Data-Driven Robust Taxi Dispatch under Demand Uncertainties

Fei Miao, Shuo Han, Shan Lin, Qian Wang, John Stankovic, Abdeltawab Hendawi, Desheng Zhang, Tian He, George J. Pappas

***Abstract*—In modern taxi networks, large amounts of taxi occupancy status and location data are collected from networked in-vehicle sensors in real-time. They provide knowledge of system models on passenger demand and mobility patterns for efficient taxi dispatch and coordination strategies. Such approaches face new challenges: how to deal with uncertainties of predicted customer demand while fulfilling the system's performance requirements, including minimizing taxis' total idle mileage and maintaining service fairness across the whole city; how to formulate a computationally tractable problem. To address this problem, we develop a data-driven robust taxi dispatch framework to consider spatial-temporally correlated demand uncertainties. The robust vehicle dispatch problem we formulate is concave in the uncertain demand and convex in the decision variables. Uncertainty sets of random demand vectors are constructed from data based on theories in hypothesis testing, and provide a desired probabilistic guarantee level for the performance of robust taxi dispatch solutions. We prove equivalent computationally tractable forms of the robust dispatch problem using the minimax theorem and strong duality. Evaluations on four years of taxi trip data for New York City show that by selecting a probabilistic guarantee level at 75%, the average demand-supply ratio error is reduced by 31.7%, and the average total idle driving distance is reduced by 10.13% or about 20 million miles annually, compared with non-robust dispatch solutions.**

I. Introduction

Modern transportation systems are equipped with various sensing technologies for passenger and vehicle tracking, such as radio-frequency identification (RFID), global positioning system (GPS), and occupancy status sensing systems. Sensing data collected from transportation systems provides us opportunities for understanding spatial-temporal patterns of passenger demand. Methods of predicting taxi-passenger demand [23], [29], travel time [3], [16], [28] and traveling speed [2] according to traffic monitoring data have been developed. Simple travel time predictors were demonstrated to come close to fundamental error bounds in delay prediction [14].

This work was supported by NSF CPS-1239152, Project Title: CPS: Synergy: Collaborative Research: Multiple-Level Predictive Control of Mobile Cyber Physical Systems with Correlated Context. Shuo Han was supported in part by the NSF (CNS-1239224) and TerraSwarm, one of six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA. Part of the results of this work appeared at the 54th IEEE Conference on Decision and Control, Osaka, Japan, December 2015 [20].

F. Miao, S. Han and G. J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA 19014. Email:{miaofei, hanshuo, pappasg}@seas.upenn.edu. S. Lin is with Department of Electrical and Computer Engineering, Stony Brook University, Long Island, NY, USA 11794. Email: shan.x.lin@stonybrook.edu. Q. Wang is with China Chengxin Credit Information Co., Beijing, China, 100033. Email:qianwangchina@gmail.com. J. Stankovic and A. Hendawi are with the Department of Computer Science, University of Virginia, Charlottesville, VA, USA, 22904. Email: {stankovic, hendawi}@virginia.edu. D. Zhang and T. He are with Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Email:{tianhe, zhang}@cs.umn.edu.

Based on such rich spatial-temporal information about passenger mobility patterns and demand, many control and coordination solutions have been designed for intelligent transportation systems. Coverage control algorithms to allocate groups of vehicles [10], robotic load re-balancing policies that minimize the number of re-balancing trips for mobility-on-demand systems [25], [31], and smart parking systems that allocates resource based on a driver's cost function [15] have been proposed. Dispatch algorithms that aim to minimize customers' waiting time [18], [27] or to reduce cruising mile [30] have been developed. Miao et. al design a Receding Horizon Control (RHC) framework of dispatching taxis towards both current and predicted future demand, balancing taxi supply throughout the city and reducing idle cruising distance [21], [22]. Considering future demand when making the current dispatch decisions helps to reduce autonomous vehicle balancing costs [31] and taxis' total idle distance [21]. Strategies for resource allocation depend on the model of demand in general, and the knowledge and assumptions about the demand affect the performance of the supply-providing approaches [9], [24]. These works heavily rely on precise passenger-demand models to make dispatch decisions.

However, passenger-demand models have their intrinsic model uncertainties that result from many factors, such as weather, passenger working schedule, and city events etc. Algorithms that do not consider these uncertainties can lead to inefficient dispatch services, resulting in long waiting times of under-served passengers, imbalanced workloads, and increased taxi idle mileage. Although robust optimization aims to minimize the worst-case cost under all possible random parameters, it sacrifices average system performances [1]. For a taxi dispatch system, it is essential to address the trade-off between worst-case and the average dispatch costs under uncertain demand. A promising yet challenging approach is a robust dispatch framework with an uncertain demand model, called an uncertainty set, that captures spatial-temporal correlations of demand uncertainties and the robust optimal solution under this set provides a probabilistic guarantee for the dispatch cost (as defined in problem (12)).

In this work, we consider two aspects of a robust vehicle dispatch model given a taxi-operational records dataset: (1) how to formulate a robust resource allocation problem that dispatches vacant vehicles towards predicted uncertain demand, and (2) how to construct spatial-temporally correlated uncertain demand sets for this robust resource allocation problem without sacrificing too much average performance of the system. We have the freedom to specify a lower bound value for the probability that an actual dispatch cost under the true demand being smaller than the optimal cost of the robust dispatch solutions. The data-driven robust dispatch model we

design allows us to find a better solution for considering the trade-off between the average dispatch cost and the minimum cost under the worst-case scenario than previous methods without considering the trade-off.

We first develop the objective and constraints of a robust dispatch problem considering spatial-temporally correlated demand uncertainties. The objective of a system-level optimal dispatch solution is balancing workload of taxis in each region of the entire city with minimum total current and expected future idle cruising distance. We then design a data-driven algorithm for constructing uncertainty demand sets without assumptions about the true model of the demand vector. The constructing algorithm is based on hypothesis testing theories [6] [12] [26], however, how to apply these theories for spatial-temporally correlated transportation data and uncertainty sets of a robust vehicle resource allocation problem have not been explored before. To the best of our knowledge, this is the first work to design a robust vehicle dispatch model that provides a desired probabilistic guarantee using predictable and realistic demand uncertainty sets.

With two types of uncertainty sets — box type and second-order-cone (SOC) type, we prove equivalent convex optimization forms of the robust dispatch problem via the minimax theorem and the strong duality theorem. The robust dispatch problem formulated in this work is convex over the decision variables and concave over the constructed uncertain sets with decision variables on the denominators. This form is not the standard form (i.e., linear programming (LP) or semi-definite programming (SDP) problems) that has already been covered by previous work [4], [6], [11]. With proofs shown in this work, both system performance and computational tractability are guaranteed under spatial-temporal demand uncertainties.

The main contribution of our previous work [21] (a journal version of [22]) is to design a receding horizon control (RHC) framework that incorporates predicted demand model and real-time sensing data. The robust taxi dispatch problem formulated in [21], [22] is convex of the uncertain demand, and only solvable in polynomial time when the uncertain demand is described as a range type. In contrast, we define an approximation of the balanced vehicle objective in this work, such that the robust vehicle dispatch problem is concave of the uncertain demand and convex of the decision variables. Furthermore, we explicitly design an algorithm to build demand uncertainty set from data according to different probabilistic guarantee level for the cost. We then prove equivalent computationally tractable forms of the robust dispatch problem under both polytope type (including range) and SOC type of uncertainty demand models. The average performance of the robust taxi dispatch solutions with SOC type of uncertain demand set is better compared with that of the box (range) type of uncertainty set in the evaluations based on data. Hence, it is critical to use a more complex type of uncertainty set, the SOC type, and the corresponding robust dispatch model we design in this work. The contributions of this work are:

- We develop a robust optimization model for taxi dispatch systems under spatial-temporally correlated uncertainties of predicted demand, and define an approximation of the balanced vehicle objective. The robust optimization problem of approximately balancing vacant taxis with least total idle distance is concave of the uncertain demand, convex of the decision variables and computationally tractable under multiple types of uncertainties.
- We design a data-driven algorithm to construct uncertainty sets that provide a desired level of probabilistic guarantee for the robust taxi dispatch solutions.
- We prove that there exist equivalent computationally tractable convex optimization forms for the robust dispatch problem with both polytope and second-order-cone (SOC) types of uncertainty sets constructed from data.
- Evaluations on four years of taxi trip data in New York City show that the SOC type of uncertain set provides a smaller average dispatch cost than the polytope type. The average demand-supply ratio mismatch is reduced by 31.7%, and the average total idle distance is reduced by 10.13% or about 20 million miles annually with robust dispatch solutions under the SOC type of uncertainty set.

The rest of the paper is organized as follows. The taxi dispatch problem is described and formulated as a robust optimization problem given a closed and convex uncertainty set in Section II. We design an algorithm for constructing uncertain demand sets based on taxi operational records data in Section III. Equivalent computationally tractable forms of the robust taxi dispatch problem given different forms of uncertainty sets are proved in Section IV. Evaluation results based on a real data set are shown in Section V. Concluding remarks are provided in Section VI.

NOTATIONS

For any vector x, we denote by x^T the transpose of x, and x_i as the i-th component of x. We denote $\mathbf{1}_n \in \mathbb{R}^n$ as a vector of all ones.

II. PROBLEM FORMULATION

The goal of taxi dispatch is to direct vacant taxis towards current and predicted future requests with minimum total idle mileage. There are two objectives. One is sending more taxis for more requests to reduce mismatch between supply and demand across all regions in the city. The other is to reduce the total idle driving distance for picking up passengers in order to save cost. Involving predicted future demand when making current decisions benefits to increasing total profits, since drivers are able to travel to regions with better chances to pick up future passengers. In this section, we formulate a taxi dispatch problem with uncertainties in the predicted spatial-temporal patterns of demand. A typical monitoring and dispatch infrastructure is shown in Figure 1. The dispatch center periodically collects and stores real-time information such as GPS location, occupancy status and road conditions; dispatch solutions are sent to taxis via cellular radio.

A. Problem description

We discretize time and space to describe the demand model and dispatch decisions in the problem formulation for computational efficiency. We assume that the entire city is

Parameters of (11)	Description
n	the number of regions
τ	model predicting time horizon
$r^k \in \Delta_k$	the uncertain total number of requests at each region during time k
$W \in \mathbb{R}^{n\times n}$	weight matrix, W_{ij} is the distance from region i to region j
$P^k \in [0,1]^{n\times n}$	probability matrix that describes taxi mobility patterns during one time slot
$L^1 \in \mathbb{N}^n$	the initial number of vacant taxis at each region provided by GPS and occupancy status data
$m \in \mathbb{R}^+$	the upper bound of distance each taxi can drive idly for picking up a passenger
$\alpha \in \mathbb{R}_+$	the power on the denominator of the cost function
$\beta \in \mathbb{R}_+$	the weight factor of the objective function
Variables of (11)	
$X_{ij}^k \in \mathbb{R}_+$	the number of taxis dispatched from region i to region j during time k
$L^k \in \mathbb{R}_+^n$	the number of vacant taxis at each region before dispatching at the beginning of time k
Parameters of Algorithm 1	
$r_c \in \Delta$	the uncertain concatenated demand vector of τ consecutive time slots
$\tilde{r}_c(d_l, t, I_p)$	one sample of $r_c(t)$ according to sub-dataset I_p, records of date d_l
α_h	significance level of a hypothesis testing

Table I
PARAMETERS AND VARIABLES OF TAXI DISPATCH PROBLEM (11).

divided into n regions, and time of one day is discretized to time slots indexed by $t = 1, 2, \ldots, K$. Taxi dispatch decision is calculated in a receding horizon process, considering future demand when making the current dispatch decisions helps to reduce resource re-allocating costs [31] and taxis' total idle distance [21]. At time t, we consider the effects of current decision to the following τ time slots, and use demand model for time $(t, t+1, \ldots, t+\tau-1)$ respectively to get dispatch solutions of τ time slots. Then only the solution of $k = 1$ for time t is implemented, while the solutions for remaining time slots are not materialized. When the time horizon rolls forward by one time step from t to $(t+1)$, information about vehicle locations and occupancy status are observed and updated, $k = 1$ represents for the time slot $t+1$ during one day and we calculate a new dispatch solution for $(t+1)$.

