Max Shen. Working with them as a teaching assistant, I learned a great deal about teaching. I am very thankful to the IEOR staff, especially Ms. Anayancy Paz, Ms. Sonia Chahal, and Mr. Jay Sparks, who have been so nice and helpful.

Meanwhile, working outside campus was most rewarding. As a part-time graduate researcher at the Lawrence Berkeley National Laboratory, I would like thank my supervisor Dr. Arman Shehabi, and my colleagues Dr. Roger Sathre and Dr. William Morrow. I am also thankful to my colleagues at AOL Inc. where I completed a summer internship–Dr. Niklas Karlsson, Dr. Jianlong Zhang, and Ms. Antje Dittmer. I treasure the friendship that I have with them.

Many thanks to my friends in Berkeley–Tony, Long, Luming, Chris, Tianhu, Ricky, Mengshi, Yong, Vivian, Anthony, Amber and Sheng, just to name a few. They filled my graduate school life with so much joy.

Special thanks to my family in China, for their unconditional love.

Chapter 1

Joint Planning of Wind Energy Storage and Transmission

1.1 Introduction

Renewable energy, such as wind energy, is the key to a sustainable energy future. Drivers for the renewables include alleviated dependence on fossil-fuel power and nuclear power, reduced environmental hazards and prospective cheaper energy production ([30]). Governments around the world have widely released targets to push the adoption of renewable energy. For example, a collaborative effort has been made to explore a scenario in which wind provides 20% of U.S. electricity by 2030 ([27]); China plans its non-fuel energy to account for 15% of its total consumption by 2020 and has led the expansion of wind power capacity ([77]). In the mean time, leading companies in the IT sector, such as Apple, Google and Facebook, are taking significant steps to power their data centers with an increasing percent of renewable energy ([43]).

Nonetheless, wind energy infrastructure planners are facing major challenges when they are trying to meet these ambitious goals. Firstly, wind resources are geographically distributed. In the initial phase when multiple wind farms are approved to be built, the planners need to carefully design the transmission network that is usually more complex than one single line. Secondly, most of the high-quality wind resources in North America and Asia are not near major load centers and cannot be directly integrated into the existing transmission network ([26]; [32]; [33]; [42]). As a result, dedicated long-distance transmission lines have to be built to deliver electricity from remote wind farms. Thirdly, the intermittent nature of wind necessitates high-capacity but lightly-loaded transmission lines, which otherwise would result in significant generation curtailment. For example, Southern California Edison reported curtailed wind energy generation of about 15MW for 6-8% of the time as of 2010 due to transmission constraints ([78]).

We try to address these challenges by proposing models of transmission network planning with colocation of energy storage (ES) systems. The primary function of an ES system is to

decrease the variability of wind energy generation by absorbing/discharging electricity when wind power output mismatches the rated transmission capacity or power demand. The extensive value of colocating ES with wind energy generation has been reported in [29] and [36]. In this chapter, our goal is to develop models and solution methods to determine sizes and sites of ES systems as well as the associated topology and capacity of the transmission network. As a result, wind energy from these geographically distributed wind farms can be effectively tapped with minimum infrastructure investment cost and energy loss. In doing so, we also try to understand how to best exploit the value of using ES for future renewable energy production.

Our model and analysis are based on the following problem settings: a set of sites in a region with abundant wind resources have been selected as wind farms. These sites are located in desolated areas that have no ready access to main transmission infrastructure (see [92] for practical considerations in siting wind farms). A planner of the local government or a utility company is to design a network of energy storage systems and transmission lines (hereafter referred to as ES-transmission network) to connect the wind farms to a single load center (e.g., a town) or a substation of the region; or, an IT company aims to power its 120MW data center with 100 percent wind energy. We consider the network topology to be radial. That is, wind outputs from different farms are first transmitted to junction sites with or without ES, and then the pooled power at each junction site flows to the load center. The radial transmission network is widely adopted in practice to tap remote wind resources, such as in southern California and Atlantic offshore zones ([18] and [7]).

In addition, we assume that the region implements feed-in-tariffs (FIT) policy, which guarantees a long-term contract for renewable power producers to sell their electricity at a fixed price ([67]). As a result, wind farms have no price arbitrage incentive and it is optimal to deliver as much energy as transmission capacities permit. Among the existing policy mechanisms, FIT is particularly effective to foster initial adoption of renewable energy and fits well with the practice in most of the world's major electricity markets, where governments enforce the purchase price of electricity to be higher than its energy production cost or subsidize wind energy generation utilities to attract them to enter the market ([4] and [77]).

Our first model considers the case where ES systems are assumed to have sufficient energy capacity to accommodate intermittent surplus wind output. We first derive two optimal transmission line capacities as functions of wind characteristics for a single wind farm with and without ES being coupled, respectively. Then we use these optimal quantities to formulate a model to design an ES-transmission network. The model is in the form of a mixed-integer second-order-conic program (MISOCP), which can be efficiently solved by commercial software. Our second model considers the sizing problem of ES. We derive a closed-form upper bound of the expected energy overflow due to the capacity limit, as a function of ES and transmission capacities. In the above models, following one similar assumption of [54], we approximate the hourly and daily wind output by uniform distribution. The numerical experiments suggest that the approximation error is small. Combining these models, the infrastructure planner obtains both lower and upper bounds of the expected minimum capital and operational cost of the network. The gap between the two bounds is

reasonably small.

The contributions of this work are as follows: (1) To the best of our knowledge, this work is the first attempt to provide infrastructure planners with models and solution approaches to jointly plan the sites and the sizes of ES systems and transmission lines for distributed wind resources. (2) We develop quantitative models and managerial insights to help planners understand the value and cost of using ES. We analyze the dual effects of using ES, that is, saving transmission capacity by reducing output variability versus incurring energy loss due to friction and overflow (as an in-depth quantitative extension to the discussion in [26]). We find that, in most cases, using even small-sized ES systems can significantly reduce the total expected cost, but their marginal values diminish faster than those of the transmission lines as their capacities expand. We also identify the bottleneck cost factors for different forms of ES technologies. For example, for compressed air storage systems, it is more beneficial to improve their energy conversion efficiency than to reduce their per-unit capacity cost. These insights can be used to make long-term investment decisions as technology advancements bring down ES cost. (3) Another finding is that the layout of the ES-transmission network that we obtain is robust against uncertainties such as FIT rates adjustments, technology advancements, climate changes and construction material cost fluctuations. Hence, planners can determine the infrastructure layout long before these uncertainties are resolved, without worrying about costly reconfiguration of the network. (4) We also incorporate major wind characteristics into infrastructure planning. In particular, our models capture the nature of wind energy such as hourly and daily intermittence, spatial correlation and the variability pooling effect, which are all important factors but have not been well considered in the literature.

The remainder of this chapter is organized as follows. Section 1.2 reviews the related literature. Section 1.3 introduces notation and basic settings of our models. Section 1.4 presents the infrastructure planning model with uncapacitated ES systems. Section 1.5 derives an upper bound of the size of an ES system and incorporates it into the planning procedure. Section 1.6 demonstrates our computational results and presents managerial insights into technology impact and layout robustness. Finally, Section 1.7 concludes the chapter. In addition, numerical and theoretical analysis of model inaccuracy, additional structural properties, proposition proofs and numerical experiment settings are available in Appendix A.

1.2 Literature Review

There has only been a very limited number of studies that are related to the important problem of deploying ES systems for wind power delivery in the literature of transmission expansion planning (see [57] and [48] for comprehensive reviews, [93] for recent progress on conic approximations to alternative current (AC) transmission system planning, and [68] and [8] for some recent studies that incorporate wind resources). The closest to ours are [74] and [102], both of which formulate deterministic mixed-integer linear programs to plan

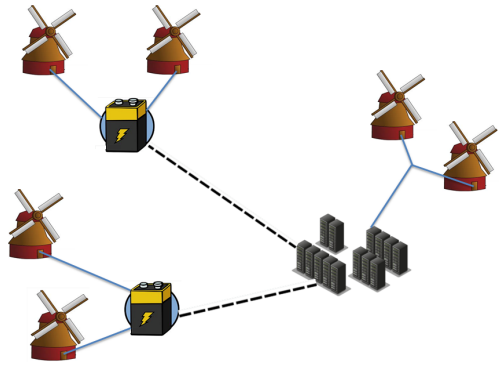

Figure 1.1: A radial ES-transmission network with economic (dashed) and ES-free (solid) lines.

ES systems in an existing power grid, without wind resources being considered. In addition, [26] study how ES saves transmission cost when it is located close to wind farms. However, none of these studies address issues such as determining ES/transmission capacities while capturing wind characteristics, which are the contributions of our work from a supply chain design perspective.

Our infrastructure planning model for wind energy generation is reminiscent of the location-inventory model proposed by [84], though they are different on some fundamental aspects. Compared to warehouses in a distribution network, the ES systems incur fixed upfront cost, variable capacity cost and nonlinear cost due to charge/discharge friction loss and overflow loss of energy. Meanwhile, whereas the retailers to be assigned to the warehouses face random custom demands, the wind farms in our problem setting are to be assigned to ES-coupled or ES-free junction sites and face intermittent wind energy outputs. For the solution approach, we formulate our planning problem in the form of a computationally-tractable second-order conic program. For more applications of conic formulations, please refer to [6] for solving the location-inventory problem and its various extensions, [64] for planning battery-swapping stations of electric vehicles, and [70] for portfolio optimization.

1.3 Model Settings

We consider a radial ES-transmission network as illustrated in Figure 1.1. A given set of geographically distributed wind farms are approved to be built. Each of these farms is to generate and deliver electrical power to a junction site. Then the power pooled at each junction site flows to a given common load center (or substation). Our objective is to jointly determine a) the assignment of the wind farms to the junction sites; b) whether to install an ES system at a selected junction site; c) energy capacities of the ES systems; and d) power capacities of the transmission lines.

We are interested in how the total expected cost relates to ES and transmission capacities as well as wind intermittence. Specifically, this cost breaks down to two parts:

(1) The *building cost* of ES (transmission), which consists of a fixed installation cost and a variable cost proportional to the ES (transmission) capacity. The fixed and the variable cost components of transmission are assumed to be both proportional to the length of the line.

(2) And the *energy loss*, which is incurred as (i) *friction loss*, owing to non-perfect roundtrip conversion efficiency when electricity is charged into and discharged from an ES system, (ii) *overflow loss*, when both an ES system and its downstream transmission line hit their maximum capacities and the surplus wind energy can be neither stored nor transmitted, or (iii) *curtailment loss*, when ES is absent and instantaneous wind power output that exceeds the downstream transmission line capacity has to be abandoned.

As for the choices of transmission capacity, we consider two types of transmission lines. First, between ES-equipped junction sites and the load center, we choose *economic lines*. An economic line and its associated ES system are complementary in reducing energy loss and their capacities need to be jointly optimized. Second, *ES-free lines* are built between the wind farms and the junction sites and from ES-free junction sites to the load center. A properly sized ES-free line strikes the balance between saving transmission building cost and reducing curtailment loss. In both cases, given the FIT instrument, wind farms have no price arbitrage incentive. It is optimal to deliver as much generated and stored energy as transmission line capacities permit.

1.4 Model with Uncapacitated Storage

In this section, we first derive an optimal transmission line capacity for a single wind farm colocated with an uncapacitated ES system. Then we derive an optimal transmission capacity of an ES-free line. These results then lead to an ES-transmission planning problem formulation for multiple distributed wind farms with uncapacitated ES systems. In Section 1.5 we will show that this uncapacitated case provides a reasonable approximation to the case with capacity limts. Table 1.1 summarizes the notation for the case of a single wind farm. Parameters and functions are denoted by lowercase letters, random variables by bold lowercase letters, matrices by uppercase Greek letters and decision variables by uppercase English letters.

A Single Wind Farm with ES

Consider a basic scenario: a single wind farm is coupled with an ES system and delivers electricity through a capacitated transmission line. In this case, it is optimal to colocate the ES system with the wind farm to avoid the cost of building ES-free transmission line between them.

Table 1.1: Summary of notation.

Systems Parameters	
α, β	Charge and discharge efficiency of ES systems, respectively.
δ	Length (in years) of each time interval. We assume $\delta = 1\text{hr} = \frac{1}{24\times365}\text{yr}$ hereafter.
$\mathbf{w}_t, w_t$	Random variable and its realization of the energy that a wind farm captures during interval $[t-1, t)$, respectively, where $t \in \mathbb{N}$ is the index of the intervals of length δ.
$\mathbf{l}_t$	Loss of energy during $[t-1, t)$.
$f_w(\cdot)$	Probability density function of $\mathbf{w}_t$.
μ, ϵ	Mean and interval length of the approximated uniform distribution of $\mathbf{w}_t$, respectively.
l	Length of a transmission line.
Price and Costs	
p	Fixed contracted electricity selling price.
r	Annualized per-kWh building cost of ES capacity.
a	Annualized building cost of a transmission line per kW per mile.
$q = al$	Annualized building cost of a transmission line per kW.
θ, η	Dimensionless capacity cost indices of an economic line and an ES-free line, respectively.
Decision Variable	
C	Maximum electrical energy that can be transmitted over a period of δ by a transmission line. C is also in the unit of power (kW) when $\delta = 1\text{hr} = \frac{1}{24\times365}\text{yr}$.

Intuitively, as the transmission line capacity increases and/or the installed ES capacity increases, the building cost increases while the energy loss decreases. To quantify this trade-off, we first assume that $r = 0$ and the ES system is large enough to incur no overflow loss almost surely. The energy loss $\mathbf{l}_t$ during $[t-1, t)$ thus consists only of the friction loss:

$$\mathbf{l}_t = \begin{cases} (\mathbf{w}_t - C)(1-\alpha\beta), & \text{if } \mathbf{w_t} - C > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{1.1}$$

At times when the wind output power exceeds the transmission line capacity, the surplus energy $(\mathbf{w}_t - C)$ is charged into and at some future time discharged from the ES system, incurring a friction loss of $(\mathbf{w}_t - C)(1-\alpha\beta)$. Otherwise, all the generated wind energy can be directly delivered. We follow the approach in [54] to assume that $\mathbf{w}_t$ is uniformly distributed: $\mathbf{w}_t \sim \text{uniform}(\mu - \frac{\epsilon}{2}, \mu + \frac{\epsilon}{2})$. We obtain the mean μ and the interval length ϵ by matching the mean and the variance of the real wind outputs. Two reasons lead to our choice of uniform distributions over others (such as normal distributions) to approximate wind outputs. First, the uniform distribution is mathematically tractable, enabling us to derive closed-form results that are key not only to the efficient planning problem formulation, but also to the managerial insights into ES value and model suboptimality. Second, the uniform distribution, with its

bounded support, is effective in characterizing wind curtailment, which results from wind turbine operations and capacitated transmission lines. This approximation is further justified by the numerical experiments later in this section and in Section A.1. Note that the wind output process $\{\mathbf{w}_t\}$ can be auto-correlated and non-stationary. When $C \geq \mu - \frac{\epsilon}{2}$, the expected energy loss in $[t-1, t)$ is given by:

$$\begin{aligned} \mathbb{E}[\mathbf{l}_t] &= \int_C^{\mu+\epsilon/2} (w_t - C)(1-\alpha\beta)\tfrac{1}{\epsilon} dw_t \\ &= \tfrac{1-\alpha\beta}{2\epsilon}(\mu + \tfrac{\epsilon}{2} - C)^2. \end{aligned} \tag{1.2}$$

Although in practice transmission capacity C can only be chosen from a finite set of discrete values, we assume C is continuous-valued for model tractability, because the discrete set of candidate capacities is considerably flexible, with various line specifications available. In addition, we assume that the variable transmission capital cost is linear in C. This linear approximation is present and justified in early literature of transmission expansion planning (e.g., [53] and [55]). Recent empirical evaluation ([65] and [66]) also suggests that transmission capital cost exhibits a significant linear relation with transmission capacity in a wide range. With these two assumptions, the expected annual variable cost due to friction loss and capital investment can be expressed as a quadratic function of C:

$$\begin{aligned} v_1(C) &= \tfrac{p\mathbb{E}[\mathbf{l}_t]}{\delta} + qC \\ &= p\tfrac{1-\alpha\beta}{2\epsilon\delta}(\mu + \tfrac{\epsilon}{2} - C)^2 + qC. \end{aligned} \tag{1.3}$$

It can be verified that $C = (\mu + \frac{\epsilon}{2}) - \frac{\epsilon\delta q}{p(1-\alpha\beta)}$ minimizes (1.3). We make an additional assumption that transmission line capacity should be greater than or equal to the average wind output power; otherwise there is no steady state distribution of the storage level. Hence, the economic transmission capacity which minimizes the expected annual variable cost v_1 is given by:

$$\begin{aligned} C^* &= \arg\min_{C>=\mu} v_1(C) \\ &= \max\{\mu, (\mu + \tfrac{\epsilon}{2}) - \tfrac{\epsilon\delta q}{p(1-\alpha\beta)}\} \\ &= \begin{cases} \mu + (\frac{1}{2} - \theta)\epsilon, & \text{if } \theta < 1/2; \\ \mu, & \text{otherwise,} \end{cases} \\ \text{where } \theta &= \tfrac{\delta q}{p(1-\alpha\beta)}. \end{aligned} \tag{1.4}$$

The dimensionless number θ captures the cost associated with building transmission capacity. For example, if q is large due to long transmission distance or high unit capacity cost, or if the ES conversion is very efficient such that $\alpha\beta$ is close to 1, then θ tends to be large, indicating that building extra transmission capacity is cost-ineffective. When $\theta \geq \frac{1}{2}$, it is favorable to construct a line that transmits at most average wind power. It is important to notice that θ is independent from the wind characteristics. When planning the entire ES-transmission network, this invariance helps pre-determine which segment of the following non-smooth cost expression (1.5) to use in formulating the network design problem, before

we know the assignment of the wind farms to the junction sites. Substituting (1.4) into (1.3) yields the optimal annual variable cost:

$$v_1^*(C^*) = \begin{cases} q\mu + q(\frac{1}{2} - \frac{1}{2}\theta)\epsilon, & \text{if } \theta < \frac{1}{2}; \\ q\mu + \frac{p(1-\alpha\beta)}{8\delta}\epsilon, & \text{otherwise.} \end{cases} \tag{1.5}$$

A Single Wind Farm without ES

We next look into another basic scenario where a single wind farm is connected with a transmission line without ES being colocated. In this case, the optimal transmission capacity can be directly expressed as the optimal quantile in the classic newsvendor model with stockout cost $(p - q\delta)$ and inventory holding cost $q\delta$. An explicit derivation resembles the steps in the case with ES. Specifically, the curtailment loss during $[t-1, t)$ is given by:

$$\mathbf{l}_t = \begin{cases} (\mathbf{w}_t - C), & \text{if } \mathbf{w_t} - C > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{1.6}$$

Applying uniform distribution approximation again, when $C \geq \mu - \frac{\epsilon}{2}$, the expected curtailment loss is given by:

$$\begin{aligned} \mathbb{E}[\mathbf{l}_t] &= \textstyle\int_C^{\mu+\epsilon/2} (w_t - C)\frac{1}{\epsilon} dw_t \\ &= \frac{1}{2\epsilon}(\mu + \frac{\epsilon}{2} - C)^2. \end{aligned} \tag{1.7}$$

The expected annual variable cost due to curtailment loss and transmission capacity investment as a function of C becomes:

$$\begin{aligned} v_2(C) &= \frac{p\mathbb{E}[\mathbf{l}_t]}{\delta} + qC \\ &= \frac{p}{2\epsilon\delta}(\mu + \frac{\epsilon}{2} - C)^2 + qC \\ &= \frac{p}{2\epsilon\delta}C^2 + (-\frac{p}{\epsilon\delta}(\mu + \frac{\epsilon}{2}) + q)C + \frac{p}{2\epsilon\delta}(\mu + \frac{\epsilon}{2})^2. \end{aligned} \tag{1.8}$$

The cost-minimizing ES-free line capacity is thus given by:

$$\begin{aligned} C^* &= \begin{cases} \arg\min v_2(C), & \text{if } \eta < 1; \\ 0, & \text{otherwise;} \end{cases} \\ &= \begin{cases} \mu + (\frac{1}{2} - \eta)\epsilon, & \text{if } \eta < 1; \\ 0, & \text{otherwise,} \end{cases} \\ \text{where } \eta &= \frac{\delta q}{p}. \end{aligned} \tag{1.9}$$

Similar to θ for an economic line, η is the dimensionless capacity cost index for an ES-free line, independent from wind characteristics. Larger η indicates higher level of line capacity restriction. When $\eta \geq 1$, building transmission capacity is no longer profitable even when the line is fully loaded all the time, so we opt not to build the line and forgo all the wind energy. The associated optimal annual cost is:

$$v_2^*(C^*) = \begin{cases} q\mu + q(\frac{1}{2} - \frac{1}{2}\eta)\epsilon, & \text{if } \theta < 1; \\ \frac{p\mu}{\delta}, & \text{otherwise.} \end{cases} \tag{1.10}$$

.

A Numerical Example

The following simple numerical example illustrates how the variable cost and the approximation error of wind output vary with respect to the transmission distance. Consider distance l ranging from $0 - 200$ miles and set $p = \$0.08$/kWh (which is projected to be the levelized electricity cost for new wind plants in 2020, estimated by [30]), $\alpha\beta = 0.72$ and $a = \$1$/kW-mile.

Figure 1.2 (a) and (b) are the area plots of the total variable cost with and without ES being colocated, respectively. We use wind output data of a modeled site from [73] (The site ID is 24648). In Figure 1.2 (a), the total variable cost generated using these data with the transmission capacity C^* prescribed by (1.4) is shown as the sum of the friction cost and the transmission capacity cost. When $l < 98$ miles, the transmission capacity cost increases in transmission distance l yet with a decreasing rate of change, as C^* decreases to partly offset the increased capacity cost, which in turn incurs more charge/discharge friction. When $l \geq 98$ miles, $C^* = \mu$, and hence the friction cost reaches standstill and the transmission capacity cost increases linearly in l. In Figure 1.2 (b), the total variable cost with C^* prescribed by (1.9) is shown as the sum of the curtailment cost and the transmission capacity cost. Similarly, longer transmission distance results in smaller transmission capacity and thus higher curtailment cost.

Figure 1.2 also suggests that the error of approximating the wind output using uniform distribution is reasonably small. We first use line search on the same data set to find the actual optimal transmission capacity and the corresponding cost. This cost, as represented by the dashed line in each plot, is close to the cost with transmission capacity C^*. In the ES-coupled (ES-free) case, the average relative gap is 3.01% (7.02%) for $l \leq 50$ miles and 0.22% (1.09%) for $l > 50$ miles. Then we use simulated data from a uniform distribution that matches the first and the second moments of the raw data. Again, the approximated total variable cost (the line with triangles) is close to the real cost, with about a 2.5% (5%) relative gap in the ES-coupled (ES-free) case. It can be verified that the bias of the cost approximation is eliminated when $C = \mu$ regardless of the real distribution of the wind output. These facts enable us to incorporate the cost terms (1.5) and (1.10) into the planning model for multiple wind farms, which we elaborate in the remainder of this section.

Multiple Farms

We next develop the planning model for multiple wind farms. Obviously, building ES systems at all the sites of the wind farms can be cost-inefficient. We instead try to economically select power junction sites (which can also be some wind farms) to aggregate wind outputs with or without ES systems being colocated.

Additional notation used for the scenario of multiple wind farms is summarized in Table 1.2. Wherever it is necessary, we add subscripts to symbols to indicate location. For instance,

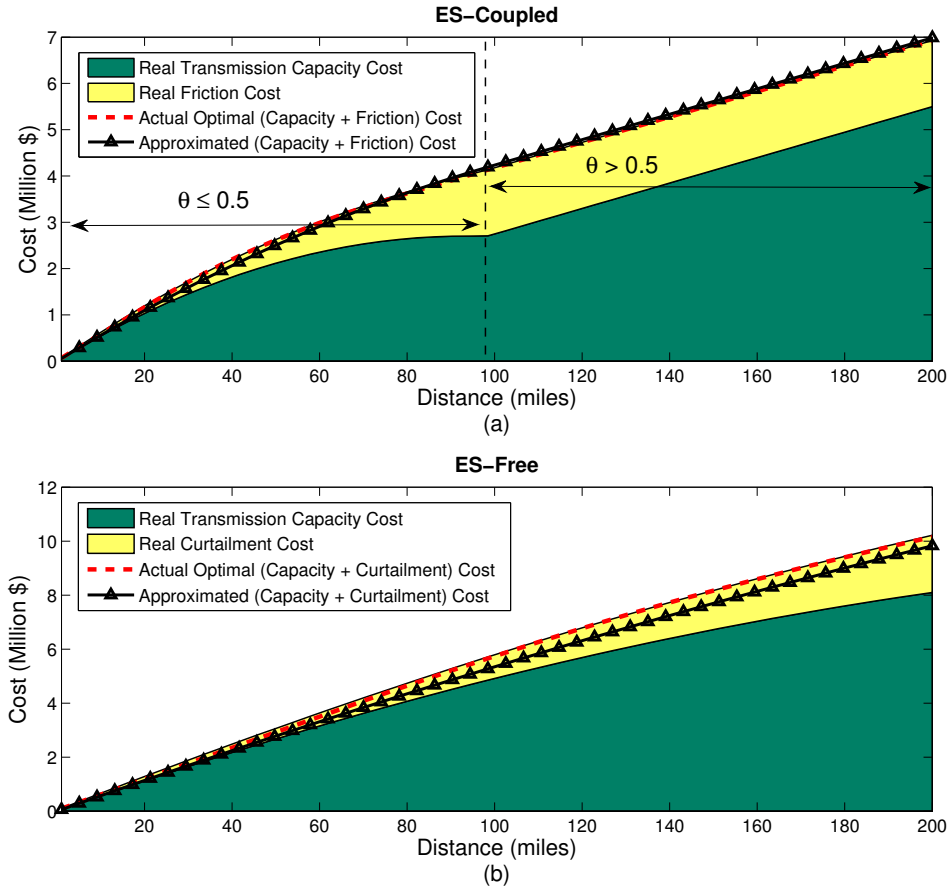

Figure 1.2: Variable costs of a single farm (a) with and (b) without ES.

q_{ij} and q_j refer to capacity per unit costs of building transmission lines from wind farm i to junction site j, and from junction site j to the load center, respectively.

We assume that ES-free lines are built between wind farms and junction sites and between ES-free junction sites and the load center, with line capacities given by (1.9). We follow the same logic as in the single-farm scenario to use a uniform distribution to approximate the probability distribution of the curtailed wind power $\mathbf{w}_{t,ij}$, which is from farm i and faced by site j. When $\eta_{ij} < 1$, this uniform distribution has mean and interval length expressed as follows (The derivation is available in Section A.1):

$$\mu_{ij} = \mu_i - \frac{1}{2}\epsilon_i \eta_{ij}^2, \tag{1.11a}$$

$$\epsilon_{ij} = \sqrt{(1-\eta_{ij})^3(1+3\eta_{ij})}\epsilon_i. \tag{1.11b}$$

Table 1.2: Summary of additional notation.

Sets	
I	Set of wind farms, indexed by $i \in I$.
J	Set of candidate junction sites pooling wind outputs, indexed by $j \in J$.
Systems Parameters	
Σ_j	Covariance matrix of wind outputs from the wind farms in I, aggregated at site j.
ρ_{ikj}	Correlation coefficient of the outputs at site j from $i, k \in I$, aggregated at site j.
Costs	
h_j	Annualized fixed upfront cost to build an ES system at junction site $j \in J$.
g_{ij}, g_j	Annualized fixed construction cost of the transmission line from wind farm $i \in I$ to site $j \in J$ and from site j to the load center (or a substation), respectively.
Decision Variables	
X_j	1 if an ES system is built on site $j \in J$, 0 otherwise.
V_j	1 if site $j \in J$ is selected as a junction site with no ES system, 0 otherwise.
Y_{ij}	1 if wind farm $i \in I$ is assigned to junction site j with an ES system, 0 otherwise.
Y_j	vector $(Y_{1j}, Y_{2j}, ..., Y_{\|I\|j})^T$.
Z_{ij}	1 if wind farm $i \in I$ is assigned to site $j \in J$ without an ES system, 0 otherwise.
Z_j	vector $(Z_{1j}, Z_{2j}, ..., Z_{\|I\|j})^T$.
E_j,$\hat{E}_j$	Interval length of the approximated uniform distribution of the pooled wind output faced by junction site j with and without an ES system, respectively.

Spatial correlation of wind speed and power has been extensively reported and used in wind forecast ([2]); therefore, it should be explicitly modeled. Each selected junction site j faces pooled and correlated wind outputs from a subset of wind farms. Again, we apply uniform distribution approximation to this pooled wind output, i.e., $\mathbf{w_{t,j}} \sim \text{Unif}(\mu_j - \frac{\epsilon_j}{2}, \mu_j + \frac{\epsilon_j}{2})$. Let Σ_j be the covariance matrix of wind outputs from the wind farms in I and aggregated at junction site j. Each entry $\Sigma_{ikj} = \epsilon_{ij}\rho_{ikj}\epsilon_{kj}$, where ρ_{ikj} is the correlation coefficient of the curtailed wind outputs from i and k at site j. Matching the first and the second moments of $\mathbf{w_{t,j}}$, we obtain:

$$\mu_j = \sum_{i \in I} Y_{ij}\mu_{ij}, \tag{1.12a}$$

$$\epsilon_j^2 = \sum_{i \in I} \epsilon_{ij}^2 Y_{ij} + \sum_{i,k \in I, i<k} 2\rho_{ikj}\epsilon_{ij}\epsilon_{kj}Y_{ij}Y_{kj} = Y_j^T \Sigma_j Y_j. \tag{1.12b}$$

Similarly, ϵ_j^2 and μ_j for site j having no ES system can be expressed by (1.12) with Y being replaced with Z. Note that it suffices to only know historical statistics of wind at individual farms as well as transmission per unit cost in order to compute the values of $\{\eta_{ij}, \mu_{ij}, \epsilon_{ij}, \rho_{ikj}\}, \forall i, k \in I, j \in J$.

For the lines between ES-equipped junction sites and the load center, the economic transmission capacity is given by (1.4). As discussed previously, by computing the dimensionless

number θ_j we can pre-determine which segment of the non-smooth cost in (1.5) to be incorporated into our planning model before knowing the assignment of the wind farms. The candidate junction sites are thus categorized into the following two subsets based on θ_j:

$$\begin{cases} J_1 = \{j \in J | \theta_j < \frac{1}{2}\}; \\ J_2 = J \setminus J_1. \end{cases}$$

With the above curtailment and pooling considerations as well as transmission capacity choices, the ES-transmission planning model is formulated as follows.

$$\begin{aligned}
\text{minimize} \quad & v_3(X,V,Y,Z,E,\hat{E}) = \sum_{j\in J}[h_j X_j + \sum_{i\in I} g_{ij}(Y_{ij}+Z_{ij}) + g_j(X_j+V_j)] \\
& + \sum_{j\in J}\sum_{i\in I}[q_{ij}\mu_i + q_{ij}(\frac{1}{2} - \frac{1}{2}\eta_{ij})\epsilon_i](Y_{ij}+Z_{ij}) \\
& + \sum_{j\in J}[q_j(\sum_{i\in I}\mu_{ij}Z_{ij} + q_j(\frac{1}{2} - \frac{1}{2}\eta_j)\hat{E}_j)] \\
& + \sum_{j\in J_1}[q_j\sum_{i\in I}\mu_i Y_{ij} + q_j(\frac{1}{2} - \frac{1}{2}\theta_j)E_j] + \sum_{j\in J_2}[q_j\sum_{i\in I}\mu_i Y_{ij} + \frac{p(1-\alpha\beta)}{8\delta}E_j] \quad (1.13a) \\
\text{subject to} \quad & \sqrt{Y_j^T \Sigma_j Y_j} \le E_j \quad \forall j \in J \quad (1.13b) \\
& \sqrt{Z_j^T \Sigma_j Z_j} \le \hat{E}_j \quad \forall j \in J \quad (1.13c) \\
& \sum_{j\in J}(Y_{ij}+Z_{ij}) = 1 \quad \forall i \in I \quad (1.13d) \\
& X_j + V_j \le 1 \quad \forall j \in J \quad (1.13e) \\
& Y_{ij} \le X_j \quad \forall i \in I, \forall j \in J \quad (1.13f) \\
& Z_{ij} \le V_j \quad \forall i \in I, \forall j \in J \quad (1.13g) \\
& Y_{ij} = 0, Z_{ij} = 0 \quad \forall (i,j) \in \{(i,j) | \eta_{ij} \ge 1\} \quad (1.13h) \\
& X_j = 0, V_j = 0 \quad \forall j \in \{j | \eta_j \ge 1\} \quad (1.13i) \\
& E_j, \hat{E}_j \ge 0 \quad \forall j \in J \quad (1.13j) \\
& X_j, V_j, Y_{ij}, Z_{ij} \in \{0,1\} \quad \forall i \in I, \forall j \in J. \quad (1.13k)
\end{aligned}$$

In the above formulation, the objective (1.13a) is to minimize the total expected annual building and operating cost of the ES-transmission network for the given set of wind farms. The three terms in the first bracket are the fixed construction cost of ES systems, the fixed cost of transmission lines from the wind farms to the junction sites, and the fixed cost of transmission lines from the junction sites to the load center, respectively. The terms in the second and the third brackets are the variable costs of the ES-free transmission lines from the wind farms to the junction sites and from ES-free junction sites to the load center,

respectively, according to (1.10). The uniform distribution parameter ϵ_j for junction site j with (without) an ES system is denoted by variable E_j ($\hat{E}_j$). The terms in the fourth and the fifth brackets are the variable costs of the economic lines from the two θ-categorized subsets of ES-equipped junction sites, J_1 and J_2, to the load center, with different cost expressions given by (1.5).

Constraints (1.13b) and (1.13c) define E_j and $\hat{E}_j$, respectively, based on equation (1.12b). Constraints (1.13d) ensure that each wind farm in set I is assigned to one and only one junction site in set J. Constraints (1.13e) suggest that each candidate junction site can only get one of the three outcomes: to be selected and equipped with an ES system, to be selected without ES or not to be selected. Constraints (1.13f) and (1.13g) require that a wind farm can only be assigned to a junction site that is selected. Constraints (1.13h) and (1.13i) exclude the potential assignment and junction site selection that would result in unprofitable transmission investment, as discussed in Section A.1.

The above planning model is formulated as an MISOCP. The right-hand side of constraints (1.13b) and (1.13c) can be converted to the standard two-norm form as $\| \tilde{\Sigma}_j Y_j \|_2 \leq E_j$ and $\| \tilde{\Sigma}_j Z_j \|_2 \leq \hat{E}_j$, respectively, $\forall j \in J$, where $\tilde{\Sigma}_j^T \tilde{\Sigma}_j = \Sigma_j$ since Σ_j is positive definite. Meanwhile, the objective function and all the other constraints are linear in the decision variables. Recent years commercial software such as CPLEX have launched standard solvers for MISOCP. We will show that this planning model can be efficiently solved in a case of practical scale in Section 1.6. Please refer to [14] and [3] for comprehensive review of convex conic programs and MISOCP.

The above modeling process introduces two sources of inaccuracy. First, we apply uniform distribution approximations to wind outputs. As discussed in the numerical example in Section A.1, this approximation at wind farms affects the sizing of the transmission lines that are upstream of the junction sites. Then we apply similar approximations to the wind outputs that are curtailed by the upstream ES-free lines and that are aggregated at the junction sites. These approximations affect the sizing of the transmission lines that are downstream of the junction sites. Second, planning model (1.13) tends to underestimate the transmission cost and thus oversize the lines. This is because the capacities of transmission lines upstream of the junction sites are evaluated based on (1.9), which overlooks transmission lines downstream of the junction sites.

However, we find that the model inaccuracy is well-contained in practical settings, as briefly summarized in Table 1.3 for the cases with and without ES being colocated. In Section A.1, each of those components of inaccuracy as well as the overall inaccuracy is quantified numerically and/or theoretically with detailed discussion. The theoretical studies also identify two more structural properties, namely the curtailment-independence of downstream transmission capacity and the decomposition of the joint optimization of upstream and downstream transmission capacities.

Planning model (1.13) can be extended in several ways to account for different investment considerations, such as maximum covering of wind farms. We omit such discussion for brevity, since those extensions are structurally similar.

Table 1.3: Overall model inaccuracy.

	ES-Free	ES-Coupled
Mean	8.25%	6.42%
Maximum	17.6%	14.8%

Table 1.4: Summary of additional notation.

