



Production, Manufacturing, Transportation and Logistics

## Traffic signal control under stochastic traffic demand and vehicle turning via decentralized decomposition approaches



Xinyu Fei<sup>a</sup>, Xingmin Wang<sup>b</sup>, Xian Yu<sup>a</sup>, Yiheng Feng<sup>c</sup>, Henry Liu<sup>b</sup>, Siqian Shen<sup>a,\*</sup>, Yafeng Yin<sup>b</sup>

<sup>a</sup> Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, USA

<sup>b</sup> Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, USA

<sup>c</sup> Lyles School of Civil Engineering, Purdue University, USA

### ARTICLE INFO

#### Article history:

Received 15 April 2022

Accepted 5 April 2023

Available online 9 April 2023

#### Keywords:

Traffic

Traffic signal control

Alternating direction method of multipliers (ADMM)

Stochastic mixed-integer programming

Benders decomposition

### ABSTRACT

Traffic congestion is a global pressing issue but can be mitigated via effective traffic signal control schemes. In this paper, based on a cell transmission model we coordinate the control of traffic signals at multiple intersections to maximize vehicle throughput on corridors or road networks, under stochastic traffic demand and vehicle turning. We formulate a two-stage stochastic mixed-integer linear program using finite samples of the uncertain parameter, and combine Benders decomposition with the alternating direction method of multipliers to develop spatially-temporally distributed algorithms for optimizing the problem. We test instances of traffic signal control on corridors and grid networks, generated based on synthetic and real-world traffic data. Our results show that (i) considering traffic uncertainty can significantly improve the signal control quality and (ii) decentralized decomposition approaches can quickly find high-quality signal plans for multiple intersections in complex road networks, and fully utilize the computation and communication technologies in smart-transportation infrastructures.

© 2023 Elsevier B.V. All rights reserved.

### 1. Introduction

In the past few decades, city scales have increased significantly and as a result, privately owned vehicles increase, resulting in rapidly growing congestion issues in cities of all sizes worldwide. According to Schrank et al. (2019), in 2017, traffic congestion caused urban Americans to travel extra 8.8 billion hours and to purchase extra 3.3 billion gallons of fuel. Taking the city of Detroit in the United States (US) as an example, in 2019, each driver lost 39 hours on the road on average due to traffic delay (Levin, 2020). Various approaches have been proposed for addressing traffic congestion issues on road networks, including congestion pricing, road expansion, and traffic signal control. Among them, traffic signal control can effectively mitigate congestion by optimizing the traffic signal timing parameters (cycle, green split, and offset) at signalized intersections without major changes to the existing infrastructure (Isa et al., 2014).

Existing traffic signal control systems can be divided into three main categories: 1) fixed-time; 2) actuated; and 3) adaptive control according to their responsiveness and flexibility. Both actuated

and adaptive are real-time control strategies that can react to time-varying traffic demand and outperform fixed-time control in most cases. However, a large proportion of the traffic intersections in the US are still controlled by fixed-time traffic signals due to the absence of detectors (Tang et al., 2019). Even with detectors on the road, fixed-time control might still be used in congested urban areas. For fixed-time traffic signals, a whole day is split into different time of day (TOD) intervals and each of the TOD uses a traffic signal timing plan with fixed parameters (cycle, split, and offset). However, traffic conditions are stochastic and can vary within the same time periods (Yin, 2008; Yu & Recker, 2006), and changing a signal timing plan is infrequent as it requires a series of labor works such as monitoring the traffic systems, tuning the parameters, and setting new parameters to the controller (Zhang et al., 2010). Therefore, a fixed-time traffic signal timing plan is required to accommodate different traffic conditions within a certain time of day.

In this paper, we focus on the problem of optimizing fixed-time traffic signal parameters of a large-scale network considering stochastic traffic input. Based on the Cell Transmission Model (CTM) (Daganzo, 1992), this problem is formulated as a mixed-integer stochastic linear programming. There are different types of network topology in the real world including isolated intersections, corridors, and grid networks (Urbanik et al., 2015). Our proposed

\* Corresponding author.

E-mail address: [siqian@umich.edu](mailto:siqian@umich.edu) (S. Shen).

method aims to deal with large-scale networks and does not limit the network topology to a certain type. Such a centralized formulation based on CTM can better incorporate the coordination among intersections even for a complicated network topology while traditional methods can only deal with traffic signals in a corridor (Urbanik et al., 2015).