Typically, it is difficult to predict a deterministic value of passenger demand of a region during specific time. With prior knowledge and data sets, we assume that the passenger demand model is described by uncertainty vectors belonging to closed and convex, or compact sets defined as

$$r^k \in \Delta_k \subset \mathbb{R}_+^n, \quad k = 1, \ldots, \tau,$$

where r_j^k is the number of total requests within region j during time k, and τ is the model predicting time horizon. Here we relax the integer constraint of $r_j^k \in \mathbb{N}$ to positive real, since the integer constraint will make the uncertainty set Δ_k non-convex and the robust dispatch problem we formulate in this section not computationally tractable. The total number of requests at region j may have similar patterns as its neighbors, for instance, during busy hours, several regions locate in downtown area may all have peak demand. This type of spatial correlations of demand across each region during the same time slot k is reflected by the correlation of each element of r^k. Meanwhile, demand during several consecutive time slots $r^k, k = 1, \ldots, \tau$ may show similar characteristics such as busy hours and be temporally correlated. We describe both spatial and temporal correlations by one set Δ for r^k, $k = 1, \ldots, \tau$, and define a concatenated demand vector as

$$r_c = \left[(r^1)^T, (r^2)^T, \cdots, (r^\tau)^T\right]^T \in \Delta \subset \mathbb{R}_+^{\tau n},$$

and each closed, convex set Δ_k is a projection of Δ

$$\Delta_k := \{r^k \mid \exists r^1, \ldots, r^{k-1}, r^{k+1}, \ldots, r^\tau, \text{ s.t. } r_c \in \Delta\}.$$

The closed and convex form of Δ depends on the method and theory applied to construct the uncertainty set, which we will describe in detail in Section III.

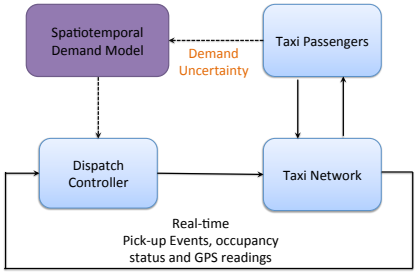

Figure 1. A prototype of the taxi dispatch system

Considered as one type of resource allocation problem, the basic idea of a robust dispatch model that balances taxis' supply in a network flow model is described in Figure 2. The dispatch framework decides the amount of vacant taxis that should traverse between each node pair according to the demand at each node according to control requirements and practical constraints. The edge weight of the graph represents the distance between two regions. Specifically, each region has an initial number of vacant taxis provided by real-time sensing information and an uncertain predicted demand.

We define a non-negative decision variable matrix $X^k \in \mathbb{R}_+^{n\times n}$, $X_{ij}^k \geq 0$, where X_{ij}^k is the number of taxis (amount of resource) dispatched from region i to region j. We relax the integer constraint of $X_{ij}^k \in \mathbb{N}$ to a non-negative real constraint, since mixed integer programming is not computational efficient for a large-scale robust optimization problem. In this work, every time when making taxi resource allocation decision, we consider the following robust optimization problem

$$\begin{aligned} &\min_{X^1} \max_{r^1\in\Delta_1} \min_{X^2} \max_{r^2\in\Delta_2} \ldots \min_{X^\tau} \max_{r^\tau\in\Delta_\tau} \\ &J = \sum_{k=1}^{\tau} (J_D(X^k) + \beta J_E(X^k, r^k)) \\ &\text{s.t.} \quad X^k \in \mathcal{D}_c, \end{aligned} \tag{1}$$

where J_D is a convex cost function for allocating or re-allocating resources, J_E is a function concave in r^k and convex in X^k that measures the service fairness of the resource allocating strategy, and $\mathcal{D}_c$ is a convex domain of the decision variables that describes the constraints of the resource

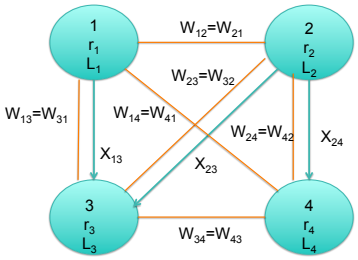

Figure 2. A network flow model of the robust taxi dispatch problem. A circle represents a region with region ID $1, 2, 3, 4$. We omit the superscript of time k since every parameter is for one time slot only. Uncertain demand is denoted by r_i, L_i is the original number of vacant taxis before dispatch at region i, and X_{ij} is a dispatch solution that sending the number of vacant taxis from region i to region j with the distance W_{ij}.

allocating strategies. We define specific formulations of the objective and constraint functions for a robust taxi dispatch problem in the rest of this section.

B. Robust taxi dispatch problem formulation

Estimated cross-region idle-driving distance: When traversing from region i to region j, taxi drivers take the cost of cruising on the road without picking up a passenger till the target region. Hence, we consider to minimize this kind of idle driving distance while dispatching taxis. We define the weight matrix of the network in Fig. 2 as $W \in \mathbb{R}^{n\times n}$, where W_{ij} is the distance between region i and region j. The across-region idle driving cost according to X^k is

$$J_D(X^k) = \sum_i \sum_j X_{ij}^k W_{ij}. \tag{2}$$

We assume that the region division method is time-invariant in this work, and W is a constant matrix for the optimization problem formulation – for instance, the value of W_{ij} represents the length of shortest path on streets from the center of region i to the center of region j [1].

The distance every taxi can drive should be bounded by a threshold parameter $m \in \mathbb{R}^+$ during limited time

$$X_{ij}^k = 0 \text{ if } W_{ij} > m,$$

which is equivalent to

$$X_{ij}^k \geqslant 0, \quad X_{ij}^k W_{ij} \leq m X_{ij}^k, \quad \forall i, j \in \{1, \ldots, n\}. \tag{3}$$

To explain this, assume the constraint (3) holds. If $W_{ij} > m$ and $X_{ij}^k > 0$, we have $X_{ij}^k W_{ij} > m X_{ij}^k$, which contradicts to (3). The threshold m is related to the length of time slot and traffic conditions on streets. For instance, with an estimated average speed of cars in one city during time $k = 1, \ldots, \tau$, and idle driving time to reach a dispatched region is required to be less than 10 minutes, then the value of m should be the distance one taxi can drive during 10 minutes with the current average speed on road (m can also be dependent on k, denoted as m_k if a different average speed during each time slot k can be monitored or predicted).

Metric of serving quality: We design the metric of service quality as a function $J_E(X^k, r^k)$ concave in r^k and convex in X^k in this work for computational efficiency [4]. Besides vacant taxis traverse to region j according to matrix X^k, we define $L_j^k \in \mathbb{R}_+$ as the number of vacant taxis at region j before dispatching at the beginning of time k, and $L^k \in \mathbb{R}_+^n$, and $L^1 \in \mathbb{R}_+^n$ is provided by real-time sensing information. We assume that the total number of vacant taxis is greater than the number of regions, i.e., $N^k \geqslant n$, and each region should have at least one vacant taxi after dispatch. Then the total number of vacant taxis at region i during time k satisfies that

$$\sum_{j=1}^n X_{ji}^k - \sum_{l=1}^n X_{ij}^k + L_i^k > 0, \tag{4}$$

$$\sum_{i=1}^n \left(\sum_{j=1}^n X_{ji}^k - \sum_{j=1}^n X_{ij}^k + L_i^k \right) = \sum_{i=1}^n L_i^k = N^k. \tag{5}$$

One service metric is fairness over all regions, or that the demand-supply ratio of each region equals to that of the whole city. A balanced distribution of vacant taxis is an indication of good system performance from the perspective that a customer's expected waiting time is short as shown by a queuing theoretic model in [31]. Meanwhile, a balanced demand-supply ratio means that regions with less demand will also get less resources, and idle driving distance will also be reduced in regions with more supply than demand if we pre-allocate possible redundant supply to those regions in need. We aim to minimize the mismatch value or the total difference between local demand-supply ratio at each region and the global demand-supply ratio of the whole city, similarly as the objective defined in [21], [22]

$$\sum_{k=1}^{\tau} \sum_{i=1}^{n} \left| \frac{r_i^k}{\sum_{j=1}^n X_{ji}^k - \sum_{j=1}^n X_{ij}^k + L_i^k} - \frac{\sum_{j=1}^n r_j^k}{N^k} \right|. \tag{6}$$

However, the function (6) is not concave in r^k for any X^k. It is worth noting we need a function $J_E(X^k, r^k)$ concave in r^k for any X^k, and convex in X^k for any r^k, to make sure the robust optimization problem is computationally tractable. Hence, we define

$$J_E(X^k, r^k) = \sum_i \frac{r_i^k}{\left(\sum_{j=1}^n X_{ji}^k - \sum_{j=1}^n X_{ij}^k + L_i^k \right)^\alpha}, \quad \alpha > 0 \tag{7}$$

as a service fairness metric to minimize. This is because we approximately minimize (6) by minimizing (7) under the constraints (4) and (5) with an α value chosen according to the desired approximation level, and the following Lemma explains this approximation.

Lemma 1: Given deterministic demand vectors $(r^1, \ldots, r^\tau)$ and initial number of vacant vehicles before dispatch $(L^1, \ldots, L^\tau)$ that satisfy $r_i^k \geqslant 1$, $L_i^k \geqslant 0$, $\sum_{i=1}^n L_i^k = N^k$, for any $\epsilon_0 > 0$, any $i \in \{1, \ldots, n\}$, $k = 1, \ldots, \tau$, there exists an $\alpha > 0$, such that the optimal solution $(X^k)^*$ by minimizing (7)

[1]For control algorithms with a dynamic region division method, the distance matrix can be generalized to a time dependent matrix W^k as well.

under constraints (4) and (5) satisfies

$$\left|\frac{r_i^k}{\sum_{j=1}^{n}(X_{ji}^k)^* - \sum_{j=1}^{n}(X_{ij}^k)^* + L_i^k} - \frac{\sum_{j=1}^{n} r_j^k}{N^k}\right| < \epsilon_0, \ \textit{and}$$

$$\sum_{k=1}^{\tau}\sum_{i=1}^{n}\left|\frac{r_i^k}{\sum_{j=1}^{n}(X_{ji}^k)^* - \sum_{j=1}^{n}(X_{ij}^k)^* + L_i^k} - \frac{\sum_{j=1}^{n} r_k}{N^k}\right| < n\tau\epsilon_0. \tag{8}$$

□

Proof: See Appendix VII-A. ■

According to the proof, we can always choose α to be small enough (or close enough to 0) in order to obtain a desired level of approximation ϵ_0. Hence, in the experiments of Section V, we numerically choose $\alpha = 0.1$ based on simulation results. Therefore, with function (7), we map the objective of balancing supply according to demand across every region in the city to a computationally tractable function that concave in the uncertain parameters and convex in the decision variables for a robust optimization problem.

The number of initial vacant taxis L_j^{k+1} depends on the number of vacant taxis at each region after dispatch during time k and the mobility patterns of passengers during time k, while we do not directly control the latter. We define P_{ij}^k as the probability that a taxi traverses from region i to region j and turns vacant again (after one or several drop off events) around the beginning of time $k+1$, provided it is vacant at the beginning of time k. Examples of getting P_{ij}^k based on data include but not limited to methods of describing trip patterns of taxis [22] and autonomous mobility on demand systems [31]. Then the number of vacant taxis within region j by the end of time k is $\left(\mathbf{1}_n^T X^k - (X^k\mathbf{1}_n)^T + (L^k)^T\right)P_{\cdot j}^k$, where $P_{\cdot j}^k$ is the j-th column of P^k, and

$$(L^{k+1})^T = (\mathbf{1}_n^T X^k - (X^k\mathbf{1}_n)^T + (L^k)^T)P^k. \tag{9}$$

Weighted-sum objective function: Since there exists a trade-off between two objectives, we define a weighted-sum with parameter $\beta > 0$ of the two objectives $J_D(X^k)$ defined in (2) and $J_E(X^k, r^k)$ defined in (7) as the objective function. Let $X^{1:\tau}$ and $L^{2:\tau}$ represent Without considering model uncertainties corresponding to r^k, a convex optimization form of taxi dispatch problem is

$$\begin{aligned}\min_{X^{1:\tau},L^{2:\tau}} \quad & J = \sum_{k=1}^{\tau}(J_D(X^k) + \beta J_E(X^k, r^k)) \\ \text{s.t} \quad & (3),\ (4),\ (9).\end{aligned} \tag{10}$$

Robust taxi dispatch problem formulation: We aim to find out a dispatch solution robust to an uncertain demand model in this work. For time $k = 1, \ldots, \tau$, uncertain demand r^k only affects the dispatch solutions of time $(k, k+1, \ldots, \tau)$, and dispatch solution at $k+\tau$ is related to uncertain demand at $(k+1, \ldots, \tau)$, similar to the multi-stage robust optimization problem in [7]. However, the control laws considered in [7] are polynomial in past-observed uncertainties; in this work, we do not restrict the decision variables to be any forms of previous-observed uncertain demands. The dispatch decisions are numerical optimal solution of a robust optimization problem. With a list of parameters and variables shown in Table I, considering both the current and future dispatch costs when making the current decisions, we define a robust taxi dispatch problem as the following

$$\begin{aligned}&\min_{X^1}\ \max_{r^1\in\Delta_1}\ \min_{X^2,L^2}\ \max_{r^2\in\Delta_2}\ \cdots\ \min_{X^\tau,L^\tau}\ \max_{r^\tau\in\Delta_\tau} \\ &J = \sum_{k=1}^{\tau}(J_D(X^k) + \beta J_E(X^k, r^k)) \\ &\quad = \sum_{k=1}^{\tau}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} X_{ij}^k W_{ij} + \frac{\beta r_i^k}{\left(\sum_{j=1}^{n} X_{ji}^k - \sum_{j=1}^{n} X_{ij}^k + L_i^k\right)^\alpha}\right) \\ &\text{s.t.} \quad (L^{k+1})^T = (\mathbf{1}_n^T X^k - (X^k\mathbf{1}_n)^T + (L^k)^T)P^k, \\ &\quad \mathbf{1}_n^T X^k - (X^k\mathbf{1}_n)^T + (L^k)^T \geqslant \mathbf{1}_n^T, \\ &\quad X_{ij}^k W_{ij} \leq m X_{ij}^k, \\ &\quad X_{ij}^k \geq 0, \quad i, j \in \{1, 2, \ldots, n\}.\end{aligned} \tag{11}$$

After getting an optimal solution $(X^1)^*$ of (11), we adjust the solution by rounding methods to get an integer number of taxis to be dispatched towards corresponding regions. It does not affect the optimality of the result much in practice, since the objective function is related to the demand-supply ratio of each region. A feasible integer solution of (11) always exists, since $X_{ij}^k = 0, \ \forall i, j, k$ is feasible.