Systems Parameters	
δ_b	Length (in hours) of each time interval of independent bulk wind energy process.
$\mathbf{w}_{b,\tau}$	Random variable of the bulk energy a wind farm captures during interval. $[\tau-1,\tau)$, where $\tau \in \mathbb{N}$ is the index of the intervals of length δ_b.
μ_b, ϵ_b	Mean and interval length of the approximated uniform distribution of $\mathbf{w}_{b,\tau}$, respectively.
$\mathbf{s}_\tau$	Storage level of an ES system at the end of interval τ.
$f_s(\cdot)$	Probability density function of $\mathbf{s}_\tau$.
$\mathbf{o}_\tau$	Energy overflow loss during interval τ.
Decision Variable	
S	Maximum amount of potential energy that can be stored in an ES system.

1.5 Capacitated Storage

In this section, we relax the assumption of uncapacitated ES and explicitly characterize the dependence of the energy overflow cost on ES and transmission capacities. Since the fluctuation of storage level is complicated and cannot be quantified in closed form, we instead derive a conservative estimate of the storage level distribution and then an upper bound for the optimal ES capacity. Moreover, we show in Sections 1.6 and A.3 that this upper bound results in near-optimal expected total cost. The additional notation is summarized in Table 1.4.

Upper Bound of Energy Overflow

We choose relatively long periods (e.g. 1 day) with indices denoted as $\tau = 1, 2, ...$ and interval length as δ_b. In this way, the bulk wind output process, $\{\mathbf{w}_{b,\tau}\}$, is much less auto-correlated than the hourly process, due to diurnal cycles of wind speed ([94]). We further assume that $\{\mathbf{w}_{b,\tau}\}$ is an independent and identical process. Such simplification causes underestimation of inter-period energy overflow loss when the storage level is nearly full. On the other hand, we implicitly assume that the wind output power within each interval τ is of constant value $\frac{\mathbf{w}_{b,\tau}}{\delta_b}$. Consequently, intra-period energy overflow is overestimated, in that, in the long run, those real sample paths of wind output that are not constant but amount to the same bulk energy $\mathbf{w}_{b,\tau}$ within an interval result in more friction loss and thus less occurrences of $\mathbf{s}_\tau$ hitting S. These counteracting inaccuracies are further discussed later in the section and

Section A.3.

We use $\mathbf{s}_\tau$ to denote the storage level at the end of each interval τ, while assuming that the base storage capacity that accommodates periodic variation of storage level within each interval has already been captured by the fixed cost h. Our goal is to derive an economic storage capacity S in addition to the base capacity by obtaining the expected energy overflow as a function of S. As the first step, the transition of $\mathbf{s}_\tau$ is modeled as follows ([54]):

$$\mathbf{s}_{\tau+\mathbf{1}} = \begin{cases} S, & \text{if } \mathbf{s}_\tau + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S; \\ \mathbf{s}_\tau + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b), & \text{if } \mathbf{w_{b,\tau+1}} - C\delta_b > 0, \mathbf{s}_\tau + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) < S; \\ \mathbf{s}_\tau - \frac{1}{\beta}(C\delta_b - \mathbf{w_{b,\tau+1}}), & \text{if } \mathbf{w_{b,\tau+1}} \leq C\delta_b < \beta\mathbf{s}_\tau + \mathbf{w_{b,\tau+1}}; \\ 0, & \text{if } C\delta_b \geq \beta\mathbf{s}_\tau + \mathbf{w_{b,\tau+1}}. \end{cases} \tag{1.14}$$

The four segments of the above piecewise linear function represent the four states of the storage level, respectively: fully charged, being charged, being discharged and out of charge.

Deriving a closed-form expression of the probability density function $f_s(\cdot)$ for $\mathbf{s}_\tau$ is mathematically challenging, and hence we construct an approximation of $f_s(\cdot)$. We first make two assumptions: (1) The storage capacity S is large enough such that the probability of storage level switching from zero state to full state (or the other way around) within one interval is negligible. In fact, suppose that the ES capacity is small such that the above assumption is violated. Then the additional cost due to ES capacity and energy overflow becomes very small and dominated by the cost of transmission capacity. We will further examine this assumption in the numerical studies in Section A.3. (2) We also assume that $f_s(s)$ is decreasing in s in the open interval $(0, S)$ when $C\delta_b$ is greater than or equal to the mean of $\mathbf{w}_{b,\tau}$. This assumption is realistic, since at certain storage levels s_τ, the probability of charge is greater than or equal to the probability of discharge. Also notice that, due to friction loss, any difference $\Delta S = |C\delta_b - \mathbf{w}_{b,\tau}|$ results in smaller magnitude of increase in s_τ when $C\delta_b > \mathbf{w}_{b,\tau}$ than the magnitude of decrease in s_τ when $C\delta_b < \mathbf{w}_{b,\tau}$ (see more detailed justification of this assumption in the proof of Proposition 1 in Section A.2). Then we obtain the following proposition:

Proposition 1. *Assume $C\delta_b \geq \mathbb{E}[\mathbf{w}_{b,\tau}]$, and suppose $\tilde{f}_s(s)$ is an approximation of $f_s(s)$ such that $\tilde{f}_s(s)$ is constant in the open interval $(0, S)$. Then*

(i) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) \geq \mathbb{P}(\mathbf{s}_\tau = S)$;

(ii) $\tilde{\mathbb{E}}[\mathbf{o}_\tau] \geq \mathbb{E}[\mathbf{o}_\tau]$,

where $\tilde{\mathbb{P}}(\cdot)$ and $\tilde{\mathbb{E}}[\cdot]$ denote probability and expectation with respect to $\tilde{f}_s$, respectively.

Proposition 1 (ii) establishes a sufficient condition to obtain an upper bound for the expected energy overflow $\mathbb{E}[\mathbf{o}_\tau]$. All proofs are given in Section A.2. Intuitively, when $C\delta_b \geq \mathbb{E}[\mathbf{w}_{b,\tau}]$, the probability of an ES system being discharged is greater than or equal to its probability of being charged. As a result, $f_s(s)$ is decreasing in the open interval $(0, S)$ and it can be shown that $\tilde{f}_s(s)$ is greater than or equal to $f_s(s)$ when s is close to S. We

also know that the expected overflow is non-decreasing in storage level once it is given. It thus can be shown that the approximated unconditioned overflow $\tilde{\mathbb{E}}[\mathbf{o}_\tau]$ is an upper bound for the real value.

To further obtain a closed-form upper bound for $\mathbb{E}[\mathbf{o}_\tau]$, we again approximate $\mathbf{w}_{b,\tau}$ by uniform distribution with mean μ_b and interval length ϵ_b, matching the first and the second moments of the real distribution of $\mathbf{w}_{b,\tau}$. The closed-form expression of $\tilde{f}_s(s)$ is given by the following Proposition:

Proposition 2. *Assume* $\mathbf{w}_{b,\tau} \sim$ *unif*$(\mu_b - \frac{\epsilon_b}{2}, \mu_b + \frac{\epsilon_b}{2})$. *Let* $A = \frac{(C\delta_b - \mu_b + \frac{\epsilon_b}{2})^2}{2\beta(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)}$ *and* $B = \frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \frac{\epsilon_b}{2})}$. *Then*

(i) For $s \in (0, S)$, $\tilde{f}_s(s) \equiv f_s^c = \frac{1}{A+S+B}$;

(ii) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = 0) = \frac{A}{A+S+B}$;

(iii) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) = \frac{B}{A+S+B}$.

Proposition 2 provides an analytical probability model for the distribution of storage level $\mathbf{s}_\tau$. Though it tends to overestimate the probability that $\mathbf{s}_\tau$ is close or equal to S, the model does capture the dependence of the storage level distribution on system parameters such as conversion efficiency and capacities of transmission and ES. In particular, Proposition 2 results in a simple analytical upper bound for the expected energy overflow, as stated in the following proposition:

Proposition 3. *Assume* $\mathbf{w}_{b,\tau} \sim$ *unif*$(\mu_b - \frac{\epsilon_b}{2}, \mu_b + \frac{\epsilon_b}{2})$. *Then the expected energy overflow of each interval of length* δ_b *is bounded from above by* $\frac{5\alpha}{24S}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$. *The derivatives of this upper bound are* $-\frac{5\alpha}{24S^2}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$ *with respect to* S *and* $-\frac{5\alpha\delta_b}{12S}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$ *with respect to* C.

An insight from Proposition 3 is that the upper bound of the expected overflow is more sensitive to C than to S. The marginal benefit in terms of overflow prevention shrinks quadratically in S. Therefore, transmission lines should be the dominant means over ES to reduce overflow. With the aid of Proposition 3, we next estimate ES capacity for a single and for multiple wind farms.

A Single Wind Farm

Our goal is to find the economic trade-offs between transmission and ES capacities to minimize the total variable cost for a single wind farm. In particular, the variable ES cost consists of energy overflow loss and capacity cost. Following Proposition 3, an estimate of the annual variable ES cost, denoted by v_4, is as follows:

$$v_4(C, S) \quad = \frac{5p\alpha}{24\delta\delta_b S}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2 + rS. \tag{1.15}$$

Table 1.5: Economic ES-transmission capacities and total variable costs.

Condition	C^*	S^*	v_5^*
$\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b} < \mu+\epsilon(\frac{1}{2}-\theta)$	$\mu+\epsilon(\frac{1}{2}-\theta)$	0	$q\mu+\epsilon q(\frac{1}{2}-\frac{1}{2}\theta)$
$\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b} \in [\mu+\epsilon(\frac{1}{2}-\theta),$ $\mu+\epsilon(\frac{1}{2}-\hat{\theta}))$	$\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b}$	0	$q(\mu+\frac{\epsilon_b}{2\delta_b})$ $+p\frac{1-\alpha\beta}{8\epsilon\delta}(\epsilon-\frac{\epsilon_b}{\delta_b})^2$
$\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b} \geq \mu+\epsilon(\frac{1}{2}-\hat{\theta}) > \mu$	$\mu+\epsilon(\frac{1}{2}-\hat{\theta})$	$\sqrt{\frac{5p\alpha}{24r\delta\delta_b}}(\frac{\epsilon_b}{2}-\epsilon\delta_b(\frac{1}{2}-\hat{\theta}))$	$q\mu+\epsilon\hat{q}(\frac{1}{2}-\frac{1}{2}\hat{\theta})+\sqrt{\frac{5pr\alpha}{6\delta\delta_b}}\frac{\epsilon_b}{2}$
$\mu+\epsilon(\frac{1}{2}-\hat{\theta}) \leq \mu$ (i.e., $\hat{\theta} \geq \frac{1}{2}$)	μ	$\sqrt{\frac{5p\alpha}{6r\delta\delta_b}}(\frac{\epsilon_b}{4})$	$q\mu+\epsilon\frac{p(1-\alpha\beta)}{8\delta}+\sqrt{\frac{5pr\alpha}{6\delta\delta_b}}\frac{\epsilon_b}{2}$

Minimizing the two terms on the right-hand side of (1.15) yields the upper bounds for economic storage capacity and the associated ES variable cost, both as functions of transmission capacity:

$$S^*(C) = \sqrt{\frac{5p\alpha}{24r\delta\delta_b}}(\mu_b+\tfrac{\epsilon_b}{2}-C\delta_b), \tag{1.16}$$

$$v_4^*(C) = \sqrt{\frac{5pr\alpha}{6\delta\delta_b}}(\mu_b+\tfrac{\epsilon_b}{2}-C\delta_b). \tag{1.17}$$

Note that (1.16) and (1.17) are valid only if $C\delta_b \in (\mu_b, \mu_b+\frac{\epsilon_b}{2})$. When $C\delta_b \geq \mu_b+\frac{\epsilon_b}{2}$, the transmission line is always able to deliver at least all the energy that are produced within the current interval by the end of the interval; as a result, no overflow occurs and $v_4^*(C)=0$. Based on (1.3), (1.17) and the above discussion, the total variable cost is given by:

$$\begin{aligned} & v_5(C) = v_1(C)+v_4^*(C) \\ = & \begin{cases} p\frac{1-\alpha\beta}{2\epsilon\delta}(\mu+\frac{\epsilon}{2}-C)^2+qC+\sqrt{\frac{5pr\alpha}{6\delta\delta_b}}(\mu_b+\frac{\epsilon_b}{2}-C\delta_b), & \text{if } C\in[\frac{\mu_b}{\delta_b},\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b}); \\ p\frac{1-\alpha\beta}{2\epsilon\delta}(\mu+\frac{\epsilon}{2}-C)^2+qC, & \text{if } C\in[\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b},\mu+\frac{\epsilon}{2}]. \end{cases} \end{aligned} \tag{1.18}$$

In (1.18), $\frac{\mu_b}{\delta_b}=\mu$ and the economic value of C is within $[\mu,\mu+\frac{\epsilon}{2}]$. In addition, we assume $\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b} < \mu+\frac{\epsilon}{2}$, since $\mathbf{w}_{b,\tau}$ aggregates hourly wind output $\mathbf{w}_t$ and is thus less variable. Function $v_5(C)$ is convex in C since it is the point-wise maximum of two convex quadratic functions of C. Therefore, the analytical expression of the economic transmission capacity C^* that minimizes (1.18), as well as the associated ES capacity S^* (from (1.16)) and the total variable cost v_5^*, can be obtained as summarized in Table 1.5 (In the table, recall $\theta=\frac{\delta q}{p(1-\alpha\beta)}$, and let $\hat{q}=q-\sqrt{\frac{5p\alpha r\delta_b}{6\delta}}$ and $\hat{\theta}=\frac{\delta\hat{q}}{p(1-\alpha\beta)}$). These quantities can be viewed as generalization from (1.4) and (1.5), with ES capacity cost and overflow being considered.

In practice, the actual optimal cost may be greater than its theoretical upper bound v_5^* due to two sources of model inaccuracy - the uniform distribution approximation and the neglect of auto-correlation of the bulk wind output process. As for the latter, the resulting

Table 1.6: Average cost gaps between the upper and the lower bounds, between the upper bound and the optimal value, and between the upper bound and the cost in the ES-free scenario.

Distance (mile)	UB - LB	|UB - Opt.|	ES-Free - UB
200	23.6%	6.6%	89.7%
120	20.2%	5.17%	81.5%
50	3.09%	6.78%	31.8%

underestimation of inter-period overflow loss may overly compensate the overestimation of intra-period overflow loss and ES capacity. Hence, v_5^* is closer to the actual optimal cost than theoretically expected. In addition, Section A.3 presents detailed numerical studies on tightness and accuracy of the upper bound model. The studies verify that the cost gaps are bounded and v_5^* is not violated in most cases. Also, as an estimate to the actual optimal cost, v_5^* is more accurate than the cost lower bound developed in Section 1.4, as shown in Table 1.6. The table also shows that ES can potentially significantly bring down the cost of otherwise building an ES-free line.

Multiple Wind Farms

We next use the results in the preceding single-farm case to plan an ES-transmission network for multiple wind farms. However, these results can not be directly incorporated into the planning model. Unlike in the scenario of uncapacitated ES where we can pre-determine which segment of the non-smooth cost expression (1.5) to use in formulating planning model (1.13), the segmentation of v_5^* depends not only on wind-independent parameter θ, but also on wind characteristics $(\mu, \epsilon, \mu_d, \epsilon_d)$, which are not available for a candidate junction site before the assignment of the wind farms to the junction site is known. We hence resort to a heuristic outlined as follows:

1. For given sets of wind farms and candidate junction sites, solve planning model (1.13), which assumes uncapacitated ES. The model outputs an assignment of the wind farms to the junction sites as well as the associated transmission capacities.

2. For each junction site j that is selected to build ES on, compute the expected variable cost $v_{5,j}^*$ using Table 1.5. If $v_{5,j}^* + h_j < q_j\mu_j + q_j(\frac{1}{2} - \frac{1}{2}\eta_j)\epsilon_j$, then building ES on site j is still economical, and we choose capacities of ES and transmission line from site j to the load center according to Table 1.5. Otherwise, we opt for an ES-free line with capacity given by (1.9).

Then we summarize the models developed in this chapter in the following proposition:

Proposition 4. *The optimal cost of the ES-transmission network is bounded from below by the optimal objective value of planning model (1.13), and bounded from above by the total cost given by the above outlined heuristic.*

Table 1.7: Costs and computational times of the ES-transmission networks.

Wind Farms	Candidate Sites	LB Cost ($)	UB Cost ($)	Cost Gap	Time (sec)
6	12	1.9556×10^8	2.1136×10^8	8.08%	0.6080
12	18	3.9703×10^8	4.2268×10^8	6.46%	1.0264
18	28	6.2031×10^8	6.7153×10^8	8.26%	2.0009
24	28	8.6438×10^9	9.4643×10^9	9.49%	2.5876

The Value of ES

All the above analysis suggests an important insight to the planners: **even small ES saves big, but the marginal value of ES diminishes fast**. Compared with the ES-free scenario, the combination of an economic line and an ES system has dual effects. The positive effect is that part of the investment in transmission capacity can be salvaged by the ES system, which accommodates short-period fluctuations of the wind output; the negative effect is the incurred energy loss due to friction and overflow. However, these dual effects respond to the size of the ES system differently. A relatively small capacity of ES should be adequate to achieve the positive effect. In contrast, additional ES capacity has much diminished value because it does not help reduce friction loss and is mainly used to hedge overflow. Moreover, Proposition 3 has suggested that the marginal benefit in terms of overflow prevention shrinks quadratically in S. As a result, building even more ES capacity would be less cost-efficient than increasing transmission capacity.

1.6 Computational Results and Insights

A Case Study

This section presents numerical studies to demonstrate the effectiveness of the proposed planning procedure and insights concerning technology impact and layout robustness.

Our first set of experiments generate ES-transmission network designs for potential wind farms near Billings, Montana. We solve planning model (1.13) to obtain network layouts as well as lower-bound total costs. Then the heuristic in Section 1.5 is applied to output upper-bound total costs as well as the associated transmission and ES capacities. The detailed settings of the experiments are described in Section A.4. The experiments are repeated with different numbers of potential wind farms to be covered. Table 1.7 shows the costs and computational times of solving model (1.13). A network layout for the 24 wind farms is depicted in Figure 1.3 (a). The correlation between wind outputs from those farms is shown in Figure 1.3 (b).

Table 1.7 shows that the gap between the lower- and the upper-bounds of the total expected costs is smaller than that implied in Table 1.6. This is mainly because the additional fixed ES and transmission costs dilute the share of the cost that ES capacity and expected overflow account for. Also notice that, in most cases, the cost gap increases as we expand

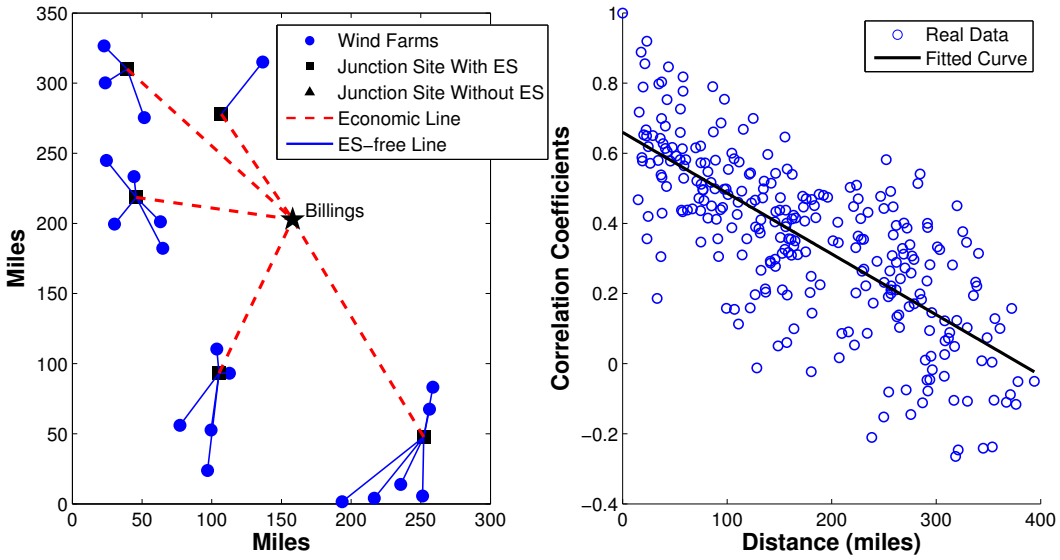

Figure 1.3: (a) Deployment of transmission lines and ES systems for 24 wind farms. (b) Correlation coefficients between wind outputs of the 24 wind farms.

the network by adding distant wind farms. Again, this is mainly because the average transmission capacity decreases in transmission distance; as a result, more ES capacity has to be added, which is not captured in the lower-bound cost. Finally, the computational times in all the instances are in magnitude of seconds. We thus conclude that, under the assumed settings of the ES-transmission network, our model is computationally efficient to solve and is able to output near-optimal ES-transmission deployment.

Technological Considerations

How do the forms of ES technologies and their advancements impact cost savings in power infrastructure planning? Various ES technologies will continue to be competing in the foreseeable future ([83]). These technologies distinguish themselves from each other by storing potential energy in different ways with different cost-efficiency parameters. For example, pumped hydro storage systems have high upfront cost, low per-unit capital cost and relatively high conversion efficiency ($75\% - 78\%$), lead-acid batteries have high capital cost with high conversion efficiency, while compressed air energy storage systems feature very low capital cost with low conversion efficiency ([72]). As characterized in our models, these technological aspects impact the need for ES via different mechanisms. Furthermore, as these technologies advance and diffuse, the ES capacity cost and conversion efficiency remain uncertain to a large extent.

Our second set of experiments evaluate how the cost-savings respond to different values of

Table 1.8: The number of junction sites and the number of ES systems (in parentheses), as a parameter varies from 10% to 500% of its baseline value or the conversion efficiency varies from 0.40 to 0.99 (the "N/A" entries correspond to the cases where the infrastructure is not economically feasible).

Perturbed Parameters	10%	50%	200%	500%
FIT Rate (p)	N/A	5(5)	5(5)	5(4)
Fixed ES Capacity Cost (h)	5(5)	5(5)	5(5)	5(4)
Variable Transmission Capacity Cost (a)	5(0)	5(4)	5(5)	N/A
Fixed Transmission Capacity Cost (g)	8(8)	5(5)	5(5)	5(5)
Mean Distance to the Load Center	5(0)	5(4)	5(5)	N/A
Perturbed Parameter	0.40	0.60	0.80	0.99
ES Round-trip Conversion Efficiency ($\alpha\beta$)	5(4)	5(5)	5(5)	5(5)

these parameters. Our major finding is that the R&D priority should be given to addressing the bottleneck cost factor in order to maximize cost-savings. For example, more investment is desirable in improving conversion efficiency in the case of compressed air energy storage systems, and in reducing the per-unit ES capacity cost for lead-acid batteries. More detailed analysis is omitted here for brevity and is available in Section A.4.

Layout Robustness

As the last issue, is our planning model robust in the cost-minimizing sense? An ES-transmission network has to be planned in the presence of uncertainties, such as FIT rates adjustments, technology advancements, building material cost fluctuations, climate changes, etc. After infrastructure is built, it is still possible to locally adjust ES and transmission capacities. However, it will be too cost-prohibitive to change the network layout as those uncertainties evolve over time. Therefore, a layout that is cost-efficient for a wide range of system parameter values is desirable.

The optimal layout may change at two levels. The topology-level reconfiguration affects the assignment of wind farms to junction sites. At the junction level is the switch between installing and not installing an ES system at a junction site, and subsequently between building an economic and an ES-free transmission line downstream of it.

We solve multiple instances of planning model (1.13) for cost-efficient layouts based on the settings of the case study, perturbing parameters one at a time. Table 1.8 summarizes how the number of junction sites and the number of deployed ES systems change accordingly. These two numbers represent the layout changes at the topology- and junction- levels, respectively. The baseline layout has five ES-coupled junction sites, as depicted in Figure 1.3 (a).

Table 1.8 shows that the baseline layout is remarkably robust. Set from 50% to 200% of the baseline value, almost no single parameter alone can affect the layout, except that one ES system is removed when either the variable transmission capital cost or the mean

distance is halved. Further parameter deviations result in more significant layout changes, but are very unlikely to occur.

The above results suggest that the cost-efficient layout of the ES-transmission network is robust against gross misestimation of model parameters. Two reasons may explain this robustness. First, the whole system intricately depends on multiple parameters. Impact of misestimation of any one of them is substantially mitigated by the others. Even if more than one cost factors deviate, their influences upon layout is more likely to counteract than reinforce each other, with the fixed and variable transmission capacity costs being an example. Second, most of the two-echelon facility location models in literature (for example, [84] and [17]) assume that the upstream facilities to be deployed are disjoint. In contrast, in our problem, all the selected junction sites are connected to a given load center. As a result, the favored junction sites not only cluster nearby wind farms, but also tend to be close to the given load center. This centrality of the junction sites substantially enhances the layout robustness.

1.7 Conclusion

In this chapter, we study the problem of planning economic energy storage systems and transmission lines for distributed wind resources. Under the FIT policy instrument, the ES operating policy is to store surplus energy that exceeds the rated transmission capacity, and release it later when the transmission line becomes available for additional loads. While saving transmission capacity cost, operating ES systems incurs both energy friction loss and overflow loss over time, in addition to its building cost. We develop models to characterize these trade-offs and determine the sites and the sizes of ES systems as well as the associated topology and capacity of the transmission network. In our first model, under the assumption that ES systems have sufficient energy capacity, we derive optimal transmission line capacities for a single wind farm with and without ES being colocated, respectively. Then we incorporate these quantities to formulate a location model as an MISOCP to determine the ES-transmission network. Our second model addresses the ES sizing problem. We derive an upper bound of the expected energy overflow, from which an overestimated economic storage size as well as the associated transmission capacity is obtained. These two models provides the lower and the upper bounds of the total expected cost. The case study, as well as the analysis of model inaccuracy in Appendix A, demonstrates that our deployment scheme is near-optimal.

Our model and analysis lead to several findings. 1) First, while utilization of ES systems saves significant amount of transmission cost, the marginal value of adding ES capacity diminishes faster than the marginal value of adding transmission capacity. This is because a relatively small amount of ES capacity is sufficient for smoothing short-term fluctuations of wind outputs; but adding more ES capacity does not help reduce energy friction loss and is inefficient in reducing energy overflow loss compared with adding transmission capacity. Therefore, it is cost-efficient to install small-sized ES systems on electricity junction sites. 2)

Advancements in ES technologies can further bring down the total cost of an ES-transmission network by improving round-trip energy conversion efficiency and reducing upfront cost and per-unit capacity cost of ES. Depending on the form of the ES technology, the priority of R&D resources should be given to addressing the bottleneck cost factor. For example, the bottleneck cost factor is the round-trip conversion efficiency for compressed air energy storage systems, and is the per-unit capacity cost for lead-acid batteries. 3) The layout of the ES-transmission network is robust against FIT rates adjustments, technology advancements, building material cost fluctuations, climate changes, etc. This is partly because the system relies on multiple parameters and partly because of the centrality of the favored junction sites towards the given load center in our problem setting.

This work is the first attempt to provide tractable methods to determine both the layout and capacity of an ES-transmission network to tap distributed wind resources. The work adopts some core ideas from the field of supply chain design. We believe that more contribution to planning renewable energy systems can be made by the supply chain research community. Two possible extensions to this work can be beneficial. Firstly, despite the prevalence of the FIT policy instrument in the initial phase, wind energy is expected to enter the energy biding market when it accounts for a considerable share of the total energy production in the future. Under such scenario, the process of wind energy prices are stochastic, and more ES capacity is desired to store energy and shift its delivery to high-price hours. The question then arises: how to make adaptive budgeting decisions on investment in transmission lines and ES systems, given the evolving wind energy market (i.e., from FIT-based to bid-based)? Secondly, although the radial topology with a single load center is a realistic setting for the initial stage of wind energy development and captures most of the trade-offs in the scenario of covering multiple wind farms, it will be more interesting (and much more challenging) to develop models that address the complexity where the grid is meshed and multiple load centers are present.

Chapter 2

Hierarchical Coordination of Charging Plug-in Electric Vehicles

2.1 Introduction

Plug-in electric vehicles (PEVs) are currently receiving worldwide attention. With widely released incentive policies and technological maturity, the adoption of PEVs will potentially reach a remarkable scale in the foreseeable future [87]. However, PEVs' charging load can be a substantial burden to the operations of existing power grid when they highly penetrate and when their charging is not properly managed. In the distribution system, voltage-drops [23], excessive increase in power losses [37] and overloading of distribution transformers [28][89] are likely to occur. As a result, coordinated PEV charging, as an effective and cost efficient way to mitigate the charging stress on power systems, has attracted wide interest.

In the literature, coordinated PEV charging strategies can be categorized into two groups, namely the centralized and the decentralized approaches. In the centralized approach, a group of PEVs are controlled by an aggregator who dictates the optimal charging decisions for each vehicle at each time slot. For example, [23] proposes two mathematical models based on quadratic and dynamic programming to coordinate the charging of multiple electric vehicles in distribution network to minimize the power losses. [88] explores the relationships between feeder losses, load factor and load variance and proposes three optimal charging control algorithms to minimize the impacts of PEVs charging on the distribution system. [91] proposes a centralized coordinated PEVs charging control framework that considers the interactions between the charging service provider, the retailer and the distribution system operator. [49] formulates a centralized charging coordination problem for a relatively small vehicle population (20-50 PEVs), and proposes a dual-ascent solution method. In [40] and [39], the authors consider a centralized supervisory scheduler to coordinate PEV charging in distributed charging stations. They use control price signals derived via utility and game theory approaches to regulate the charging demands. However, it can be difficult in reality to calibrate customers charging utility or to obtain truthful control price signals. The

computational burden may also become non-trivial if the scheduler faces a huge volume of demands. For decentralized strategies, in [80], a dynamic program is formulated to determine the optimal charging schedule for individual PEVs facing day-ahead prices. [98] develops a decentralized charging coordination framework using ADMM algorithm. [41] proposes a decentralized control algorithm to achieve the valley-filling charging profile. In each computation iteration, each PEV locally optimizes its own charging profile based on the price signals broadcasted by the utility. The global optimality is also obtained when the algorithm converges. Additionally, [63] and [69] both propose game-based, decentralized control strategy for coordination of electric vehicles charging.

We differ from these papers in that our focus is on development of centralized and distributed PEV charging coordination strategies specifically for large-scale, multiple parking decks in multi-family dwellings. While most of the papers that propose a decentralized charging scheme decompose the centralized problem with respect to each PEV, our work emphasizes coordination of distributed charging decks from the service aggregators' perspective. Occupants of large cities living in multi-family dwellings are likely to become the early adopters of PEVs. Compared with people residing in other types of areas, they are generally more aware of environmental issues that fossil fuel consumption brings about and thus more motivated to embrace this green technology. In addition, being indispensable supporting facilities for multi-family dwellings, parking decks naturally aggregate the electric vehicles and serve as an ideal space to enable coordinated charging. Moreover, in cities where time-based electricity rates are applied, charging service providers have incentive to reap profit from shifting charging loads to off-peak periods by means of charging coordination. Related literature in this subject is relatively scarce. For instance, [90] applies an intelligent EDA algorithm to manage charging of large number of PEVs parked at a municipal parking station. Through early work in [101], it is found that in absence of coordination across multiple distributed parking decks under a single distribution feeder, coordinated operation within only individual parking decks, in response to electricity price variability and mere local constraints, may still not be adequate. In some cases, the occurrence of unexpected overlapped charging hours in multiple parking decks could even cause more severe damage to the power grid than uncoordinated charging.

Based on the early work [101], this chapter presents a hierarchical optimal control framework to achieve charging coordination between multiple PEV parking decks. We first formulate a centralized finite-horizon control problem to maximize the utility of charging service provider. Then we develop distributed control strategy based on Lagrangian relaxation (LR) techniques by decoupling the proposed centralized model with respect to each parking deck. The major advantage of this approach is that only a set of sub-problems of much smaller scales need to be solved in parallel by the local parking deck operators. And a major contribution of this work is our design of an effective heuristic algorithm to search for feasible solutions that generate tight lower bounds of the optimal objective value. Simulation results show that valley-filling can be efficiently achieved 1) under the distributed charging coordination framework that we propose and 2) with a simple flat charging price, a pricing scheme that we believe is more applicable to local residential communities than the variable

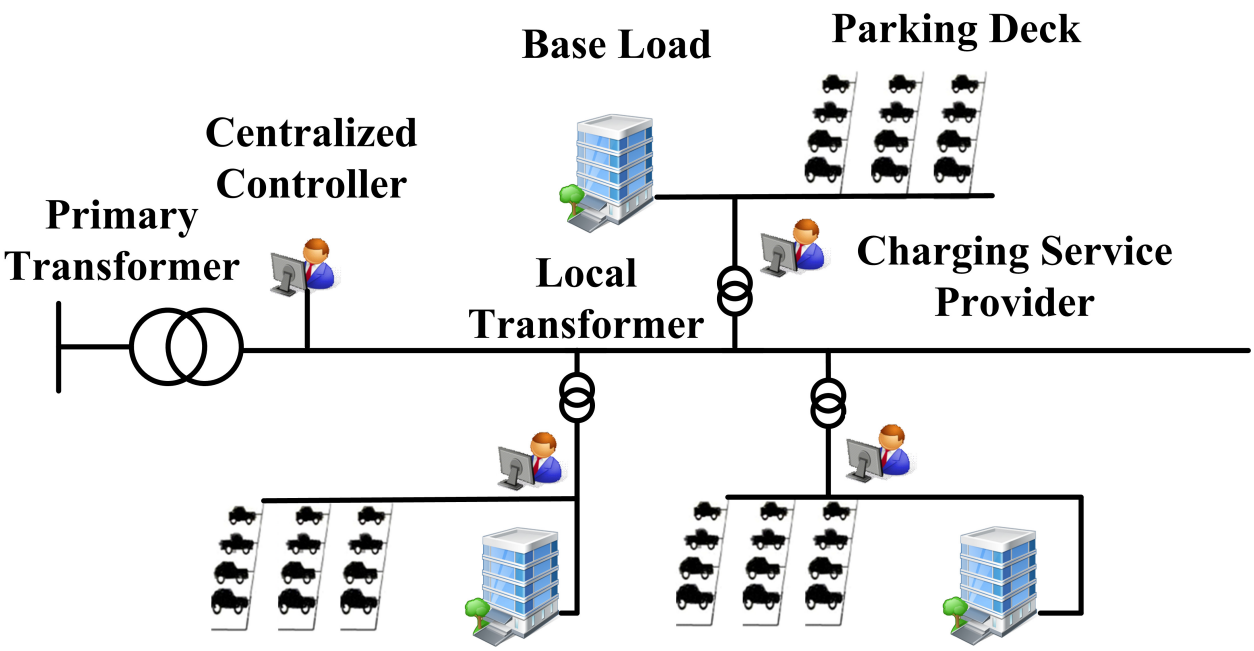

Figure 2.1: Schematic illustration of a distributed charging system in an urban area.

price scheme.

The remainder of this chapter is organized as follows. System description is presented in Section 2.2. The problem formulation and solution algorithm are described in Section 2.3. We carry out simulations for distributed charging systems in Section 2.4, and finally, conclude this chapter in Section 2.5.

2.2 Description of Distributed Charging Systems

A typical topology of a distributed charging system in an urban area is illustrated in Fig.2.1. Power voltage from the transmission side is stepped down first by the primary distribution transformer and then by the local transformers before the power is used in each of the M local residential communities. Each community i ($i = 1, 2, ..., M$) consists of one or multiple dwellings (typically multi-storied residence halls) and a PEV charging deck with $|N_i|$ charging ports (N_i denotes the set of charging ports of parking deck i). The charging load and the base load (i.e., loads excluding the charging load in a community) are directly connected to the local transformers. Both the primary transformer and the local transformers are capacitated.

The M distributed charging decks can be owned by either a single or multiple charging service providers. Each charging deck purchases electricity from the utility at time-of-use (TOU) rates (which vary in different hours during a day) and then sells it to the customers at retail price.

We assume that the PEV arrival and departure processes and the initial SOC of each PEV are all stochastic but with known probability distributions. In fact, this type of information can be adaptively learned from the historical charging behaviors of customers. Once a PEV connects to the n_ith charging port ($n_i = 1, 2, ..., |N_i|$) at charging deck i, its battery capacity B_{n_i} and initial/pre-charge state of charge (SOC), $SOC^A_{n_i}$, are obtained through the battery management system (BMS) on board. In addition, the customer is required to inform the charging system of the expected parking duration, T_{n_i}, and the desired SOC after charge, $SOC^D_{n_i}$. We assume that the TOU electricity rates are exogenously given. When PEVs become prevalent, it's likely that their large-scale aggregate charging demand will influence the TOU prices in a way that the prices during the peak charging hours (e.g. from midnight to 6:00 am) will increase. However, despite the highly stochastic charging behavior of each individual PEV, the patterns of aggregate charging demand and thus the TOU rates in a certain region should be relatively stable and publicly known over a long period. The problem setting in our work is also applicable to the scenario where exogenous day-head market electricity prices are charged by service providers. Lastly, based on the historical load data, the day-ahead forecast of both the aggregate and the local base loads are available to the charging system. Based on above information, the coordinated charging system schedules PEV charging at each port by repeatedly solving an optimal charging control problem, which we elaborate on in the next section.