III. Algorithm For Constructing Uncertain Demand Sets

With many factors affecting taxi demand during different time within different areas of a city, explicitly describing the model is a strict requirement and errors of the model will affect the performance of dispatch frameworks. Considering future demand uncertainties benefits for minimizing worst-case demand-supply ratio mismatch error and idle distance described as shown in [21], [22]. However, the uncertainty set constructed by only using a standard deviation range [21], [22] cannot tell how possible the true real-world cost is smaller than the optimal cost. Hence, with a large amount of taxi operational records data, it is essential to construct a model that captures the spatial-temporal demand uncertainties and provides a probabilistic guarantee about the true possible values of costs by solving robust dispatch problem (11).

Then given a dataset, the algorithm for constructing uncertainty sets includes three main steps—getting a sample set of r_c from the original dataset and partition the sample set, bootstrapping a threshold for the test statistics according to the requirement of the probability guarantee, and calculating the model of uncertainty sets based on the thresholds. In this section, we explain each step, summarize the process in Algorithm 1, and discuss factors to consider for choosing parameters of the algorithm. Numerical examples are shown in Section V.

A. An uncertainty set with probabilistic guarantee

For convenience, we concisely denote all the variables of the taxi dispatch problem as x. Assume that we do not have knowledge about the true distribution $\mathbb{P}^*(r_c)$ of the random demand vector r_c. When the uncertainty parameter is included in the objective function $J(r_c, x)$ of problem (11), the probabilistic guarantee for the event that the true dispatch cost being smaller than the optimal dispatch cost is described by the following chance constrained problem

$$\begin{aligned} \min_{x} \quad & M \\ \text{s.t} \quad & P_{r_c \sim \mathbb{P}^*(r_c)}(f(r_c, x) = J(r_c, x) - M \leqslant 0) \geqslant 1 - \epsilon. \end{aligned} \tag{12}$$

The constraint f and objective function J are concave in r_c for any x, and convex in x for any r_c. Without loss of generality about the objective and constraint functions, equivalently we aim to find solutions of the following form of chance constrained problem

$$\begin{aligned} \min_{x} \quad & J(r_c, x) \\ \text{subject to} \quad & P_{r_c \sim \mathbb{P}^*(r_c)}(f(r_c, x) \leqslant 0) \geqslant 1 - \epsilon. \end{aligned} \tag{13}$$

When it is difficult to explicitly estimate $\mathbb{P}^*(r_c)$, given constraints $f(r_c, x)$ that concave in r_c for any x, we solve the following robust optimization problem such that optimal solutions of (14) satisfy the probabilistic guarantee of constraints for problem (13)

$$\min_{x} \max_{r_c \sim \Delta} \quad J(r_c, x), \quad \text{subject to} \quad f(r_c, x) \leqslant 0, \tag{14}$$

Then r_c of problem (14) can be any vector in the uncertainty set Δ instead of a random vector in problem (13), and we require that by solving an optimization problem with this constrained uncertain set performance of optimal solutions is guaranteed for $r_c \sim \mathbb{P}^*$. Another requirement is that the robust optimization problem is computationally tractable problem with this uncertainty set. Hence, the uncertainty set for problem (14) that keeps the optimal solution of (14) satisfying the constraints of problem (13) is defined as the following:

Problem 1: Construct an uncertainty set $\Delta, r_c \in \Delta$, given a value of $0 < \epsilon < 1$ and a data set of random vectors r_c, such that

(P1). The robust constraint (14) is computationally tractable.

(P2). The set Δ implies a probabilistic guarantee for the true distribution $\mathbb{P}^*(r_c)$ of a random vector r_c at level ϵ, that is, for any optimal solution $x^* \in \mathbb{R}^k$ and for any function $f(r_c, x)$ concave in r_c, we have the implication:

$$\begin{aligned} &\text{If } f(r_c, x^*) \leq 0, \quad \text{for } \forall r_c \in \Delta, \\ &\text{then } \mathbb{P}^*_{r_c \sim \mathbb{P}^*(r_c)}(f(r_c, x^*) \leqslant 0) \geq 1 - \epsilon. \end{aligned} \tag{15}$$

The given probabilistic guarantee level ϵ is related to the degree of conservativeness of the robust optimization problem. The trade-off between the average cost of robust optimal solutions and the probabilistic level is shown by evaluations in Section V.

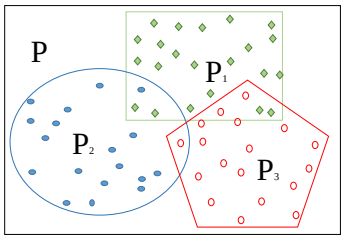

Figure 3. Intuition for partitioning the whole dataset. When the data set includes data from three distributions P_1, P_2, P_3, without prior knowledge, we can build a larger uncertainty set that describes the range of all samples in the dataset. The problem is that the uncertainty set is not accurate enough.

B. Aggregating demand and partition the sample set

We assume that one day is discretized as K time slots in total, and the demand of each region during one time slot is described as $r^k, k = 1, \ldots, K$. Then every τ discretized time slots of r^k, $k = t, \ldots, t + \tau$ are concatenated to a vector $r_c(t)$. The first step is to transform the original dataset of taxi operational records to a dataset of sampled vector $\tilde{r}_c(d, t)$ of different dates d for each index t. For instance, assume we choose the length of each time slot as one hour, and the dataset records all trip information of taxis during each day. According to the start time and GPS coordinate of the pick-up position of each trip, we aggregate the total number of pick up events during one hour at each region to get samples $\tilde{r}_c(d, t)$. It is computationally efficient to process the original data for obtaining a sample set of r_c in general, though the amount of available taxi trips or trajectory information is large – the time complexity is $O(N_{record})$ of the number of total records N_{record}. By only passing through the raw data once, we are able to group each pick up and drop off events to a specific discretized time slot and region. We assume that the dataset contains independent samples of the random vector r_c, and we do not impose any prior knowledge of the true distribution $\mathbb{P}^*(r_c)$. It is always possible to describe the support of the distribution of the entire dataset, even when all samples contained in the dataset do not follow the same distribution, as explained in Figure 3. When there is prior knowledge or categorical information such that the dataset can be partitioned into several subsets according to some feature space, we get a more accurate uncertainty set according to each sub-dataset to provide the same probabilistic guarantee level compared with the uncertainty set from the entire dataset.

Clustering algorithms with categorical information [17] is applicable for dataset partition when information besides pick up events is available in the dataset, such as weather or traffic conditions. It is worth noting that if the uncertainty sets are built for a categorical information set $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \ldots\}$, then for the robust dispatch problems, we require the same set of categories is available in real-time, hence we apply the uncertainty set built for $\mathcal{I}_1$ to find solutions when the current situation is considered as $\mathcal{I}_1$. For instance, when there is additional information like weather or traffic condition for each trip provided by the taxi operational records, these types of information can be used as categorical information for clustering. The dataset applied in the evaluations of Section V does not have additional categorical information of trips that

available for a clustering algorithm such as [17], hence, we partition the dataset as demand during weekdays and demand during weekends. Even with this simple and intuitive partition process, we shrink the area of an uncertainty for the same probabilistic guarantee level. Then during weekdays (weekends) we use uncertainty sets built from weekdays (weekends) data to calculate robust dispatch solutions.

C. Uncertainty Modeling

The basic idea to define an uncertainty set is to find a threshold for a hypothesis testing that is acceptable with respect to the given dataset and a required probabilistic guarantee level, and the formula of an uncertainty set is related to the threshold value of a an acceptable hypothesis testing. Given original operational records data, the null hypothesis H_0, α_h, and the test statistics T, we need to find a threshold that accepts H_0 at significance value α_h for each subset of sampled demand vectors. Since we do not assume that the marginal distribution for every element of vector r_c is independent with each other, we apply two models without any assumptions about the true distribution $\mathbb{P}^*(r_c)$ in the robust optimization literature [6] [12] [26] on the spatial-temporally correlated demand data.

1) Uncertainty demand sets built from marginal samples: One intuitive description about a random vector is to define a range for each element.

For instance, David and Nagaraja [12] considered the following multivariate hypothesis holds simultaneously for $i = 1, 2, \ldots, \tau n$ with given thresholds $\bar{q}_{i,0}, \underline{q}_{i,0} \in \mathbb{R}$

$$\begin{aligned} H_{0,i} : & inf\{t : \mathbb{P}(r_{c,i} \leqslant t) \geqslant 1 - \frac{\epsilon}{\tau n}\} \geqslant \bar{q}_{i,0} \\ & inf\{t : \mathbb{P}(-r_{c,i} \leqslant t) \geqslant 1 - \frac{\epsilon}{\tau n}\} \geqslant -\underline{q}_{i,0}. \end{aligned} \tag{16}$$

Assume that we have N_B random samples for each component $r_{c,i}$ of r_c, ordered in increasing value as $r_{c,i}^{(1)}, r_{c,i}^{(2)}, \ldots, r_{c,i}^{(N_B)}$ no matter what is the original sampling order. We define the index s by

$$s = min\left\{k \in \mathbb{N} : \sum_{j=k}^{N_B} \binom{N_B}{j} \left(\frac{\epsilon}{\tau n}\right)^{N_B - j} \left(1 - \frac{\epsilon}{\tau n}\right)^{j} \leqslant \frac{\alpha_h}{2\tau n}\right\}, \tag{17}$$

and let $s = N_B + 1$ if the corresponding set is empty. The test H_0 is rejected if

$$r_{c,i}^{(s)} \geqslant \bar{q}_{i,0} \text{ or } -r_{c,i}^{(N_B - s + 1)} \geqslant -\underline{q}_{i,0},$$

To construct an uncertainty set, we need an accepted hypothesis test. Hence, we set $\bar{q}_{i,0} = r_{c,i}^{(s)}$ and $\underline{q}_{i,0} = r_{c,i}^{(N_B - s + 1)}$, then $H_{0,i}$ (16) is always accepted. The following uncertainty set is then applied in this work based on the range hypothesis testing (16).

Proposition 1 ([6], [12]): If s defined by equation (17) satisfies that $N_B - s + 1 < s$, then, with probability at least $1 - \alpha_h$ over the sample, the set

$$\mathcal{U}_\epsilon^M(r_c) = \left\{r_c \in \mathbb{R}^{\tau n} : r_{c,i}^{(N_B - s + 1)} \leqslant r_{c,i} \leqslant r_{c,i}^{(s)}\right\} \tag{18}$$

implies a probabilistic guarantee for $\mathbb{P}^*(r_c)$ at level ϵ.

2) Uncertainty set motivated by moment hypothesis testing: Though the box type of uncertainty set reflects the spatial-temporal correlations by varying range values with different dimensions of r_c (value of τn), it is not easy to tell directly from the uncertainty set (18) when the range of one component changes how will others be affected. To construct an uncertainty set that directly shows the spatial-temporal correlations of the demand model, we consider to apply hypothesis testing related to both the first and second moments of the true distribution $\mathbb{P}^*(r_c)$ of the random vector [26].

$$H_0 : \mathbb{E}^{\mathbb{P}^*}[r_c] = r_0 \text{ and } \mathbb{E}^{\mathbb{P}^*}[r_c r_c^T] - \mathbb{E}^{\mathbb{P}^*}[r_c]\mathbb{E}^{\mathbb{P}^*}[r_c^T] = \Sigma_0, \tag{19}$$

where r_0 and Σ_0 are the (unknown) true mean and covariance of r_c, $\mathbb{E}^{\mathbb{P}^*}[r_c]$ and $\mathbb{E}^{\mathbb{P}^*}[r_c r_c^T]$ are the estimated mean and covariance from data. Without knowledge of r_0 and Σ_0, H_0 is rejected when the difference among the estimation of mean or covariance according to multiple times of samples is greater than the threshold, i.e.,

$$\begin{aligned} & \|\mathbb{E}^{\mathbb{P}}[\tilde{r}_c] - \hat{r}_c\|_2 > \Gamma_1^B \text{ or} \\ & \|\mathbb{E}^{\mathbb{P}}[\tilde{r}_c \tilde{r}_c^T] - \mathbb{E}^{\mathbb{P}}[\tilde{r}_c]\mathbb{E}^{\mathbb{P}}[\tilde{r}_c^T] - \hat{\Sigma}\|_F > \Gamma_2^B, \end{aligned}$$

where $\mathbb{E}^{\mathbb{P}}[\tilde{r}]$ is the estimated mean value of one experiment, $\hat{r}_c$ and $\hat{\Sigma}$ are the estimated mean and covariance from multiple experiments, Γ_1^B and Γ_2^B are the thresholds. The remaining problem is then to find the values of the thresholds such that hypothesis testing (19) holds given the dataset. In the following Section III, the detailed steps of calculating the thresholds Γ_1^B and Γ_2^B at a desired significance value α_h and probabilistic guarantee level ϵ based on the given dataset is described[2].

The uncertainty set derived based on the moment hypothesis testing is defined in the following proposition.

Proposition 2 ([6], [26]): With probability at least $1 - \alpha_h$ with respect to the sampling, the following uncertainty set $\mathcal{U}_\epsilon^{CS}(r_c)$ implies a probabilistic guarantee level of ϵ for $\mathbb{P}^*(r_c)$

$$\begin{aligned} \mathcal{U}_\epsilon^{CS}(r_c) = & \{r_c \geqslant \mathbf{0}, \hat{r}_c + y + C^T w : \exists y, w \in \mathbb{R}^{n\tau} \ s.t. \\ & \|y\|_2 \leqslant \Gamma_1^B, \|w\|_2 \leqslant \sqrt{\frac{1-\epsilon}{\epsilon}}\}, \end{aligned} \tag{20}$$

where $C^T C = \hat{\Sigma} + \Gamma_2^B \mathbf{I}$ is a Cholesky decomposition.
When one component of r_c increases or decreases, we have an intuition how it affects the value of other components of r_c by the expression (20).

D. Algorithm

With a threshold of the test statistics calculated via the given dataset, we then apply the formula (18) for constructing a box type of uncertainty set, and the formula (20) for an SOC type of uncertainty set, respectively. The following Algorithm 1 describes the complete process for constructing uncertain demand sets based on the original dataset.

We do not restrict the method of estimating mean $\hat{r}_c(t, I_p)$ and covariance $\hat{\Sigma}(t, I_p)$ matrices of a subset $\mathcal{S}(t, I_p)$ in step 2,

[2]Bootstrapped thresholds and theoretic bounds proposed by work [19] are compared in [6]. The bootstrapped thresholds result in a smaller uncertainty set in general, hence reduces the ambiguity in $\mathbb{P}^*$. In this work, we apply the bootstrapped thresholds Γ_1^B and Γ_2^B based on the dataset.

Algorithm 1 Algorithm for constructing uncertain demand sets

Input: A dataset of taxi operational records

1. Demand aggregating and sample set partition

Aggregate demand to get a sample set $\mathcal{S}$ of the random demand vector r_c from the original dataset. Partition the sample set $\mathcal{S}$ and denote a subset $\mathcal{S}(I_p) \subset \mathcal{S}$, $p = 1, \ldots, P$ as the subset partitioned according to either prior knowledge or categorical information I_p. Denote the partitioned sample subset for each time index t as $\mathcal{S}(t, I_p)$.

2. Bootstrapping thresholds for test statistics

for each subset $\mathcal{S}(t, I_p)$ **do**

Initialization: Testing statistics T, a null-hypothesis H_0, the probabilistic guarantee level ϵ, a significance level $0 < \alpha_h < 1$, the number of bootstrap time $N_B \in \mathbb{Z}_+$.

Estimate the mean $\hat{r}_c(t, I_p)$ **and covariance** $\hat{\Sigma}(t, I_p)$ **for vector** r_c **based on subset** $\mathcal{S}(t, I_p)$**.**

for $j = 1, \ldots, N_B$ **do**

(1). Re-sample $\mathcal{S}^j(t, I_p) = \{\tilde{r}_c(d_1, t, I_p), \ldots, \tilde{r}_c(d_{N_B}, t, I_p)\}$ data points from $\mathcal{S}(t, I_p)$ with replacement for each t.

(2). Get the value of the test statistics based on $\mathcal{S}^j(t, I_p)$.

end for

(3). Get the thresholds of the α significance level for H_0.

end for

3. Calculate the model of uncertainty sets

Get the box type and the SOC type of uncertainty sets according to (18) and (20), respectively, for each t and I_p.

Output: Uncertainty sets for problem (11)

and bootstrap is one method for this step. The estimations of this step are considered as the true mean and covariance for calculating Γ_1^B and Γ_2^B in the following repeated sampling process. For step $2.(2)$, the process for the box type of uncertainty sets is: calculate index s that satisfies (17) with the given ϵ, sort each component of sampled vectors $r_c(d_l, t, I_p)$, and get the order statistics $r_{c,i}^{(N_B-s+1)}(j, t, I_p)$, $r_{c,i}^{(s)}(j, t, I_p)$ of the j-th sample set $\mathcal{S}^j(t, I_p)$. For the SOC type, we calculate the mean and covariance of the samples of the vector according to the subset $\mathcal{S}^j(t, I_p)$ as $\hat{r}_c(j, t, I_p)$ and $\hat{\Sigma}(j, t, I_p)$, respectively.

In step $2.(3)$, the α_h level thresholds for the box type of uncertainty sets are the $\lceil N_B(1-\alpha_h)\rceil$-th largest value of the upper bound $r_{c,i}^{(s)}(j, t, I_p)$ and the $\lceil N_B\alpha_h\rceil$-th largest value of the lower bound $r_{c,i}^{(N_B-s+1)}(j, t, I_p)$ for the i-th component of each t and I_p. For the SOC type of uncertainty sets, we calculate the mean and covariance of $r_c(t, I_p)$ for the N_B times bootstrap as $\hat{r}_c(t, I_p)$ and $\hat{\Sigma}(t, I_p)$, and get

$$\Gamma_1(j, t, I_p) = \|\hat{r}_c(j, t, I_p) - \hat{r}_c(t, I_p)\|_2,$$
$$\Gamma_2(j, t, I_p) = \|\hat{\Sigma}(j, t, I_p) - \hat{\Sigma}(t, I_p)\|_2.$$

Denote the $\lceil N_B(1-\alpha_h)\rceil$-th largest value of $\Gamma_1(j, t, I_p)$ and $\Gamma_2(j, t, I_p)$as $\Gamma_1^B(t, I_p)$ and $\Gamma_2^B(t, I_p)$, respectively.

Remark 1: The process of constructing uncertainty sets only requires that the hypothesis test is accepted for i.i.d. samples of the random vector. We accept the hypothesis test when there is not enough evidence to reject it, which does not mean the claim of H_0 is true. This property is very important for constructing the uncertainty demand set of the robust dispatch problem, since the true distribution function of a demand model can be complex and we only have datasets of taxi operational records instead of ground truth knowledge of the distribution function. Hence, even without enough knowledge of the true, high-dimensional demand model, based on the dataset and an accepted hypothesis test, we are able to construct an uncertainty set with probabilistic guarantee for the robust taxi dispatch problem. □

It is worth noting that the above Algorithm 1 provides a valid estimation of uncertain sets based on hypothesis testing and bootstrapped thresholds for the robust resource allocation problem when the sampled data set is consistent with the real world scenario. For demand missed in the dataset, for instance, some customer might leave the request queue after waiting for a long time and the operational records did not show the event of picking up the customer, we are not able to get the exact rate of missed customers. However, missed requests are only part of the historical requests, and this type of events is also random – for instance, even for the same time length of waiting, some customers were more patient and finally got a taxi. By constructing an uncertainty set to describe the demand model based on occurred records of the original dataset, we involve the effect of random missing events better than only applying a deterministic model from this perspective. More properties of each type of uncertainty set and application level problems, such as how to choose the number of samples N_B for the hypothesis testing with high dimensional r_c will be discussed in evaluations of Section V.

In summary, to construct a spatial-temporal uncertain demand model for the robust taxi dispatch (11), in this section, we consider the taxi operational record of each day as one independent and identically distributed (i.i.d.) sample for the concatenated demand vector r_c. By partitioning the entire dataset to several subsets according to categorical information such as weekdays and weekends, we are able to build uncertainty sets for each subset of data without additional assumptions about the true distribution of the spatial-temporal demand profile. Then we design Algorithm 1 to construct a box type and an SOC type of uncertainty sets based on data. The key advantage of the data-driven approach we propose is that we do not rely on prior knowledge of the true distribution of the random demand vector to provide a desired probabilistic guarantee of robust solutions. Furthermore, theories proved for i.i.d. datasets are applicable to construct uncertainty sets that reflect the spatial-temporal correlations of the demand model.

IV. Computationally Tractable Formulations

We build equivalent computationally tractable formulations of problem (11) with different definitions of uncertain sets calculated by Algorithm 1 in this section. Hence, the robust taxi dispatch problem considered in this work can be solved efficiently. Computational tractability of a robust linear programming problem for ellipsoid uncertainty sets is discussed in [4]. The process is to reformulate constraints of the original problem to its equivalent convex constraints that must hold given the uncertainty set. The objective function of problem (11) is concave of the uncertain parameters r^k, convex

of the decision variables X^k, L^k with the decision variables on the denominators, not standard forms of linear programming (LP) or semi-definite programming (SDP) problems that already covered by previous work [4], [6]. Hence, we prove one equivalent computationally tractable form of problem (11) for each uncertainty set constructed in Section III.

Only the J_E components of objective functions in (11) include uncertain parameters, and the decision variables of the function are in the denominator of the function J_E. The box type uncertainty set defined as (18) is a special form of polytope, hence, we first prove an equivalent standard form of convex optimization problem for (11) for a polytope uncertainty set as the following.

Theorem 1: (Next step dispatch) If the uncertainty set of problem (11) when $\tau = 1$ is defined as the following non-empty polytope

$$\Delta := \{r \geq 0, Ar \leq b\},$$

and we omit the superscripts k for variables and parameters without confusion. Then problem (11) with $\tau = 1$ is equivalent to the following convex optimization problem

$$\begin{aligned}
\underset{X\geq 0,\lambda\geq 0}{\text{minimize}} \quad & \sum_i \sum_j X_{ij}W_{ij} + b^T\lambda \\
\text{subject to} \quad & A^T\lambda - \beta \begin{bmatrix} \frac{1}{\left(\sum_{j=1}^n X_{j1} - \sum_{j=1}^n X_{1j} + L_1\right)^\alpha} \\ \vdots \\ \frac{1}{\left(\sum_{j=1}^n X_{jn} - \sum_{j=1}^n X_{nj} + L_n\right)^\alpha} \end{bmatrix} \geq 0, \\
& \mathbf{1}_n^T X - X\mathbf{1}_n + L^T \geqslant 1, \\
& X_{ij}W_{ij} \leq mX_{ij}, \\
& X_{ij} \geq 0, \quad \forall i,j \in \{1,\ldots,n\}.
\end{aligned} \tag{21}$$

□

Proof: See Appendix VII-B. ■

To directly use the demand uncertainty set that describes the spatial-temporal correlation of $(r^1,\ldots,r^\tau)$, for instance, sets (18) and (20) for the concatenated demand r_c in problem (11), we first consider to group the maximization over each r^k together to save the process of projection $r_c \in \Delta$ for individual $r^k \in \Delta_k$. Furthermore, we can find the dual (a minimizing problem) of the maximizing cost problem over $r_c \in \Delta$, and then numerically efficiently solve the robust taxi resource allocation problem (11) that minimizes the total cost during time $(1,2,\ldots,\tau)$ under uncertain concatenated demand r_c. Hence, we first prove that the minimax equality holds for the maximin problem over each pair of stages k and $k+1$ for problem (11), and (11) is equivalent to the robust optimization problem shown in the following lemma. The computationally tractable convex optimization forms of problem (11) are then proved based on this theorem.