2.3 Coordinated Charging Strategies

Control System Overview

In this section, we design coordinated PEV charging control system for multiple charging decks under a primary distribution transformer. We propose to design the charging control system to be model-based and forward-looking in making control decisions, and that is able to explicitly take into account the optimality considerations and handle state and input constraints [76]. Specifically, the optimal control problem is solved at discrete time instants $t_k = t_0 + k\Delta_t, k = 0, 1, ...$, with t_0 being the initial time and Δ_t being the interval length. The charging power $P^*_{n_i}$ at each port n_i is kept constant within each interval and is regarded as a semi-continuous decision variable which can either be zero or take value between a threshold value and the maximum charging power P_{max} due to the limits of the regulation capability of power electronics in charging port. Only the first segment of each power trajectory (i.e., the charging power for interval $(t_k, t_{k+1}]$) is sent to the corresponding port for implementation. At the end of interval $(t_k, t_{k+1}]$, the charging control system receives updated information of each port (e.g., being vacant/occupied and the SOC of the accommodated PEV) as well as the forecasts of upcoming charge demand process and base load over the planning horizon. Then it repeats the above procedure to make and implement the charging decisions for the next interval $(t_{k+1}, t_{k+2}]$ and so on. The structure of the control system is depicted in Fig.2.2.

Note that the forecast of the upcoming demand process consists of sample paths of the

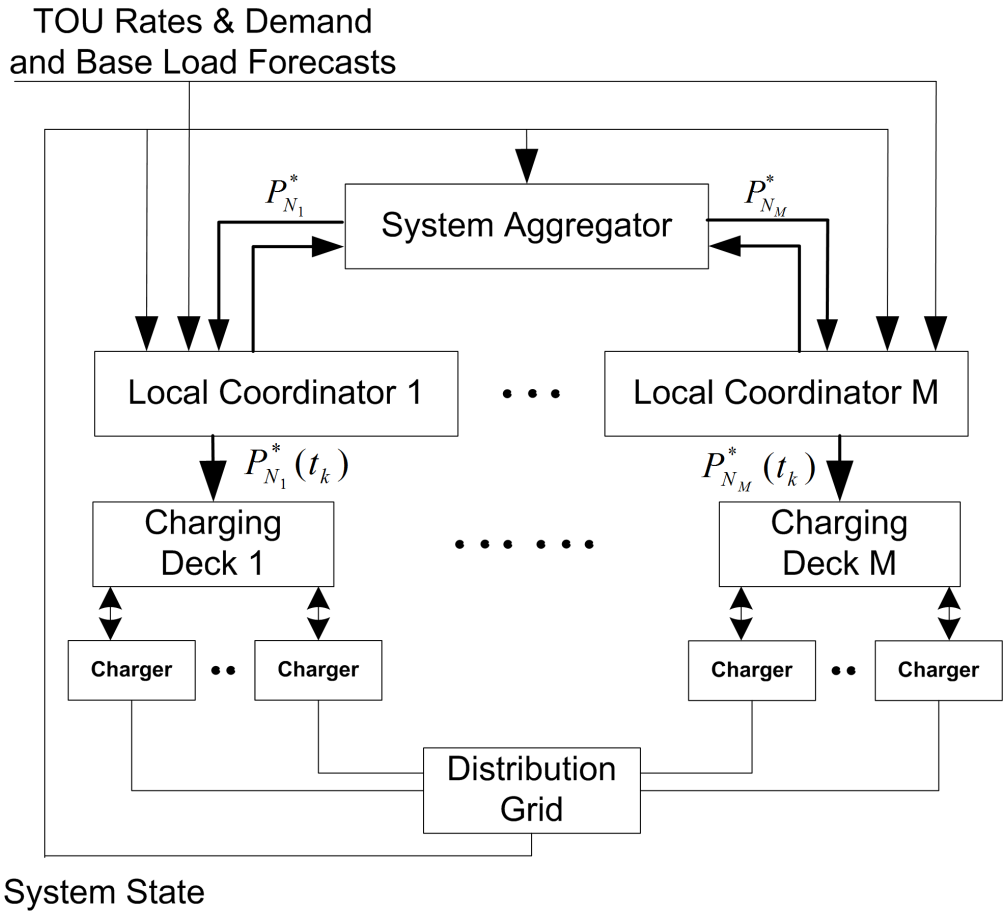

Figure 2.2: Structure of the coordinated charging control system.

upcoming PEV arrival/departure processes and of the initial/desired SOC associated with each upcoming PEV. The charging control system generates these sample paths based on the probability distributions of these stochastic processes and random variables. The realized demand process for each charging port does not necessarily resemble the forecast precisely; however, the pooled charging power can thus be reasonably estimated when each charging deck accommodates sufficient number of PEVs.

With the aid of the forecast sample paths, the controller is able to estimate the aggregated charging load across the ports over the planning horizon and schedule charging for each port accordingly.

Centralized Coordinated Charging

We first formulate the centralized coordinated control problem. In the centralized scenario, a single control system, or the aggregator make charging schedules for each port of the M distributed charging decks by solving a centralized optimization problem at each interval.

For coordination purposes, we simply regard the charging power during each 15-min control interval as constant. The electro-thermal dynamic response of distribution transformer during the charging power regulation process is not discussed. Comparing to the length of a control interval, which is normally selected as 15 minutes, the duration of transient processes of distribution transformer is insignificant. Mathematically, we propose the following objective function to be maximized at t_k.

$$\begin{aligned} & J(t_k) = \\ & \sum_{i\in I}\sum_{n_i\in N_i}\sum_{j=0}^{H-1}\gamma^j C_{n_i}(t_{k+j})\cdot P_{max}\Delta_t\cdot[p(t_{k+j})-c(t_{k+j})] \\ & -\sum_{i\in I}\sum_{n_i\in \bar{N}_i\bigcup\hat{N}_i} G(\alpha_{n_i}|SOC^A_{n_i}, SOC^D_{n_i}, B_{n_i}) \\ & +\sum_{i\in I}\sum_{n_i\in N_i}\sum_{j=0}^{H-1}\kappa\cdot(H-j)\cdot C_{n_i}(t_{k+j}), \end{aligned} \tag{2.1}$$

where:
I is the set of distributed electric vehicle parking decks.
N_i is the set of the charging ports in parking deck $i \in I$.
$\bar{N}_i$ is the set of the charging ports which become occupied by EV arrivals during the current time interval in charging deck $i \in I$.
$\hat{N}_i$ is the set of charging ports which are predicted to be occupied by forecast EV arrivals within the planning horizon in charging deck $i \in I$.
H is the planning horizon and is also the maximum number of intervals within which a PEV should be charged to the desired SOC.
γ is a time discount factor, $0 < \gamma \leq 1$.
P_{max} is the maximum charging power of each charging port.
$C_{n_i}(t_{k+j})$, the decision variable, is the fraction of P_{max} that the control system charges the PEV at port n_i at $(k+j)$th interval. The charging power $P^*_{n_i} = C_{n_i}P_{max}$.
$p(t_{k+j})$ and $c(t_{k+j})$ are the retail and the purchase (TOU) electricity prices at $(k+j)$th interval, repectively.
α_{n_i} is the fraction of battery capacity that will not be fulfilled for the PEV which has just arrived during the current interval or is predicted to arrive within the planning horizon at port n_i. This variable is used to modify the service level in case of excessive charging demand (e.g. high energy demand in extremely short duration) to ensure the feasibility of the optimization problem. The resulting choice of α_{n_i} can also help the system operator determine the maximum possible departure SOC that can be satisfied.
$G(\cdot)$ is the dollar valued penalty of not fully satisfying customer demand. Given information of $SOC^A_{n_i}$, $SOC^D_{n_i}$ and B_{n_i} for PEV at port n_i, $G(\cdot)$ is an increasing function of α_{n_i}.
κ is a positive weight factor associated with the earliness consideration which is explained in the next paragraph.

The first term of the objective function is to maximize the projected total profit over the planning horizon. The second term penalizes unfulfilled charging demands and is regarded as the dollar valued loss of customer goodwill. The last term implies our preference to early charging, as the weighting factors associated with the charging powers decrease in time. The charging system thus tends to charge PEVs early so as to prepare for unexpected early departure of PEVs or unexpected mass arrivals. Following this objective function, we formulate the following centralized coordinated charging optimization problem to be solved at t_k.

$$\max_{C_{n_i},u_{n_i},\alpha_{n_i}} \quad J(t_k)$$

$$\text{s.t.} \sum_{i\in I}\sum_{n_i\in N_i} C_{n_i}(t)P_{max} \le A_T\xi_T(t)\lambda,$$
$$\forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}, \tag{2.2a}$$
$$\sum_{n_i\in N_i} C_{n_i}(t)P_{max} \le A_i\xi_i(t)\lambda,$$
$$\forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}, \forall i \in I, \tag{2.2b}$$
$$\sum_{t=t_k}^{t_k+T_{n_i}} \rho C_{n_i}(t)\cdot P_{max}\Delta_t + SOC_{n_i}^A B_{n_i} \ge (SOC_{n_i}^D - \alpha_{n_i})B_{n_i},$$
$$\forall n_i \in N_i, \forall i \in I, \tag{2.2c}$$
$$\sum_{t=t_k}^{t_k+T_{n_i}} \rho C_{n_i}(t)\cdot P_{max}\Delta_t + SOC_{n_i}^A B_{n_i} \le B_{n_i}$$
$$\forall n_i \in N_i, \forall i \in I, \tag{2.2d}$$
$$C_{n_i}(t) \le S_{n_i}(t),$$
$$\forall t \in \{t_k, t_{k+1}, ..., t_k + T_{n_i}\}, \forall n_i \in N_i, \forall i \in I, \tag{2.2e}$$
$$\theta_{th}u_{n_i}(t) \le C_{n_i}(t) \le u_{n_i}(t),$$
$$\forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}, \forall n_i \in N_i, \forall i \in I, \tag{2.2f}$$
$$0 \le \alpha_{n_i} \le SOC_{n_i}^D - SOC_{n_i}^A \quad \forall n_i \in \bar{N}_i \bigcup \hat{N}_i, \forall i \in I, \tag{2.2g}$$
$$\alpha_{n_i} = 0, \quad \forall n_i \in N_i \backslash (\bar{N}_i \bigcup \hat{N}_i), \forall i \in I, \tag{2.2h}$$
$$u_{n_i}(t) \in \{0, 1\}, \quad \forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\},$$
$$\forall n_i \in N_i, \forall i \in I, \tag{2.2i}$$

where:
A_T is capacity of the primary distribution transformer which serves the area covering all the charging decks in set I.

A_i is capacity of the local distribution transformer which serves the local community with charging deck i.
$\xi_T(t)$ is the proportion of the available capacity of the primary distribution transformer that can be used for PEV charging at time t (The remaining capacity are used to satisfy base load).
$\xi_i(t)$ is the proportion of the available capacity of the local distribution transformer that can be used for EV charging at deck i at time t.
λ is the average power factor of charging load.
ρ is the charging efficiency.
T_{n_i} is the expected parking duration of the PEV at port n_i. We assume that the charging should be finished within the planning horizon, that is, $T_{n_i} \leq H\Delta_t, \forall n_i \in N_i, \forall i \in I$.
$S_{n_i}(t)$ is the state of charging port n_i at time t. 1 if it is occupied by a PEV, 0 otherwise. At the beginning of each interval, PEVs that have just arrived during the last interval are directed to the unoccupied charging ports.
$u_{n_i}(t)$ is a binary decision variable, 1 if charging at port n_i is activated at time t, 0 otherwise.
θ_{th} is the minimum ratio of the actual charging power to the rated charging power P_{max} because of the limits in power electronics.

In the above optimization problem, constraints (2.2a) impose the limit on capacity availability of the primary distribution transformer. At any interval within the planning horizon, the sum of the total base load and the total charging load of all the charging decks in set I should not exceed the rated capacity of the primary distribution transformer. Constraints (2.2b) are the local versions of constraints (2.2a), requiring that the sum of the local base load and the load of the corresponding charging deck should not exceed the rated capacity of the local distribution transformer. In practice, considering the load diversity, the rated capacity of the primary transformer is normally lower than the sum of all the rated capacities of local transformers. Constraints (2.2c) and (2.2d) make sure that each PEV will be charged to the promised state before the specified departure time. Constraints (2.2e) ensure that port n_i is on the charging mode at time t only if it is occupied by an PEV. Constraints (2.2f) ensure that each port n_i is either on the standby mode or the charging mode where the charging power cannot be lower than θ_{th} of the maximum charging power P_{max}. Constraints (2.2g) specify the range of unfulfillment for each EV that has just arrived during the current interval or is predicted to arrive within the planing horizon. Constraints (2.2h) enforce that the promised charging service level for the PEVs which arrive before the current time interval cannot be compromised. Constraints (2.2i) ensure that the decision on the mode of each port at each time is binary.

Optimization problem (2.2) is a mixed integer linear program when $G(\cdot)$ in the second term of the objective function (2.1) is linear in α_{n_i} as follows:

$$G(\alpha_{n_i}|SOC^A_{n_i}, SOC^D_{n_i}, B_{n_i}) = \mu\alpha_{n_i}, \forall n_i \in N_i, \forall i \in I, \tag{2.3}$$

where μ is a positive weight factor. The problem is in general NP-hard and the computational burden increases exponentially in the length of the planning horizon and in the number of

PEVs connected to the charging system.

Distributed Coordinated Charging

An alternative to the centralized coordinated charging strategy is to develop a distributed charging coordination system in which local controllers schedule for their respective local charging decks in parallel and iterate to achieve the common objective. The solution approach that we design is a distributed sub-gradient method (See [71] for more theoretical discussion). Specifically, following the Lagrangian Relaxation method [86], we relax the primary capacity constraints (2.2a) and add them as penalty terms to the objective function (2.1). The objective and the optimization problem thus become:

$$\begin{aligned} &J_{LR}(t_k|\vec{\eta}) = J(t_k) \\ &+ \sum_{j=0}^{H-1} \eta_{t_{k+j}} (\frac{A_T \xi_T(t_{k+j})\lambda}{P_{max}} - \sum_{i\in I} \sum_{n_i \in N_i} C_{n_i}(t_{k+j})), \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} &\max_{C_{n_i}, u_{n_i}, \alpha_{n_i}} J_{LR}(t_k|\vec{\eta}) \\ &\text{s.t. } (2.2b) - (2.2i), \end{aligned} \tag{2.5}$$

where $\vec{\eta} = (\eta_{t_k}, \eta_{t_{k+1}}, ..., \eta_{t_{k+H-1}})$ is the vector of nonnegative Lagrangian multipliers. Then given $\vec{\eta}$, we decompose optimization problem (2.2) into M subproblems with respect to each charging deck. For charging deck i, the local coordinated charging optimization problem at t_k becomes:

$$\begin{aligned} &\max_{C_{n_i}, u_{n_i}, \alpha_{n_i}} J_i(t_k|\vec{\eta}) \\ &\text{s.t. } (2.2b) - (2.2i), \text{ for the fixed } i, \end{aligned} \tag{2.6}$$

where we define

$$\begin{aligned} &J_i(t_k|\vec{\eta}) \doteq \\ &\sum_{n_i \in N_i} \sum_{j=0}^{H-1} C_{n_i}(t_{k+j}) \{\gamma^j P_{max} \Delta_t [p(t_{k+j}) - c(t_{k+j})] \\ &+ \kappa \cdot (H-j) - \eta_{t_{k+j}}\} \\ &- \sum_{n_i \in \bar{N}_i \bigcup \hat{N}_i} G(\alpha_{n_i}|SOC^A_{n_i}, SOC^D_{n_i}, B_{n_i}). \end{aligned} \tag{2.7}$$

It can be easily shown that

$$J_{LR}(t_k|\vec{\eta}) = \sum_{i\in I} J_i(t_k|\vec{\eta}) + \sum_{j=0}^{H-1} \eta_{t_{k+j}} \frac{A_T \xi_T(t_{k+j})\lambda}{P_{max}}.$$

To find a close-to-optimal solution, we iteratively solve the relaxed optimization problems (2.6) with the Lagrangian multipliers updated by the following subgradient method. At iteration r, the optimal solution to the relaxed optimization problems (2.6) gives an upper bound $J^*_{LR}(t_k|\vec{\eta^r})$ for the optimal objective value $J^*(t_k)$ of the centralized optimization problem (2.2). If the optimal solution to (2.6) is infeasible to (2.2) (i.e., if it violates at least one of the primary transformer capacity constraints (2.2a)), we propose a heuristic algorithm to find a feasible solution to (2.2), thus giving a lower bound LB^r for $J^*(t_k)$. In order to update the Lagrangian multipliers to reduce the gap between $J^*_{LR}(t_k|\vec{\eta^r})$ and LB^r, we first compute the step size as follows:

$$\Delta^r = \frac{\beta^r(J^*_{LR}(t_k|\vec{\eta^r}) - LB)}{\sum_{t=t_k}^{t_k+\max\{T_{n_i}\}}(\frac{A_T\xi_T(t)\lambda}{P_{max}} - \sum_{i\in I}\sum_{n_i\in N_i} C_{n_i}(t))^2}, \tag{2.8}$$

where $\beta^r > 0$ is a step size coefficient. Formula (2.8) satisfies $\Delta^r \to 0$ as $r \to \infty$ and $\sum_{r=0}^{\infty}\Delta^r = \infty$, which guarantees asymptotic convergence of $J^*_{LR}(t_k|\overrightarrow{\eta}^r)$ to the optimal objective value $J^*(t_k)$ [47].

Then the multipliers for the next iteration $r+1$ are set as:

$$\begin{gathered}\eta_t^{r+1} = \max\{0, \eta_t^r - \Delta^r(\frac{A_T\xi_T(t)\lambda}{P_{max}} - \sum_{i\in I}\sum_{n_i\in N_i} C_{n_i}(t))\}, \\ \forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}.\end{gathered} \tag{2.9}$$

We propose the following heuristic algorithm to find a feasible solution and consequently a lower bound LB^r for $J^*(t_k)$ at iteration r.

1. For each instant $t, t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}$, compute $P_i^{LR}(t)$, the power consumption at charging deck i suggested by the optimal solutions to the decomposed optimization problems (2.6). Compute the total power consumption $P_T^{LR}(t) = \sum_{i\in I} P_i^{LR}(t)$.

2. Tighten the local capacity constraints (2.2b) by shrinking $\xi_i(t)$, the proportion of the local transformer capacity that is available for its affiliated PEV charging deck, in the following scheme:

$$\xi_i^s(t) = \min\{\xi_i^{s,0}(t), \xi_i(t)\}, \tag{2.10}$$

   where

$$\xi_i^{s,0}(t) = \begin{cases} \frac{P_i^{LR}(t)}{P_T^{LR}(t)}\frac{A_T\xi_T(t)}{A_i}, & \text{if } P_T^{LR}(t) > 0; \\ \frac{A_T\xi_T(t)}{\sum_{i\in I}A_i}, & \text{if } P_T^{LR}(t) = 0. \end{cases} \tag{2.11}$$

3. Solve the decomposed optimization problems (2.6) with $\xi_i^s(t)$ in replace of $\xi_i(t)$ in the local transformer capacity constraints (2.2b), $\forall t \in \{t_k, t_{k+1}, ..., t_{k+H-1}\}$.

The solutions obtained from the above heuristic algorithm are guaranteed to satisfy the primary transformer capacity constraints (2.2a) and are subsequently feasible to the centralized optimization problem (2.2). In fact, it can be easily verified that Eq. (2.11) ensures $\sum_{i\in I}\xi_i^{s,0}(t)A_i\lambda = A_T\xi_T(t)\lambda$ irrespective with $P_T^{LR}(t)$ being greater than or equal to zero; since $\xi_i^s(t) \leq \xi_i^{s,0}, \forall i \in I$ by Eq. (2.10), we guarantee that the local transformer capacity constraints (2.2b) dominate the primary transformer capacity constraints (2.2a) by replacing $\xi_i(t)$ with $\xi_i^s(t)$. Furthermore, Eq. (2.11) implies that the constraint tightening scheme tends to shrink $\xi_i(t)$ less significantly when $P_i^{LR}(t)$ accounts for a larger proportion of $P_T^{LR}(t)$ than when it accounts for less. This differentiation in constraint tightening results in a feasible solution which does not deviate significantly from the solution to the decomposed problem (2.6); we thus expect that the gap between the lower bound LB^r and the upper bound $J_{LR}^*(t_k|\vec{\eta^r})$ can be small from the initial iteration, which helps us find the optimal solution within few iterations.

We summarize the implementation steps of the distributed coordinated charging strategy as follows:

1. At each time instant t_k, each charging deck i updates the realized and the forecast information of charging demands as well as the amount of available power for PEV charging from the local transformer. Then each local charging coordination system solves the distributed optimization problem (2.6), with initial Lagrangian multipliers $\vec{\eta}^0 = \vec{0}$.

2. Based on the optimal solution obtained in step 1), each charging deck i reports to the upper-level aggregator its projected power consumption P_i^{LR} at each time interval over the planning horizon. The aggregator checks if the global transformer capacity constraints (2.2a) are satisfied:

   a) If YES, the distributed solution is centrally feasible and optimal. Skip to step 4).

   b) Otherwise, the distributed solution is centrally infeasible. The aggregator dictates updated Lagrangian multipliers $\vec{\eta}^1$ to the charging decks following the preceding subgradient method. Then it broadcasts the updated $\{\xi_i^s\}$ to charging deck i following the heuristic algorithm in search of a feasible solution.

3. Repeat step 2 until one of following termination conditions is satisfied at iteration r:

   a) The gap between LB^r and $J_{LR}^*(t_k|\vec{\eta^r})$ is smaller than a specified tolerance value.

   b) The iteration number exceeds a specified maximum iteration number r_{max}.

   c) The evaluation time exceeds a specified maximum time T_{max}.

4. Implement the charging schedule for interval $(t_k, t_{k+1}]$ based on the best possible feasible solution obtained from step 1) or step 3).

5. When the time proceeds to the next instant t_{k+1}, the above process from step 1) is repeated to determine the charging schedule for interval $(t_{k+1}, t_{k+2}]$.

2.4 Case Studies

Parameter Settings

Assume $M = 5$ residential communities get electrical power supply through a common primary distribution transformer in the city of Beijing, China. According to the empirical study on PEV charging behavior [62] and the 2011 Beijing transportation report [15], the initial SOC of private PEVs arriving at a residential charging deck follows the normal distribution $N(0.6, 0.1^2)$, the arrival time follows the normal distribution $N(7:00pm, (1.5hr)^2)$ and the departure time can be described by the normal distribution $N(7:45am, (1.0hr)^2)$. ($N(a, b^2)$ denotes a normal distribution with mean a and standard deviation b.) Battery of BMW ActiveE is selected in this study [12]. The rated charging power $P_{max} = 7$kW and the PEV battery capacity is 32kWh. In ideal conditions, the rated operation range of ActiveE is 160km. Considering the relatively long driving range of ActiveE, customers are not likely to charge their PEVs every day. In our study, it is suggested that a dweller charges his/her PEV for a certain day with probability 0.7. Following the central limit theorem, when the number of PEVs of a local community is substantial, the daily number of arrivals with charging needs can be approximately normally distributed, which we assume to be $N(60, 6^2), N(80, 8^2), N(100, 10^2), N(120, 12^2)$ and $N(120, 12^2)$ for each charging deck, respectively. In addition, we assume that the desired SOC is fixed at 1 for simplicity. Using these parameter values, we generate the forecast and the realized demand profiles independently in the simulations. As to the charging system parameters, the rated capacities of the five local transformers are 400kVA, 500kVA, 750kVA, 750kVA and 750kVA, respectively, and the rated capacity of the primary distribution transformer is 2520kVA, which is 80% of the sum of the local transformer capacities. The average charging power factor is assumed to be $\lambda = 1$, the charging efficiency $\rho = 0.92$ and the minimum ratio of the actual charging power to the rated charging power $\theta_{th} = 0.3$. The step-size coefficient β^r in Eqn. (2.8) is tuned to be 250 for fast convergence. As to the termination conditions described in the preceding section, we set the maximum tolerated gap between LB^r and $J^*_{LR}(t_k|\vec{\eta^r})$ to be 5% of $J^*_{LR}(t_k|\vec{\eta^r})$, the maximum number of iteration $r_{max} = 20$, and the maximum evaluation time $T_{max} = 5$min. The values of these termination parameters are chosen specifically for the case that we discuss in this work, respecting various aspects of the problem, such as its scale, topology, data accuracy and realistic service requirements.

Fig. 2.3 shows the time-varying ratios of the base load to the rated transformer capacity for the whole distributed communities and for each local community. Also shown in Fig. 2.3 is the industrial TOU electricity rates in Beijing [11], being \$0.0584/kWh, \$0.1099/kWh and \$0.139/kWh for off-peak, shoulder and on-peak periods, respectively. The electricity retail prices for PEV charging is assumed uniform, as \$0.1608/kWh. It can be seen that the fluctuations of the base load and of the TOU rates positively correlate to each other; therefore, the charging system tends to charge PEVs during the off-peak (i.e. 0:00am - 6:00am) periods, when it is more profitable and there is also more available power for charging than in other periods.

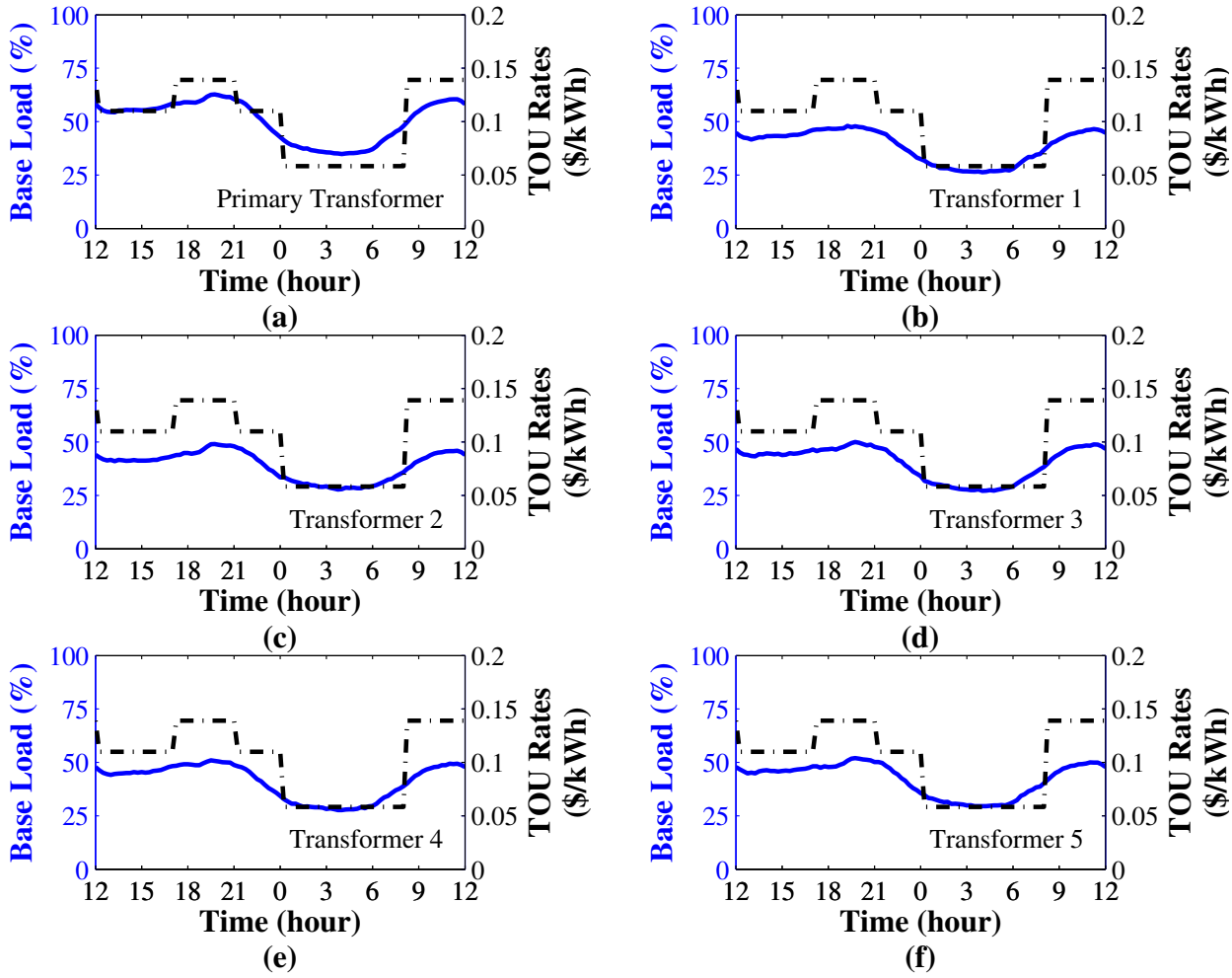

Figure 2.3: Ratios of the base load to the rated transformer capacity (solid line) and TOU electricity rates (dashed line), (a) with respect to the primary distribution transformer, and (b)-(f) with respect to local transformers 1-5, respectively.

We assume that the sampling and decision interval $\Delta_t = 15$min. At each instant t_k, the forecasts of PEV arrivals during the next four intervals (i.e. 1 hour) are taken into account and the maximum charging planning horizon for each PEV is 15hours. Therefore, the planning horizon $H = 1 + 4 + 4 \times 15 = 65$. In the optimization problems, we use Eqn. (2.3) as the linear penalty term for unfulfilled demands. Other parameter values are taken as follows: $\gamma = 1$, $\mu = 0.1608$ and $\kappa = 1.608 \times 10^{-6}$. All the parameter values are chosen or tuned based on reality and optimization considerations. The simulations run on a PC with Intel Core i5 CPU (2.27 Ghz) and 4 GB RAM.

Less Coordinated Charging Strategies

In order to demonstrate the effectiveness of the centralized and the distributed coordinated charging strategies, we also consider two less coordinated strategies for comparison purposes.

1) The no-control strategy imposes no control on PEV charging. That is, any PEV will start being charged at the rated charging power immediately after it enters a parking deck and gets plugged in. The full-speed charging lasts until the PEV departs or its battery becomes fully charged.

2) Another strategy, which we name "myopic" strategy, enables coordination only within each charging deck. In other words, each local charging control system solves the optimization problem (2.6) for only one iteration and then schedules PEV charging independently at each interval.

Simulation Results

We carry out simulations of 24-hour operations of the five distributed charging decks starting from 12:00pm, following different charging strategies. The simulated charging systems start from a stationary state after multiple days of operations. Figs. 2.4 and 2.5 show the time evolution of the charging powers of the local charging decks and of the whole distributed charging system, respectively. In the figures, we refer to the distributed, the centralized, the myopic and the no-control charging strategies as DCTR, CTR, MYO and NoC, respectively, and refer to the available charging power (i.e transformer capacity less the base load) as AVL.

Fig. 2.4 shows that, following the distributed, the centralized and the myopic strategies, local charging decks choose to charge PEVs during the off-peak period 12:00am-6:00am when TOU electricity rates are the lowest in a day; in addition, while all stratifying the local power availability constraints (2.2b), the decks start charging PEVs immediately at 12:00am owing to the earliness incentive that we incorporate into the objective function (2.1). Comparing these three coordinated charging strategies during the busy charging period, we observe that the distributed and the centralized charging strategies consume strictly less power than available power most of time, while the myopic strategy hits the local charging power limit almost all the time. To explain this difference, note that the capacity constraints on the primary transformer (2.2a) are tighter than the local power availability constraints (2.2b), so the latter are non-binding during the busy period when the centralized and the distributed strategies are adopted. In contrast, the functionality of charging coordination across the distributed charging decks is not available to the myopic strategy; consequently, the myopic charging schedule violates the primary capacity constraints (2.2a) during the peak hours, which is clearly shown in Fig. 2.5. Fig. 2.4 also shows that much smoother local charging power over time results from the distributed coordination, compared with the fluctuating charging power following the centralized coordination, as the latter strategy is indifferent to the affiliation of the charging ports to the charging decks while shifting the loads. The value of charging coordination is also demonstrated by the inferior performance that the

Table 2.1: Overall system performance following different charging strategies.

	Profit ($\$/day$)	Transformer Capacity Constraint Violated	Valley Filling
DCTR	706.1	None	Yes
CTR	706.1	None	Yes
MYO	706.1	Primary	No
NoC	208.7	Primary & Local	No

no-control strategy presents in Figs. 2.4 and 2.5. Its power consumption in both the local and the primary levels follows the PEV arrival process and violates the capacity constraints.

We summarize the overall system performance following different charging strategies in Table 2.1. It shows that the total profits reaped in a day following the distributed and the centralized strategies are the same as the profit by the myopic charging strategy; this is because the TOU rates are fixed within the off-peak period and thus shifting loads within this period does not decrease the revenue. It is also shown that these profits are 3.4 times as large as the profit earned by the no-control strategy. This remarkable gap in profit demonstrates the value of load shifting from peak periods to off-peak periods. Moreover, while all the four strategies are able to satisfy charging demands to almost 100% level, only the distributed and the centralized charging strategies are feasible both within and across the distributed charging decks, in the perspective of transformer overload protection. Finally, only under distributed and centralized charging strategies, the valley filling target is achieved by responding to TOU incentives. Due to its uncoordinated nature, myopic charging strategy causes "rebound effect" (i.e. it creates an excessive overlapping charging peak), and does not effectively achieve valley filling.

We next investigate the computational efficiency of the four charging strategies. The coordinated PEV charging problem needs to be solved in real time and is thus demanding efficient online computing algorithms. The evaluation times of making charging decisions at each interval are summarized in Table 2.2. It can be seen that, due to dimension reduction in the optimization problems, the distributed and the myopic charging strategies consume less than one third of the mean/maximum/minumum evaluation times that are required by the centralized strategy. The distributed strategy consumes roughly the same amount of the minimum evaluation time as the myopic strategy, but consumes more than twice the amount of the maximum evaluation time. This is because the distributed strategy is able to find the optimal solution within one iteration during the periods when transformer capacity constraints are not binding, thus being computationally identical with the myopic strategy; on the other hand, at some intervals during the busy charging period, the distributed strategy needs additional computation to find a globally feasible (and optimal) solution. In most cases, the convergence is reached within two iterations. This is because the feasible charging power schedules slightly shift charging load among adjacent intervals. Consequently, the expected charging cost does not change under the block-wise TOU tariffs.

To further demonstrate the computational advantage of the distributed charging strategy,

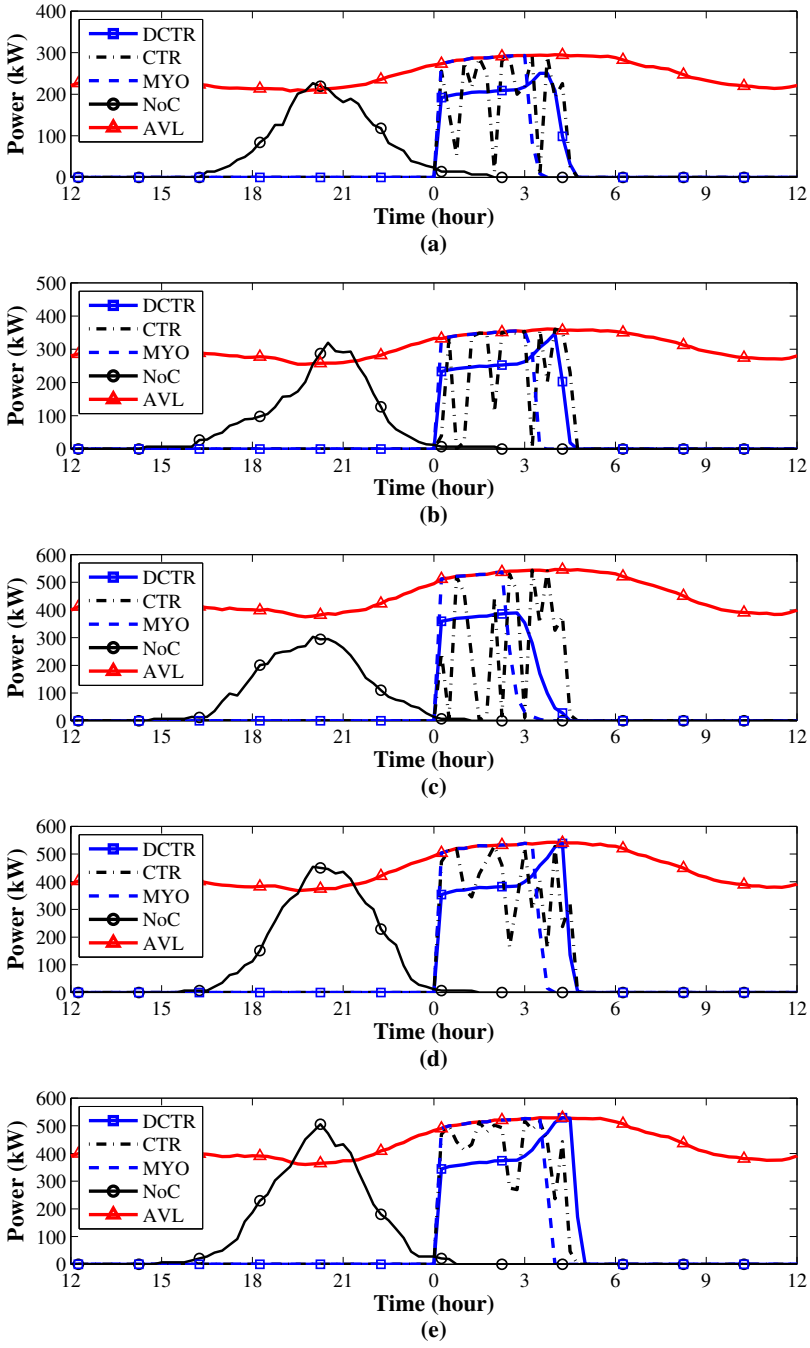

Figure 2.4: Charging power following distributed, centralized, myopic and no-control charging strategies. (a)-(e) correspond to charging decks 1-5, respectively.