Lemma 2: (Minimax equality) Given the assumption that the definition of the uncertainty sets $r_c \in \Delta$ and $r^k \in \Delta_k$ are compact (closed and convex), the robust dispatch problem (11) is equivalent to the following robust dispatch problem

$$\begin{aligned}
\min_{X^{1:\tau},L^{2:\tau}} \max_{r_c\in\Delta} \quad & J = \sum_{k=1}^{\tau}(J_D(X^k) + \beta J_E(X^k, r^k)) \\
\text{s.t.} \quad & \text{constraints of (11)}, k = 1,\ldots,\tau.
\end{aligned} \tag{22}$$

□

Proof: See Appendix VII-C. ■

For the robust optimization problem (11), the computationally tractable convex form depends on the definition of uncertainty sets.When conditions of Lemma 2 hold, equivalent convex optimization forms of problem (11) are derived based on problem (22). For a multi-stage robust optimization problem that restricts the near-optimal control input of linear dynamical systems to be a certain degree of polynomial of previous observed uncertainties, an approximated semidefinite programming method for calculating the time dependent control input is proposed in [7]. The method does not require minimax equality holds for the robust optimal control problem.

The box type uncertainty set (18) is a special form of polytope, that the uncertain demand model during different time of a day is described separately. The process of converting problem (11) to an equivalent computationally tractable convex form is similar to that of the one-stage robust optimization problem. The result is described as the following lemma.

Lemma 3: If the uncertain set for $r^k, k = 1,\ldots,\tau$ describes each demand vector r^k separately as a non-empty polytope with the form

$$\Delta_k := \{r^k \geq 0, A_k r^k \leq b_k\}, \quad k = 1,\ldots,\tau, \tag{23}$$

problem (11) is equivalent to the following convex optimization problem

$$\begin{aligned}
\min_{X^k,\lambda^k,L^k\geq 0} \quad & \sum_{k=1}^{\tau}(\sum_i \sum_j X_{ij}^k W_{ij} + b_k^T\lambda^k) \\
\text{subject to} \quad & A_k^T\lambda^k - \beta \begin{bmatrix} \frac{1}{\left(\sum_{j=1}^n X_{j1}^k - \sum_{j=1}^n X_{1j}^k + L_1^k\right)^\alpha} \\ \vdots \\ \frac{1}{\left(\sum_{j=1}^n X_{jn}^k - \sum_{j=1}^n X_{nj}^k + L_n^k\right)^\alpha} \end{bmatrix} \geq 0, \\
& \text{other constraints of (11)}, k = 1,\ldots,\tau.
\end{aligned} \tag{24}$$

□

Proof: See Appendix VII-D1. ■

For a more general case that the uncertainty sets for $r^1,\ldots,r^\tau$ are temporally correlated, the following theorem and proof describe the equivalent computationally tractable convex form of (11).

Theorem 2: When Δ is defined as the following non-empty polytope set

$$\Delta := \{(\Delta_1,\ldots,\Delta_\tau) | A_1 r^1 + \cdots + A_\tau r^\tau \leq b, r^k \geq 0\}, \tag{25}$$

problem (11) is equivalent to the following convex optimization problem

$$\begin{aligned}\min_{X^k,L^k,\lambda\geq 0}\quad & \sum_{k=1}^{\tau}(\sum_i\sum_j X_{ij}^k W_{ij}) + b^T\lambda \\ \text{subject to}\quad & A_k^T\lambda - \beta\begin{bmatrix}\frac{1}{\left(\sum_{j=1}^n X_{j1}^k - \sum_{j=1}^n X_{1j}^k + L_1^k\right)^\alpha}\\ \vdots \\ \frac{1}{\left(\sum_{j=1}^n X_{jn}^k - \sum_{j=1}^n X_{nj}^k + L_n^k\right)^\alpha}\end{bmatrix} \geq 0, \\ & \text{constraints of (11)}, \quad k=1,\ldots,\tau.\end{aligned} \tag{26}$$

□

Proof: See Appendix VII-D2. ■

With an uncertain demand model defined as (20) for concatenated $r^1,\ldots,r^\tau$, the following theorem derive the equivalent computationally tractable form of problem (11).

Theorem 3: When the uncertainty set for $r^1,\ldots,r^\tau$ is defined as the SOC form of (20), problem (11) is equivalent to the following convex optimization problem (27).

$$\begin{aligned}\min_{X^k,L^k,z}\quad & \sum_{k=1}^{\tau}\sum_i\sum_j X_{ij}^k W_{ij} \\ & + \beta\left(\hat{r}_c^T z + \Gamma_1^B\|z\|_2 + \sqrt{\frac{1}{\epsilon}-1}\|Cz\|_2\right) \\ \text{subject to}\quad & c_l(X) \leqslant z, \\ & \text{constraints of (11)}, \quad k=1,\ldots,\tau,\end{aligned} \tag{27}$$

where $c_l(X)\in\mathbb{R}^{\tau n}$ is the concatenation of $c(X^1),\ldots,c(X^\tau)$.

□

Proof: See Appendix VII-E. ■

It is worth noting that any optimal solution for problem (10) has a special form between any pair of regions (i,q).

Proposition 3: Assume $X^{1*},\ldots,X^{\tau*}$ is an optimal solution of (10), then any X^{k*} satisfies that for any pair of (p,q), at least one value of the two elements X_{qi}^{k*} and X_{iq}^{k*} is 0.

Proof: We prove by contradiction. Assume that one optimal solution has the form X^k such that $X_{qi}^k>0$ and $X_{iq}^k>0$. Without loss of generality, we assume that $X_{qi}^k \geq X_{iq}^k$, and let

$$X_{qi}^{k*} = X_{qi}^k - X_{iq}^k, X_{iq}^{k*} = 0,$$

other elements of X^{k*} equal to X^k. Then

$$\mathbf{1}_n^T X_{\cdot i}^k - X_{i\cdot}^k\mathbf{1}_n + L_i^k = \mathbf{1}_n^T X_{\cdot i}^{k*} - X_{i\cdot}^{k*}\mathbf{1}_n + L_i^k,$$

because

$$\begin{aligned}&\sum_j X_{ji}^k - \sum_l X_{il}^k = X_{qi}^k - X_{iq}^k + \sum_{j\neq q} X_{ij}^k - \sum_{l\neq q} X_{qi}^k \\ =&X_{qi}^{k*} + 0 + \sum_{j\neq q} X_{ij}^{k*} - \sum_{l\neq q} X_{qi}^{k*} = \sum_j X_{ji}^{k*} - \sum_l X_{il}^{k*}.\end{aligned}$$

Hence, we have $J_E(X^k,r^k) = J_E(X^{k*},r^k)$. All constraints are satisfied and X^{k*} is also a feasible solution for (11).

Next, we compare $J_D(X^k)$ and $J_D(X^{k*})$. With $X_{qi}^k > X_{iq}^k > 0$, and $X_{qi}^{k*} = X_{qi}^k - X_{iq}^k \geq 0$, we have

$$X_{qi}^k > X_{qi}^{k*},\ X_{qi}^k W_{qi} + X_{iq}^k W_{iq} > X_{qi}^{k*} W_{qi} + X_{iq}^{k*} W_{iq}.$$

Thus the partial cost $J_D(X^k) > J_D(X^{k*})$, which contradicts with the assumption that X^k is an optimal solution. To summarize, we show that an optimal solution cannot have $X_{qi}^k>0, X_{iq}^k>0$ at the same time, and at least one of X_{qi}^{k*} and X_{iq}^{k*} should be 0. ■

With equivalent convex optimization forms under different uncertainty sets, robust taxi dispatch problem (11) is computationally tractable and solved efficiently.

V. Data-Driven Evaluations

We conduct data-driven evaluations based on four years of taxi trip data of New York City [13]. A summary of this data set is shown in Table II. In this data set, every record represents an individual taxi trip, which includes the GPS coordinators of pick up and drop off locations, and the date and time (with precision of seconds) of pick-up and drop-off locations. One region partition example according to the map of Manhattan of New York City is shown in Figure 4 where we visualize the density of taxi passenger demand with the data we used for our large-scale data-driven evaluation. The lighter the region, the higher the daily demand density. As we can see in the figure, the middle regions typically have higher density than the uptown and downtown regions in Manhattan. We construct uncertainty sets according to Algorithm 1, discuss factors that affect modeling of the uncertainty set, and compare optimal costs of the robust dispatch formulation (11) and the non-robust optimization form (10) in this section.

How vacant taxis are balanced across regions with different α values: Figure 5 shows mismatch between supply and demand defined as (6) for different optimal solutions of minimizing J_E defined in (7) for $\alpha\in(0,1]$. With α closer to 0, the optimal value of (6) is smaller. We choose $\alpha=0.1$ for calculating optimal solutions of (11) and (10) in this section.

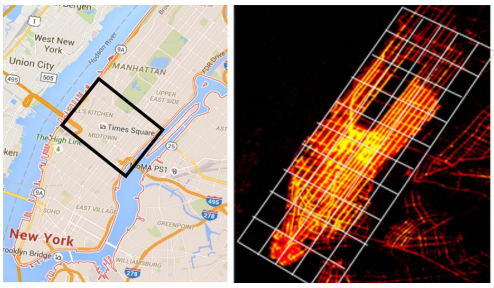

Figure 4. Map of Manhattan area in New York City.

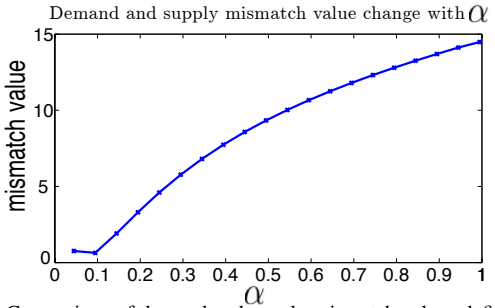

Figure 5. Comparison of demand and supply mismatch values defined as (6) with different solutions for minimizing J_E defined in (7) with α in range $(0,\ 1]$. The value of function (6) under an optimal solution of J_E is smaller with an α closer to 0, which means the dispatch solution tends to be more balanced throughout the entire city.

Taxi Trip Data set			Format		
Collection Period	Data Size	Record Number	ID	Trip Time	Trip Location
01/01/2010-12/31/2013	$100GB$	about 7 million	Date	Start and end time	GPS coordinates of start and end

Table II
NEW YORK CITY DATA IN THE EVALUATION SECTION.

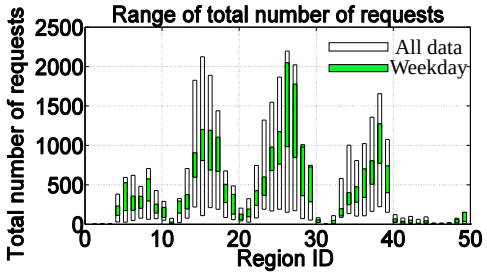

Figure 6. Comparison of box type of uncertainty sets constructed from all data and those constructed only based on trip records of weekdays. When keeping all parameters the same, by applying data of weekdays, the range of uncertainty set for each $r_{c,i}$ is smaller than that based on the whole dataset.

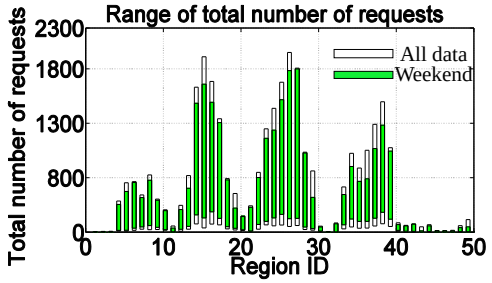

Figure 7. Comparison of box type of uncertainty sets constructed from all data and uncertainty sets constructed only based on trip records of weekends.

A. Box type of uncertainty set

For all box type of uncertainty sets shown in this subsection with the model described in Subsection III-C1, we set the confidence level of hypothesis testings as $\alpha_h = 10\%$, bootstrap time as $N_b = 1000$, number of randomly sampled data (with replacement) for each time of bootstrap as $N_B = 10000$.

Partitioned dataset compared with non-partitioned dataset: We show the effects of partitioning the trip record dataset by weekdays and weekends in Figure 6 and 7. The whole city is partitioned into 50 regions, the prediction time horizon is $\tau = 4$, where one time instant means one hour, $\epsilon = 0.3$, and every $r_c \in \mathbb{R}^{200\times 1}$. Figures 6 and 7 show the lower and upper bounds of each region during one time slot of (18). By applying data of weekdays and weekends separately, the range $[r_{c,i}^{(s)},\ r_{c,i}^{(N_B-s+1)}]$ of each component is reduced. To get a measurement of the uncertainty level, we defined the sum of range of every component for r_c as the following

$$U(r_c) = \sum_{i=1}^{\tau n} (r_{c,i}^{(s)} - r_{c,i}^{(N_B-s+1)}).$$

For the box type of uncertainty sets, when values of the dimension of r_c, i.e., τn, α_h and ϵ are fixed, a smaller $U(r_c)$ means a smaller area of the uncertainty set, or a more accurate model. We denote $U(r_c)$ calculated via records of weekdays and weekends as $U_{wd}(r_c)$ and $U_{wn}(r_c)$ respectively, compared with $U(r_c)$ constructed from the complete dataset, we have

$$\frac{U(r_c) - U_{wd}(r_c)}{U(\hat{r}_c)} = 52\%, \quad \frac{U(r_c) - U_{wn}(r_c)}{U(\hat{r}_c)} = 28\%.$$

N_B	α_h	ϵ	n	τ	s
10000	0.1	0.2	50	2	9992
10000	0.1	0.5	50	2	9970
10000	0.3	0.2	50	2	9991
10000	0.1	0.2	1000	2	9999
10000	0.1	0.5	1000	2	9999

Table III
VALUE OF INDEX s FOR THE BOX TYPE UNCERTAINTY SET (17). FOR LARGE τn, N_B NEED TO BE LARGE, OR s IS TOO CLOSE TO N_B THAT THE RANGE COVERS VALUES OF ALMOST ALL SAMPLES.

Data type	Weekdays	Weekends	Non partitioned
Γ_1^B	10.53	13.84	17.96
Γ_2^B	2576.94	2923.35	3864.47

Table IV
COMPARING THRESHOLDS WITH AND WITHOUT DISCRIMINATING WEEKDAYS AND WEEKENDS DATA. WHEN Γ_1^B OR Γ_2^B IS SMALLER, THE VOLUME OF THE UNCERTAINTY SET IS SMALLER. HERE $n = 1000$, $\tau = 3$, $N_B = 1000$, $\epsilon = 0.3$, $\alpha_h = 0.2$.

This result shows that when by constructing an uncertainty set for each subset of partitioned data, we reduce the range of uncertainty sets to provide the same level of probabilistic guarantee for the robust dispatch problem. This is because samples contained in each subset of data do not follow the same distribution and can be categorized as two clusters.

Choose an appropriate N_B for high-dimensional r_c: It is worth noting that the index s affects the range selection for every component $r_{c,i}$, hence, for different values of $\alpha_h, \epsilon, \tau, n$, we should adjust the number of samples N to get an accurate estimation of the marginal range. As shown in Table V, N need to be large enough for a large τn value, or s is too close to N and the upper and lower bounds $r_{c,i}^{(N_B-s+1)}$, $r_{c,i}^{(s)}$ cover almost the whole range of samples. Hence, the box type uncertainty set is not a good choice for large τn value, though the computational cost of solving problem (26) is smaller than that of (27) with the same size of τn.

B. SOC type of uncertainty set

The SOC type of uncertainty set is a high-dimensional convex set that is not able to be plotted. The bootstrapped thresholds for the hypothesis testing to construct the SOC uncertainty sets based on partitioned and non-partitioned data are summarized in Table IV. Similarly as the box type of uncertainty sets, when we separate the dataset and construct an uncertainty demand model for weekdays and weekends respectively, the sets are smaller compared to the uncertain demand model for all dates. When α and ϵ values are fixed, with smaller Γ_1^B and Γ_2^B, the demand model $\mathcal{U}_\epsilon^{CS}$ is more accurate to guarantee that with at least probability $1-\epsilon$, the constraints of the robust dispatch problems are satisfied. Numerical results of this conclusion are shown in Table IV.

How n and τ affect the accuracy of uncertainty sets: For a box type of uncertainty set, when τn is a large value, the bootstrap sample number N_B should be large enough such

	Γ_1^B	Γ_2^B
$n = 50, \tau = 1$	42.37	1.52×10^5
$n = 50, \tau = 3$	52.68	4.29×10^4
$n = 50, \tau = 6$	107.35	8.23×10^5
$n = 10, \tau = 3$	71.35	3.56×10^5
$n = 1000, \tau = 3$	10.53	2576.94

Table V
COMPARING THRESHOLDS OF SOC UNCERTAINTY SETS FOR DIFFERENT DIMENSIONS r_c, BY CHANGING EITHER THE REGION PARTITION NUMBER n OR THE PREDICTION TIME HORIZON τ.

that index s is not too close to N. Without a large enough sample set, we choose to construct an SOC type of uncertainty set (such as $\tau n = 1000, N_B = 10000$ in Table V). Since SOC captures more information about the second moment properties of the random vector compared with the box type uncertainty set, some uncorrelated components of r_c will be reflected by the estimated covariance matrix, and the volume of the uncertainty set will be reduced. We show the value of Γ_1^B and Γ_2^B with different dimensions of r_c or τn values in table V. When increasing the value of τn, values of Γ_1^B and Γ_2^B are reduced, which means the uncertainty set is smaller. However, it is not helpful to reduce the granularity of region partition to a smaller than street level, since we construct the model for a robust dispatch framework and a too large n is not computationally efficient for the dispatch algorithm.

C. Compare robust solutions with non-robust solutions

In the experiments, the idle geographical distance of one taxi between a drop-off event of one passenger and the following pick-up event is approximately as the one norm distance between the 2D geographical coordinates (provided as longitude and latitude values of GPS data in the trip dataset) of the two points. Then the corresponding idle miles on ground is converted from the geographical distance according to the geographical coordinates of New York City. For testing the quality of the uncertainty sets applied in the robust dispatch problems, we use the idea of cross-validation from machine learning. The dataset is separated as a training set for building the uncertain demand model, and a testing set for comparing the results of the dispatch solutions. The customer demand models applied in the robust and non-robust optimization problems are different. For the non-robust dispatch problem, the demand prediction r^k is a deterministic vector. For instance, in this work we use the average or mean of the bootstrapped value of the training dataset. The non-robust dispatch solution for each time k is calculated by solving the convex optimization form of dispatch problem formulated in work [21], [22] with deterministic demand model. For all the experiments, we let $\beta = 10$, $\alpha = 0.1$ in problem (11) to calculate the optimal solutions.

In the robust dispatch problem, the part that directly includes the uncertain demand r^k is the penalty function for violating a balanced demand-supply ratio requirement. For each testing data r^k, we denote the demand-supply ratio mismatch error of a dispatch solution as (6). We then compare the value of (6) of robust dispatch solutions with the SOC type of uncertainty set constructed in this work with the value of (6) of non-robust solutions of testing samples. The distribution of values

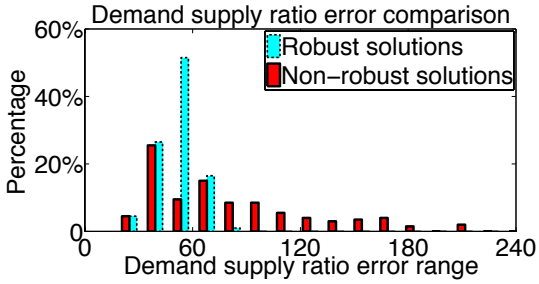

Figure 8. Demand-supply ratio error distribution of the robust optimization solutions with the SOC type of uncertain demand set ($\epsilon = 0.25$, or probabilistic guarantee level 75%) and non-robust optimization solutions. The demand-supply ratio error of robust solutions is smaller than that of the non-robust solutions, that the average demand-supply ratio error is reduced by 31.7%.

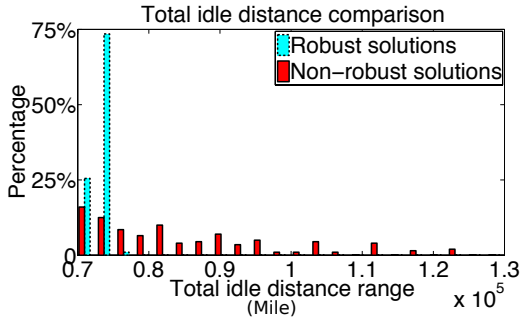

Figure 9. Total idle distance comparison of robust optimization solutions with the SOC type of uncertain demand set ($\epsilon = 0.25$, or probabilistic guarantee level 75%) and non-robust optimization solutions. The average total idle distance is reduced by 10.13%. For all samples used in testing, the robust dispatch solutions result in no idle distance greater than 0.8×10^5, and non-robust solutions has 48% of samples with idle distance greater than 0.8×10^5. The number of total idle distance shown in this figure is the direct calculation result of the robust dispatch problem, and we convert the number to an estimated value of corresponding miles in one year, the result is a total reduction of 20 million miles in NYC.

are shown in Figure 8. The average demand-supply ratio error is reduced by 31.7% with robust solutions. We compare the cost distribution of total idle distance in Figure 9. It shows the average total idle distance is reduced by 10.13%. For all testing, the robust dispatch solutions result in no idle distance greater than 0.8×10^5, and non-robust solutions has 48% of samples with idle distance greater than 0.8×10^5. The cost of robust dispatch (11) is a weighted sum of both the demand-supply ratio error and estimated total idle driving distance, and the average cost is reduced by 11.8% with robust solutions. It means that the performance of the system is improved when the true demand deviates from the average historical value considering model uncertainty information in the robust dispatch process. It is worth noting that the number of total idle distance shown in this figure is the direct calculation result of the robust dispatch problem. When we convert the number to an estimated value of corresponding miles in one year, the result is a total reduction of 20 million miles in NYC.

Check whether the probabilistic level ϵ is guaranteed: Theoretically, the optimal solution of the robust dispatch problems with the uncertainty set should guarantee that with at least the probability $(1 - \epsilon)$, when the system applies the robust dispatch solutions, the actual dispatch cost under a true demand is smaller than the optimal cost of the robust dispatch problem. Figures 10 and 11 show the cross-validation testing result that the probabilistic guarantee level is reached

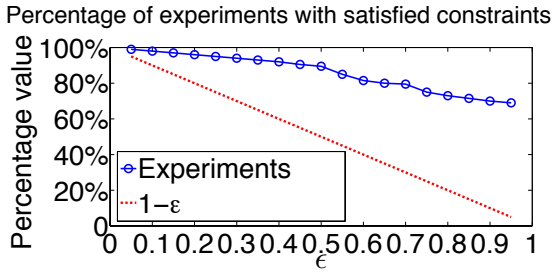

Figure 10. The percentage of tests that have a smaller true dispatch cost than the optimal cost of the robust dispatch problem with the box type uncertainty set constructed from data. When $1-\epsilon$ decreases, the percentage value also decreases, but always greater than $1-\epsilon$.

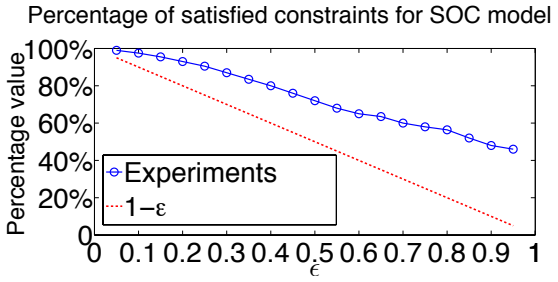

Figure 11. The percentage of tests that have a smaller true dispatch cost than the optimal cost of the robust dispatch problem with the SOC type of uncertainty set. When $1-\epsilon$ decreases, the percentage value also decreases, but always greater than $1-\epsilon$. The true percentage value is closer to the value of $1-\epsilon$ compared with the solution given a box type uncertainty set.