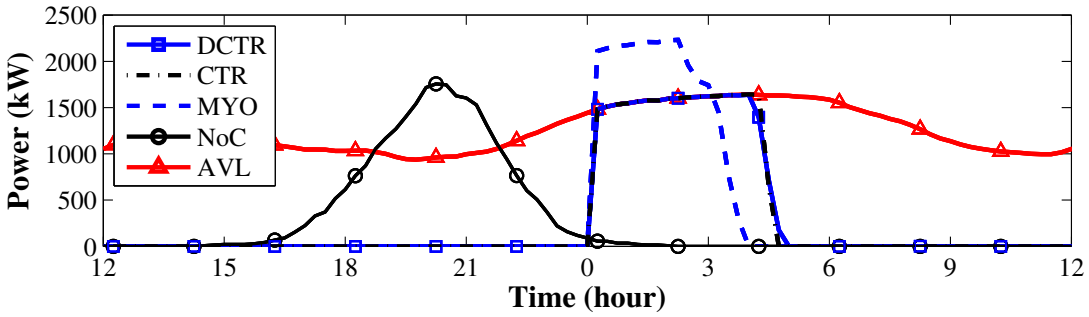

Figure 2.5: Total charging power of the five charging decks following distributed, centralized, myopic and no-control charging strategies.

Table 2.2: Evaluation times of making charging decisions following different charging strategies.

	Mean (sec)	Maximum (sec)	Minimum (sec)
Distributed	3.7653	8.8229	1.6194
Centralized	11.2769	24.5647	7.5939
Myopic	2.4770	3.9164	1.6744

we scale down/up the problem size to see how the evaluation times change with respect to the problem size. By scaling the size of the problem, we multiply the transformer capacities, the base load and the number of PEVs arrivals at each charging deck by a common scaling factor. In these numerical tests, we place no restriction on the evaluation time, but keep all other termination conditions unchanged. As shown in Fig. 2.6, both the maximum and the mean evaluation times are smaller under the distributed control than under the centralized control of all scales being considered. The evaluation times for solving centralized/distributed scheduling problems in each iterations increase; but the convergence under the distributed control can still be achieved within very few iterations. Moreover, the evaluation time grows approximately linearly in the problem scale under the distributed control, whereas it grows superlinearly under the centralized control. Therefore, we expect that the distributed charging strategy is more computationally viable than the centralized charging strategy when being implemented in a large scale.

2.5 Conclusion

This chapter investigates the plug-in electric vehicles charging coordination problem for multiple parking decks in multi-family dwellings. A centralized finite-horizon optimization model is first proposed to maximize the total utility of charging service providers (i.e. total profits less the dollar value of penalty for failing to meet customer charging needs). By exploiting the structure of the centralized control problem, the original model is then decomposed into

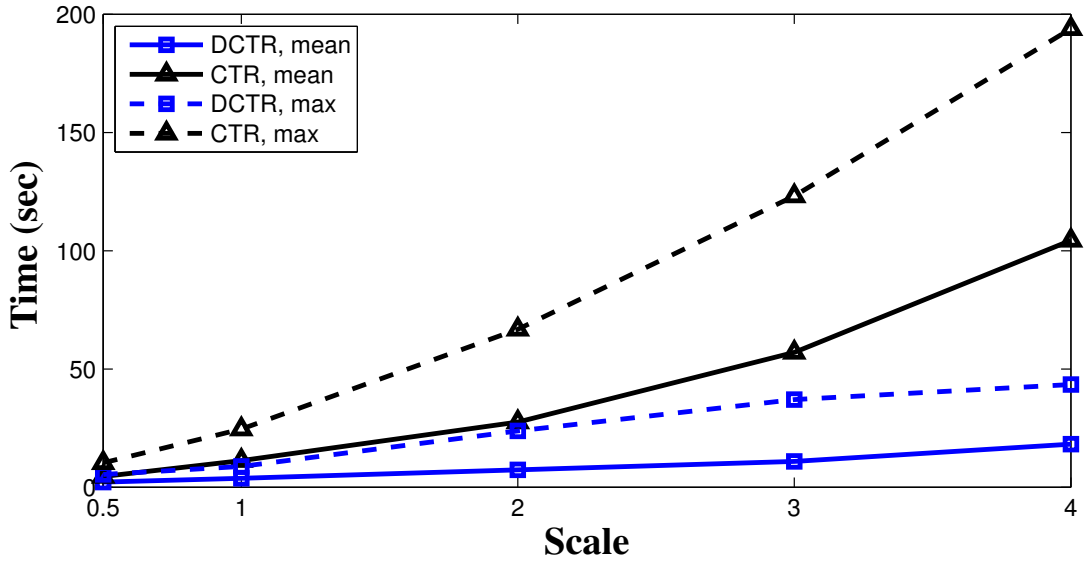

Figure 2.6: Mean and maximum evaluation times with different system scales following the distributed and the centralized coordination strategies.

several sub-problems that can be solved locally in parallel iteratively with updated multipliers information broadcasted by the centralized controller. We then propose an efficient heuristic method to find the feasible solution to accelerate the convergence. Simulations in a scenario with 5 parking decks are carried out and computational results are compared. The simulation results demonstrate the effectiveness of centralized, distributed and myopic control strategies in shifting charging load to off-peak periods by responding to TOU electricity tariffs. Besides, the proposed distributed approach has the following advantages:

1) The distributed approach successfully achieves charging coordination between multiple parking decks and prevents undesirable charging rebound, which happens when charging coordination is only implemented within each charging deck separately.

2) Compared with the centralized approach, the distributed approach reduces the problem dimension by decoupling the centralized charging problem into several subproblems with respect to each parking deck. Therefore, this approach is able to be effectively applied online even when the number of electric vehicles and parking decks controlled are large.

3) Besides, at each iteration in the distributed algorithm, only aggregate charging power demand of each parking deck is required to be publicized. This is applicable for the scenario when the controlled parking decks are operated or owned by different entities who are not willing to disclose all of its customers charging information to the centralized controller. What's more, the reduced volume of data exchange between the aggregator and the local controllers enables the proposed hierarchical distributed framework to be less demanding on communication infrastructure than the centralized approach.

4) Finally, though we do not explicitly investigate this point in this work, the distributed

approach is more likely to be robust against system faults. In case of system failure at the centralized controller, which may otherwise cause a total collapse of the whole system, each parking deck can switch to myopic mode and still effectively shift the charging load to off-peak periods. In other cases, when fault occurs to any of the subsystems at parking decks, the fault can be easily isolated.

There are a number of extensions worth further investigations. For example, it is interesting to investigate the problem where power flow and power losses of distribution network are also considered. Also, a distributed control framework coordinating PEV charging/discharging and renewable energy generation can be an interesting and important extension of this work. PEVs and renewable energy can jointly enable the sustainable energy future that is featured by distributed energy generation and storage [50]. In a scenario where multi-family dwellings are mounted with renewable energy (e.g., wind/solar energy) generation units, the proposed distributed PEV charging systems are also able to provide energy reserve to offset intermittence of wind/solar energy generation. The distributed charging service providers can also participate in ancillary service by means of vehicle to grid support to earn more profit [80].

Chapter 3

Integration of Shared Mobility for Home Delivery Services

California, along with many other regions in the world, is experiencing two socioeconomic transformations: 1) Sharing economy, in which people share access to goods and services, is prospering. The market size of sharing economy has the potential for increasing global revenues from roughly $15 billion today to around $335 billion by 2025 ([75]). One of the key sectors that is being impacted is transportation, with leading companies (e.g., Uber, Lyft and Car2Go) disrupting the traditional business of people-riding transit. 2) Meanwhile, retail e-commerce sales is expanding rapidly worldwide, with an annual growth rate of around 20%, and is projected to reach $2.4 trillion in 2017 according to [34]. To advance their logistics competitiveness, some major players have kept piloting new delivery services (e.g., Walmart To Go, Google Express and Amazon Fresh, etc.) during the past several years.

A question that naturally arises from these phenomena but has not been well addressed is: Is it possible for retail e-commerce to rely on crowdsourced shared mobility for delivery services on a large scale? [13] anticipates that the power of such a mode is that *"...it does not require the asset-heavy infrastructure of warehouses, vehicle fleets, fuel costs and employed drivers that traditional logistics companies have to pay for and manage... it's an asset-light model, akin to the likes of Uber and Airbnb, with low overheads meaning it can scale relatively fast depending on demand."* Can this marriage of sharing economy and e-commerce really have all these claimed merits? Until recently, such practice has remained in its infancy. On the one hand, most of the existing leading companies and promising start-ups in the courier industry primarily rely on professional people and dedicated fleets (e.g., 95% of Zipments' couriers are professionals). On the other hand, some companies (e.g., DHL, Walmart and start-ups including Posmates and Deliv) have begun to experiment with crowdsourced delivery ([81]).

The goal of this chapter is to evaluate whether and/or how to foster the large-scale adoption of this prospective sharing mode of delivery services in urban areas. While business objectives in general do not directly address the interests of external stakeholders, governments and logistics services providers are responsible for assessing the potentially substantial

impacts of this mode shift with the ultimate goal of balancing societal and environmental concerns with economic objectives as the Triple Bottom Line ([31]). This responsibility, for example, is particularly relevant to the California state government, given its strategies to develop multi-model integrated freight systems and to create car sharing programs ([22]) as well as for its commitment to a 15% reduction (from 2010 levels) of GHG emissions to achieve 1990 levels by 2020 as dictated in [5]. According to the California Environmental Protection Agency ([21]), transportation was accountable for 37% of the total GHG emissions in the state for 2013, dwarfing all other sectors.

The first part of this chapter develops planning models for this prospective mode of sharing logistics. The models are based on a one-transshipment logistics setting: a fleet of short-haul trucks or vans are dispatched from a depot and unload goods at terminals of service zones. Passenger cars nearby with available mobility are attracted to each terminal, each picking up a ration of goods and delivering them within the service zone to their destinations, which are assumed to be uniformly distributed. Contrasting the conventional mode where dedicated vehicles travel closed-loop routes, a salient feature of the shared mobility is its one-way nature: a car starts an outbound trip by approaching the closest destination and the service ends once it drops off the last package. Numerically solving $9,600$ different instances of this open vehicle routing problem reveals the structural properties of the route patterns. To characterize these properties, a continuous approximation (CA) model is developed. Combining this model with the CA model of short-haul truck routes results in strategic service zone deployment in the form of a closed-form expression of the optimal density of zones as well as the associated optimal cost.

The second part of this chapter studies the operating costs and GHG emissions implications of the sharing logistics paradigm. Based on the analytical solutions to the service zone designs, three logistics scenarios are evaluated: two prospective shared-mobility scenarios that minimize operating costs and GHG emissions, respectively, are evaluated; then, the conventional truck-only home delivery services scenario is evaluated as the benchmark. Case studies calibrate these optimal designs with detailed empirical parameter estimates. Comparing operating costs and GHG emissions across these three scenarios leads to a discussion of what conditions under which it is worthwhile (and how) to facilitate the transition into this prospective sharing mode. As an extension, a non-linear driver payment scheme is derived, which induces efficient supply of shared mobility.

The contributions of this work are as follows. 1) To the best of the authors' knowledge, this work is the first attempt to *design and analyze the logistics system that features large-scale integration of shared mobility for home delivery services, using analytical models and empirical parameter estimates*. In particular, the asymmetry of one-way passenger cars' routes invalidates existing CA models for vehicle routing problems (VRP); the CA model in this work, along with the underlying insights into the open-loop routes properties, fills this gap in literature. 2) A major finding of this work is that *adopting shared mobility for home delivery services is likely to be either asset-light or cost-efficient, but not both, unless additional policy instruments exist*. Under normal operating conditions, the prospective sharing logistics system increases operating costs by 2.73% and increases GHG emissions by

3.95%, compared with the conventional truck-only scenario. Having not been fully identified in literature, this potential cost inefficiency can be a major explanation for the sharing logistics paradigm not having taken place on a large scale. On the other hand, this paradigm has the potential to create considerable economic and environmental benefits if additional investment in heavy-duty fleet assets and policy instruments (such as fuel-efficiency standards and subsidy incentives) are possible. 3) Another finding is that, upon entering this paradigm, *even exclusively minimizing operating costs does not significantly increase emissions* relative to the minimum level of emissions. This is a more robust result than in other retail logistics contexts such as the one considered in [17], in which retail store operations may cause a more significant tension between the environmental and financial preferences. 4) Finally, governments and delivery services companies potentially need to carefully design *the driver compensation schemes to cost-effectively induce supply of shared mobility.* The proposed non-linear pricing scheme serves this purpose.

The remainder of this chapter is organized as follows: Section 3.1 reviews the related literature; Section 3.2 introduces the basic model settings; Section 3.3 presents the shared-mobility logistics planning model; Section 3.4 analyzes operating costs and GHS emissions implications based on empirical estimates; finally, Section 3.5 concludes the chapter. Additional technical proof is provided in Appendix B.

3.1 Literature Review

Large-scale adoption of shared mobility for retail e-commerce is an important prospect, but it has not received wide attention in literature. The vast majority of the studies concerning the impact of sharing economy on the transportation sector have focused on the transportation of passengers instead of packages. One group of these studies investigates vehicle sharing, which refers to the case where people access mobility services by temporary usage without ownership of the vehicles. For example, [10] study the economic and environmental implications of the car sharing business and find that its overall environmental impact can be negative due to aggregate vehicle usage. [85] and [52] study the deployment and operations problems arising in bike-sharing systems, employing network flow models and empirical methods, respectively. [46] notice the recent development of one-way electric vehicles sharing (e.g. Car2Go) and formulate a service region design problem. Another stream of literature is concerned with ride sharing, which refers to a different mode of transportation where a passenger accesses a vehicle for a trip with its driver already en route. See [38] for a comprehensive review of ride sharing management and [1] for a review of an important operations problem, namely the dynamic matching of shared rides.

In contrast, studies on goods logistics with shared mobility are few and mostly qualitative. [19] identify different types of logistics coexisting in the sharing economy based on an exploratory analysis of 32 cases. [81] examine 18 businesses. They argue that the two major obstacles to the development of crowdsourced delivery are trust building and a "chicken-and-egg" dilemma: a critical mass of couriers is needed to insure quality delivery and to attract

customers, but a critical mass of customers is needed to attract couriers. However, the analysis in this chapter shows that a major obstacle may be high operating costs, which contradicts the "low overheads" claim by [13]. Recently, [59] and [58] propose a conceptual taxi-based people-freight sharing system. They formulate a routing problem as a mixed integer linear program and propose a neighborhood search heuristic solution algorithm. In comparison, this work develops continuous approximation models for a home delivery logistics system, with emphasis on its overall economic and environmental implications.

Logistics planning with environmental sustainability considerations has been studied in contexts other than sharing economy. [17] models the dependence of operating costs and GHG emissions on the density of retail stores and finds that improving consumer fuel efficiency can be an effective way of mitigating emissions. [20] find that household-level economies of scale in transportation (i.e., a person performs many errands in a single trip) may actually increase the overall carbon footprint relative to the case of home delivery services. The service region deployment problem considered in this work is based on CA models in a one-transshipment setting. CA methods for one-transshipment logistics systems are investigated in [100] with detailed discussion on the approximation error, and in [82] with environmental considerations. Again, the key distinction of this work is the inclusion of shared mobility.

3.2 Model Settings

Topology: Consider a one-transshipment logistics system. A *depot* is located at the center of a *service region* of interest. At the inbound stage, a fleet of short-haul trucks or vans (hereafter referred to as trucks) are dispatched from the depot, each carrying a truck load of packages. A truck stops at one or multiple *terminals*, where it unloads packages to be delivered in *service zones*. Each terminal is located at the center of a service zone. Figure 3.1 illustrates the route of a truck that visits three service zones. At the outbound stage, passenger cars nearby with available mobility (hereafter referred to as cars) are attracted to each terminal. Then they pick up a ration of goods and deliver them to their *destinations* within a service zone. Such a transshipment setting captures the fundamental tradeoff: To carry a truck load of goods to travel a certain distance, a truck is typically more cost efficient than a cohort of cars of equal capacity; however, to carry one kilogram of goods to travel a certain distance, a car is more cost efficient than a truck.

Moreover, assume that the whole service region is of a squared diamond shape with area a. The service zones partition the service region into z squared diamonds of equal size. Each zone has one vertex pointing toward the depot (Figure 3.2 in Section 3.3 depicts a service zone and outbound car routes). Manhattan metric of distance (l-one norm) is used throughout the chapter. As a result, all points on the boundary of the service region are equally distant to the depot; so are all points on the boundary of a zone to its terminal. The above topological assumptions are made for simplicity of analysis. Layout robustness in Section 3.4 implies that locally relaxing those restrictions (i.e., in terms of tilings and

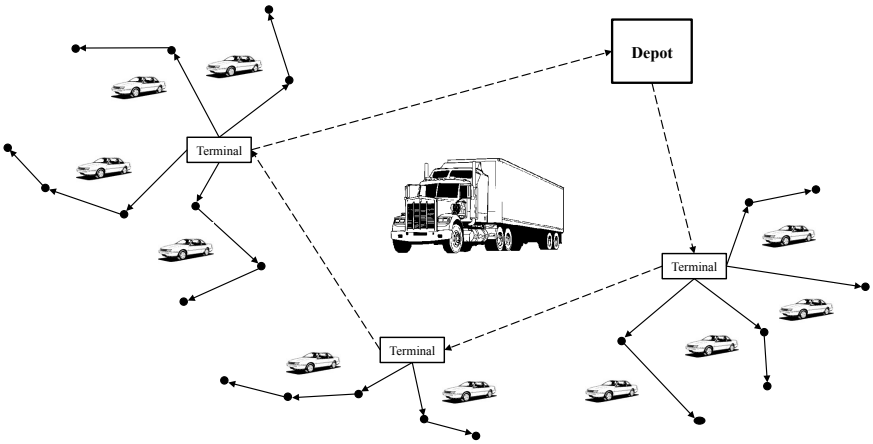

Figure 3.1: An illustration of home delivery services with shared mobility of passenger vehicles–one truck route with three service zones.

distance metrics) do not significantly affect the results.

Demands and mobility availability: Assume that the demand destinations constitute a Poisson point process–the destinations are uniformly and independently distributed over the service region. The total number of destinations per unit area for one dispatch, N, is a Poisson random variable, i.e., $N \sim Poisson(n)$, where n is the mean density of the destinations. The random package weight G for each destination is independently and identically distributed with mean g. Also assume that available cars are uniformly and independently distributed in the region with density m. A summary of parameter notation and estimates is available in Table 3.1 in Section 3.4. Symbols throughout this chapter are subscripted with a vehicle type ("t" for trucks and "c" for cars) wherever necessary.

The shared mobility logistics planning problem thus boils down to determining the number of the service zones, z, which implies other decisions such as the zone size and the average number of cars to serve each zone. The CA model in Section 3.3, based on the above setting, can be readily extended to the case where the distribution of demands, destinations and cars are non-uniform, as long as the non-uniformity is modest.

3.3 The Shared-Mobility Logistics Planning Problem

This section describes CA models of different components of this transshipment logistics system with shared mobility. Striking a balance among the costs of these components yields a closed-form expression of the optimal number of service zones and the associated optimal cost, which can be compared with the optimal cost of the conventional truck-only scenario.

Terminal-bound Car Trips & Transshipment

Each service zone on average requires m_z nearest cars to distribute packages within the zone. m_z is equal to $\frac{ang}{v_c z}$, since $\frac{ang}{z}$ is the average weight of packages of a zone and the cars are assumed to be loaded to capacity v_c for simplicity. These cars are uniformly scattered in a squared diamond of area $\frac{m_z}{m}$ centering the terminal. Their mean distance to the terminal can be approximated by the mean distance between a point in this diamond and the center. This l-1 distance is computed by integration as follows:

$$d_a = \int_0^{\frac{1}{2}\sqrt{\frac{m_z}{m}}} \sqrt{2} r \frac{8r}{\frac{m_z}{m}} dr = \frac{\sqrt{2}}{3}\sqrt{\frac{m_z}{m}}. \tag{3.1}$$

At a terminal, a freight truck unloads packages weighing $m_z v_c$, which m_z cars then pick up. This transshipment process incurs the following cost:

$$c_f + c_{h_0} m_z v_c + c_{h_1} m_z = c_f + c_h m_z, \tag{3.2}$$

where c_f represents the fixed charge. c_{h_0} is per kilogram cost of unloading packages from the truck and c_{h_1} is the cost of loading packages into each of the m_z cars. $c_h = c_{h_0} v_c + c_{h_1}$ is the combined variable handling cost per car.

Outbound Car Trips

Each car's outbound trip breaks down into two portions. In the *line-haul* portion, the car travels the shortest path from the terminal to the destination that is the closest to the terminal among those on its route. After that, the car proceeds along the *detour* portion of the route to visit all the remaining destinations. Aggregating all the trips in a zone, an approximation of the total outbound trip distance consists of the following two parts:

$$d_o = d_l(n, \frac{v_c}{g}) m_z + \lambda\sqrt{an}, \tag{3.3}$$

where the two terms on the right-hand side represent the total line-haul and detour trip lengths, respectively. In particular, $d_l(n, \frac{v_c}{g})$ is the average of individual line-haul trip lengths. d_l primarily depends on the demand density n as well as the number of destinations that each car visits, as approximately measured by $\frac{v_c}{g}$. The detour portion is asymptotically structurally similar to a traveling salesman problem (TSP) tour as n goes to infinity. Hence, the total detour length is approximated to be proportional to $\sqrt{an}$ with a coefficient λ.

The appropriate function form of $d_l(n, \frac{v_c}{g})$ and the value of λ should fit with the optimal routes, which are solutions to the open vehicle routing problems (OVRPs). As a variant of the classic, closed-loop VRP, the OVPP is defined so that vehicles do not return to the origin once they finish delivering all the packages, and is no easier to solve. State-of-the-art commercial solvers are only able to consistently solve instances with up to 50 nodes. Fortunately, various local-search heuristics developed in the past 15 years have demonstrated

superior solution accuracy and efficiency. In this work, a record-to-record heuristic embedded in the open-source library VRPH ([44]) is applied to solve 96,000 randomly generated OVRP instances with different values of n, $\frac{v_c}{g}$, variation coefficient of G and random realizations of destinations ([60] demonstrate the high accuracy of this type of heuristic). Figure 3.2 depicts four OVRP routes with different demand densities and car capacities. Figure 3.3 shows the average individual and the total line-haul lengths with different demand densities and car capacities. Figure 3.4 shows the values of the detour coefficient λ and the total detour lengths.

Figures 3.2-3.4 reveal the following properties of the outbound car trips: 1) From Figure 3.3 (a), d_l approaches a value near $\frac{\sqrt{2}}{3} = 0.471$, the mean distance from a point in a zone of unit area to its terminal, when $n \to 0$. 2) Within the parameter value range of interest, either increasing n or $\frac{v_c}{g}$ decreases the line-haul trip length, since the first destination on a one-way route thus becomes more likely to be closer to the terminal. In contrast, the line-haul portion of a standard VRP route is less sensitive to n or car capacity due to its closed-loop symmetry. This observation suggests that a region with sparser demand destinations or smaller cars will gain more savings by not having to travel the return line-haul trip, which would be relatively long. 3) Numerical experiments also show that the variability of package weight G, though affecting the mean number of destinations a car visits (i.e., $\mathbb{E}[\frac{v_c}{G}]$), has negligible effect on the average line-haul length. 4) Comparing Figure 3.3(b) and Figure 3.4(b), the detour portion dominates the whole route length when car capacity is relatively large (e.g., $\frac{v_c}{g} = 14, 23$). In those cases, Figure 3.4(a) shows that the detour coefficient λ is nearly constant. When $\frac{v_c}{g} = 5$, the detour and the line-haul portions are of similar length and λ increases in n. However, heavy packages for e-commerce home delivery are rare in reality, and the ratio $\frac{v_c}{g}$ is likely to be greater than five.

Based on the these observations, the following function form is proposed to approximate the average individual line-haul trip length:

$$d_l(n, \frac{v_c}{g}) = (\frac{\sqrt{2}}{3} - \alpha(\frac{v_c}{g})^{\beta}(1 - \frac{1}{\gamma\sqrt{\frac{an}{z}}}))\sqrt{\frac{a}{z}} \tag{3.4}$$

where $\alpha = 0.055$, $\beta = 0.470$ and $\gamma = 0.374$ are parameters that minimize the sum of squared approximation errors of the problem instances. The dashed lines in Figure 3.3(a) represent the approximations given by this model. Despite the certain degree of misestimation (up to 10.7% when $n = 70$ and $\frac{v_c}{g} = 23$), the model and the estimates capture the aforementioned one-way routing structural properties and are effective for a wide range of parameter settings. Further validation with real out-of-sample data is beyond the scope of this work. Similarly, the minimum-squared-error estimate of λ is 1.005, which is close to 0.97 as [51] estimates for TSP tour with the Manhattan metric.

Inbound Truck Trips

At the inbound stage, assume that each truck carries a full truck load (i.e., v_t kilograms) of packages, unloads them at $y = \frac{v_t}{m_z v_c} = \frac{v_t z}{ang}$ terminals on average and then returns to the

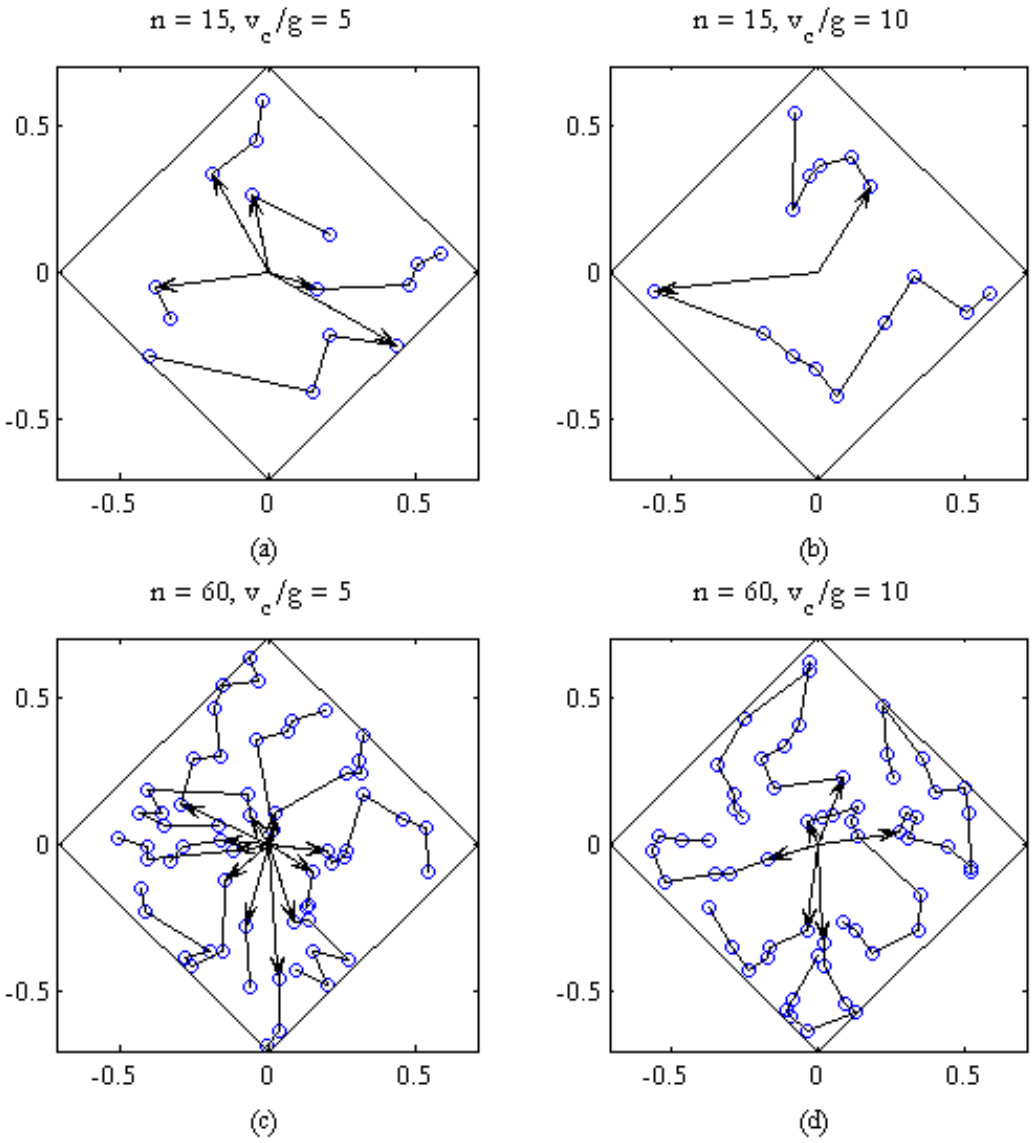

Figure 3.2: Outbound routes with different demand densities and car capacities in a squared-diamond service zone of unit area. Arrowed line segments represent the line-haul portion of routes. Blue circles represent destinations.

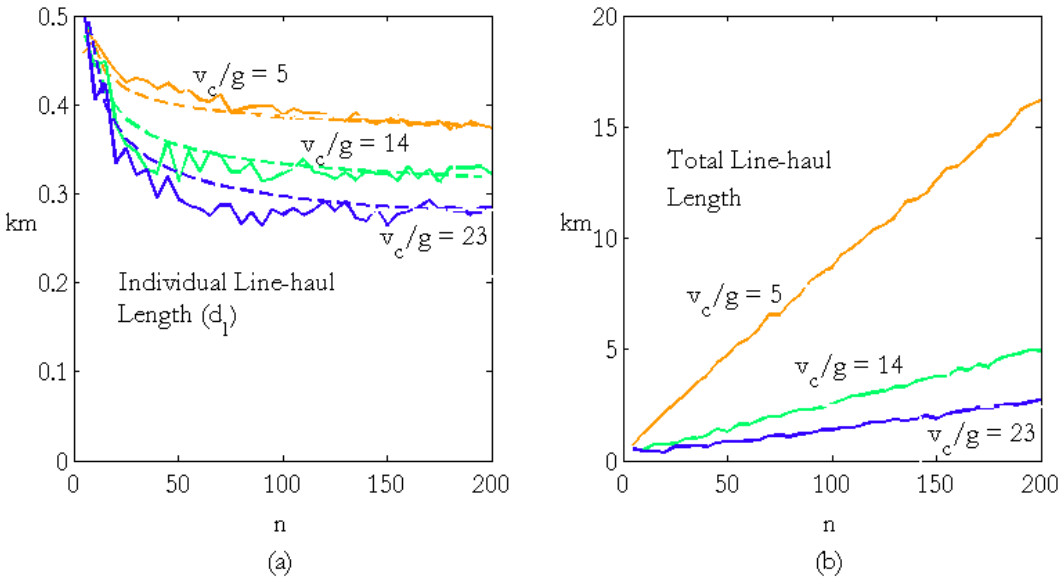

Figure 3.3: Line-haul trip lengths in a squared-diamond service zone of unit area. (a) Average of 100 individual trip lengths (solid line) and its approximation given by Eqn. (3.4) (dashed line). (b) Total length.

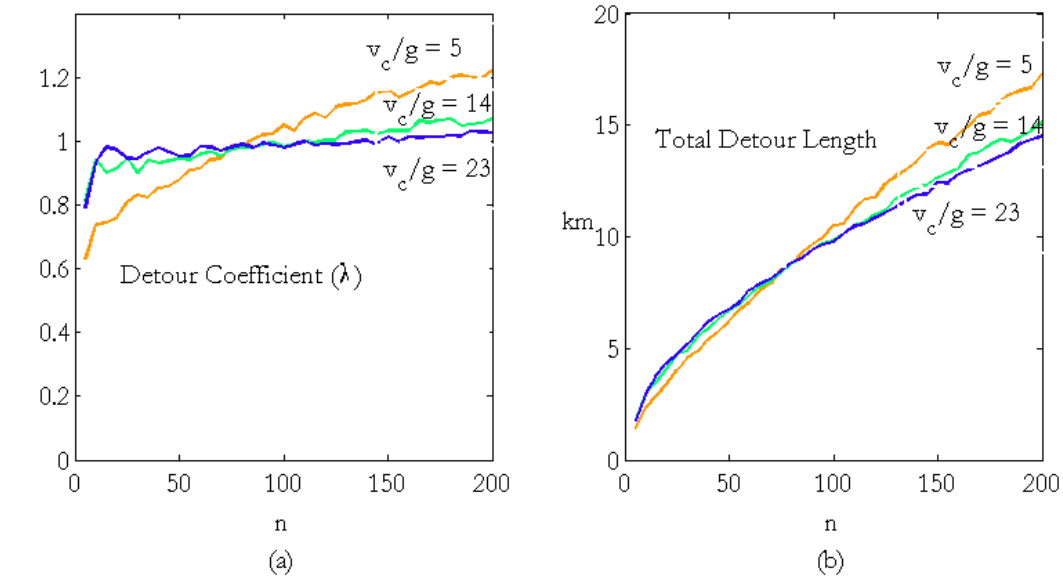

Figure 3.4: Detour trip lengths in a squared-diamond service zone of unit area. (a) Average of 100 individual detour coefficients. (b) Total length.

depot. The detour portion is thus a trip on s lattices and has a fixed length $d_{t,d} = y\sqrt{\frac{2a}{z}}$. The line-haul portion consists of the trips both before and after the detour trip. Following a similar discussion in [25] for this routing scenario, to minimize the total line-haul truck trip length $d_{t,l}$ in the region, arrange y zones for each truck into an area such that this area is elongated towards the depot. A truck first visits the terminal that is the nearest to the depot, and then moves forward to visit the rest of the first $\lceil \frac{y}{2} \rceil$ terminals. On its return trip, the truck visits the remaining terminals. Consequently, the average departing/returning line-haul trip plus a one-way trip halfway into the elongated area of y zones approximates an average trip from the depot to a point in the region, i.e., $\frac{d_{t,l}}{2} + \frac{1}{2}\frac{d_{t,d}}{2} \approx \frac{\sqrt{2a}}{3}$. Hence, let $d_{t,l} = \frac{2\sqrt{2a}}{3} - \frac{d_{t,d}}{2}$. The total inbound trip length becomes:

$$d_i = \frac{ang}{v_t}(d_{t,l} + d_{t,d}) = \frac{ang}{v_t}(\frac{2\sqrt{2a}}{3} + \frac{y\sqrt{\frac{2a}{z}}}{2}) = \frac{ang}{v_t}\frac{2\sqrt{2a}}{3} + \frac{\sqrt{2az}}{2}, \tag{3.5}$$

where $\frac{ang}{v_t}$ is the number of trucks in one dispatch. As discussed in [24], this approximation is particularly appropriate for large-scale logistics systems with large numbers of trucks and services zones.

Optimal Number of Service Zones

The strategic deployment of delivery service zones is represented by the optimal number of service zones, z^*, in the homogenous setting of this work, striking a balance between the truck mode and the shared mobility mode with a goal of minimizing total operating costs and/or GHG emissions. The following procedure solves for z^*.