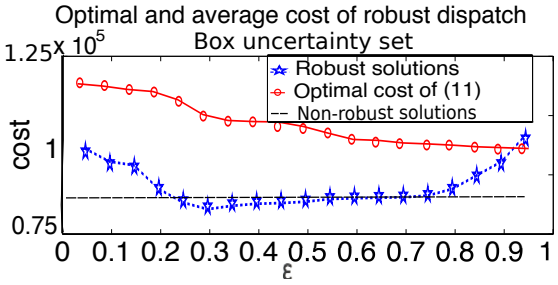

Figure 12. Comparison of the optimal cost of the robust dispatch problem with box type of uncertainty set and the average cost when applying the robust solutions for the test subset of sampled r_c. When $\epsilon = 0.3$ the average cost is the smallest.

for both box type and SOC type of uncertainty sets via solving (26) and (27), respectively. Comparing these two figures, one key insight is that the robust dispatch solution with an SOC type uncertainty set provides a tighter bound on the probabilistic guarantee level that can be reached under the true random demand compared with solutions of the box type uncertainty set. It shows the advantage of considering second order moment information of the random vector, though the computational cost is higher to solve problem (27) than to solve problem (26).

How probabilistic guarantee level affects the average cost: There exists a trade-off between the probabilistic guarantee level and the average cost with respect to a random vector r_c. Selecting a value for ϵ is case by case, depending on whether a performance guarantee for the worst case scenario is more important or whether the average cost performance is more important. For a high probabilistic guarantee level or a large $1-\epsilon$ value, the average cost may not be good enough

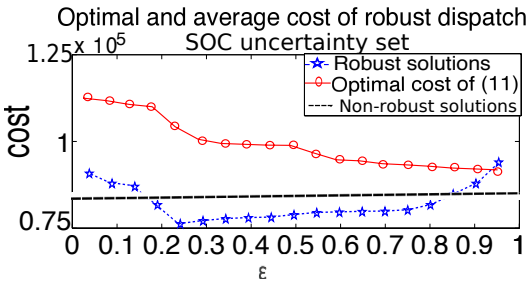

Figure 13. Comparison of the optimal cost of the robust dispatch problem with SOC type of uncertainty set and the average cost when applying the robust solutions for the test subset of sampled r_c. When $\epsilon = 0.25$ the average cost is the smallest.

since we minimize a worst case that rarely happens in the real world. When the $1-\epsilon$ value is relatively small, the average cost can also be large since many possible values of the random vector are not considered.

We compare the optimal cost of robust solutions and average cost of empirical tests for two types of uncertainty sets via solving (26) and (27) in Figure 12 and 13, respectively. The optimal cost of the robust dispatch framework shows that the result of minimized worst case scenario for all possible r_c included in the uncertainty set, and the average cost of empirical tests show the real world scenario when we applying the optimal solution to dispatch taxis under random demand r_c. The horizontal line shows the average cost of non-robust solutions since this cost is not related to ϵ. The ϵ values that provide the best average costs are not exactly the same for different types of uncertainty sets according to the experiments. For the box type of uncertainty set shown in Figure 12, $\epsilon = 0.3$ provides the smallest average experimental cost, and for SOC type of uncertainty set shown in Figure 13, $\epsilon = 0.25$ provides the smallest average experimental cost. The minimum average cost of an SOC robust dispatch solution is smaller than that of a box type. It indicates that the second order moment information of the random variable should be included for modeling the uncertainty set and calculating robust dispatch solutions for the dataset we use in this section, though its computational cost is higher.

VI. Conclusion

In this paper, we develop a multi-stage robust optimization model considering demand model uncertainties in taxi dispatch problems. We model spatial-temporal correlations of the uncertainty demand by partitioning the entire data set according to categorical information, and applying theories without assumptions on the true distribution of the random demand vector. We prove that an equivalent computationally tractable form exist with the constructed polytope and SOC types of uncertainty sets, and the robust taxi dispatch solutions are applicable for a large-scale transportation system. A robust dispatch formulation that purely minimizes the worst-case cost under all possible demand usually sacrifices the average system performance. The robust dispatch method we design allows any probabilistic guarantee level for a minimum cost solution, considering the trade-off between the worst-case cost and the average performance. Evaluations show that under the

robust dispatch framework we design, the average demand-supply ratio mismatch error is reduced by 31.7%, and the average total idle driving distance is reduced by 10.13% or about 20 million miles in total in one year. In the future, we will enhance problem formulation considering more uncertain characteristics of taxi network model, like traffic conditions.

REFERENCES

[1] S. Ali, A. Maciejewski, H. Siegel, and J.-K. Kim. Measuring the robustness of a resource allocation. *IEEE Transactions on Parallel and Distributed Systems*, 15(7):630–641, July 2004.

[2] M. Asif, J. Dauwels, C. Goh, A. Oran, E. Fathi, M. Xu, M. Dhanya, N. Mitrovic, and P. Jaillet. Spatiotemporal patterns in large-scale traffic speed prediction. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):797–804, 2014.

[3] R. K. Balan, K. X. Nguyen, and L. Jiang. Real-time trip information service for a large taxi fleet. In *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, pages 99–112, 2011.

[4] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[5] D. Bertsekas, A. Nedi, A. Ozdaglar, et al. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[6] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *Operations Research*, 7(arXiv: 1401.0212), 2015.

[7] D. Bertsimas, D. Iancu, and P. Parrilo. A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Transactions on Automatic Control*, 56(12):2809–2824, Dec 2011.

[8] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[9] G. Chasparis, M. Maggio, E. Bini, and K.-E. Årzén. Design and implementation of distributed resource management for time sensitive applications. *Automatica*, 2015. Accepted for publication.

[10] J. Cortes, S. Martinez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.

[11] F. A. Cuzzola, J. C. Geromel, and M. Morari. An improved approach for constrained robust model predictive control. *Automatica*, 38(7):1183–1189, 2002.

[12] H. David and H. Nagaraja. *Order statistics*. Wiley Online Library, 1970.

[13] B. Donovan and D. B. Work. Using coarse gps data to quantify city-scale transportation system resilience to extreme events. In *presented at the 2015 Transportation Research Board Annual Meeting*, July 2015.

[14] R. Ganti, M. Srivatsa, and T. Abdelzaher. On limits of travel time predictions: Insights from a new york city case study. In *2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS)*, pages 166–175, June 2014.

[15] Y. Geng and C. Cassandras. New "smart parking" system based on resource allocation and reservations. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1129–1139, 2014.

[16] J. Herrera, D. Work, R. Herring, X. Ban, Q. Jacobson, and A. Bayen. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C*, 18(4):568–583, 2010.

[17] Z. Huang. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3):283–304, Sept. 1998.

[18] D.-H. Lee, R. Cheu, and S. Teo. Taxi dispatch system based on current demands and real-time traffic conditions. *Transportation Research Record:Journal of the Transportation Research Board*, 8(1882):193–200, 2004.

[19] E. Lehmann and J. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics, 2010.

[20] F. Miao, S. Han, S. Lin, and G. J. Pappas. Robust taxi dispatch under model uncertainties. In *54th IEEE Conference on Decision and Control (CDC)*, pages 2816–2821, Dec 2015.

[21] F. Miao, S. Han, S. Lin, J. A. Stankovic, D. Zhang, S. Munir, H. Huang, T. He, and G. J. Pappas. Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach. *IEEE Transactions on Automation Science and Engineering*, 13(2):463–478, April 2016.

[22] F. Miao, S. Lin, S. Munir, J. A. Stankovic, H. Huang, D. Zhang, T. He, and P. G. J. Taxi dispatch with real-time sensing data in metropolitan areas — a receding horizon control approach. In *6th International Conference of Cyber-Physical Systems*, 2015.

[23] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas. Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1393–1402, Sept 2013.

[24] A. Pantoja and N. Quijano. A population dynamics approach for the dispatch of distributed generators. *IEEE Transactions on Industrial Electronics*, 58(10):4559–4567, Oct 2011.

[25] M. Pavone, S. L. Smith, E. Frazzoli, and D. Rus. Robotic load balancing for mobility-on-demand systems. *Int. J. Rob. Res.*, 31(7):839–854, June 2012.

[26] J. Shawe-Taylor and N. Cristianini. Estimating the moments of a random estimating the moments of a random vector with applications. In *GRETSI Conference*, pages 47–52, 2003.

[27] K. I. Wong and M. G. H. Bell. The optimal dispatching of taxis under congestion: A rolling horizon approach. *Journal of Advanced Transportation*, 40:203–220, 2006.

[28] D. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. Bayen. A traffic model for velocity data assimilation. *Applied Research Mathematics eXpress (ARMX)*, pages 1–35, 2010.

[29] D. Zhang, T. He, S. Lin, S. Munir, and J. Stankovic. Dmodel: Online taxicab demand model from big sensor data in a roving sensor network. In *2014 IEEE International Congress on Big Data (BigData Congress)*, pages 152–159, June 2014.

[30] D. Zhang, T. He, S. Lin, S. Munir, and J. Stankovic. Online cruising mile reduction in large-scale taxicab networks. *IEEE Transactions on Parallel and Distributed Systems*, 26(11):3122–3135, Nov 2015.

[31] R. Zhang and M. Pavone. Control of robotic mobility-on-demand systems: a queueing-theoretical perspective. In *Proceedings of Robotics: Science and Systems*, July 2014.

VII. APPENDIX

A. Proof of Lemma 1

Proof: We first consider the problem of minimizing $\sum_{i=1}^{n}\left|\frac{r_i^k}{\sum_{j=1}^{n}X_{ji}^k-\sum_{j=1}^{n}X_{ij}^k+L_i^k}-\frac{\sum_{j=1}^{n}r_j^k}{N^k}\right|$ for one time slot k. To simplify notation, let $r_i^k=a_i\geqslant 1$,

$$\sum_{j=1}^{n}X_{ji}^k-\sum_{j=1}^{n}X_{ij}^k+L_i^k=b_i,\ i=1,\dots,n. \tag{28}$$

Given a vector L^k that satisfies $L_i^k\geqslant 0$, $\sum_{i=1}^{k}=N^k$, we have $\sum_{i=1}^{n}b_i=\sum_{i=1}^{n}L_i^k=N^k$, since balancing vacant vehicles does not change the total number of vacant vehicles in the city. It is worth noting that given any values of $b_1>0,\dots,b_n>0$, $L_1^k\geqslant 0,\dots,L_n^k\geqslant 0$ that satisfies $\sum_{i=1}^{n}b_i=\sum_{i=1}^{n}L_i^k$, the equation set (28) has a feasible solution for $n\times n$ variables of the matrix X^k. This can be easily checked by vectorizing matrix X^k to a vector $Y^k\in\mathbb{R}^{n^2}$ and transforming equation set (28) to a new equation set of Y^k. We get a homogeneous equation set with n equations in $n\times n$ variables in Y^k and always has a feasible solution.

To explain how (7) approximates (6) under constraints (4) and (5), consider the following problem given $a_1,\dots,a_n$, $N^k=c$:

$$\underset{b>0,\sum_i b_i=c}{\text{minimize}}\ \sum_i\frac{a_i}{b_i^\alpha},\quad c\text{ is a constant.} \tag{29}$$

We substitute $b_n=c-b_1\cdots-b_{n-1}$ into (29), and take partial derivatives of $\sum_i\frac{a_i}{b_i^\alpha}$ over $b_i, i=1,\dots,n-1$. When the

minimum of (7) is achieved, each partial derivative should be 0, namely

$$-\alpha\frac{a_i}{b_i^{\alpha+1}}-\alpha(-1)\frac{a_n}{(c-b_1\cdots-b_{n-1})^{\alpha+1}}=0,$$

which is equivalent to

$$\frac{a_1}{b_1^{\alpha+1}}=\cdots=\frac{a_{n-1}}{b_{n-1}^{\alpha+1}}=\frac{a_n}{b_n^{\alpha+1}}.$$

Let $\frac{a_1}{b_1^{\alpha+1}}=\cdots=\frac{a_{n-1}}{b_{n-1}^{\alpha+1}}=\frac{a_n}{b_n^{\alpha+1}}=c_0$, $\gamma=\frac{1}{\alpha+1}$, when $\alpha>0$, $0<\gamma<1$. Assume that $\sum_{i=1}^{n}a_i=a$, then

$$a_1^\gamma=b_1c_0,\quad\ldots,\quad a_n^\gamma=b_nc_0,$$

$$\sum_{i=1}^{n}a_i^\gamma=(b_1+\cdots+b_n)c_0=cc_0\Rightarrow c_0=\frac{1}{c}\sum_{i=1}^{n},$$

$$a_i^\gamma=\frac{b_i}{c}\sum_{j=1}^{n}a_j^\gamma,\quad \frac{a_i}{b_i}=\frac{a_i^{1-\gamma}}{c}\sum_{j=1}^{n}a_j^\gamma$$

We would like to prove that for any $\epsilon_0>0$, any $i\in\{1,\ldots,n\}$, there exists a $0<\gamma<1$, such that

$$\left|\frac{a_i^{1-\gamma}}{c}\sum_{j=1}^{n}a_j^\gamma-\frac{a}{c}\right|<\epsilon_0,\quad \sum_{i=1}^{n}\left|\frac{a_i^{1-\gamma}}{c}\sum_{j=1}^{n}a_j^\gamma-\frac{a}{c}\right|<n\epsilon_0. \tag{30}$$

To prove (30), it is worth noting that for any given values of $a_i\geqslant 1, i=1,\ldots,\tau, c>0$, function $f_i(\gamma)=\frac{a_i^{1-\gamma}}{c}\sum_{j=1}^{n}a_j^\gamma$ is a continuous function of γ, and $f_i(\gamma=1)=\frac{a}{c}$ for any i. Then for any $\epsilon_0>0$ and any i, there exists a $\delta_i>0$, such that

$$|\gamma-1|<\delta_i\Rightarrow\left|\frac{a_i^{1-\gamma}}{c}\sum_{j=1}^{n}a_j^\gamma-\frac{a}{c}\right|<\epsilon_0.$$

Considering the case for all time slots k, similarly to the above proof, we conclude that for each index k and i, there exists a $\delta_i^k>0$, such that $|\gamma-1|<\delta_i^k$ indicates the ϵ_0-bound holds. Then let $\delta=\min\{\delta_1^1,\delta_2^1,\ldots,\delta_n^k\}$ (when ϵ_0 is small δ also indicates a small range, so we have $\delta<1$), then for all γ in the range $1-\delta<\gamma<1$, or equivalently all α in the range $0<\alpha<\frac{\delta}{1-\delta}$, the inequality (8) holds.