To calculate the total trip length of cars, d_c, combine models (3.1) and (3.3) for terminal-bound and outbound trips, respectively, and substitute the line-haul formula (3.4) as well as $m_z = \frac{ang}{v_c z}$ into the model. Mathematically,

$$d_c = z(m_z d_a + d_o) = \Phi_c \frac{1}{\sqrt{z}} + \Psi_c, \tag{3.6}$$

in which

$$\Phi_c = (\frac{\sqrt{2}}{3} - \alpha(\frac{v_c}{g})^\beta)\frac{a^{1.5}ng}{v_c} + \frac{\sqrt{2}}{3}\frac{1}{\sqrt{m}}(\frac{ang}{v_c})^{1.5}, \tag{3.7a}$$

$$\Psi_c = \alpha(\frac{v_c}{g})^\beta \frac{a\sqrt{n}g}{\gamma v_c} + \lambda a\sqrt{n}. \tag{3.7b}$$

Similarly, rewrite model (3.5) of the total inbound truck trip length as:

$$d_t = \Phi_t\sqrt{z} + \Psi_t, \tag{3.8}$$

in which

$$\Phi_t = \frac{\sqrt{2a}}{2}, \tag{3.9a}$$

$$\Psi_t = \frac{ang}{v_t}\frac{2\sqrt{2a}}{3}. \tag{3.9b}$$

Let c_c and c_t denote the per-km cost incurred by trucks and cars, respectively, where the cost refers to operating costs, GHG emissions, or a monetized combination of both. Meanwhile, assume that the cost of package handling in the transshipment process as characterized in (3.2) is negligible, with three reasons: 1) First, omitting the term associated with the fixed charge c_f leads to a closed-form expression of z^*. 2) This cost, which is mainly due to fuel consumption of vehicles while idling and the time consumption of drivers while loading/unloading packages, is small relative to the en route cost. 3) In fact, the estimates of c_c and c_t in Section 3.4 already takes into account different states (i.e., traveling and idling) of both vehicles and drivers. Subsequently, the total cost of one dispatch is given by:

$$c(z) = \Phi_c c_c \frac{1}{\sqrt{z}} + \Psi_c c_c + \Phi_t c_t \sqrt{z} + \Psi_t c_t, \tag{3.10}$$

which immediately implies the optimal number of zones and the minimum cost, respectively:

$$z^* = \frac{\Phi_c c_c}{\Phi_t c_t}, \tag{3.11a}$$

$$c(z^*) = 2\sqrt{\Phi_c c_c \Phi_t c_t} + \Psi_c c_c + \Psi_t c_t. \tag{3.11b}$$

The above closed-form quantities lead to several observations on systems configuration: 1) Both z^* and $c(z^*)$ are proportional to the service region area a, as expected. 2) Scale economy of household demand does exist, but in a weak way. Specifically, when demand scales up, $\sqrt{\Phi_c \Phi_t}$ and Ψ_c in Eqn. (3.11b) increase by less than the proportional increase in both demand density n and intensity g, but Ψ_t is proportional to n and g. Therefore, when demand becomes extremely dense and intense, the truck mode component $\Psi_t c_t$ dominates other terms in (3.11b) and the scale economy tapers off. 3) Similarly, the discussion as to whether to adopt shared mobility is most meaningful when demand is moderate such that the car-related terms in (3.11b) are significant (however, this statement may not hold when more general performance considerations are taken into account, such as agility and social trust building, which are beyond the scope of this work). 4) Finally, increasing the density of available shared mobility m is not an effective means to directly reduce the total cost, as m only appears in the denominator of the second term of Φ_c and has much diminished marginal value. Nevertheless, increasing the supply of shared mobility may indirectly reduce the total cost by lowering the price of delivery services. (3.11a) and (3.11b) also result in economic and environmental analysis, which is the focus of Section 3.4.

Benchmark: the Truck-only Scenario

The benchmark for evaluating the logistics system with shared mobility is a scenario where only trucks fulfill home delivery services. Following the approach in [24], the following VRP CA model approximates the total truck travel length in this scenario:

$$d_b = \frac{2\sqrt{2}}{3}\sqrt{a}\frac{ang}{v_t} + \frac{2}{\sqrt{6}}a\sqrt{n}, \tag{3.12}$$

where the first term on the right-hand side approximates the total line-haul trip length and the second term approximates the detour length. Hence, the total cost is $c_t d_b$.

3.4 Analysis

This section begins with parameter estimates for the logistics planning problem in Section 3.3. The first part of analysis is a comparison between logistics system designs with two different objectives: to minimize operating costs and to minimize GHG emissions, within the shared-mobility scenario. The second part of analysis is a comparison between the shared-mobility scenario and the benchmark truck-only scenario. These comparisons lead to discussion on possible government incentives. The section ends with an extended discussion on designing an efficient car driver payment scheme.

Parameter Estimates

Cost parameters: Cost parameter values are calibrated based on empirical evidence. To consider both the economic and the environmental objectives, the total cost refers to either operating costs or GHG emissions, denoted by c^o and c^e, respectively:

1. Operating costs consist of vehicle costs and driver wages:

    a) The vehicle costs break down into components of fuel consumption, maintenance, and distance-related depreciation. On September 14, 2015, the average prices of diesel and gasoline in the U.S. were p_t = \$0.664 L^{-1} and p_c = \$0.663 L^{-1}, respectively (http://www.eia.gov/petroleum/gasdiesel/). For trucks, according to [56], the fuel efficiency of a UPS delivery van (Freightliner P70) is 10.6 miles per gallon of diesel, or f_t = 0.222 L/km. The truck fuel cost estimate is thus 0.664×0.222 = \$0.147 km^{-1}. From the same report, the maintenance cost is \$0.130 mile^{-1}, or \$0.081 km^{-1}. From [9], distance-related depreciation cost of a van under city driving conditions is \$0.081 mile^{-1}, or \$0.050 km^{-1}. Collectively, adjusting for annual inflation of 2.5%, the 2015 vehicle cost of a truck is $o_t = f_t p_t + u_t + l_t = 0.147 + 0.081 \times 1.025^3 + 0.050 \times 1.025^{12}$ = \$0.301 km^{-1}, where u and l denote per-km costs of maintenance and depreciation, respectively.

For cars, the average fuel efficiency of passenger cars in the U.S. is 23.4 miles per gallon of gasoline according to the [97]. Since a car for delivery services in an urban area typically has more frequent stops than under normal operating conditions, estimate the fuel efficiency to be 18 miles per gallon, or $f_c = 0.157$ L/km. The car fuel cost is thus $0.663 \times 0.157 = \$0.104$ km^{-1}. [9] estimate the maintenance cost of a passenger car to be \$0.047 mile^{-1}, or \$0.029 km^{-1}, and the depreciation cost to be \$0.074 mile^{-1}, or \$0.046 km^{-1}. Collectively, the 2015 vehicle cost of a car is $o_c = f_c p_c + u_c + l_c = 0.104 + 0.029 \times 1.025^{12} + 0.046 \times 1.025^{12} = \$0.205$ km^{-1}.

b) To consider driver wages, a truck driver's wage is $b_t = \$30$ hour^{-1} as [9] assume. In the benchmark truck-only scenario, the average speed of a UPS delivery van is 20.8 miles per hour, or $s_t = 33.5$ km/hour, according to [56]. The per-km wage is thus $w_t = 30/33.5 = \$0.896$ km^{-1}. In the shared-mobility scenario, trucks do not stop at demand destinations. Hence, assume the average speed to be 29.9 miles per hour, or $s_t = 48.3$ km/hour, which is the simple average of driving speeds in 50 U.S. cities (http://infinitemonkeycorps.net/projects/cityspeed/). The wage is thus $w_t = 30/48.3 = \$0.621$ km^{-1}.

For car drivers, they participate in delivery services only if the expected payout is no smaller than can be earned otherwise. Therefore, assume the car driver wage to be $b_c = \$19.0$ hour^{-1}, which is the average earnings per hour of Uber drivers according to [45]. Also assume that the average speed of cars for delivery services is 18 miles per hour, or $s_c = 29.0$ km/hour, which is smaller than the aforementioned average speed of a delivery van, considering that the relatively slow last-mile delivery accounts for a greater share in car trips than in truck trips. The car driver wage is thus $w_c = 19.0/29.0 = \$0.655$ km^{-1}.

c) Combining vehicle costs and driver wages, the per-km operating costs for trucks are $c_t^o = o_t + w_t = \$1.20$ km^{-1} in the truck-only scenario and \$0.922 km^{-1} in the shared-mobility scenario. The per-km operating cost for cars is $c_c^o = o_c + w_c = \$0.860$ km^{-1}.

2. When the cost refers to GHG emissions, the per-km estimates are $c_t^e = e_t f_t = 0.597$ kg/km for trucks and $c_c^e = e_c f_c = 0.369$ kg/km for cars, where [35] estimates that CO2 emissions are $10,180$ grams from a gallon of diesel (i.e., $e_t = 2.69$ kg/L) and $8,887$ grams from a gallon of gasoline (i.e., $e_c = 2.35$ kg/L).

System parameters: [96] reports that a similar P70 delivery van has capacity of up to 20.8 m^3 with 110 kg m^{-3}, which amounts to $2,288$ kg. Hence, let truck capacity $v_t = 2,000$ kg. For cars, interior space is not a realistic estimate of the loading capacity for home delivery services. Instead, suppose the car capacity is $v_c = g \cdot 15 = 150$ kg, where $g = 10$ kg is assumed to be the mean weight of goods demanded at each designation. This value of g is smaller than the estimate of 18 kg in [17] for the amount of goods that the average

Table 3.1: Parameter estimates.

Logistic setting	Region area a 100 km^2	Demand density n 100 km^{-2}	Mean demand g 10 kg	Car density m 50 km^{-2}
Aggregate constants	Φ_c 2.42×10^3	Ψ_c 1.04×10^3	Φ_t, Φ_h 7.07, 7.07	Ψ_t, Ψ_h 471, 157
Per-km costs	Operating cost $c^o = o + w$	Vehicle cost $o = fp + u + l$	Driver wages $w = b/s$	GHG emissions $c^e = fe$
Trucks	\$0.923 (1.20)†	\$0.302	\$0.621 (0.896)	0.597 kg
Heavy-duty	\$1.02 (1.30)	\$0.402	\$0.621 (0.896)	1.00 kg
Cars	\$0.860	\$0.205	\$0.655	0.369 kg
Vehicle specifications	Fuel efficiency f	Emissions rate e	Load capacity v	Average speed s
Trucks	0.222 L/km	2.69 kg/L	2,000 kg	48.3 (33.5) km/hr
Heavy-duty	0.373 L/km	2.69 kg/L	6,000 kg	48.3 (33.5) km/hr
Cars	0.157 L/km	2.35 kg/L	150 kg	29.0 km/hr
Basic costs	Fuel price p	Maintenance u	Depreciation l	Hourly wages b
Trucks	\$0.664 L^{-1}	\$0.0872 km^{-1}	\$0.0672 km^{-1}	\$30 hr^{-1}
Cars	\$0.663 L^{-1}	\$0.0390 km^{-1}	\$0.0619 km^{-1}	\$19 hr^{-1}

† Parameter values for the benchmark truck-only scenario are either in parentheses or the same as for the shared-mobility scenario.

consumer carries with each shopping trip. Additionally, assume that the service region area $a = 100$ m^2, the demand density $n = 100$ km^{-2} and the density of available cars $m = 50$ km^{-2}. Substituting these numbers into formulas (3.7) and (3.9) yields the values of Φ_c, Ψ_c, Φ_t and Ψ_t, respectively.

For analysis later in this section, also consider a rare scenario where heavy-duty trucks are in place of P70 delivery vans. From [16], the Freightliner M2-106 has load capacity of about 6,000 kg and fuel efficiency of 550 g/mile (6.3 miles per gallon). Repeating the above procedure yields other parameter estimates for this heavy-duty model. Their symbols are subscripted by "h". Table 3.1 summarizes the parameter estimates. Parameters that most significantly impact the logistics system design are analyzed later in this section.

Minimizing Operating Costs vs. Minimizing Emissions

Suppose that the service region adopts shared mobility for home delivery services. To what extent does the design that minimizes operating costs compromise the goal of GHG emissions abatement? It turns out that *this logistics system has remarkably small tension between these two preferences*: Minimizing operating costs while ignoring GHG emissions results in a service zone deployment that increases emissions from the minimum level by a negligible

Table 3.2: Ratios of the number of service zones, operating costs and GHG emissions in the operating costs minimizing system to those in the emissions minimizing system.

Scenarios	Zone number z_1/z_2	Operating costs c_1^o/c_2^o	Emissions c_1^e/c_2^e
Baseline†	1.51	0.997	1.00
Demand density doubles ($n \to 2n$)	1.51	0.997	1.00
Trucks are heavy-duty	2.28	0.985	1.02
Car fuel efficiency doubles ($f_c \to \frac{1}{2} f_c$)	2.82	0.980	1.02
Car driver wage doubles ($w_c \to 2w_c$)	2.66	0.984	1.02
Car driver wage is halved ($w_c \to \frac{1}{2} w_c$)	0.934	1.00	1.00

† In the baseline scenario, $z_1 = 319$, $c_1^o = \$1.56 \times 10^3$ and $c_1^e = 790$ kg.

percentage (e.g., 0.25% in the baseline scenario); this compatibility of objectives is robust with a wide range of parameter settings. Nonetheless, the optimal number of service zones can significantly vary when the objective switches. Table 3.2 shows the ratios of the number of service zones, operating costs and GHG emissions in the system that minimizes operating costs to those in the system that minimizes emissions, in the baseline and alternative scenarios. Symbols for these two systems are subscripted by "1" and "2", respectively.

To understand the large ratio (i.e., significantly greater than 1) of z_1/z_2 in those scenarios, notice from Eqn. (3.11a) that the optimal number of zones is proportional to $\frac{c_c}{c_t}$, the per-km cost ratio of moving a car to moving a truck. This ratio is larger when the cost refers to operating costs than when the cost refers to emissions, i.e., $\frac{c_c^o}{c_t^o} > \frac{c_c^e}{c_t^e}$, or $z_1/z_2 = \frac{c_c^o}{c_t^o} / \frac{c_c^e}{c_t^e} = \frac{f_c p_c + u_c + l_c + w_c}{f_t p_t + u_t + l_t + w_t} / \frac{f_c e_c}{f_t e_t} > 1$. In fact, the fuel-related ratios $\frac{f_c p_c}{f_t p_t} \approx \frac{f_c e_c}{f_t e_t} \approx \frac{f_c}{f_t} < 1$, but the ratio of the non-fuel portion $\frac{u_c + l_c + w_c}{u_t + l_t + w_t} > \frac{f_c}{f_t}$. Therefore, whether the service zones are denser in the operating cost minimizing design than in the emissions minimizing design is irrelevant to the fact that cars have higher per-km fuel efficiency than trucks (given $e_c \approx e_t$ and $p_c \approx p_t$), but hinges on whether the fuel consumption is more efficient than the non-fuel cost when comparing cars with trucks. Put another way, if the car driver wage w_c is low enough so that $\frac{u_c + l_c + w_c}{u_t + l_t + w_t} < \frac{f_c}{f_t}$ (the last scenario shown in Table 3.2), then the operating cost minimizing design chooses sparser service zones to exploit more car trips because cars now have higher operating cost efficiency than emissions efficiency relative to those of trucks.

Minimizing operating costs and minimizing emissions are compatible and robust because of two structural properties. First, the service zone number z related terms in the total cost model (3.10), $\Phi_c c_c \frac{1}{\sqrt{z}} + \Phi_t c_t \sqrt{z}$, constitute a remarkably flat function of z in a large neighborhood of z^*. Such structure is common to transshipment systems in other contexts (see detailed discussion in [24] and [17]). Specifically, the emissions inefficiency in the operating

cost minimizing design relative to the minimum emissions level is bounded as follows:

$$\begin{aligned} \frac{c_1^e - c_2^e}{c_2^e} &= \frac{(\Phi_c c_c^e \frac{1}{\sqrt{z_1}} + \Phi_t c_t^e \sqrt{z_1}) - 2\sqrt{\Phi_c c_c^e \Phi_t c_t^e}}{2\sqrt{\Phi_c c_c^e \Phi_t c_t^e} + \Psi_c c_c^e + \Psi_t c_t^e} \\ &< \frac{(\Phi_c c_c^e \frac{1}{\sqrt{z_1}} + \Phi_t c_t^e \sqrt{z_1}) - 2\sqrt{\Phi_c c_c^e \Phi_t c_t^e}}{2\sqrt{\Phi_c c_c^e \Phi_t c_t^e}} \\ &= \frac{1}{2}\left(\frac{\sqrt{c_c^e/c_c^o}}{\sqrt{c_t^e/c_t^o}} + \frac{\sqrt{c_t^e/c_t^o}}{\sqrt{c_c^e/c_c^o}}\right) - 1, \end{aligned} \tag{3.13}$$

in which $z_1 = \sqrt{\frac{\Phi_c c_c^o}{\Phi_t c_t^o}}$. This upper bound depends only on the discrepancy in the ratio of per-km emissions to per-km operating costs between cars and trucks (i.e., c_c^e/c_c^o and c_t^e/c_t^o), rather than on these two ratio values themselves or on logistic setting parameters (e.g., n and a). Moreover, this upper bound is small and robust. In the baseline scenario, $c_c^e/c_c^o = 0.429$ and $c_t^e/c_t^o = 0.647$, resulting in an upper bound of only $\frac{1}{2}(\frac{\sqrt{0.429}}{\sqrt{0.647}} + \frac{\sqrt{0.647}}{\sqrt{0.429}}) - 1 = 2.17\%$. Significantly enlarging the discrepancy between c_c^e/c_c^o and c_t^e/c_t^o (e.g., by using heavy-duty trucks, doubling car fuel efficiency or doubling car driver wages as shown in Table 3.2) can only slightly increase this upper bound.

Second, the objectives compatibility is further enhanced by the fact that z-independent terms in (3.10), $\Psi_c c_c$ and $\Psi_t c_t$, dominate the total cost. In the baseline scenario, for example, these two components account for 85.1% of the total operating costs and 84.4% of the total emissions. Consequently, emissions inefficiency is even much smaller than its upper bound in (3.13), as the inequality results from dropping $\Psi_c c_c + \Psi_t c_t$ from the denominator. A closer look into formula (3.7b) shows that the dominant component of Ψ_c is $\lambda a\sqrt{n}$, which represents the total detour trip length of cars; formula (3.9b) shows that $\Psi_c = \frac{ang}{v_t}\frac{2\sqrt{2a}}{3}$ represents the total line-haul trip length of trucks. These two parts account for a large share of routes because the downstream mode of transport in a transshipment system tends to be more engaged in the last-mile delivery to exploit its cost-efficiency in transporting small loads, whereas the upstream mode of transport tends be more engaged and efficient in bulk trucking. Therefore, adjusting service zone deployment only affects parts of the routes, namely the line-haul portion of car trips and the detour portion of truck trips. In contrast, this insensitivity is not the case in [17], where the density of retail stores significantly affects trip lengths of both the retail truck and shoppers' cars.

Shared Mobility vs. Conventional Mobility

The objectives compatibility identified above makes it legitimate to use the operating costs minimizing design to address a more fundamental question: Is it worthwhile to transition to the shared-mobility business model for home delivery services? In particular, would adopting shared mobility be both "asset-light" and "with low overheads," as [13] claims?

Table 3.3 lists the increases in operating costs, GHG emissions and total vehicle trip length due to transition from the benchmark truck-only mode to the shared-mobility mode, in the baseline and alternative scenarios. The table and the following analysis show that,

Table 3.3: Increases (by percent) in operating costs, GHG emissions and trip length due to transition from the benchmark truck-only mode to the shared-mobility mode.

Scenarios	Operating costs $\frac{(c_t^o - c_b^o)}{c_b^o}$	Emissions $\frac{(c_t^e - c_b^e)}{c_b^e}$	Trip length $\frac{(d_c + d_t - d_b)}{d_b}$
Baseline	2.73	3.95	39.5
Trucks are heavy-duty†	−14.1	−6.87	15.2
Demand density doubles ($n \to 2n$)	0.475	4.61	35.9
Demand density is halved ($n \to \frac{1}{2}n$)	4.71	3.27	42.8
Car density doubles ($m \to 2m$)	1.76	2.92	38.2
Car density is halved ($m \to \frac{1}{2}m$)	4.01	5.30	41.3
Car fuel efficiency doubles ($f_c \to \frac{1}{2}f_c$)	−1.29	−26.1	39.6
Truck fuel efficiency doubles ($f_c \to \frac{1}{2}f_c$)	6.35	61.0	39.5
Car driver wage doubles ($w_c \to 2w_c$)	52.3	3.95	40.2
Car driver wage is halved ($w_c \to \frac{1}{2}w_c$)	−22.8	3.95	40.4
Car capacity increases ($v_c/g \to 20$)	−0.642	0.438	35.0
Car capacity decreases ($v_c/g \to 10$)	8.55	10.0	47.4

† Scenarios where shared mobility is likely to be favorable are in boldface.

under normal operating conditions, *shared mobility is likely to have either of those two merits, but not both, unless additional policy instruments exist.*

(a) The baseline scenario represents an "asset-light" scenario, in which the final stage of delivery is directly outsourced to shared mobility with no major investments in vehicle fleets and other assets. However, Table 3.3 shows that adopting shared mobility in the baseline scenario increases both operating costs (by 2.73%) and emissions (by 3.95%). Having not been fully identified in literature, this potential cost inefficiency can be a major explanation for the sharing logistics paradigm not having taken place on a large scale. In fact, the efficiency gain of using cars is twofold: their open-loop routes are in general shorter than conventional closed-loop routes for cars of the same capacity, and their per-km operating costs and emissions are lower than those of trucks. However, these efficiency gains are overly offset in the baseline scenario (as well as other non-boldfaced scenarios in Table 3.3) by a demerit: cars are of much smaller loading capacity than trucks and consequently the share-mobility logistics system incurs over 35% longer total trip distance than the truck-only system.

(b) If heavy-duty trucks replace regular-size delivery vans as the upstream mode of transport, Table 3.3 shows that adopting shared mobility can reduce operating costs by 14.1% and reduce emissions by 6.87%. This scenario is thus "with low overheads." However, it also typically implies a potentially large investment in fleet assets.

From the total cost expression (3.11b), heavy-duty trucks are able to save cost in a shared

mobility logistics system if and only if the following inequality holds:

$$2\sqrt{\Phi_c c_c \Phi_h c_h} + \Psi_c c_c + \Psi_h c_h < 2\sqrt{\Phi_c c_c \Phi_t c_t} + \Psi_c c_c + \Psi_t c_t. \tag{3.14}$$

Rearranging terms and substituting $\Phi_t c_t = \Phi_h c_h = \frac{\sqrt{2a}}{2}$ into the above inequality yields:

$$\frac{\Psi_t c_t - \Psi_h c_h}{\sqrt{c_h} - \sqrt{c_t}} > \sqrt{2a\Phi_c c_c}. \tag{3.15}$$

This condition requires that $\Psi_h = \frac{ang}{v_h}\frac{2\sqrt{2a}}{3}$ should be small enough such that the lefthand side ratio is greater than the righthand side threshold. In other words, the load capacity of heavy-duty trucks (v_h) must be high enough to overcome the per-km cost inefficiency (i.e., $c_h > c_t$). Meanwhile, the righthand side threshold is increasing in $\sqrt{\Phi_c c_c}$. Intuitively, greater value of $\Phi_c c_c$ leads to denser service zones deployment and thus requires higher bulk-trucking cost efficiency to offset the inefficiency that prolonged truck detour incurs.

(c) Another two scenarios in favor of adopting shared mobility are where car fuel efficiency is high and where car driver wage is low, both implying smaller value of c_c. However, these scenarios are unlikely to occur in the near future, unless policy instruments are in place. For example, a government may launch subsidy incentives to passenger car drivers for fulfilling delivery services and/or mandate stringent fuel-efficiency standards to those cars. Nonetheless, to justify those policy instruments requires justifying other social benefits that shared mobility can potentially create, such as social trust building or unemployment rate reduction, which are beyond the scope of this work.

The above discussion collectively suggests that the value of shared mobility is not direct savings of operating costs or GHG emissions, but rather the potential it creates for further cost reduction by means of additional fleet asset investment and/or policy instruments.

Payment Scheme

The last part of the analysis provides logistics services providers with a scheme of payment to car drivers to induce sufficient supply of shared mobility for home delivery services. Preceding analysis assumes simple hourly wages for analytical convenience. However, a car driver is willing to participate in a home delivery service only if the payment is at least the amount that he or she can otherwise expect to earn by providing ride share to passengers. Therefore, the cost-efficient payment should be based on the expected revenue out of ride share services, as derived as follows–

Packages uploaded at each terminal are to be delivered along open-loop car routes, which are offered to nearby idling car drivers. Consider an offer of payment $w(d_o)$ that entails d_o km of outbound trip distance. If a car driver who is r km away from the terminal decides to accept this offer, the estimated total time of fulfilling the delivery service is $x = \frac{r}{s} + \frac{d_o}{s_c}$, where s denotes the normal cruise speed of a car and s_c denotes the speed during delivery services ($s_c < s$). Alternatively, if the driver is engaged in ride sharing services during the time window $(0, x)$, then the driver is either carrying passengers or waiting for a new passenger.

When carrying passengers, the driver receives payments from a ride-share company such as Uber. Typically, the ride fare consists of three parts: base fare, per-mile fare and per-minute fare. A passenger ride of duration τ thus generates income of $w_b + w_m\tau$, where w_b is the base fare and w_m is in terms of dollars per minute but incorporates both the per-minute fare and the per-km fare, since the travel time is assumed to be proportional to travel distance (this is valid for a traffic network with no congestion). Also assume that the waiting time and passenger ride time are exponentially distributed with means μ and υ, respectively. The driver's expected income $w_p(x)$ over $(0, x)$ is given by (the proof is available in the Appendix):

$$w_p(x) = \frac{\mu(w_b + w_m/\upsilon)}{\mu + \upsilon}\Big(\upsilon x + \frac{\mu}{\mu + \upsilon}(1 - \exp(-(\mu + \upsilon)x))\Big). \tag{3.16}$$

A driver's choice between delivering packages and waiting for ride requests boils down to comparing $w_p(x)$ and $w(d_o)$. Since $x = \frac{r}{s} + \frac{d_o}{s_c}$, a driver will accept the offer if $w(d_o) \geq w_p(\frac{r}{s} + \frac{d_o}{s_c})$. Therefore, for each pair of (r, d_o), the minimum payment is $w_p(\frac{r}{s} + \frac{d_o}{s_c})$.

Depending on practical settings, two different payment schemes can be implemented. The first scheme sends offers sequentially to the nearest individual drivers. Each offer with outbound trip distance d_o is with a payment $w_p(\frac{r}{s} + \frac{d_o}{s_c})$, assuming that a driver's location in terms of r is known to the logistics services provider. In contrast, the second scheme does not differentiate drivers based on their locations and broadcasts offers simultaneously to nearby drivers. Each offer is independent from r with a payment $w_p(\frac{r_c}{s} + \frac{d_o}{s_c})$, where r_c is the critical range to design. Drivers accept the offer if being within r_c km from the terminal, or reject the offer otherwise. Given r_c, the total number of drivers attracted to the terminal is $m\int_0^{r_c} 8r dr = 4m(r_c)^2$. To meet the demand for $m_z = \frac{ang}{v_c z}$ drivers requires $4m(r_c)^2 \geq m_z$. Thus the minimal payment is determined by solving the following problem:

$$w^*(d_o) = \min_{r_c}\ w_p(\frac{r_c}{s} + \frac{d_o}{s_c}),\ \text{subject to } 4m(r_c)^2 \geq m_z. \tag{3.17}$$

It follows immediately that the optimal $r_c^* = \frac{1}{2}\sqrt{\frac{m_z}{m}}$ when the constraint is binding and

$$w^*(d_o) = \frac{\mu r_c^*(w_b\upsilon + w_m)}{s(\mu + \upsilon)} + \frac{\mu(w_b\upsilon + w_m)}{s_c(\mu + \upsilon)}d_o + \frac{\mu^2(w_b\upsilon + w_m)}{\upsilon(\mu + \upsilon)^2}\big(1 - \exp(-(\mu + \upsilon)\big(\frac{r_c^*}{s} + \frac{d_o}{s_c}))\big). \tag{3.18}$$

This formula specifies the payment scheme to induce sufficient drivers for delivery services in the second scenario. The first term represents the base fare for the delivery offer; the second term is proportional to d_o and the third term is a nonlinear fare as a function of d_o.

Numerical results: Based on the data from Uber official website, set $w_b = \$1.65$ and $w_m = \$40.85$ hour^{-1}, considering that Uber takes 25% commissions (https://www.uber.com/cities/san-francisco). Table 3.4 shows $w_p(1)$, the average income of an Uber driver in one hour, with different combinations of μ and υ (which vary from city to city and time to time). As expected, increasing μ or decreasing υ results in higher driver income. The calculated range of income is consistent with the actual estimate from Hall and Krueger (2015), which varies from \$16.20 to \$30.35.

Table 3.4: Average income of Uber drivers in one hour.

(μ, υ): hour^{-1}	(3,6)	(3,5)	(4,6)	(4,5)	(5,6)	(5,5)
$w_p(1)$: \$	17.86	19.79	21.65	23.76	24.82	27.00

Table 3.5: Delivery payments with different densities of available cars.

m: km^{-1}	10	20	50	100	150	200
$w^*(d_o)$: \$	11.11	11.06	11.01	10.99	10.98	10.98

From the preceding estimates, $s = 29.9$ miles per hour and $s_c = 18.0$ miles per hour. When $v_c/g = 15$ and $n = 100$, numerical studies in Section 3.3 indicate that the average outbound trip distance is 12 km. Assuming $\mu = 4$ hour^{-1} and $\upsilon = 5$ hour^{-1}, the delivery payment for $d_o = 12$ km with different densities of available cars is listed in Table 3.5.

Table 3.5 shows that, when density m decreases, r_c^* increases and so does payment $w^*(d_o)$ to drivers (since drivers are likely to be more distant from the terminal). Nonetheless, when m decreases from 200 km^{-1} to 10 km^{-1}, the payment only increases by 1.2%. This is because the time for a driver traveling to the terminal is on average much less than the time for outbound delivery services, i.e. $\frac{r_c^*}{s} \ll \frac{d_o}{s_c}$. Thus, the m-independent delivery time is the dominant factor of driver payment. This result, along with the discussion on m at the end of Section 3.3, suggests that abundance of shared mobility is not a key driver of decisions on whether to foster sharing logistics.

3.5 Conclusion

This chapter studies the problem of logistics systems planning that integrates shared mobility for home delivery services, given the rapid development of sharing economy and retail e-commerce in recent years. The logistics setting is a one-transshipment system: a fleet of short-haul trucks are dispatched from a depot and unload packages at terminals of service zones. Passenger cars with available mobility are attracted to the terminals to pick up and deliver packages to demand destinations. Such a logistics system exploits both the efficiency of trucks in bulk-transport and the per-km efficiency of cars in hauling small loads of goods. The open-loop nature of the shared vehicle routes incurs less trip length than conventional closed-loop routes of vehicles of the same capacity. The downside of such a system, however, is the prolonged total trip length due to the small load capacity of passenger cars.

The first part of this chapter presents planning models for this sharing logistics system. Numerically solving open vehicle routing problem instances reveals the structural properties of the shared-mobility routes, which a CA problem is proposed to characterize. With the assumption of homogenous distributions of demands and available cars, the optimal service zone design is in the form of a closed-form expression of service zone density and the associated cost. The second part of the chapter analyzes the operating costs and GHG emissions implications of the sharing logistics system. Empirical estimates are provided to calibrate

the models. Comparing three systems, one that minimizes operating costs, one that minimizes emissions, and one as the benchmark that involves only trucks, this chapter discusses the economic and environmental viability of the prospective sharing mode under different logistics and policy conditions. Finally, a non-linear car driver payment scheme is provided to induce efficient supply of shared mobility.

The models and analysis lead to several findings: 1) First, contrary to the belief that adopting shared mobility for home delivery services is both "asset-light" and "with low overheads," such practice may not create immediate savings in operating costs and emissions. On the other hand, in the long run, sharing logistics is able to generate a considerable amount of economic and environmental benefit, because this new business model allows more room to optimize truck fleet assets and policy instruments. 2) Second, upon entering this sharing paradigm, even exclusively minimizing operating costs does not significantly increase emissions relative to the minimum level of emissions. This compatibility of economic and environmental preferences results from two system characteristics: i) The cost objective function is remarkably flat, being insensitive to even large discrepancy in the ratio of per-km emissions to per-km operating costs between cars and trucks. ii) The line-haul portion of truck trips and the detour portion of car trips account for a large share of total vehicle routes but are independent of service zone deployment. As a result, a system planner is able to focus on the cost minimizing logistics design. 3) Finally, the CA model that characterizes open-loop vehicle routes and the non-linear model of car driver payments can be useful for governments and logistics services providers to evaluate the routing behavior and cost-effectiveness of this logistics mode under various conditions.

This work is the first attempt to design and analyze the prospective sharing logistics system based on analytical models and empirical parameter estimates. To further investigate the impacts that this combination of sharing economy and retail e-commerce can potentially bring about, two directions can be explored. First, home delivery services can utilize shared mobility in different ways with additional complexity to analyze. For example, car drivers can provide ride sharing services to passengers along the trip of delivering packages, or the logistics setting can be with nonhomogeneous distributions of demands, available cars and traffic conditions. Second, whereas this work focuses on operating costs and GHG emissions as performance measures, it is worth considering other implications of sharing logistics, such as social trust building, service agility and unemployment rate, which may or may not further justify additional government regulation and incentives.

Appendix A

Supporting Results for Chapter 1

A.1 Analysis of Model Inaccuracy in Section 1.4

We have shown in the main text that the modeling process described in Section 1.4 incorporates major wind characteristics and leads to efficient ES-transmission planning problem formulation (1.13). However, these merits come at the cost of three sources of model inaccuracy: (i) The use of uniform distribution approximation of wind outputs at individual wind farms affects the sizing of the transmission lines that are upstream of the junction sites. (ii) The similar approximation applied to the aggregated wind outputs at the junction sites affects the sizing of the downstream transmission lines. (iii) Myopically sizing the upstream transmission lines without considering additional capacity cost downstream of the junction sites creates model suboptimality. The first source of model inaccuracy has been demonstrated in Section 1.4 to be reasonably small. In this section, we quantify and explain the second and the third components of model inaccuracy using numerical and/or theoretical analysis. The objectives of this analysis are to show that the model inaccuracy is reasonably small in most practical settings and to gain more insights into the nature of this infrastructure planning problem.

Errors of Approximating Aggregated Curtailed Wind Output at a Junction Site

We first investigate the approximation errors of the aggregated wind outputs and show that the errors are reasonably small under practical parameter settings. The numerical example in Section 1.4 has demonstrated that approximating wind output by uniform distribution incurs reasonably small cost error in the single-farm scenario. In the multiple-farm scenario, however, even if uniformly distributed, the wind outputs at the individual farms get first curtailed by the ES-free transmission lines and then aggregated, resulting in non-uniform wind outputs faced by the junction sites. In order to characterize those aggregated wind outputs as closed-form expressions (1.12a) and (1.12b), we applied uniform distribution

approximation again - first to the individually curtailed and then to the aggregated outputs.

To investigate this approximation error, we consider the following scenario: Three wind farms generate power flows that are correlated, have the same probability distribution and travel the same distance before getting aggregated at a junction site. This simplified setting is adequate for us to observe the effect of energy curtailment and aggregation. In the meantime, it minimizes the number of parameters to perturb. Given wind power data, each instance of the numerical test can be fully characterized by three parameters: the downstream transmission distance, the correlation coefficient between the wind outputs and the single value of upstream curtailment factor η as defined in (1.9), which has one-one correspondence to the upstream transmission distance.

Figure A.1 (a) shows the distribution of the real wind output at a single farm. The spike of the high power results from the rated power limit of wind turbines. The outputs at the other two farms are generated by re-sampling from the same data to have the same probability distribution and the specified correlation. Figures A.1 (b)-(f) are the histograms of the distributions of the aggregated wind outputs (in dark bars) and their uniform distribution approximations (in light bars) based on (1.12a) and (1.12b). We observe that, as expected, either decreasing the inter-farm correlation or increasing the upstream curtailment results in more non-uniform profile of the aggregated output distribution. Notice that the uniform distribution may poorly approximate the real profile if the upstream curtailment is extremely high, as represented by Figure A.1 (f).

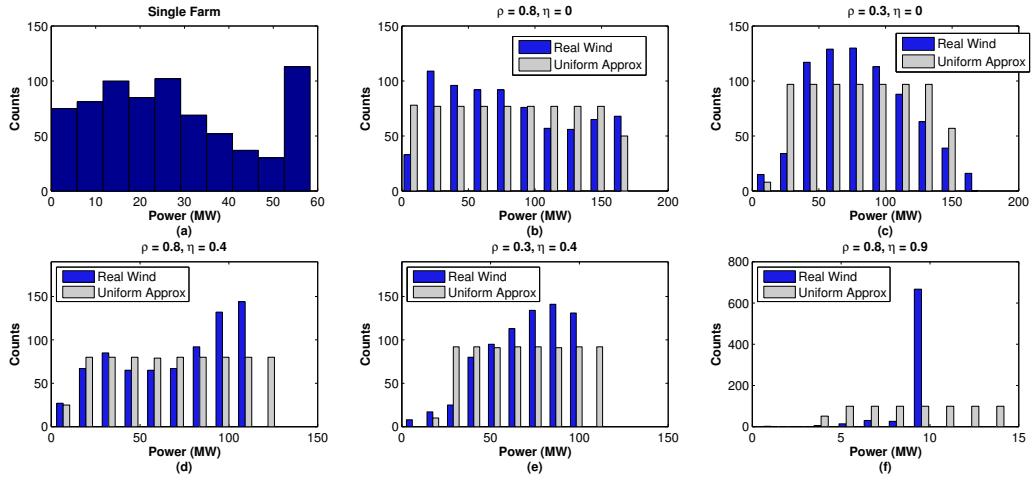

Figure A.1: Aggregated wind outputs with different degrees of correlation and curtailment.

We next define metrics of the approximation error. We first evaluate the downstream variable cost, which is the sum of the transmission capacity cost and the friction (curtailment) cost that are downstream of the ES-coupled (ES-free) junction site, in three different ways: (i) The real cost is generated with real wind output data and with downstream transmission capacity prescribed by (1.4) ((1.9)). (ii) The actual optimal cost is generated with the same

real data and with the downstream transmission capacity found by line-search that leads to the minimum downstream variable cost. (iii) The approximated cost is given by equation (1.10) ((1.5)). Then we use the following two error metrics: (i) The real error refers to the relative error between the real cost and the actual optimal cost; (ii) and the model error refers to the relative error between the approximated cost and the actual optimal cost. These two metrics, from the cost perspective, are consistent with most of error analysis in the main text. They also isolate the approximation error in the sense that the inaccuracy due to myopically sizing the upstream transmission lines becomes irrelevant, since the upstream transmission capacity is fixed in each instance as given by (1.9).

For the case where the junction site is co-located with ES, Figure A.2 shows the cost errors of 300 numerical test instances with different parameter values. One can see that the real cost errors are less than 3% in almost all the instances, which is a strong indication that the prescribed downstream transmission line capacity is near-optimal. When the junction-load distance is long and the degree of the upstream curtailment is high, the real error tends to be zero, since both the prescribed and the actual optimal line capacity tends to be the average wind power. The model errors are in general greater than the real errors, but are still contained within $\pm 10\%$ in most instances.

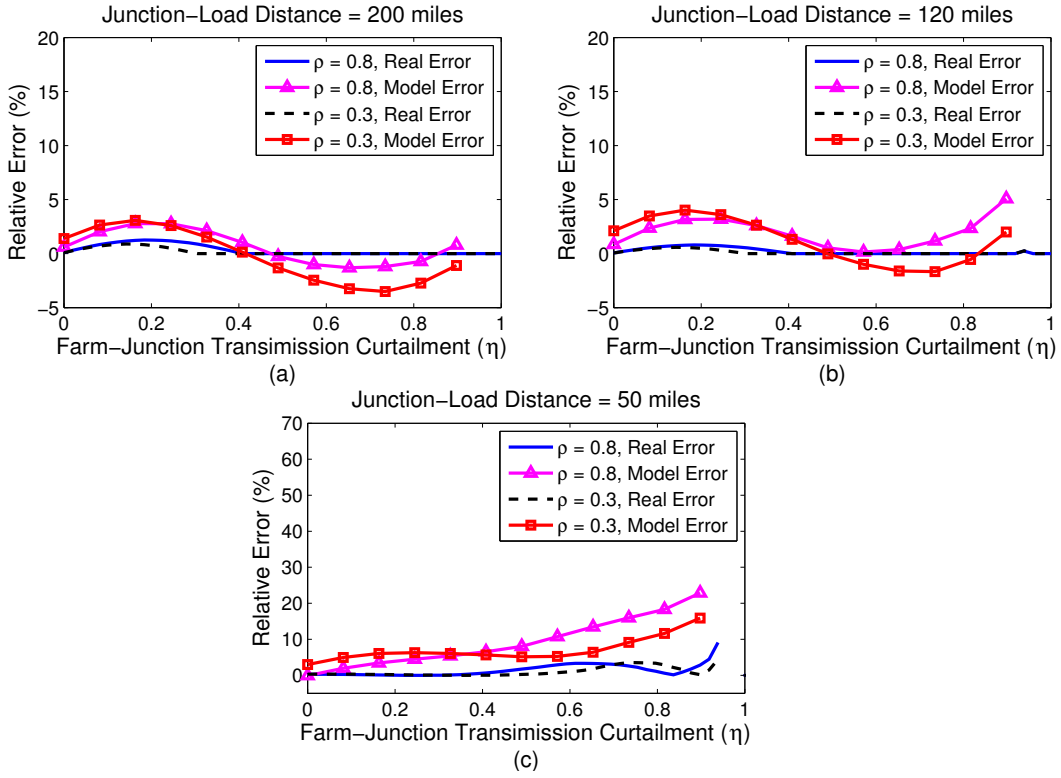

Figure A.2: Real and model errors of uniform distribution approximation with an ES-coupled junction site.

For the case where the junction site has no ES, Figure A.3 shows the cost errors of another 300 numerical test instances, from which we make the following observations: (i) Overall, the cost errors are greater than their counterparts in the ES-coupled case. This is because the absence of ES leads to the choice of higher downstream transmission capacity. The uniform distribution approximation is less effective when the line capacity is closer to the maximum wind power, since the curtailment loss, as represented by the right tail of the real wind output distribution, becomes more non-uniform and thus more difficult to approximate. (ii) Instances with lower correlation between individual wind outputs have smaller real and model errors, as a result of the evener output profile as discussed previously. It suggests that spatial pooling enhances the accuracy of uniform distribution approximation by reducing the wind power variability. (iii) The approximation errors (particularly the real errors) are still moderate when the upstream curtailment $\eta < 0.5$. Although we consider the full range of values of η for demonstration purposes, in reality, however, farm-junction transmission distance should be relatively short to save the ES-free transmission line cost and prevent substantial curtailment loss. For example, η is only 0.07 when the farm-junction distance is 50 miles in our numerical test setting; the associated approximation error is very small.

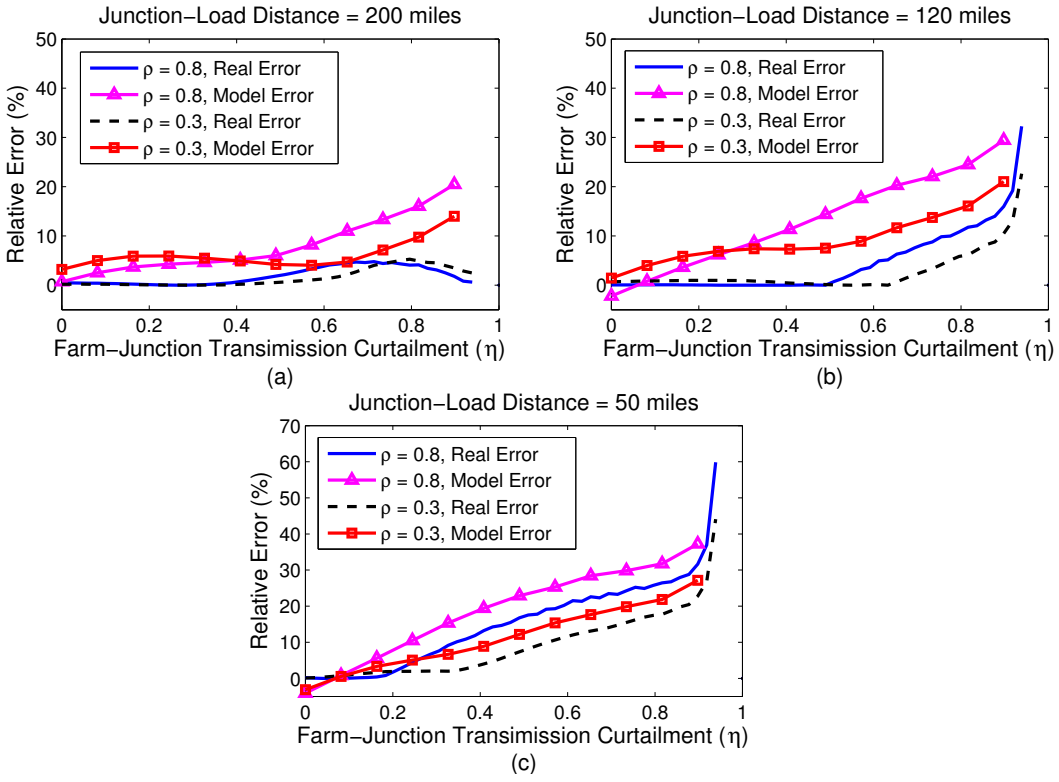

Figure A.3: Real and model errors of uniform distribution approximation with an ES-free junction site.

With the above numerical studies, we conclude that the error of approximating wind output distribution using uniform distribution can be reasonably small in practical settings. Spatial pooling of wind outputs and the deployment of ES systems both help enhance this approximation accuracy.

Sub-optimality Due to Myopically Sizing a Transmission Line Upstream of a Junction Site

Another source of model inaccuracy stems from the oversizing of the transmission lines. We next describe an analytical approach to quantifying this inaccuracy, and then use numerical experiments to justify that the inaccuracy is indeed minor.

For tractability reason, planning model (1.13) incorporates (1.9) developed in the single-farm scenario as the farm-junction transmission line capacities. These capacity quantities do not take into account the additional transmission capacity cost downstream of the junction sites, and thus tend to be larger-than-optimal. Consequently, with less curtailment loss from upstream, the transmission lines downstream of the junction sites expect to deliver more-than-optimal wind power, so their capacities prescribed by the model is no-less-than-optimal.

To analyze this sub-optimality, we use the following problem setting: Uniformly distributed power is generated from a single wind farm, curtailed and transmitted by an ES-free line down to an ES-coupled (ES-free) junction site that is l_u miles away, and then flows another $(l - l_u)$ miles through an economic (ES-free) line to get to the load center. This simplified setting not only enables us to analytically quantify the sub-optimality, but also excludes the model inaccuracy that is caused by approximating wind outputs at different stages.

The sub-optimality can be measured as the relative error between the expected real total cost and the expected optimal total cost. To compute the expected real total cost, the capacity of the transmission line upstream of the junction site is given by (1.9), while that capacity downstream the junction site is the minimizer of the expected total downstream curtailment (friction) and capacity cost. On the other hand, the expected optimal total cost is found by jointly optimizing the upstream and the downstream transmission capacities.

The ES-free Case

For the scenario where the junction site has no ES, the following proposition gives an upper bound of the sub-optimality.

Proposition 5. *In the above problem setting with a single wind farm generating power* $\mathbf{w_t} \sim$ *Unif*$(\mu - \frac{\epsilon}{2}, \mu + \frac{\epsilon}{2})$*, suppose* $\mu - \frac{\epsilon}{2} \geq 0$ *and there is no ES system at the junction site. Then an upper bound of the sub-optimality due to myopically sizing the upstream transmission line is given by*

(i) $\frac{1}{3}\frac{\eta}{(2-\eta)}$*, if* $l_u \in [0, \frac{l}{2})$*;*

(ii) $\frac{1}{4}\frac{\eta}{(2-\eta)}$*, if* $l_u \in [\frac{l}{2}, l]$*;*

where $\eta = \frac{al\delta}{p} < 1$.

Proof of Proposition 5: First consider the scenario of optimally sizing the transmission lines upstream and downstream of the junction site. Obviously, it is never optimal to assign different capacities to these two segments. Therefore, the problem is reduced to the single-farm case without ES in Section 1.4, where a single wind farm delivers power through a single line to the load center. The expected optimal total cost is given by (1.10), which we explicitly rewrite here:

$$v^* = al\mu + al(\frac{1}{2} - \frac{1}{2}\frac{\delta al}{p})\epsilon. \tag{A.1}$$

As to the case of myopically sizing the upstream transmission line, its capacity and the expected upstream cost are given by (1.9) and (1.10), respectively, as follows:

$$C_u = \mu + (\frac{1}{2} - \eta_u)\epsilon, \tag{A.2a}$$

$$v_u = al_u\mu + al_u(\frac{1}{2} - \frac{1}{2}\eta_u)\epsilon, \tag{A.2b}$$

where $\eta_u = \frac{\delta al_u}{p}$ indicates the degree of curtailment of the upstream line. Consequently, the junction site faces curtailed wind output $\check{\mathbf{w}}_\mathbf{t}$ that has the following cumulative density function:

$$F_{\check{\mathbf{w}}_\mathbf{t}}(w) = \begin{cases} 0, & \text{if } w < \mu - \epsilon/2; \\ \frac{w-(\mu-\epsilon/2)}{\epsilon}, & \text{if } \mu - \epsilon/2 \le w < C_u; \\ 1, & \text{if } w = C_u. \end{cases} \tag{A.3}$$

Choosing the quantity of downstream line capacity with uncertain wind power is a newsvendor type of problem. The overage cost is the per-unit capacity cost $\delta a(l - l_u)$ and the underage cost is the per-unit profit $p - \delta a(l - l_u)$. Therefore, the newsvendor critical fractile is $\frac{p-\delta a(l-l_u)}{p}$. Note that $F_{\check{\mathbf{w}}_\mathbf{t}}(\cdot)$ is not left-continuous at $w = C_u = \mu + (\frac{1}{2} - \eta_u)\epsilon$, since $F_{\check{\mathbf{w}}_\mathbf{t}}(C_u^-) = 1 - \eta_u < F_{\check{\mathbf{w}}_\mathbf{t}}(C_u) = 1$. This jump discontinuity leads to the following two cases:

Case (i): $l_u < \frac{1}{2}l$. This is equivalent to the critical fractile $\frac{p-\delta a(l-l_u)}{p} < 1 - \frac{\delta al_u}{p} = 1 - \eta_u$. Then the downstream transmission capacity is given by:

$$\begin{aligned} C_d &= F^{-1}_{\check{\mathbf{w}}_\mathbf{t}}(\tfrac{p-\delta a(l-l_u)}{p}) \\ &= \mu - \epsilon/2 + \tfrac{p-\delta a(l-l_u)}{p}\epsilon \\ &= \mu + (\tfrac{1}{2} - \tfrac{a(l-l_u)\delta}{p})\epsilon. \end{aligned} \tag{A.4}$$

The expected downstream curtailment loss is calculated as follows:

$$\begin{aligned} \mathbb{E}[\mathbf{l}_d] &= \textstyle\int_{C_d}^{C_u}(w_t - C_d)\frac{1}{\epsilon}dw_t + \mathbb{P}(\mathbf{w_t} \ge C_u)(C_u - C_d) \\ &= \textstyle\int_{\mu+(\frac{1}{2}-\frac{a(l-l_u)\delta}{p})\epsilon}^{\mu+(\frac{1}{2}-\frac{al_u\delta}{p})\epsilon}(w_t - \mu - (\frac{1}{2} - \frac{a(l-l_u)\delta}{p})\epsilon)\frac{1}{\epsilon}dw_t \\ &+ \tfrac{al_u\delta}{p}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - \mu - (\tfrac{1}{2} - \tfrac{a(l-l_u)\delta}{p})\epsilon) \\ &= \tfrac{\epsilon}{2}(\tfrac{a\delta}{p})^2(l^2 - 2l_ul). \end{aligned} \tag{A.5}$$

Adding the cost terms together yields the expected real total cost:

$$\begin{aligned} v &= v_u + \tfrac{p}{\delta}\mathbb{E}[\mathbf{l}_d] + C_d a(l - l_u) \\ &= al\mu + [\tfrac{al}{2} - \tfrac{a^2\delta}{2p}(3l_u^2 - 2l_u l + l^2)]\epsilon. \end{aligned} \tag{A.6}$$

Comparing (A.1) and (A.6), the absolute sub-optimality is bounded as follows:

$$\begin{aligned} v - v^* &= [\tfrac{al}{2} - \tfrac{a^2\delta}{2p}(3l_u^2 - 2l_u l + l^2)]\epsilon - al(\tfrac{1}{2} - \tfrac{1}{2}\tfrac{\delta al}{p})\epsilon \\ &= \tfrac{\delta a^2}{2p}\epsilon(2l_u l - 3l_u^2) \\ &\leq \tfrac{\delta a^2}{6p}\epsilon l^2, \end{aligned} \tag{A.7}$$

since $(2l_u l - 3l_u^2)$ reaches its maximum, $\frac{1}{3}l^2$, when $l_u = \frac{1}{3}l$. Therefore, the relative suboptimality in case (i) is bounded as follows:

$$\begin{aligned} \tfrac{v - v^*}{v^*} &\leq \tfrac{\frac{\delta a^2}{6p}\epsilon l^2}{al\mu + al(\frac{1}{2} - \frac{1}{2}\frac{\delta al}{p})\epsilon} \\ &\leq \tfrac{\frac{\delta a^2}{6p}\epsilon l^2}{al\frac{\epsilon}{2} + al(\frac{1}{2} - \frac{1}{2}\frac{\delta al}{p})\epsilon} \quad (\text{since } \mu - \epsilon/2 \geq 0) \\ &= \tfrac{\frac{\delta al}{p}}{6(1 - \frac{\delta al}{2p})} \\ &= \tfrac{1}{3}\tfrac{\eta}{2 - \eta}, \end{aligned} \tag{A.8}$$

where $\eta = \frac{\delta al}{p}$. We thus complete the proof of Proposition 5 (i).

Case (ii): $l_u \geq \frac{1}{2}l$. This is equivalent to the critical fractile $\frac{p - \delta a(l - l_u)}{p} \geq 1 - \frac{\delta a l_u}{p} = 1 - \eta_u$. Since $F_{\check{\mathbf{w}}_\mathbf{t}}(C_d) < 1 - \eta_u \ \forall C_d < C_u$, we know that the choice of downstream line capacity $C_d = C_u = \mu + (\frac{1}{2} - \eta_u)\epsilon$. Therefore, no curtailment occurs downstream of the junction site. The expected real total cost can thus be calculated as:

$$\begin{aligned} v &= v_u + C_d a(l - l_u) \\ &= al_u\mu + al_u(\tfrac{1}{2} - \tfrac{1}{2}\eta_u)\epsilon + [\mu + (\tfrac{1}{2} - \eta_u)\epsilon]a(l - l_u) \\ &= al\mu + [al(\tfrac{1}{2} - \tfrac{\delta a l_u}{p}) + \tfrac{p}{2\delta}(\tfrac{\delta a l_u}{p})^2]\epsilon. \end{aligned} \tag{A.9}$$

Comparing (A.1) and (A.9), the absolute sub-optimality is bounded as follows:

$$\begin{aligned} v - v^* &= [al(\tfrac{1}{2} - \tfrac{\delta a l_u}{p}) + \tfrac{p}{2\delta}(\tfrac{\delta a l_u}{p})^2]\epsilon - al(\tfrac{1}{2} - \tfrac{1}{2}\tfrac{\delta al}{p})\epsilon \\ &= \tfrac{\delta a^2\epsilon}{2p}(l - l_u)^2 \\ &\leq \tfrac{\delta a^2\epsilon}{8p}l^2, \end{aligned} \tag{A.10}$$

where the last inequality is due to the fact that $(v - v^*)$ is monotonically decreasing in l_u for $l_u \in [\frac{1}{2}l, l]$ and is thus maximized when $l_u = \frac{1}{2}l$. Therefore, the relative suboptimality in case (ii) is bounded as follows:

$$\begin{aligned}\frac{v-v^*}{v^*} &\leq \frac{\frac{\delta a^2 \epsilon}{8p} l^2}{al\mu + al(\frac{1}{2} - \frac{1}{2}\frac{\delta al}{p})\epsilon} \\ &\leq \frac{\frac{\delta a^2 \epsilon}{8p} l^2}{al\frac{\epsilon}{2} + al(\frac{1}{2} - \frac{1}{2}\frac{\delta al}{p})\epsilon} \quad (\text{since } \mu - \epsilon/2 \geq 0) \\ &= \frac{\frac{\delta al}{p}}{8(1 - \frac{\delta al}{2p})} \\ &= \frac{1}{4}\frac{\eta}{2-\eta}.\end{aligned} \tag{A.11}$$

We thus complete the proof of Proposition 5 (ii). □

The ES-coupled Case

Similar analysis of sub-optimality can be applied to the scenario where the junction site is coupled with an ES system. However, complete analysis turns out to involve excessive case discussion and complicated high-order terms in the bounds expressions, revealing little insight into the problem. Therefore, below we discuss only one case, which is common and is the most representative of the problem, as defined by three conditions: (i) The downstream economic transmission line capacity is greater than the mean of the upstream curtailed wind output. Otherwise the downstream line capacity would have to be the equal to the mean upstream output, which loses the information about how the parameters affect the capacity choice. (ii) The upstream transmission distance accounts for a limited proportion of the entire transmission distance. Otherwise the downstream transmission capacity would tend to be equal to the upstream transmission capacity and the problem would be reduced to the ES-free scenario. (iii) The ES system is not severely inefficient. The conditions and the result of the analysis are formalized in the following proposition:

Proposition 6. *In the above problem setting, suppose* $\mu - \frac{\epsilon}{2} \geq 0$ *and there is an ES system at the junction site. Under the conditions (i)* $\mu + \epsilon(\frac{1}{2} - \frac{a(l-l_u)\delta}{p(1-\alpha\beta)}) \geq \mu - \frac{1}{2}\epsilon(\frac{al_u\delta}{p})^2$*, (ii)* $l_u < \frac{l}{2-\alpha\beta}$*, and (iii)* $\alpha\beta \in [0.5, 1)$*, an upper bound of the sub-optimality due to myopically sizing the upstream transmission line is given by* $\frac{1}{9}\frac{\theta}{(2-\theta)}$*, where* $\theta = \frac{al\delta}{(1-\alpha\beta)p}$*.*

Proof of Proposition 6: Proving Proposition 6 requires the closed-form expressions of the expected real total cost and the expected optimal total cost, which we summarize in the following two lemmas and prove them at the end of this proof:

Lemma 1. *Under conditions (i) and (ii) in Proposition 6, the expected real total cost is* $v = al(\mu + \epsilon/2) - \frac{a^2\delta(1-\alpha\beta)l_u^2 + 2\delta a^2(l-l_u)^2}{2p(1-\alpha\beta)}\epsilon$*.*

Lemma 2. *Under conditions (ii) and (iii) in Proposition 6, the expected optimal total cost is* $v^* = al(\mu + \epsilon/2) - \frac{a^2\delta[\alpha\beta l(l-2l_u) + l_u^2]}{2p\alpha\beta(1-\alpha\beta)}\epsilon$*.*

Given these two lemmas, the absolute sub-optimality is calculated as follows:

$$\begin{aligned} v - v^* &= al(\mu+\epsilon/2) - \tfrac{a^2\delta(1-\alpha\beta)l_u^2+2\delta a^2(l-l_u)^2}{2p(1-\alpha\beta)}\epsilon - al(\mu+\epsilon/2) + \tfrac{a^2\delta[\alpha\beta l(l-2l_u)+l_u^2]}{2p\alpha\beta(1-\alpha\beta)}\epsilon \\ &= \tfrac{a^2\delta\epsilon}{2\alpha\beta(1-\alpha\beta)p}[-\alpha\beta(1-\alpha\beta)l_u^2 - 2\alpha\beta(l-l_u)^2 + \alpha\beta l(l-2l_u) + l_u^2] \\ &= \tfrac{a^2\delta\epsilon}{2\alpha\beta(1-\alpha\beta)p}[-\alpha\beta l^2 + 2\alpha\beta l l_u + (1-3\alpha\beta+\alpha^2\beta^2)l_u^2] \\ &= \tfrac{a^2\delta\epsilon}{2\alpha\beta(1-\alpha\beta)p}[l_u^2(1-\alpha\beta)^2 - \alpha\beta(l_u-l)^2], \end{aligned} \tag{A.12}$$

which is clearly increasing in l_u when $l_u < l$. Following condition (ii), letting $l_u = \frac{l}{2-\alpha\beta}$ generates the following upper bound of the absolute sub-optimality:

$$\begin{aligned} v - v^* &< \tfrac{a^2\delta\epsilon}{2\alpha\beta(1-\alpha\beta)p}[(\tfrac{1-\alpha\beta}{2-\alpha\beta})^2 l^2 - \alpha\beta(\tfrac{\alpha\beta-1}{2-\alpha\beta})^2 l^2] \\ &= \tfrac{a^2\delta\epsilon}{2\alpha\beta p}(\tfrac{1-\alpha\beta}{2-\alpha\beta})^2 l^2. \end{aligned} \tag{A.13}$$

We next derive a lower bound of v^*. This is equivalent to finding the value of l_u that maximizes $m(l_u) = \alpha\beta l(l-2l_u) + l_u^2$, which is the part of the expression of v^* that involves l_u. Being a quadratic function of l_u, $m(l_u)$ is deceasing for $l_u \in [0, \alpha\beta)$ and increasing for $l_u \in [\alpha\beta, \frac{l}{2-\alpha\beta})$. Following condition (iii) that $\alpha\beta \in [0.5, 1)$, it can be verified that $\alpha\beta > \frac{1}{2}\frac{l}{2-\alpha\beta}$. Therefore, $m(0) = \alpha\beta l^2$ is greater than $m(\frac{l}{2-\alpha\beta})$ and is thus the maximum of $m(l_u)$ for $l_u < \frac{l}{2-\alpha\beta}$. Subsequently, we obtain the following lower bound of v^*:

$$\begin{aligned} v^* &\geq al(\mu+\epsilon/2) - \tfrac{a^2\delta m(0)}{2p\alpha\beta(1-\alpha\beta)}\epsilon \\ &= al(\mu+\epsilon/2) - \tfrac{a^2\delta l^2}{2p(1-\alpha\beta)}\epsilon \\ &\geq al\epsilon - \tfrac{a^2\delta l^2}{2p(1-\alpha\beta)}\epsilon, \end{aligned} \tag{A.14}$$

where the last inequality is due to the assumption that $\mu - \epsilon/2 \geq 0$. Combining inequalities (A.13) and (A.14), we obtain the following upper bound of the suboptimality:

$$\begin{aligned} \tfrac{v-v^*}{v^*} &< \frac{\frac{a^2\delta\epsilon}{2\alpha\beta p}(\frac{1-\alpha\beta}{2-\alpha\beta})^2 l^2}{al\epsilon - \frac{a^2\delta l^2}{2p(1-\alpha\beta)}\epsilon} \\ &= \frac{(\frac{1-\alpha\beta}{2-\alpha\beta})^2}{2\alpha\beta}\frac{\frac{a\delta l}{p}}{1-\frac{a\delta l}{2(1-\alpha\beta)p}} \\ &= \frac{(\frac{1-\alpha\beta}{2-\alpha\beta})^2}{2\alpha\beta}\frac{(1-\alpha\beta)\theta}{1-\frac{\theta}{2}}, \end{aligned} \tag{A.15}$$

where $\theta = \frac{al\delta}{(1-\alpha\beta)p}$. While (A.15) already presents an upper bound of the suboptimality, we further simplify it to exclude $\alpha\beta$. Using condition (iii) that $\alpha\beta \in [0.5, 1)$, we have $\frac{1-\alpha\beta}{2-\alpha\beta} = 1 - \frac{1}{2-\alpha\beta} \leq 1 - \frac{1}{2-0.5} = \frac{1}{3}$ and $1 - \alpha\beta \leq 0.5$. Therefore, the upper bound can be relaxed and simplified as

$$\tfrac{v-v^*}{v^*} < \frac{(\frac{1}{3})^2}{1}\frac{\frac{1}{2}\theta}{1-\frac{\theta}{2}} = \frac{1}{9}\frac{\theta}{(2-\theta)}. \quad \square \tag{A.16}$$

Proof of Lemma 1: To derive the expected real total cost, we first immediately know from (1.9) that the upstream transmission capacity is

$$C_u = \mu + (\frac{1}{2} - \frac{al_u\delta}{p})\epsilon, \tag{A.17}$$

and from (1.10) that the expected real cost upstream of the junction is

$$v_u = al_u\mu + al_u(\frac{1}{2} - \frac{1}{2}\frac{al_u\delta}{p})\epsilon. \tag{A.18}$$

Given C_u and the consequent wind output that is curtailed at C_u and uniformly distributed in $[\mu - \frac{1}{2}\epsilon, C_u)$, the expected downstream friction loss as a function of the downstream transmission capacity C_d becomes

$$\begin{aligned}\mathbb{E}[\mathbf{l}_d] &= \int_{C_d}^{C_u}(w_t - C_d)(1-\alpha\beta)\tfrac{1}{\epsilon}dw_t + \mathbb{P}(\mathbf{w_t} \geq C_u)(C_u - C_d)(1-\alpha\beta) \\ &= \int_{C_d}^{\mu+(\frac{1}{2}-\frac{al_u\delta}{p})\epsilon}(w_t - C_d)(1-\alpha\beta)\tfrac{1}{\epsilon}dw_t + \tfrac{al_u\delta}{p}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - C_d)(1-\alpha\beta) \\ &= \tfrac{1-\alpha\beta}{2\epsilon}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - C_d)^2 + \tfrac{al_u\delta}{p}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - C_d)(1-\alpha\beta).\end{aligned} \tag{A.19}$$

The expected real downstream cost becomes

$$\begin{aligned}v_d(C_d) &= a(l-l_u)C_d + \tfrac{p}{\delta}\mathbb{E}[\mathbf{l}_d] \\ &= a(l-l_u)C_d + \tfrac{1-\alpha\beta}{2\delta\epsilon}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - C_d)^2 + pal_u(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon - C_d)(1-\alpha\beta) \\ &= \tfrac{1-\alpha\beta}{2\delta\epsilon}C_d^2 + [-\tfrac{p(1-\alpha\beta)(\mu+(\frac{1}{2}-\frac{al_u\delta}{p})\epsilon)}{\delta\epsilon} - pal_u(1-\alpha\beta) + a(l-l_u)]C_d \\ &+ \tfrac{1-\alpha\beta}{2\delta\epsilon}(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon)^2 + pal_u(\mu + (\tfrac{1}{2} - \tfrac{al_u\delta}{p})\epsilon)(1-\alpha\beta).\end{aligned} \tag{A.20}$$

Applying the first-order condition, we obtain the minimizer of $v_d(C_d)$

$$C_d^* = \mu + \epsilon(\frac{1}{2} - \frac{a\delta(l-l_u)}{p(1-\alpha\beta)}). \tag{A.21}$$

Condition (i) imposes that this quantity is greater than the mean of the upstream wind output (which is given by (A.31)); condition (ii) is equivalent to $\mu + \epsilon(\frac{1}{2} - \frac{a\delta(l-l_u)}{p(1-\alpha\beta)}) < \mu + (\frac{1}{2} - \frac{al_u\delta}{p})\epsilon$, or $C_d^* < C_u$, so that ES is necessary. Therefore, C_d^* is the downstream transmission capacity. Substituting C_d with C_d^* in (A.20) yields the expected real downstream cost:

$$\begin{aligned}v_d^* &= a(l-l_u)(\mu + \tfrac{1}{2}\epsilon - \tfrac{\delta\epsilon a(l-l_u)}{p(1-\alpha\beta)}) + pal_u[\tfrac{\delta\epsilon a}{p}(\tfrac{l-l_u}{1-\alpha\beta} - l_u)](1-\alpha\beta) \\ &+ \tfrac{p(1-\alpha\beta)}{2\delta\epsilon}[\tfrac{\delta\epsilon a}{p}(\tfrac{l-l_u}{1-\alpha\beta} - l_u)]^2.\end{aligned} \tag{A.22}$$

Then the expected real total cost is the sum of the expected upstream and downstream real costs as given in (A.18) and (A.22):

$$\begin{aligned}v &= v_u + v_d^* \\ &= al(\mu + \tfrac{\epsilon}{2}) - \tfrac{a^2\delta(1-\alpha\beta)l_u^2 + 2\delta a^2(l-l_u)^2}{2p(1-\alpha\beta)}\epsilon. \qquad \square\end{aligned} \tag{A.23}$$

Proof of Lemma 2:

The evaluation of the expected real total cost needs joint optimization of transmission capacities both upstream and downstream of the junction site, denoted by C_u and C_d, respectively. From (1.8) we know that the expected upstream cost is

$$v_u(C_u) = \frac{p}{2\epsilon\delta}(\mu + \frac{\epsilon}{2} - C_u)^2 + al_uC_u. \tag{A.24}$$

Given C_u, C_d and the wind output that is curtailed at C_u and uniformly distributed in $[\mu - \frac{1}{2}\epsilon, C_u)$, the expected downstream friction loss becomes

$$\begin{aligned}\mathbb{E}[\mathbf{l}_d] &= \textstyle\int_{C_d}^{C_u}(w_t - C_d)(1-\alpha\beta)\frac{1}{\epsilon}dw_t + \mathbb{P}(\mathbf{w_t} \geq C_u)(C_u - C_d)(1-\alpha\beta) \\ &= \tfrac{1-\alpha\beta}{2\epsilon}(C_u - C_d)^2 + \tfrac{\mu+\epsilon/2-C_u}{\epsilon}(C_u - C_d)(1-\alpha\beta) \\ &= -\tfrac{1-\alpha\beta}{2\epsilon}C_u^2 + \tfrac{1-\alpha\beta}{2\epsilon}C_d^2 + \tfrac{(\mu+\epsilon/2)(1-\alpha\beta)}{\epsilon}(C_u - C_d).\end{aligned} \tag{A.25}$$

The expected downstream cost is

$$\begin{aligned}v_d(C_u, C_d) &= a(l-l_u)C_d + \tfrac{p}{\delta}\mathbb{E}[\mathbf{l}_d] \\ &= -\tfrac{p(1-\alpha\beta)}{2\epsilon\delta}C_u^2 + \tfrac{p(\mu+\epsilon/2)(1-\alpha\beta)}{\epsilon\delta}C_u + \tfrac{p(1-\alpha\beta)}{2\epsilon\delta}C_d^2 + [a(l-l_u) - \tfrac{p(\mu+\epsilon/2)(1-\alpha\beta)}{\epsilon\delta}]C_d.\end{aligned} \tag{A.26}$$

Then the expected total cost becomes

$$\begin{aligned}v(C_u, C_d) &= v_u(C_u) + v_d(C_u, C_d) \\ &= \tfrac{\alpha\beta p}{2\delta\epsilon}C_u^2 + \tfrac{(2a\delta\epsilon l_u - 2\alpha\beta\mu p - \alpha\beta p\epsilon)}{2\delta\epsilon}C_u \\ &+ \tfrac{p(1-\alpha\beta)}{2\epsilon\delta}C_d^2 + [a(l-l_u) - \tfrac{p(\mu+\epsilon/2)(1-\alpha\beta)}{\epsilon\delta}]C_d + \tfrac{p}{2\epsilon\delta}(\mu + \tfrac{\epsilon}{2})^2.\end{aligned} \tag{A.27}$$

□

Note that the bilinear term C_uC_d cancels out in (A.25). As a result, the optimal upstream and downstream transmission capacities, C_u^* and C_d^*, do not depend on each other. Applying the first-order condition to (A.27) with respect to C_u and C_d, respectively, we obtain

$$C_u^* = \mu + \epsilon(\frac{1}{2} - \frac{a\delta l_u}{\alpha\beta p}), \tag{A.28a}$$

$$C_d^* = \mu + \epsilon(\frac{1}{2} - \frac{a\delta(l-l_u)}{(1-\alpha\beta)p}). \tag{A.28b}$$

Plugging (A.28a) and (A.28b) into (A.27) yields the expected optimal total cost given as follows:

$$v^* = v(C_u^*, C_d^*) = al(\mu + \epsilon/2) - \tfrac{a^2\delta[\alpha\beta l(l-2l_u)+l_u^2]}{2p\alpha\beta(1-\alpha\beta)}\epsilon. \tag{A.29}$$

Remarks

In most of practical settings of parameter values, η and θ are both much less than one, suggesting that the model suboptimality as bounded in Propositions (5) and (6) are well-contained. Moreover, these propositions as well as their proofs lead to the following observations about the properties of the problem:

Curtailment-independence of downstream transmission capacity. After myopically sizing the upstream transmission lines in both the ES-free and ES-coupled scenarios, we may choose the downstream transmission capacity as if there were no upstream curtailment at all (though it should not exceed the upstream line capacity). To see this point, notice that, under proper conditions, the downstream transmission capacities prescribed in (A.4) and (A.21) are identical with those prescribed by (1.9) and (1.4), respectively, where there assumes to be no upstream curtailment. This effect of curtailment-independence might be against the intuition that higher degree of upstream curtailment would result in more conservative sizing of the downstream line. However, it can be understood by noticing that the power that has been curtailed by the upstream line, no matter how much it is, is forgone and irrelevant to the downstream power transmission. In other words, upstream curtailment does not affect the profile of the wind power that can be transmitted via the downstream line.

This property may help simplify planning model (1.13). Specifically, we may replace those farm-junction-specific wind characteristics (e.g., moments given by (11)) simply with wind characteristics at each wind farm. This simplification frees us from approximating curtailed wind power distribution by uniform distribution, which may also help reduce the approximation error.