When a feasible solution violates the non-negative constraint of X_{ij}^k, just compare the value of X_{ij}^k and X_{ji}^k, without loss of generality we assume that $X_{ij}^k>X_{ji}^k$, then let the final feasible solution be $(X_{ij}^k)'=X_{ij}^k-X_{ji}^k$, $(X_{ji}^k)'=0$, the equation set (28) still holds and we have a non-negative solution of X_{ij}^k, X_{ji}^k.

It is worth noting that when ϵ_0 is small and γ_0 is close to 1, α is close to 0. ■

B. Proof of Theorem 1

Proof: For any fixed X, the maximum part of the objective function is equivalent to

$$\max_{r\in\Delta}\quad J_D(X)+\beta J_E(X,r)=J_D(X)+c^T(X)r$$
$$[c(X)]_i=\beta\frac{1}{(\sum_{j=1}^{n}X_{ji}-\sum_{j=1}^{n}X_{ij}+L_i)^\alpha},J_D(X)=\sum_i\sum_j X_{ij}W_{ij}. \tag{31}$$

The Lagrangian of problem (31) with the Lagrangian multipliers $\lambda\geq 0, v\geq 0$ is

$$\mathcal{L}(X,r,\lambda,v)=J_D(X)+b^T\lambda-(A^T\lambda-c(X)-v)^Tr,$$

where $(A^T\lambda-c(X)-v)^Tr$ is a linear function of r, and the upper bound exists only when $A^T\lambda-c(X)-v=0$. The objective function of the dual problem is

$$g(X,\lambda,v)=\sup_{r\in\Delta}\mathcal{L}(X,r,\lambda,v)$$
$$=\begin{cases}J_D(X)+b^T\lambda & \text{if}\quad A^T\lambda-c(X)-v=0.\\ \infty & \text{otherwise}\end{cases}$$

With $v\geq 0$, the constraint $A^T\lambda-c(X)-v=0$ is equivalent to $A^T\lambda-c(X)\geq 0$. Strong duality holds for problem of (31) since it satisfies the refined Slater's condition for affine inequality constraints as stated in book [8, Chapter 5.2.3]—the primal problem is convex, $c^T(X)r$ is affine of r, and by the definition of the uncertainty set, the non-empty affine inequality constraint of r is feasible. The primal convex problem is feasible with affine inequality constraints. The dual problem of (31) is

$$\begin{aligned}\underset{\lambda\geq 0}{\text{minimize}}\quad & J_D(X)+b^T\lambda\\ \text{subject to}\quad & A^T\lambda-c(X)\geq 0.\end{aligned} \tag{32}$$

Hence, problem (11) with $\tau=1$ can be solved as the convex optimization problem defined in (21). ■

C. Proof of Lemma 2

Proof: Now consider the maximin problem over stage k and $k+1$, $1\leqslant k\leqslant\tau-1$ of problem (11)

$$\max_{r^k\in\Delta_k}\min_{X^{k+1},L^{k+1}}\quad J=\sum_{k=1}^{\tau}(J_D(X^k)+\beta J_E(X^k,r^k))$$
$$\text{s.t.}\quad \text{constraints of (11).} \tag{33}$$

The domain of problem (33) satisfies that $X^{k+1}, L^{k+1}, \lambda$ is compact, and the domain of r^k is compact. The objective function is a closed function convex over X^{k+1}, L^{k+1} and concave over r^k. According to Proposition 2.6.9 with condition (1) of [5], when the objective and constraint functions are convex of the decision variables, concave of the uncertain parameters, and the domain of decision variables and uncertain parameters are compact, the set of saddle points for the maximin problem at time k and $k+1$, i.e., $\max_{r^k\in\Delta_k}\min_{X^{k+1},L^{k+1}}J$ with the objective function and constraints of problem (33) is nonempty. It means that the minimax equality holds for problem (33) at time k and $k+1$, and we have:

$$\max_{r^k\in\Delta_k}\min_{X^{k+1},L^{k+1}}J=\min_{X^{k+1},L^{k+1}}\max_{r^k\in\Delta_k}J.$$

Repeat the above proof process from $k=\tau-1$ backwards to $k=1$, we get a minimax form of robust optimization problem $\min_{X^{1:\tau},L^{2:\tau}}\max_{r^1\in\Delta_1,\ldots,r^\tau\in\Delta_\tau}J$. By the definition of demand uncertainty sets, δ_k is a projection of Δ, and the minimax problem is equivalent to $\min_{X^{1:\tau},L^{2:\tau}}\max_{r_c\in\Delta}J$. We then get the conclusion of this lemma. ■

D. Proof of Lemma 3 and Theorem 2

1) Proof of Lemma 3: *Proof:* With the polytope form of uncertainty set (23), the domain of each r^k is closed and convex, i.e., is compact, and Lemma 2 holds. Considering the maximizing part of problem (22)

$$\max_{r^1\in\Delta_1,\ldots,r^\tau\in\Delta_\tau} J, \quad \text{s.t.} \quad \text{constraints of (11)}, \tag{34}$$

the Lagrangian of (34) with multipliers $\lambda^k \geq 0, v^k \geq 0$ is

$$\begin{aligned}&\mathcal{L}(X^k, r^k, \lambda^k, v^k)\\ =&\sum_{k=1}^{\tau}(J_D(X^k) + b_k^T\lambda^k - (A_k^T\lambda^k - c(X^k) - v^k)^T r^k),\end{aligned} \tag{35}$$

Hence, based on the proof of Theorem 1, we take partial derivative of the Lagrangian (35) for every $r^k \in \Delta_k$. The inequality constraint of $r^k \in \Delta_k$ defined as (23) is affine of r^k and feasible (non-empty), $c^T(X^k)r^k$ is affine of r^k, and problem (34) is convex with feasible affine inequality constraints. Hence, refined Slater's condition for affine constraints is satisfied and strong duality holds for problem (34). An equivalent form of (11) under uncertainty set (23) is defined as (24). ∎

2) Proof of Theorem 2: *Proof:* With uncertain set defined as (25), the domain of each r^k is compact and Lemma 2 holds. We consider the equivalent problem (22) of problem (11), and first derive the Lagrangian of the maximum part of the objective function (22) with constraint $\lambda \geq 0, v_k \geq 0$

$$\begin{aligned}&\mathcal{L}(X^k, r^k, \lambda, v_k)\\ =&b^T\lambda - \sum_{k=1}^{\tau}((A_k^T\lambda - c(X^k) - v_k)^T r^k - J_D(X^k)),\end{aligned} \tag{36}$$

Similarly as the proof of Theorem 1, we take the partial derivative of (36) over each r^k, the objective function of the dual problem is

$$\begin{aligned}g(X^k, L^k, \lambda, r^k) &= \sup_{r^k\in\Delta_k} \mathcal{L}(X^k, r^k, \lambda, v_k)\\ &= \begin{cases}\infty & \text{if } \exists k \text{ s.t. } A_k^T\lambda - c(X^k) - v_k \neq 0,\\ \sum_{k=1}^{\tau} J_D(X^k) + b^T\lambda & \text{o.w.}\end{cases}\end{aligned}$$

Since the inequality constraint of the uncertainty set defined as (25) is affine of each r^k and feasible (non-empty uncertainty set), $c^T(X^k)r^k$ is affine of r^k, and problem (34) is convex with feasible affine inequality constraints, refined Slater's condition with affine inequality constraints is satisfied. Then strong duality holds, problem (26) is a equivalent to the computationally tractable convex optimization form (11) under uncertain set (25). ∎

E. Proof of Theorem 3

Proof: Under the definition of uncertainty set (20) for concatenated r^k, the domain of each r^k is compact, and problem (11) is equivalent to (22). We now consider the dual form for the objective function $\sum_{k=1}^{\tau} J_E(X^k, r^k)$ that relates to r^k. By the definition of inner product, we have

$$\sum_{k=1}^{\tau} c^T(X^k)r^k = c_l^T(X)r_c, c_l(X) = [c^T(X^1) \ \ldots \ c^T(X^\tau)]^T.$$

When the uncertainty set of r_c is an SOC defined as (20), problem (22) is equivalent to

$$\begin{aligned}\min_{X^k, L^k}\ \max_{r_c\geqslant 0}\ &\left(c_l^T(X)r_c + \sum_{k=1}^{\tau}\sum_i\sum_j X_{ij}^k W_{ij}\right)\\ \text{subject to}\ \ &r_c = \hat{r}_c + y + C^T w,\\ &\|y\|_2 \leqslant \Gamma_1^B, \|w\|_2 \leqslant \sqrt{\frac{1}{\epsilon} - 1},\\ &\text{constraints of (11)}.\end{aligned} \tag{37}$$

We first consider the following minimax problem related to the uncertainty set

$$\begin{aligned}\max_{r_c\geqslant 0}\ \ &c_l^T(X)r_c\\ \text{subject to}\ \ &r_c = \hat{r}_c + y + C^T w,\\ &\|y\|_2 \leqslant \Gamma_1^B, \|w\|_2 \leqslant \sqrt{\frac{1}{\epsilon} - 1}.\end{aligned} \tag{38}$$

The constraints of problem (38) have a feasible solution $r_c = \hat{r}_c$, $y = 0$ and $w = 0$, such that $\|y\|_2 < \Gamma_1^B, \|w\|_2 < \sqrt{\frac{1}{\epsilon} - 1}$, and $c_l^T(X)r_c$ is affine of r_c. Hence, Slater's condition is satisfied and strong duality holds.

To get the dual form of problem (38), we start from the following Lagrangian with $v \geqslant 0$

$$\mathcal{L}(X, r_c, z, v) = c_l^T(X)r_c + z^T(\hat{r}_c + y + C^T w - r_c) + v^T r_c.$$

By taking the partial derivative of the above Lagrangian over r_c, we get the supreme value of the Lagrangian as

$$\sup_{r_c}\mathcal{L}(X, r_c, z, v) = \begin{cases} z^T(\hat{r}_c + y + C^T w) & \text{if } c_l(X) \leqslant z\\ \infty & \text{o.w.}\end{cases}$$

Then with the norm bound of y and w, we have

$$\begin{aligned}&\sup_{\|y\|_2\leqslant\Gamma_1^B, \|w\|_2\leqslant\sqrt{\frac{1}{\epsilon}-1}} (z^T(\hat{r}_c + y + C^T w))\\ =&\hat{r}_c^T z + \Gamma_1^B\|z\|_2 + \sqrt{\frac{1}{\epsilon} - 1}\|Cz\|_2.\end{aligned}$$

Hence, the objective function of the dual problem for (38) is

$$\begin{aligned}g(X, r_c, z) &= \sup_{r_c\in\mathcal{U}_\epsilon^{CS}} \mathcal{L}(X, r_c, z)\\ &= \begin{cases}\hat{r}_c^T z + \Gamma_1^B\|z\|_2 + \sqrt{\frac{1}{\epsilon} - 1}\|Cz\|_2, & \text{if } c_l(X) \leqslant z\\ \infty & \text{o.w..}\end{cases}\end{aligned}$$

Together with the objective function $J_D(X^k)$ and other constraints that do not directly involve r_c, an equivalent convex form of (11) given the uncertainty set (20) is shown as (27). ∎