Decoupling of the joint optimization of upstream and downstream transmission capacities. Proof of Lemma 2 shows that, under the given conditions, the jointly optimal upstream and downstream transmission capacities given by (A.28) can actually be obtained in a decoupled fashion, since the cross term of these two quantities cancels out in the expression of the expected total cost (A.27). The explanation of the curtailment-independence of downstream transmission capacity also applies here.

More interestingly, two consequences of this decoupling effect presents more insights into our planning model. First, the jointly optimal downstream transmission capacity given by (A.28b) turns out to be identical to that capacity given by (A.21) following the sub-optimal approach, which is employed in our planning problem formulation. It suggests that, the downstream transmission capacity, which accounts for a major part of the total cost in reality, can be prescribed to be near-optimal by our planning model.

On the other hand, the jointly optimal upstream transmission capacity given by (A.28a) differs from the myopically determined quantity (A.17) in the degree of curtailment by a factor of $\frac{1}{\alpha\beta}$. In this way, the optimal sizing of upstream transmission takes the ES conversion efficiency into account, tending to curtail more power if it anticipates more friction loss downstream. Therefore, we may partly offset the suboptimality of planning model (1.13) by replacing $\eta = \frac{\delta q}{p}$ with $\tilde{\eta} = \frac{\delta q}{\alpha\beta p}$ in the terms of upstream variable cost of the objective function (1.13a). More specifically, the terms in the second bracket in (1.13a) can be replaced with the following terms:

$$\sum_{j\in J}\sum_{i\in I}[q_{ij}\mu_i + q_{ij}(\frac{1}{2} - \frac{1}{2}\tilde{\eta}_{ij})\epsilon_i]Y_{ij} + \sum_{j\in J}\sum_{i\in I}[q_{ij}\mu_i + q_{ij}(\frac{1}{2} - \frac{1}{2}\eta_{ij})\epsilon_i]Z_{ij}. \tag{A.30}$$

Numerical Tests

We next carry out numerical tests for further validation. Note that the optimal total cost is found by doing two-dimensional line-search over the upstream and downstream transmission capacities given $2,000$ hours of uniformly distributed wind power from the wind farm.

In the ES-free scenario, we generate $10,000$ test instances for different values of $l \in [10, 460]$ and $l_u \in [0, l]$. In the ES-coupled scenario, we generate $100,000$ test instances for different values of $l \in [10, 460]$, $l_u \in [0, l]$ and $\alpha\beta \in \{0.3, 0.8\}$. The resulting suboptimality as the relative error of the real total cost against the optimal total cost is summarized in Table A.1. Note that the maximum suboptimality in the ES-free scenario, 33.6%, is close to the upper bound $\frac{1}{3}$ given by Proposition (5) when $\eta = \frac{\delta a l}{p} = 1$. When only considering more practical cases where $\eta \leq 0.5$ (i.e., $l < 230$ miles), this suboptimality becomes much smaller in both cases.

The resulting suboptimality averages 2.48% with maximum 25.3%. Again, when only considering those instances with $\eta \leq 0.5$, the suboptimality becomes much smaller, averaging 0.41% with maximum 4.97%.

Table A.1: Model suboptimality due to myopically sizing an upstream transmission line.

	ES-Free		**ES-Coupled**	
	Mean	Maximum	Mean	Maximum
$\eta \in (0, 1)$	5.68%	33.6%	1.34%	5.18%
$\eta \in (0, 0.5]$	2.48%	25.3%	0.41%	4.97%

Overall Model Inaccuracy

The overall model inaccuracy consists of the model suboptimality and the distribution approximation errors of the wind outputs at individual farms and at the junction sites. To numerically investigate the overall model inaccuracy, we again use the same three-farm scenario that is used to investigate the approximation error of aggregated wind outputs. However, for the purpose of incorporating model suboptimality, we also evaluate the upstream transmission capacity, both myopically using formula (1.9) and optimally using line-search. The overall model inaccuracy is then measured as the relative error between the real total cost and the optimal total cost.

In the ES-free scenario, we generate $5,000$ instances for different values of $l \in [20, 460]$, $l_u \in [0, l]$ and $\rho \in \{0.3, 0.8\}$. In the ES-coupled scenario, we generate $200,000$ instances for different values of $l \in [20, 460]$, $l_u \in [0, l]$, $\rho \in \{0.3, 0.8\}$ and $\alpha\beta \in [0.5, 0.99]$. The resulting overall model inaccuracy is summarized in Table A.2. Through the numerical experiments,

we observe, again, that large errors occur in the less realistic scenario where the whole transmission distance is too long to be profitable. For those instances with $\eta \leq 0.5$ in both the ES-free and ES-coupled scenarios, the maximum error is significantly smaller as shown in Table A.2.

Table A.2: Overall model inaccuracy.

	ES-Free		**ES-Coupled**	
	Mean	Maximum	Mean	Maximum
$\eta \in (0, 1)$	10.7%	39.4%	7.08%	30.7%
$\eta \in (0, 0.5]$	8.25%	17.6%	6.42%	14.8%

In summary, we have demonstrated that the model inaccuracy due to uniform distribution approximation of wind outputs and myopically sizing upstream transmission lines is reasonably small in most practical settings. Under some less realistic conditions, however, the model inaccuracy can be substantial. We also provide closed-form upper bounds of model suboptimality and present more insights that may potentially improve our planning model.

A.2 Proofs

Derivation of Equations (1.11a) and (1.11b)

In Section 1.4 we use uniform distribution to approximate the probability distribution of the curtailed wind power $\mathbf{w}_{t,ij}$, which is from farm i and faced by site j. When $\eta_{ij} < 1$, we claim that this uniform distribution has mean and interval length as follows:

$$\mu_{ij} = \mu_i - \frac{1}{2}\epsilon_i \eta_{ij}^2, \tag{1.11a}$$

$$\epsilon_{ij} = \sqrt{(1-\eta_{ij})^3(1+3\eta_{ij})}\epsilon_i. \tag{1.11b}$$

To see (1.11a), note that the expected wind power that is curtailed by a transmission line of capacity C can be expressed as:

$$\begin{aligned} \mu_c &= \int_{\mu-\frac{\epsilon}{2}}^{C} w \tfrac{1}{\epsilon} \mathrm{d}w + C\mathbb{P}(\mathbf{w_t} \geq C) \\ &= \tfrac{1}{\epsilon}\tfrac{1}{2}[C^2 - (\mu - \tfrac{\epsilon}{2})^2] + \tfrac{C}{\epsilon}(\mu + \tfrac{\epsilon}{2} - C) \\ &= C - \tfrac{(C-\mu+\frac{\epsilon}{2})^2}{2\epsilon}. \end{aligned} \tag{A.32}$$

Substituting C with the optimal quantity $\mu + (\frac{1}{2} - \eta)\epsilon$ as given by (1.9) yields the expression of the expected curtailed wind power given by (1.11a).

As for (1.11b), the second moment of the wind power curtailed by a line with capacity C is:

$$\begin{aligned}\mathbb{E}[\mathbf{w}_{t,c}^2] &= \int_{\mu-\frac{\epsilon}{2}}^{C} w^2 \tfrac{1}{\epsilon}\mathrm{d}w + C^2\mathbb{P}(\mathbf{w_t} \geq C) \\ &= \tfrac{1}{\epsilon}\tfrac{1}{3}[C^3 - (\mu - \tfrac{\epsilon}{2})^3] + \tfrac{C^2}{\epsilon}(\mu + \tfrac{\epsilon}{2} - C).\end{aligned} \tag{A.33}$$

The variance of this curtailed wind power becomes:

$$\begin{aligned}Var(\mathbf{w}_{t,c}) &= \mathbb{E}[\mathbf{w}_{t,c}^2] - \mu_c^2 \\ &= \tfrac{(2C-2\mu+\epsilon)^3(-6C+6\mu+5\epsilon)}{192\epsilon^2}.\end{aligned} \tag{A.34}$$

Again substituting C with the optimal quantity $\mu + (\frac{1}{2} - \eta)\epsilon$ into the above equation yields:

$$Var(\mathbf{w}_{t,c}) = \frac{(1-\eta)^3(1+3\eta)\epsilon^2}{12}. \tag{A.35}$$

The interval length of this uniform distribution becomes:

$$\epsilon_c = \sqrt{12Var(\mathbf{w}_{t,c})} = \sqrt{(1-\eta)^3(1+3\eta)}\epsilon, \tag{A.36}$$

which proves (1.11b). □

Proof of Proposition 1

Proposition (1). *Assume $C\delta_b \geq \mathbb{E}[\mathbf{w}_{b,\tau}]$, and suppose $\tilde{f}_s(s)$ is an approximation of $f_s(s)$ such that $\tilde{f}_s(s)$ is constant in the open interval $(0, S)$. Then*
(i) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) \geq \mathbb{P}(\mathbf{s}_\tau = S)$;
(ii) $\tilde{\mathbb{E}}[\mathbf{o}_\tau] \geq \mathbb{E}[\mathbf{o}_\tau]$,
where $\tilde{\mathbb{P}}(\cdot)$ and $\tilde{\mathbb{E}}[\cdot]$ denote probability and expectation with respect to $\tilde{f}_s$, respectively.

Proof of Proposition 1(i).

Throughout the proofs of Propositions 1, 2 and 3, we use $\mathbb{P}_w(\cdot)$ and $\mathbb{E}_w[\cdot]$ to denote probability and expectation with wind output $\mathbf{w_{b,\tau+1}}$ being the underlying random variable, respectively.

In order to prove Proposition 1(i), we first express $\mathbb{P}(\mathbf{s}_\tau = S)$ and $\mathbb{P}(\mathbf{s}_\tau = 0)$ in terms of $f_s(s)$. From (1.14), which describes the transition of $\mathbf{s}_\tau$, the recursive formulae of $\mathbb{P}(\mathbf{s}_\tau = S)$ and $\mathbb{P}(\mathbf{s}_\tau = 0)$ in the long run are given as follows:

$$\begin{aligned}\mathbb{P}(\mathbf{s}_\tau = S) &= \int_{0^+}^{S^-} \mathbb{P}_w(s + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S) f_s(s)\mathrm{d}s \\ &+ \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \geq C\delta_b)\mathbb{P}(\mathbf{s}_\tau = S) + \mathbb{P}_w(\alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S)\mathbb{P}(\mathbf{s}_\tau = 0),\end{aligned} \tag{A.37}$$

$$\begin{aligned}\mathbb{P}(\mathbf{s}_\tau = 0) &= \int_{0^+}^{S^-} \mathbb{P}_w(\beta s \leq C\delta_b - \mathbf{w_{b,\tau+1}}) f_s(s)\mathrm{d}s \\ &+ \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \leq C\delta_b)\mathbb{P}(\mathbf{s}_\tau = 0) + \mathbb{P}_w(\beta S \leq C\delta_b - \mathbf{w_{b,\tau+1}})\mathbb{P}(\mathbf{s}_\tau = S).\end{aligned} \tag{A.38}$$

The last term on the RHS of (A.37) and the last term on the RHS of (A.38) are zero, since $\mathbb{P}_w(\alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S) = \mathbb{P}_w(\beta S \leq C\delta_b - \mathbf{w_{b,\tau+1}}) = 0$, according to the first assumption that we make prior to Proposition 1. Rearranging the terms in the above equations yields:

$$\mathbb{P}(\mathbf{s}_\tau = S) = \frac{1}{1 - \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \geq C\delta_b)} \int_{0^+}^{S^-} \mathbb{P}_w(s + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S) f_s(s)\mathrm{d}s, \quad \text{(A.39)}$$

$$\mathbb{P}(\mathbf{s}_\tau = 0) = \frac{1}{1 - \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \leq C\delta_b)} \int_{0^+}^{S^-} \mathbb{P}_w(\beta s \leq C\delta_b - \mathbf{w_{b,\tau+1}}) f_s(s)\mathrm{d}s. \quad \text{(A.40)}$$

For notational simplicity, let

$$A(s) = \frac{\mathbb{P}_w(s + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S)}{1 - \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \geq C\delta_b)}, \quad \text{(A.41)}$$

$$B(s) = 1 + \frac{\mathbb{P}_w(\beta s \leq C\delta_b - \mathbf{w_{b,\tau+1}})}{1 - \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \leq C\delta_b)}. \quad \text{(A.42)}$$

We thus have $\mathbb{P}(\mathbf{s}_\tau = S) = \int_{0^+}^{S^-} A(s) f_s(s)\mathrm{d}s$ and $\mathbb{P}(\mathbf{s}_\tau = 0) = \int_{0^+}^{S^-} (B(s) - 1) f_s(s)\mathrm{d}s$. Notice that, in the interval $(0, S)$, $A(s)$ is always positive and increasing in s, whereas $B(s)$ is always positive and decreasing in s. Then substituting (A.39)-(A.42) into the identity $\mathbb{P}(\mathbf{s}_\tau = S) + \int_{0^+}^{S^-} f_s(s)\mathrm{d}s + \mathbb{P}(\mathbf{s}_\tau = 0) = 1$, we obtain

$$\int_{0^+}^{S^-} (A(s) + B(s)) f_s(s)\mathrm{d}s = 1. \quad \text{(A.43)}$$

We repeat the above procedure from (A.37) to (A.42), with $\mathbb{P}(\cdot)$ and $f_s(\cdot)$ being replaced with $\tilde{\mathbb{P}}(\cdot)$ and $\tilde{f}_s(\cdot)$, respectively. Notice that both $A(s)$ and $B(s)$ are functions of probabilities of wind outputs, which are independent from any probability distribution of storage level. Therefore, the resulting expressions of $A(s)$ and $B(s)$ are identical to those given by (A.41) and (A.42). Since $\tilde{f}_s(s)$ is also defined as pdf of the storage level, we thus obtain

$$\int_{0^+}^{S^-} (A(s) + B(s)) \tilde{f}_s(s)\mathrm{d}s = 1. \quad \text{(A.44)}$$

Comparing Eqns.(A.43) and (A.44) gives

$$\int_{0^+}^{S^-} (A(s) + B(s))[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s = 0. \quad \text{(A.45)}$$

We have assumed that $f_s(s)$ is decreasing in s on $(0, S)$. This assumption can be reasoned as follows: we have chosen the interval length δ_b such that $\{\mathbf{w}_{b,\tau}\}$ is an i.i.d process. And notice that $C\delta_b$ is greater than or equal to the mean of $\mathbf{w}_{b,\tau}$. Therefore, at certain storage

level s_τ, the probability of charge is greater than or equal to the probability of discharge. Also notice that $\alpha, \beta \in (0,1)$, indicating that any difference $\Delta S = |C\delta_b - \mathbf{w}_{b,\tau}|$ results in smaller magnitude of increase in s_τ when $C\delta_b > \mathbf{w}_{b,\tau}$ than the magnitude of decrease in s_τ when $C\delta_b < \mathbf{w}_{b,\tau}$ because of the friction loss. We thus know that $f_s(s_\tau + \Delta S|s_\tau) < f_s(s_\tau - \Delta S|s_\tau), \forall s_\tau, \Delta S$ satisfying $0 < s_\tau - \Delta S < s_\tau < s_\tau + \Delta S < S$. We do not establish a similar inequality of conditional probability given $s_\tau = 0$ or S. However, since S is assumed to be large enough and real wind output distribution is smooth and continuous, unconditioned density function $f_s(s)$ is tested to be decreasing in s on the open interval $(0, S)$.

With the above monotonic assumption, since $\tilde{f}_s(s)$ is constant, and $A(s), B(s) > 0$ on $(0, S)$, we further know from (A.45) that $f_s(s) - \tilde{f}_s(s)$ is decreasing in s and cross zero once at the point denoted by $s_{mid} \in (0, S)$.

To show $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) \geq \mathbb{P}(\mathbf{s}_\tau = S)$, it is equivalent to showing $\int_{0^+}^{S^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \leq 0$. Suppose, in order to derive a contradiction, that $\int_{0^+}^{S^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s > 0$, which also implies that $\int_{0^+}^{S^-} B(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s < 0$ from (A.45). Construct a constant function $\bar{A}(s) \equiv A(s_{mid})$. Then $\bar{A}(s) > A(s), \forall s \in (0, s_{mid})$ and $\bar{A}(s) < A(s), \forall s \in (s_{mid}, S)$. But we also know that $f_s(s) - \tilde{f}_s(s) > 0, \forall s \in (0, s_{mid})$ and $f_s(s) - \tilde{f}_s(s) < 0, \forall s \in (s_{mid}, S)$. It thus follows that:

$$\begin{array}{rl} & \int_{0^+}^{S^-} \bar{A}(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\ > & \int_{0^+}^{s_{mid}^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s + \int_{s_{mid}}^{S^-} \bar{A}(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\ > & \int_{0^+}^{s_{mid}^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s + \int_{s_{mid}}^{S^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\ = & \int_{0^+}^{S^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\ > & 0. \end{array} \tag{A.46}$$

Since $\bar{A}(s) \equiv A(s_{mid}) > 0$, the above inequality immediately implies that $\int_{0^+}^{S^-}[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s > 0$. However, following the similar approach, we also obtain that $\int_{0^+}^{S^-}[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s < 0$ from the hypothesis $\int_{0^+}^{S^-} B(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s < 0$. Therefore, a contradiction is derived, which finishes the proof that $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) \geq \mathbb{P}(\mathbf{s}_\tau = S)$. ■

Proof of Proposition 1(ii).

Let Λ denote the event that overflow occurs in the next interval. Then from Eqn.(A.41) we have $\mathbb{P}_w(s + \alpha(\mathbf{w_{b,\tau+1}} - C\delta_b) \geq S) = \mathbb{P}_w(\Lambda|s)$ and $A(s) = \frac{\mathbb{P}_w(\Lambda|s)}{1 - \mathbb{P}_w(\mathbf{w_{b,\tau+1}} \geq C\delta_b)}$.

Conditioning on whether the storage level is at or below S, $\mathbb{E}[\mathbf{o}_\tau]$ and $\tilde{\mathbb{E}}[\mathbf{o}_\tau]$ can be expressed as follows:

$$\mathbb{E}[\mathbf{o}_\tau] = \mathbb{E}_w[\mathbf{o}_\tau|S]\mathbb{P}(\mathbf{s}_\tau = S) + \int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]\mathbb{P}_w(\Lambda|s) f_s(s)\mathrm{d}s, \tag{A.47}$$

$$\tilde{\mathbb{E}}[\mathbf{o}_\tau] = \mathbb{E}_w[\mathbf{o}_\tau|S]\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) + \int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]\mathbb{P}_w(\Lambda|s) \tilde{f}_s(s)\mathrm{d}s. \tag{A.48}$$

In order to show $\tilde{\mathbb{E}}[\mathbf{o}_\tau] \geq \mathbb{E}[\mathbf{o}_\tau]$ by comparing the right-hand sides of (A.47) and (A.48), first notice that $\mathbb{E}_w[\mathbf{o}_\tau|S]\mathbb{P}(\mathbf{s}_\tau = S) \leq \mathbb{E}_w[\mathbf{o}_\tau|S]\tilde{\mathbb{P}}(\mathbf{s}_\tau = S)$. Therefore, it suffices to show $\int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]\mathbb{P}_w(\Lambda|s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \leq 0$, or, equivalently, $\int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \leq 0$. To see why the latter statement is true, notice that $\mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]$ is positive and non-decreasing in s, and $A(s)$ is positive. Therefore, $\mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]A(s) \leq \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s), \forall s \in (0, s_{mid})$ and $\mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]A(s) \geq \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s), \forall s \in (s_{mid}, S)$. Again we already have $f_s(s) - \tilde{f}_s(s) > 0, \forall s \in (0, s_{mid})$ and $f_s(s) - \tilde{f}_s(s) < 0, \forall s \in (s_{mid}, S)$. It thus follows that:

$$
\begin{aligned}
& \int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\
\leq\ & \int_{0^+}^{s_{mid}^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s + \int_{s_{mid}}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\
\leq\ & \int_{0^+}^{s_{mid}^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s + \int_{s_{mid}}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\
=\ & \int_{0^+}^{S^-} \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}]A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\
=\ & \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}] \int_{0^+}^{S^-} A(s)[f_s(s) - \tilde{f}_s(s)]\mathrm{d}s \\
\leq\ & 0.
\end{aligned}
\tag{A.49}
$$

where the last inequality is due to $\mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s_{mid}] > 0$ and the proof by contradiction in (i). Subsequently, we obtain $\tilde{\mathbb{E}}[\mathbf{o}_\tau] \geq \mathbb{E}[\mathbf{o}_\tau]$. □

Proof of Proposition 2

Proposition (2). *Assume* $\mathbf{w}_{b,\tau} \sim$ *unif*$(\mu_b - \frac{\epsilon_b}{2}, \mu_b + \frac{\epsilon_b}{2})$. *Let* $A = \frac{(C\delta_b - \mu_b + \frac{\epsilon_b}{2})^2}{2\beta(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)}$ *and* $B = \frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \frac{\epsilon_b}{2})}$. *Then*

(i) For $s \in (0, S)$, $\tilde{f}_s(s) \equiv f_s^c = \frac{1}{A+S+B}$;
(ii) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = 0) = \frac{A}{A+S+B}$;
(iii) $\tilde{\mathbb{P}}(\mathbf{s}_\tau = S) = \frac{B}{A+S+B}$.

Proof of Proposition 2.

Given $\mathbf{w}_{b,\tau} \sim \mathrm{unif}(\mu_b - \frac{\epsilon_b}{2}, \mu_b + \frac{\epsilon_b}{2})$ and $\mathbf{w}_{b,\tau}$ is i.i.d, we assume $C\delta_b < \mu_b + \frac{\epsilon_b}{2}$ and $C\delta_b >= \mu_b$ to be nontrivial. Following the transition model (1.14), we express the conditional probability distribution of the storage level as follows:

$$
\begin{aligned}
\mathbb{P}_w\{\mathbf{s}_{\tau+\mathbf{1}} = 0|s_\tau\} =& \mathbb{P}_w\{C\delta_b - \mathbf{w}_{\mathbf{b},\tau+\mathbf{1}} \geq \beta s_\tau\} \\
=& \mathbb{P}_w\{\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}} \leq C\delta_b - \beta s_\tau\} \\
=& \begin{cases} 0, & \text{if } C\delta_b - \beta s_\tau < \mu_b - \frac{\epsilon_b}{2}; \\ \frac{1}{\epsilon_b}[C\delta_b - \beta s_\tau - (\mu_b - \frac{\epsilon_b}{2})], & \text{if } \mu_b - \frac{\epsilon_b}{2} \leq C\delta_b - \beta s_\tau < \mu_b + \frac{\epsilon_b}{2}. \end{cases}
\end{aligned}
\tag{A.50}
$$

$$\begin{aligned}\mathbb{P}_w\{\mathbf{s}_{\tau+\mathbf{1}}=S|s_\tau\}=&\mathbb{P}_w\{s_\tau+(\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}}-C\delta_b)\alpha\geq S\}\\ =&\mathbb{P}_w\{\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}}\geq\frac{S-s_\tau}{\alpha}+C\delta_b\}\\ =&\begin{cases}0, & \text{if } \frac{S-s_\tau}{\alpha}+C\delta_b>\mu_b+\frac{\epsilon_b}{2};\\ \frac{1}{\epsilon_b}[\mu_b+\frac{\epsilon_b}{2}-\frac{S-s_\tau}{\alpha}-C\delta_b], & \text{if } \mu_b-\frac{\epsilon_b}{2}\leq\frac{S-s_\tau}{\alpha}+C\delta_b\leq\mu_b+\frac{\epsilon_b}{2}.\end{cases}\end{aligned} \tag{A.51}$$

To derive $f_{\mathbf{s}_{\tau+\mathbf{1}}|s_\tau}(u|s_\tau)$, for infinitesimal storage level increment δu,

$$\begin{aligned}&\mathbb{P}_w\{\mathbf{s}_{\tau+\mathbf{1}}\in[u,u+\delta u]|s_\tau\}\\ =&\begin{cases}\mathbb{P}_w\{s_\tau+\alpha(\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}}-C\delta_b)\in[u,u+\delta u]\} & \text{if } u\geq s_\tau\\ \mathbb{P}_w\{s_\tau-\frac{1}{\beta}(C\delta_b-\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}})\in[u,u+\delta u]\} & \text{if } u<s_\tau\end{cases}\\ =&\begin{cases}\mathbb{P}_w\{\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}}\in\frac{[u,u+\delta u]-s_\tau}{\alpha}+C\delta_b\}, & \text{if } u\geq s_\tau;\\ \mathbb{P}_w\{\mathbf{w}_{\mathbf{b},\tau+\mathbf{1}}\in([u,u+\delta u]-s_\tau)\beta+C\delta_b\}, & \text{if } u<s_\tau.\end{cases}\end{aligned} \tag{A.52}$$

Therefore,

$$f_{\mathbf{s}_{\tau+\mathbf{1}}|s_\tau}(u|s_\tau)=\begin{cases}\frac{1}{\alpha\epsilon_b}, & \text{if } u\in[s_\tau,\min\{S,(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)\alpha+s_\tau\});\\ \frac{\beta}{\epsilon_b}, & \text{if } u\in(\max\{0,\frac{\mu_b-\frac{\epsilon_b}{2}-C\delta_b}{\beta}+s_\tau\},s_\tau);\\ 0, & \text{otherwise.}\end{cases} \tag{A.53}$$

To obtain the approximated unconditional probability distribution of S_τ, notice that $f_{\mathbf{s}_{\tau+\mathbf{1}}}(u)=f_{s_\tau}(u)$ when τ is sufficiently large. Hence, in the case of uniformly distributed wind energy $\mathbf{w}_{b,\tau}$ as well as the constant probability density $\tilde{f}_{s_\tau}(s)\equiv f_s^c,\forall s\in(0,S)$, we have

$$\begin{aligned}&\tilde{\mathbb{P}}(S_\tau=0)=\tilde{\mathbb{P}}(\mathbf{s}_{\tau+\mathbf{1}}=0)\\ =&\int_{0^+}^{\frac{C\delta_b-\mu_b+\frac{\epsilon_b}{2}}{\beta}}\frac{C\delta_b-\beta s-\mu_b+\frac{\epsilon_b}{2}}{\epsilon_b}\tilde{f}_{s_\tau}(s)\mathrm{d}s+\tilde{\mathbb{P}}(S_\tau=0)\frac{C\delta_b-\mu_b+\frac{\epsilon_b}{2}}{\epsilon_b}\\ =&\frac{f_s^c(C\delta_b-\mu_b+\frac{\epsilon_b}{2})^2}{2\beta\epsilon_b}+\tilde{\mathbb{P}}(S_\tau=0)\frac{C\delta_b-\mu_b+\frac{\epsilon_b}{2}}{\epsilon_b}\\ &\Rightarrow\quad\tilde{\mathbb{P}}(S_\tau=0)=f_s^c\frac{(C\delta_b-\mu_b+\frac{\epsilon_b}{2})^2}{2\beta(\mu_b+\epsilon_b/2-C\delta_b)},\end{aligned} \tag{A.54}$$

and

$$\begin{aligned}&\tilde{\mathbb{P}}(S_\tau=S)=\tilde{\mathbb{P}}(\mathbf{s}_{\tau+\mathbf{1}}=S)\\ =&\int_{S-\alpha(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)}^{S^-}\frac{\mu_b+\frac{\epsilon_b}{2}-\frac{S-s}{\alpha}-C\delta_b)}{\epsilon_b}\tilde{f}_{s_\tau}(s)\mathrm{d}s+\tilde{\mathbb{P}}(S_\tau=S)\frac{\mu_b+\frac{\epsilon_b}{2}-C\delta_b}{\epsilon_b}\\ =&\frac{f_s^c\alpha(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)^2}{2\epsilon_b}+\tilde{\mathbb{P}}(S_\tau=S)\frac{\mu_b+\frac{\epsilon_b}{2}-C\delta_b}{\epsilon_b}\\ &\Rightarrow\quad\tilde{\mathbb{P}}(S_\tau=S)=f_s^c\frac{\alpha(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)^2}{2(C\delta_b-\mu_b+\epsilon_b/2)}.\end{aligned} \tag{A.55}$$

Substituting (A.54) and (A.55) into the identity $\tilde{\mathbb{P}}(S_\tau = 0) + Sf_s^c + \tilde{\mathbb{P}}(S_\tau = S) = 1$, we obtain

$$f_s^c[\frac{(C\delta_b - \mu_b + \frac{\epsilon_b}{2})^2}{2\beta(\mu_b + \epsilon_b/2 - C\delta_b)} + S + \frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \epsilon_b/2)}] = 1.$$

Let $A = \frac{(C\delta_b - \mu_b + \frac{\epsilon_b}{2})^2}{2\beta(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)}$ and $B = \frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \frac{\epsilon_b}{2})}$.

We thus obtain

$$f_s^c = \frac{1}{A + B + S}, \tag{A.56}$$

which, together with (A.54) and (A.55), finishes the proof of Proposition 2. □

Proof of Proposition 3

Proposition (3). *Assume* $\mathbf{w}_{b,\tau} \sim unif(\mu_b - \frac{\epsilon_b}{2}, \mu_b + \frac{\epsilon_b}{2})$. *Then the expected energy overflow of each interval of length* δ_b *is bounded from above by* $\frac{5\alpha}{24S}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$. *The derivatives of this upper bound are* $-\frac{5\alpha}{24S^2}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$ *with respect to* S *and* $-\frac{5\alpha\delta_b}{12S}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2$ *with respect to* C.

Proof of Proposition 3.

From Proposition 1(ii) we already know that $\mathbb{E}[\mathbf{o}_\tau]$ is bounded from above by $\tilde{\mathbb{E}}[\mathbf{o}_\tau]$. In the case of uniformly distributed wind energy $\mathbf{w}_{b,\tau}$, we next compute the upper bounds for the terms on the right-hand side of (A.48), respectively.

For the first term,

$$\mathbb{E}_w[\mathbf{o}_\tau|S] = \int_{C\delta_b}^{\mu_b + \frac{\epsilon_b}{2}} (w - C\delta_b)\frac{1}{\epsilon_b}\mathrm{d}w = \frac{1}{2\epsilon_b}(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2. \tag{A.57}$$

From Proposition 2,

$$\begin{array}{rl} & \tilde{\mathbb{P}}(\mathbf{s}_\tau = S) \\ = & f_s^c \frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \epsilon_b/2)} \\ < & \frac{1}{S}\frac{\alpha(\mu_b + \frac{\epsilon_b}{2} - C\delta_b)^2}{2(C\delta_b - \mu_b + \epsilon_b/2)} \\ < & \frac{1}{S}\frac{\alpha(\epsilon_b/2)^2}{2\epsilon_b/2} \quad (\text{since } \ C\delta_b \in [\mu_b, \mu_b + \epsilon_b/2]) \\ = & \frac{\alpha\epsilon_b}{4S}. \end{array} \tag{A.58}$$

For the second term, when $s \in [S - (\mu_b + \epsilon_b/2 - C\delta_b)\alpha, S)$,

$$\begin{array}{rl} & \mathbb{E}_w[\mathbf{o}_\tau|\Lambda, s]\mathbb{P}_w(\Lambda|s) = \mathbb{E}_w[\mathbf{o}_\tau|s] \\ = & \int_{C\delta_b + (S-s)/\alpha}^{\mu_b + \epsilon_b/2} [w - C\delta_b - (S-s)/\alpha]\frac{1}{\epsilon_b}\mathrm{d}w \\ = & \frac{[\mu_b + \epsilon_b/2 - C\delta_b - (S-s)/\alpha]^2}{2\epsilon_b}. \end{array} \tag{A.59}$$

Otherwise $\mathbb{E}_w[\mathbf{o}_\tau|\Lambda,s]\mathbb{P}_w(\Lambda|s)=0$. It thus follows that

$$\begin{aligned}
& \int_{0^+}^{S^-}\mathbb{E}_w[\mathbf{o}_\tau|\Lambda,s]\mathbb{P}_w(\Lambda|s)\tilde{f}_s(s)\mathrm{d}s \\
= & \int_{S-(\mu_b+\epsilon_b/2-C\delta_b)\alpha}^{S^-}\frac{[\mu_b+\epsilon_b/2-C\delta_b-(S-s)/\alpha]^2}{2\epsilon_b}\tilde{f}_s(s)\mathrm{d}s \\
< & \int_{S-(\mu_b+\epsilon_b/2-C\delta_b)\alpha}^{S^-}\frac{[\mu_b+\epsilon_b/2-C\delta_b-(S-s)/\alpha]^2}{2\epsilon_b}\frac{1}{S}\mathrm{d}s \\
= & \frac{\alpha(\mu_b+\epsilon_b/2-C\delta_b)^3}{6\epsilon_b S} \\
\leq & \frac{\alpha}{6\epsilon_b S}\frac{\epsilon_b}{2}(\mu_b+\epsilon_b/2-C\delta_b)^2 \quad (\text{since } \ C\delta_b\geq\mu_b) \\
= & \frac{\alpha(\mu_b+\epsilon_b/2-C\delta_b)^2}{12S}.
\end{aligned} \tag{A.60}$$

Substituting (A.57), (A.58) and (A.60) into (A.48), we obtain

$$\tilde{\mathbb{E}}[\mathbf{o}_\tau]<\frac{5\alpha}{24S}(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)^2.$$

which provides an upper bound for $\mathbb{E}[\mathbf{o}_\tau]$ as a quadratic function of C. The derivatives of this upper bound are $-\frac{5\alpha}{24S^2}(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)^2$ with respect to S and $-\frac{5\alpha\delta_b}{12S}(\mu_b+\frac{\epsilon_b}{2}-C\delta_b)^2$ with respect to C. □

Proof of Proposition 4

Proposition (4)**.** *The optimal cost of the ES-transmission network is bounded from below by the optimal objective value of planning model (1.13), and bounded from above by the total cost given by the heuristic in Section 1.5.*

Proof of Proposition 4. To justify the first half of Proposition 4, note that the real optimal ES-transmission deployment is a feasible solution to model (1.13). Therefore, its total cost v_3 in the uncapacitated scenario given by (1.13a) is bounded from below by the optimal objective value of model (1.13). But v_3 is still less than the real optimal cost, since the latter also incorporates the cost of additional ES capacity and the cost of energy overflow. As for the second half, note that the heuristic outlined in Section 1.5 generates a feasible deployment but also a total cost that includes overestimated cost of additional ES capacity and the cost of energy overflow. Therefore, this total cost must be greater than the real optimal cost. □

A.3 Further Inspection of the Upper Bound of ES Capacity in Section 1.5

In Section 1.4 and Section A.1, we have analyzed the inaccuracy of the model developed in Section 1.4, where a lower bound of the total cost results from the assumption that the ES systems are uncapacitated. In this section, our goal is to examine the tightness and the accuracy of the cost upper bound as given in Table 1.5 developed in Section 1.5, which derives an upper bound for ES capacity. Again we focus on a single wind farm scenario. The

cost error analysis for multiple farms is similar and has been presented in the case study in Section 1.6.

To set up the experiment, we consider those 24 sites in the case study and look into their wind output data for each of the twelve months in 2006. The cycle length is set to be $\delta_b = 24$hr. The theoretical lower and upper bounds for the investment cost and energy loss are calculated based on the models in Sections 1.4 and 1.5, respectively. The economic transmission and the overestimated ES capacities are computed based on Table 1.5. In order to obtain the actual optimal total cost, we simulate 2000hr system operations for multiple runs, each run with different values of C and S. Then we calculate the average total cost for each run, and select the pair of (C, S) that results in the smallest average total cost. The simulations are carried out with different per-kWh ES capacity costs and with different distances between the wind farm and the load center. Table A.3 summarizes the average relative cost gaps between the upper and the lower bounds, between the upper bound and the optimal value, and between the upper bound and the cost in the ES-free scenario. Figure A.4 shows the profiles of site 26784 in March as typical instances. From these results we make the following observations.

Table A.3: Average cost gaps between the upper and the lower bounds, between the upper bound and the optimal value, and between the upper bound and the cost in the ES-free scenario.

Distance (mile)	UB - LB	|UB - Opt.|	ES-Free - UB
200	23.6%	6.6%	89.7%
120	20.2%	5.17%	81.5%
50	3.09%	6.78%	31.8%

- Firstly, as to the tightness of the cost upper bound, we first observe from Table A.3 and Figures A.4 (a), (c) and (e) that the optimal cost in most cases is much closer to the upper bound than to the lower bound, suggesting that the cost upper bound and its associated transmission and ES capacities are more effective estimates of the optimal values. In some cases, the actual optimal cost even exceeds the upper bound, as observed in Figure A.4 (e). This is due to the inaccuracy of the upper bound which we investigate later in this section. Also notice that, in order to focus on the variables of interest, we exclude the fixed transmission and ES installation costs, which, if included, would account for a significant share of the total cost and further narrow the relative gaps. Moreover, the cost gaps narrow as the distance between the wind farm and the load center decreases, because less distance results in cheaper transmission, larger transmission capacity to reduce energy overflow loss and thus less dependence on ES.

- Secondly, the cost gaps are bounded. The cost upper bound increases and then plateaus as the per-kWh ES capacity cost increases, as shown in Figures A.4 (a), (c) and (e). This is because increase in ES cost makes the investment in transmission capacity more

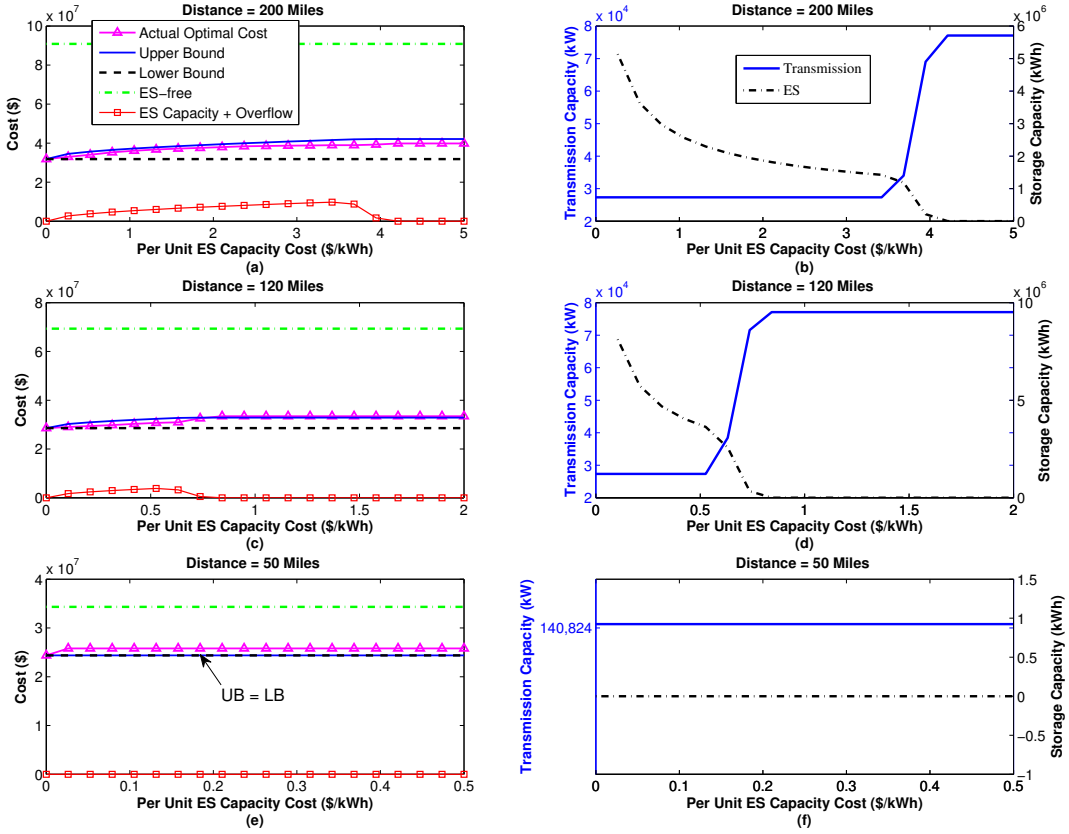

Figure A.4: Total variable costs, (a), (c) and (e), and economic transmission and (overestimated) ES capacities, (b),(d) and (f), with different distances between a wind farm and a load center.

cost-effective than in ES capacity; eventually, the transmission capacity becomes large enough to deliver all the daily wind energy and thus eliminate the use of additional ES capacity for overflow prevention. This ES-transmission interrelation is clearly demonstrated in Figure A.4(b). When the ES capacity cost is small ($r \leq \$3.4$/kWh), the transmission capacity C remains at its lowest level of the average wind power μ, since the costly long-distance transmission investment dominates the other cost components and the ES capacity S can be large to hedge against energy overflow. As the ES capacity cost increases, S has to decrease and it becomes more favorable to invest in transmission until C reaches $\frac{\mu_b+\frac{\epsilon_b}{2}}{\delta_b}$ and no additional ES capacity is required when $r \geq \$4.2$/kWh. Hence, as shown in the square-dotted line in Figure A.4(a), the cost of ES capacity investment and the expected energy overflow first increases and then diminishes to zero.

- Thirdly, the model of cost upper bound involves two sources of inaccuracy. First, the uniform distribution approximation of the daily wind output overlooks the tail of the real output distribution beyond the upper support of the approximated uniform distribution. Therefore, the approximated energy overflow is curtailed. The second source of inaccuracy is the assumption that the wind outputs are independent across days. This assumption is the key to the derivation of the closed-form upper bound expression for ES capacity, which helps deliver managerial insights. However, while choosing long intervals (such as one day) significantly reduces the auto-correlation of the wind output process, ignoring this auto-correlation still causes underestimation of energy overflow loss when storage level is nearly full. Collectively, the underestimation of inter-period overflow loss offsets the overestimation of intra-overflow loss and ES capacity to varying degrees. The resulting cost upper bound is closer to the actual optimal cost than theoretically expected on one hand, but on the other hand, this upper bound can be exceeded.

  Figures A.5 (a)-(c) are the histograms showing the distributions of the lag-one auto-correlation of the daily wind output processes at the 24 sites of twelve months with different transmission distances. The figures also show the proportion of instances where the actual optimal cost exceeds its theoretical upper bound by a certain margin, as represented by the light-colored part of each bin. As expected, the larger the auto-correlation is, the more likely the actual optimal cost is to exceed its upper bound.

  We also observe tradeoffs between accuracy and tightness of the cost upper bound. When the transmission distance is as short as 50 miles, Figure A.5 (c) shows that the upper bounds are exceeded by over 8% in nearly half of all the instances. This is because transmission investment is cheap enough and its capacity is prescribed as $\mu+\rho(\frac{1}{2}-\theta)$ by both the models with and without ES capacity limit. The ES capacity S in those instances also becomes zero. Consequently, the gap between the upper and the lower bounds is zero, but it leaves the underestimation of the overflow dominant and the actual optimal cost above the upper bound. In fact, $S = 0$ violates our assumption in Section 1.5 that S should be large enough such that the probability of the storage level switching from 0 to full state (or the other way round) is negligible. Fortunately, $S = 0$ represents the cases where the investment in ES capacity as well as the impact of this inaccuracy is small. If we only consider those instances with $S > 0$ due to cheap ES capacity cost, almost no cost upper bounds are violated, as shown in Figure A.5 (d).

  In contrast, Figures A.5 (a) and (b) and Table A.3 show that, when the transmission distance $l \geq 120$ miles, the cost upper bounds are not violated in most of the instances, at the cost of relatively large gaps between the upper and the lower bounds.

- Last but not the least, Table A.3 also shows that the combination of economic transmission line and ES can potentially significantly bring down the cost of building an ES-free line, as represented by the dash-dotted line in Figures A.4 (a), (b) and (c).

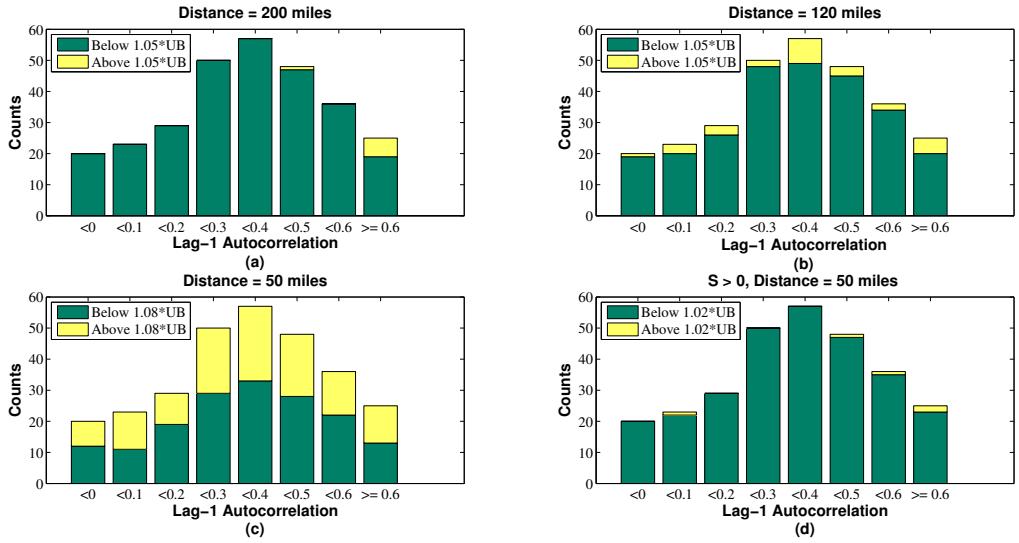

Figure A.5: Distribution of lag-one autocorrelation and the validity of the cost upper bounds.

However, this cost-saving also depends on the fixed installation cost of ES, which we do not incorporate in these experiments. .

The above results and analysis suggest that the total variable cost and its associated transmission and ES capacity given by Section 1.5 are more effective estimates of the optimal values than those given by Section 1.4. This is because the former explicitly consider ES capacity cost and involve modeling inaccuracies that are canceling to each other. We thus also justify the heuristic approach to jointly planning ES-transmission network, as outlined in Section 1.5.

A.4 Supporting Information for Section 1.6

Experiment Settings of the Case Study in Section 1.6

To generate the problem instances for the case study in Section 1.6, we choose the sites of the potential wind farms identified in [73], which also includes the data of wind outputs. The set of candidate junction sites J is the union of the set of the wind farms I and a set of additional locations in this region. We repeat the experiments by adding wind farms that are increasingly distant from Billings. In addition to the preceding parameters, we set the per-kW fixed transmission line cost $g = \$4.5 \times 10^5$/mile, fixed ES installment cost $h = \$1 \times 10^6$ and variable ES cost $r = \$0.5$/kWh. The experiments run on a Macbook Pro with 2.27 Ghz dual-core Intel Core i5 CPU and 8 GB memory. All instances are solved using CPLEX 12.3 called by YALMIP modeling language ([61]).

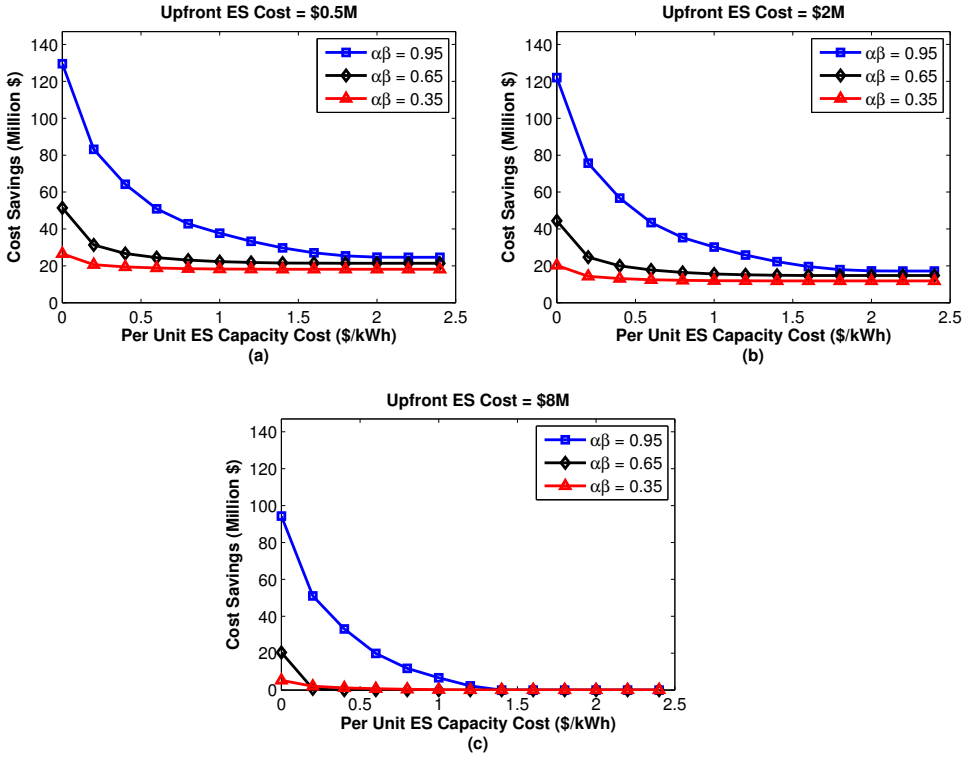

Figure A.6: Impact of different ES technologies and their advancements on cost-savings.

Analysis of Technology Considerations

The economic feasibility of each ES technology is significantly impacted by its cost-efficiency parameters, such as round-trip energy conversion efficiency, upfront construction cost and per-unit capacity cost. In Figure A.4 in Section A.3 we have already shown that systems cost as well as economic ES and transmission capacities are very sensitive to per-kWh cost. In what follows, we evaluate how the cost-savings respond to the different values of these cost-efficiency parameters in our case study with 24 wind farms and 28 candidate junction sites. The cost-savings are benchmarked against a scenario without ES co-location. The results are shown in Figure A.6, from which we draw several observations.

- Firstly, cost savings increase when (i) the round-trip conversion efficiency $\alpha\beta$ increases, (ii) per-unit ES capacity cost decreases and (iii) the fixed upfront ES cost decreases, as expected.

- Secondly, given upfront ES cost, when the per-unit ES capital cost is very high (e.g, $r \geq 1.6$ in our numerical test), cost-saving is small and cannot be effectively increased by enhancing conversion efficiency. This suggests that high per-unit capacity cost discourages adding ES capacity and it is not worth investing in improving the conversion

efficiency, as in the case of lead-acid battery. However, when the per-unit ES capacity cost decreases (e.g, $r < 1.4$ in our numerical test), the cost-saving is more sensitive to reduction in per-unit ES capital cost when energy conversion is efficient (large $\alpha\beta$) than when it is inefficient (small $\alpha\beta$). As shown in Figure A.6(b), for instance, when the per-unit ES capital cost decreases from \$0.6/kWh to \$0.4/kWh, the cost saving increases by 30.1% with $\alpha\beta = 0.95$, compared with the cost saving increased by only 12.7% with $\alpha\beta = 0.65$. This suggests that more investment is desirable in improving conversion efficiency when the per-unit ES capacity cost has already been low, as in the case of compressed air energy storage systems.

- Thirdly, as for the upfront ES cost, it is interesting to see that we can achieve cost saving even with low conversion efficiency and high per-unit ES capacity cost (as shown in Figure A.6(a) and (b)), until the upfront ES cost is too large to justify building any ES systems (as shown in Figure A.6(c)). In the former case, the positive cost saving is mainly achieved by a small base ES capacity that salvages transmission capacity cost by smoothing short-term wind-out fluctuations.

The above observations collectively suggest that the cost-efficiency parameters jointly determine the effectiveness of using ES systems. Therefore, the planners need to carefully weigh these cost factors to properly size the ES systems, and the priority of ES R&D should be given to addressing the bottleneck cost factor, be it conversion efficiency or capital cost.

Appendix B

Supporting Results for Chapter 3

Proof of (3.16) Model the passenger riding process as an alternate renewal process, in which the *on* state is waiting and the *off* state is riding. The driver is initially on for a time Z_1 and then remains off for a time Y_1; then he/she goes on for a time Z_2; then off for a time Y_2; then on, and so forth. The sequence of two random variables $\{Z_i\}$ and $\{Y_i\}$, $i \geq 1$, are independent and identically distributed as exponentials with μ and υ. A renewal cycle $\{V_i = Z_i + Y_i\}$ consists of an on period and an off period. The process starts over again when the driver picks up a new passenger. Let $N(x)$ denote the number of complete renewal cycle up to time x and S_n be the completing time of nth renewal. They have a straightforward relationship that $N(x) = \max\{n : \ S_n \leq x\}$. $m(t)$ is the mean of $N(x)$, i.e. $m(x) = E(N(x))$.

During renewal cycle i, the driver receives a reward $R_i = w_b + w_m Y_i$. It follows that $E(R_i) = w_b + w_m E(Y_i) = w_b + w_m/\upsilon = E(R)$. The objective is to calculate the expected total rewards during $(0, x)$. In the stochastic setting, the total rewards is differentiated in two cases: (1) When the driver is on at time x, i.e., the driver has finished $N(x)$ rides and the expected total rewards is $E(\sum_{i=1}^{N(x)} R_i)$; (2) When the driver is off at time x, the expected total rewards has two components: the reward from the $N(x)$ complete rides and the reward from the last incomplete ride. Since in practice the ride is always completed, the reward for the last ride is assumed to be complete, that is, the driver gets full rewards. Then the expected total reward is $E(\sum_{i=1}^{N(x)+1} R_i)$ in this case. Hence, combining these two cases, the expected total rewards is given by

$$\begin{aligned} w_p(x) &= E(\sum_{i=1}^{N(x)} R_i| \text{ on at } x)P(\text{on at } x) + E(\sum_{i=1}^{N(x)+1} R_i| \text{ off at } x)P(\text{off at } x) \\ &= E(\sum_{i=1}^{N(x)+1} R_i) - E(R_{N(x)+1}| \text{ on at } x)P(\text{on at } x) \\ &= (m(x)+1)E(R) - E(R_{N(x)+1}| \text{ on at } x)P(\text{on at } x), \end{aligned} \tag{B.1}$$

where the first two equalities use the law of total probability and the last equality utilizes

the Wald's equation by observing that $N(x)+1$ is a stopping time for $\{R_i\}$, $i \geq 1$ ([79]). First, $m(x)$ is known as ([99])

$$m(x) = \frac{\mu\upsilon}{\mu+\upsilon}\Big(x - \frac{1}{\mu+\upsilon}(1-\exp(-(\mu+\upsilon)x)\Big). \tag{B.2}$$

Second, $P(\text{on at } x)$ is given as ([95])

$$P(\text{on at } x) = \frac{\upsilon}{\mu+\upsilon} + \frac{\mu}{\mu+\upsilon}\exp(-(\mu+\upsilon)x). \tag{B.3}$$

And

$$\begin{aligned} E(R_{N(x)+1}|\text{ on at } x) &= \int_0^\infty E(R_{N(x)+1}|\ S_{N(x)} = y, \text{on at } x) dF_{S_{N(x)}|\text{ on at } x} \\ &= \int_0^\infty E(R_{N(x)+1}|\ Z_{N(x)+1} > x-y) dF_{S_{N(x)}|\text{ on at } x} \\ &= \int_0^\infty E(R) dF_{S_{N(x)}|\text{ on at } x} = E(R). \end{aligned} \tag{B.4}$$

Combine the above equations and the desired result follows

$$w_p(x) = \frac{\mu(w_b + w_m/\upsilon)}{\mu+\upsilon}\Big(\upsilon x + \frac{\mu}{\mu+\upsilon}(1-\exp(-(\mu+\upsilon)x))\Big). \quad \square \tag{B.5}$$

Bibliography

[1] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303, 2012.

[2] M.C. Alexiadis, P.S. Dokopoulos, and H.S. Sahsamanoglou. Wind speed and power forecasting based on spatial correlation models. *Energy Conversion, IEEE Transactions on*, 14(3):836–842, 1999.

[3] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95:3–51, 2001.

[4] S. Alizamir, F. de Véricourt, and P. Sun. Efficient feed-in-tariff policies for renewable energy technologies. Working paper, Yale University, New Haven, MA, 2012.

[5] Assembly Bill 32. The California Global Warming Solutions Act of 2006, 2006.

[6] A. Atamtürk, G. Berenguer, and Z.J.M. Shen. A conic integer programming approach to stochastic joint location-inventory problems. *Operations Research*, 60(2):366–381, 2012.

[7] AWC. Atlantic wind connection project, 2014. [http://atlanticwindconnection.com/awc-projects/project-phases/New-jersey-energy-link; accessed on 15-January-2014].

[8] L. Baringo and A.J. Conejo. Transmission and wind power investment. *Power Systems, IEEE Transactions on*, 27(2):885–893, 2012.

[9] G. Barnes and P. Langworthy. The per-mile costs of operating automobiles and trucks. Report, Minnesota Department of Transportation, Office of Research Services, St. Paul., 2003.

[10] I. Bellos, M. Ferguson, and L.B. Toktay. To sell and to provide? The economic and environmental implications of the auto manufacturer's involvement in the car sharing business, 2013. Under Review at *Manufacturing and Service Operations Management.*

[11] BME ActiveE. [Online; http://www.bmw.com.cn/cn/zh/newvehicles/1series/activee/2011/showroom/features_specs/engine_data.html].

[12] R. Botsman. Crowdshipping: using the crowd to transform delivery. *AFR Boss Magazine*, 2014.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. New York: Cambridge University Press, 2004.

[14] Beijing Transportation Research Center. Beijing transportation development annual report, 2011. [Online; http://www.bjtrc.org.cn/].

[15] J. Burton, K. Walkowicz, P. Sindler, and A.S. Duran. In-use and vehicle dynamometer evaluation and comparison of class 7 hybrid electric and conventional diesel delivery trucks. Technical Report, SAE Technical Paper, 2013.

[16] G.P. Cachon. Retail store density and the cost of greenhouse gas emissions. *Management Science*, 60(8):1907–1925, 2014.

[17] 2012-2013 ISO transmission plan, 2013. [Revised draft; https://www.caiso.com/planning/Pages/TransmissionPlanning/2012-2013TransmissionPlanningProcess.aspx; accessed on 14-Janurary-2014].

[18] V. Carbone, A. Rouquet, and C. Roussat. "Carried away by the crowd": what types of logistics characterise collaborative consumption?, 2015. [http://www.piggybee.com/upload/press/UniversiteitUtrecht.pdf; accessed on 27-September-2015].

[19] J. Carlsson and R. Devuapalli. Household level economies of scale in transportation., 2014. Working paper, University of Southern California, Los Angeles, CA.

[20] CEPA. California greenhouse gas emission inventory - 2015 edition. California Environmental Protection Agency, 2015. [http://www.arb.ca.gov/cc/inventory/data/data.htm; accessed on 8-November-2015].

[21] CFMP. California freight mobility plan, 2014. [http://www.dot.ca.gov/hq/tpp/offices/ogm/cfmp.html; accessed on 28-September-2015].

[22] K. Clement-Nyns, E. Haesen, and J. Driesen. The impact of charging plug-in hybrid electric vehicles on a residential distribution grid. *Power Systems, IEEE Transactions on*, 25(1):371–380, 2010.

[23] C.F. Daganzo. The distance traveled to visit N points with a maximum of C stops per vehicle: An analytic model and an application. *Transportation Science*, 18(4):331–350, 1984.

[24] C.F. Daganzo. *Logistics systems analysis*. New York: Springer Science & Business Media, 2005.

[25] P. Denholm and R. Sioshansi. The value of compressed air energy storage with wind in transmission-constrained electric power systems. *Energy Policy*, 37(8):3149–3158, 2009.

[26] DOE. 20 percent wind energy by 2030 - Increasing wind energy's contribution to U.S. electricity supply. US Department of Energy, 2008. [http://www.nrel.gov/docs/fy08osti/41869.pdf; accessed on 5-June-2013].

[27] L. Dow, M. Marshall, L. Xu, J.R. Agüero, and H.L. Willis. A novel approach for evaluating the impact of electric vehicles on the power distribution system. In *Power and Energy Society General Meeting, 2010 IEEE*, pp 1–6, 2010.

[28] EAC. Bottling electricity: Storage as a strategic tool for managing variability and capacity concerns in the modern grid. Electricity Advisory Committee, 2008. [http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/final-energy-storage_12-16-08.pdf; accessed on 5-June-2013].

[29] EIA. Annual energy outlook 2014. U.S. Energy Information Administration, 2014.

[30] J. Elkington. Partnerships from cannibals with forks: The triple bottom line of 21st-century business. *Environmental Quality Management*, 8(1):37–51, 1998.

[31] D. Elliott, M. Schwartz, G. Scott, S. Haymes, D. Heimiller, and R. George. Wind energy resource atlas of Mongolia. Technical Report, National Renewable Energy Lab., Golden, CO (US), 2001.

[32] D.L. Elliott, L.L. Wendell, and G.L. Gower. An assessment of the available windy land area and wind energy potential in the contiguous United States. Technical Report, Pacific Northwest Lab., Richland, WA., 1991.

[33] eMarketer. Global B2C ecommerce sales to hit $1.5 trillion this year driven by growth in emerging markets. 2014. [http://www.emarketer.com/Article/Global-B2C-Ecommerce-Sales-Hit-15-Trillion-This-Year-Driven-by-Growth-Emerging-Markets/1010575; accessed on 28-September-2015].

[34] EPA. Greenhouse gas emissions from a typical passenger vehicle, 2014. [http://www3.epa.gov/otaq/climate/documents/420f14040a.pdf; accessed on 19-November-2015].

[35] EPRI-DOE. Handbook of Energy Storage for Transmission & Distribution Applications. US Department of Energy, 2003.

[36] L.P. Fernandez, San Tomás G. Román, R. Cossent, C.M. Domingo, and P. Frias. Assessment of the impact of plug-in electric vehicles on distribution networks. *Power Systems, IEEE Transactions on*, 26(1):206–213, 2011.

[37] M. Furuhata, M. Dessouky, F. Ordóñez, M.E. Brunet, X. Wang, and S. Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46, 2013.

[38] M.D. Galus, C. Dobler, R.A. Waraich, and G. Andersson. *Predictive, distributed, hierarchical charging control of PHEVs in the distribution system of a large urban area incorporating a multi agent transportation simulation*. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme, 2011.

[39] M.D. Galus, R. Waraich, F. Noembrini, K. Steurs, G. Georges, K. Boulouchos, K.W. Axhausen, and G. Andersson. Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control. *Smart Grid, IEEE Transactions on*, 3(2):934–949, 2012.

[40] L. Gan, U. Topcu, and S.H. Low. Optimal decentralized protocol for electric vehicle charging. *Power Systems, IEEE Transactions on*, 28(2):940–951, 2013.

[41] J.B. Greenblatt, S. Succar, D.C. Denkenberger, R.H. Williams, and R.H. Socolow. Baseload wind energy: modeling the competition between gas turbines and compressed air energy storage for supplemental generation. *Energy Policy*, 35(3):1474–1492, 2007.

[42] Greenpeace. Clicking clean: how companies are creating the green internet. Greenpeace Inc., 2014. [http://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/clickingclean.pdf; accessed on 5-November-2014]

[43] C. Groër, B. Golden, and E. Wasil. A library of local search heuristics for the vehicle routing problem. *Mathematical Programming Computation*, 2(2):79–101, 2010.

[44] J.V. Hall and A.B. Krueger. An analysis of the labor market for Uber's driver-partners in the United States, 2015. [http://dataspace.princeton.edu/jspui/handle/88435/dsp010z708z67d; accessed on 19-November-2015].

[45] L. He, H.Y. Mak, Y. Rong, and Z.J.M. Shen. Service region design for urban electric vehicle sharing systems. Working paper, University of California, Berkeley, CA, 2015.

[46] M. Held, P. Wolfe, and H.P. Crowder. Validation of subgradient optimization. *Mathematical Programming*, 6(1):62–88, 1974.

[47] R. Hemmati, R.A. Hooshmand, and A. Khodabakhshian. State-of-the-art of transmission expansion planning: Comprehensive review. *Renewable and Sustainable Energy Reviews*, 23:312–319, 2013.

[48] R. Hermans, M. Almassalkhi, and I. Hiskens. Incentive-based coordinated charging control of plug-in electric vehicles at the distribution-transformer level. In *American Control Conference (ACC)*, Montréal, Canada, pp 264–269, June 2012.

[49] A. Ipakchi and F. Albuyeh. Grid of the future. *Power and Energy Magazine, IEEE*, 7(2):52–62, 2009.

[50] P. Jaillet. A priori solution of a traveling salesman problem in which a random subset of the customers are visited. *Operations Research*, 36(6):929–936, 1988.

[51] A. Kabra, E. Belavina, and K. Girotra. Bike-share systems: Accessibility and availability, 2015. Under Review at *Management Science*.

[52] J.C. Kaltenbach, J. Peschon, and E.H. Gehrig. A mathematical optimization technique for the expansion of electric power transmission systems. *Power Apparatus and Systems, IEEE Transactions on*, 89(1):113–119, 1970.

[53] J.H. Kim and W.B. Powell. Optimal energy commitments with storage and intermittent supply. *Operations Research*, 59(6):1347–1360, 2011.

[54] K.J. Kim, Y.M. Park, and K.Y. Lee. Optimal long term transmission expansion planning based on maximum principle. *Power Systems, IEEE Transactions on*, 3(4):1494–1501, 1988.

[55] M. Lammert and K. Walkowicz. Thirty-six month evaluation of UPS diesel hybrid electric delivery vans. Technical Report, National Renewable Energy Laboratory, 2012.

[56] G. Latorre, R.D. Cruz, J.M. Areiza, and A. Villegas. Classification of publications and models on transmission expansion planning. *Power Systems, IEEE Transactions on*, 18(2):938–946, 2003.

[57] B. Li, D. Krushinsky, H.A. Reijers, and T. Van Woensel. The share-a-ride problem: People and parcels sharing taxis. *European Journal of Operational Research*, 238(1):31–40, 2014.

[58] B. Li, D. Krushinsky, T. Van Woensel, and H.A. Reijers. An adaptive large neighborhood search heuristic for the share-a-ride problem. *Computers & Operations Research*, 66:170–180, 2016.

[59] F. Li, B. Golden, and E. Wasil. The open vehicle routing problem: Algorithms, large-scale test problems, and computational results. *Computers & Operations Research*, 34(10):2918–2930, 2007.

[60] J. Löfberg. Yalmip : a toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, September 2004.

[61] Z. Luo, Z. Hu, Y. Song, Z. Xu, and H. Lu. Optimal coordination of plug-in electric vehicles in power grids with cost-benefit analysis—part ii: A case study in china. *Power Systems, IEEE Transactions on*, 28(4):3556–3565, 2013.

[62] Z. Ma, D.S. Callaway, and I. Hiskens. Decentralized charging control of large populations of plug-in electric vehicles. *Control Systems Technology, IEEE Transactions on*, 21(1):67–78, 2013.

[63] H.Y. Mak, Y. Rong, and Z.J.M. Shen. Infrastructure planning for electric vehicles with battery swapping. *Management Science*, 59(7):1557–1575, 2013.

[64] T. Mason, T. Curry, and D. Wilson. Capital costs for transmission and substations: recommendations for WECC transmission expansion planning. *Black & Veatch*, 2012.

[65] T. Mason, T. Curry, and D. Wilson. Transmission capital cost for WECC-TEPPC, 2012. [Unpublished; accessed on 4-April-2015].

[66] M. Mendonca, D. Jacobs, and B. K. Sovacool. *Powering the Green Economy: The Feed-in Tariff Handbook*. Earthscan, 2009.

[67] M. Moeini-Aghtaie, A. Abbaspour, and M. Fotuhi-Firuzabad. Incorporating large-scale distant wind farms in probabilistic transmission expansion planning - part i: Theory and algorithm. *Power Systems, IEEE Transactions on*, 27(3):1585–1593, 2012.

[68] A.H. Mohsenian-Rad, V.W.S. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia. Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *Smart Grid, IEEE Transactions on*, 1(3):320–331, 2010.

[69] K. Natarajan, M. Sim, and J. Uichanco. Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, 20(4):695–731, 2010.

[70] A. Nedić and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *Automatic Control, IEEE Transactions on*, 54(1):48–61, 2009.

[71] NREL. Western wind power resources database. National Renewable Energy Laboratory, 2008.

[72] NREL. Lifecycle cost analysis of hydrogen versus other technologies for electrical energy storage. National Renewable Energy Laboratory, 2012.

[73] H. Oh. Optimal planning to include storage devices in power systems. *Power Systems, IEEE Transactions on*, 26(3):1118–1128, 2011.

[74] PWC. The sharing economy - consumer intelligence series. PricewaterhouseCoopers LLP., 2015. [https://www.pwc.com/us/en/technology/publications/assets/pwc-consumer-intelligence-series-the-sharing-economy.pdf; accessed on 25-September-2015].

[75] W. Qi, J. Liu, X. Chen, and P.D. Christofides. Supervisory predictive control of standalone wind/solar energy generation systems. *Control Systems Technology, IEEE Transactions on*, 19(1):199–207, 2011.

[76] Renewables 2011 global status report. Renewable Energy Policy Network for the 21st Century, 2011.

[77] J. Rogers, S. Fink, and K. Porter. Examples of wind energy curtailment practices. *Subcontract Report NREL/SR-550*, 48737, 2010.

[78] S.M. Ross. *Stochastic processes*, volume 2. New York: John Wiley & Sons, 1996.

[79] N. Rotering and M. Ilic. Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets. *Power Systems, IEEE Transactions on*, 26(3):1021–1029, 2011.

[80] J.F. Rougès and B. Montreuil. Crowdsourcing delivery: New interconnected business models to reinvent delivery. In *First International Physical Internet Conference*, 2014.

[81] B. Schaefer and D. Konur. Economic and environmental considerations in a continuous review inventory control system with integrated transportation decisions. *Transportation Research Part E: Logistics and Transportation Review*, 80:142–165, 2015.

[82] S.M. Schoenung and W.V. Hassenzahl. Long-vs. short-term energy storage technologies analysis. A life-cycle cost study. A study for the DOE Energy Storage Systems program. Sandia National Laboratories, 2003.

[83] Z.J.M. Shen, C. Coullard, and M.S. Daskin. A joint location-inventory model. *Transportation Science*, 37(1):40–55, 2003.

[84] J. Shu, M.C. Chou, Q. Liu, C.P. Teo, and I.L. Wang. Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research*, 61(6):1346–1359, 2013.

[85] L.V. Snyder and Z.J.M Shen. *Fundamentals of supply chain theory*. New York: John Wiley & Sons, 2011.

[86] Y. Song, X. Yang, and Z. Lu. Integration of plug-in hybrid and electric vehicles: experience from china. In *Power and Energy Society General Meeting, 2010 IEEE*, Detroit, MI, pp 1–6, July 2010.

[87] E. Sortomme, M.M. Hindi, S.D.J. MacPherson, and S.S. Venkata. Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses. *Smart Grid, IEEE Transactions on*, 2(1):198–205, 2011.

[88] P.T. Staats, W.M. Grady, A. Arapostathis, and R.S. Thallam. A procedure for derating a substation transformer in the presence of widespread electric vehicle battery charging. *Power Delivery, IEEE Transactions on*, 12(4):1562–1568, 1997.

[89] W. Su and M.Y. Chow. Performance evaluation of an eda-based large-scale plug-in hybrid electric vehicle charging algorithm. *Smart Grid, IEEE Transactions on*, 3(1):308–315, 2012.

[90] O. Sundström and C. Binding. Flexible charging optimization for electric vehicles considering distribution grid constraints. *Smart Grid, IEEE Transactions on*, 3(1):26–37, 2012.

[91] I. Talinli, E. Topuz, E. Aydin, and S. Kabakci. A holistic approach for wind farm site selection by using FAHP, 2011. [http://www.intechopen.com/books/wind-farm-technical-regulations-potential-estimation-and-siting-assessment/a-holistic-approach-for-wind-farm-site-selection-by-using-fahp; accessed on 17-September-2013].

[92] J.A. Taylor and F.S. Hover. Conic AC transmission system planning. *Power Systems, IEEE Transactions on*, 28(2):952–959, 2013.

[93] G.C. Thomann and M.J. Barfield. The time variation of wind speeds and wind farm power output in kansas. *Energy Conversion, IEEE Transactions on*, 3(1):44–49, 1988.

[94] H.C. Tijms. *A first course in stochastic models*. New York: John Wiley and Sons, 2003.

[95] TOU (Time-of-use) power prices in Beijing (Summer). [Online; http://www.bjpc.gov.cn/tztg/200911/P020091121010684247161.xls].

[96] UPS. See more, hear less: New UPS vehicles in Germany. UPS official blog., 2013. [http://blog.ups.com/2013/01/09/see-more-hear-less-new-ups-vehicles-in-germany/; accessed on 19-November-2015].

[97] U.S. Department of Transportation. National transportation statistics. Table 4-23: Average fuel efficiency of U.S. light duty vehicles., 2015. [http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_04_23.html; accessed on 18-November-2015].

[98] C.K. Wen, J.C. Chen, J.H. Teng, and P. Ting. Decentralized plug-in electric vehicle charging selection algorithm in power systems. *Smart Grid, IEEE Transactions on*, 3(4):1779–1789, 2012.

[99] L.C. Wolstenholme. *Reliability modeling: a statistical approach*. Florida: CRC Press, 1999.

[100] W. Xie and Y. Ouyang. Optimal layout of transshipment facility locations on an infinite homogeneous plane. *Transportation Research Part B: Methodological*, 75:74–88, 2015.

[101] Z. Xu, Z. Hu, Y. Song, Z. Luo, K. Zhan, and J. Wu. Coordinated charging strategy for pevs charging stations. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8. IEEE, 2012.

[102] F. Zhang, Z. Hu, and Y. Song. Mixed-integer linear model for transmission expansion planning with line losses and energy storage systems. *IET Generation, Transmission and Distribution*, 7(8):919–928, 2013.