For the traffic demand, our proposed method takes the link flow and turning ratios as the direct input and both parameters can be stochastic. Even without detectors, link flow and turning ratios can be estimated from other data sources. For example, vehicle trajectory data that can be collected through different resources (e.g., ride-hailing services and navigation systems) is an ideal alternative, which is more scalable and almost available at every signalized intersection. Limited by the low penetration rate, it can hardly support real-time applications and but is more applicable for offline parameter estimation (Wang et al., 2022a; Zhan et al., 2016; Zhao, 2019; Zhao et al., 2019b; Zheng & Liu, 2017). Other than the vehicle trajectory data, if origin-destination (OD) demand is given, an equilibrium model can be used to obtain the link flow and turning ratios (Prashker & Bekhor, 2004; Sheffi, 1985). As aforementioned, traffic demand is stochastic and time-varying and all measurements are also prone to errors and noises (Comert & Cetin, 2011; Yin, 2008). By taking a stochastic traffic demand into consideration, the resulting traffic signal timing plan is more robust and works well for more different cases.

Our goal is to establish an optimization-based control paradigm to speed up solutions to large-scale network-level traffic signal control in a distributed manner. We demonstrate the results by testing a diverse set of instances, generated using synthetic data and also real-world road networks and traffic data. Via out-of-sample tests, we show that the signal control plans produced by our methods perform consistently better than plans solved using off-the-shelf optimization solvers, or models that do not consider uncertainty. Our decomposition and decentralized algorithms can significantly reduce computational time and produce reliable traffic signal plans for moderate-sized city networks.

The remainder of the paper is organized as follows. In Section 2, we review the most relevant literature on traffic signal control and optimization methods used in this paper. In Section 3, we formulate the traffic signal control problem under uncertain demand and turning ratios as a two-stage stochastic mixed-integer program (MIP). In Section 4, we combine the alternating direction method of multipliers (ADMM) and Benders decomposition algorithms to develop spatially distributed algorithms for solving the stochastic MIP in a decentralized manner. In Section 5, we present numerical results of diverse instances with different uncertainty settings. In Section 6, we conclude the paper and state future research directions.

## 2. Literature review

Traffic flow models have different scales from microscopic to macroscopic (see, e.g., Hoogendoorn & Bovy, 2001; Li & Ioannou, 2004). A microscopic model simulates the behavior of individual vehicles using dynamic variables to describe states such as location and velocity of vehicles (Pipes, 1953; Wu & Brilon, 1999). A meso- or macro-scale model simulates traffic conditions using variables that describe overall states of all vehicles, including traffic density, volume, and average speed (Hoogendoorn & Bovy, 2001). Daganzo (1992) first develops a cell transmission model (CTM) to model traffic flows on highways. As a mesoscopic traffic model, the CTM divides roads into homogeneous sections and considers states of each section across discrete time steps. We refer interested readers to Adacher & Tiriolo (2018) for a comprehensive review of different CTM variants and appropriate situations for using them. Compared to solving microscopic traffic mod-

els, computing macroscopic models is much more efficient because of the aggregated traffic states and the resulting fewer number of variables and dynamic equations. In this paper, our optimization framework is based on the CTM for traffic condition modeling. Next, in Section 2.1, we review the most relevant work on distributed traffic signal control, especially with respect to ADMM. Section 2.2 is a literature review of stochastic traffic signal control and Section 2.3 includes contributions of this paper.

### 2.1. Distributed traffic signal control

Many existing studies formulate the traffic signal control problem to optimize the network performance based on a certain traffic flow model. Lo (1999) builds an MIP based on the CTM for deriving traffic signal control policies but does not consider traffic demand or vehicle turning uncertainties. The model is only tested on corridors instead of general road networks, mainly due to the exponentially increased number of variables and constraints in the latter case. Indeed, solving an MIP requires gathering global traffic information in a road network and becomes extremely difficult as the network size increases. On the other hand, each intersection is physically isolated, making it natural to consider decentralized traffic signal control. Advancements in parallel computing, sensing, and wireless communication technologies allow efficient implementations of decentralized control at individual intersections, leading to more research focusing on how to design parallel and distributed traffic signal control schemes. For example, Al Islam & Hajbabaie (2017) present a distributed-coordinated approach for signal timing optimization in connected urban street networks. They first, formulate the traffic signal control problem at each intersection as an MIP. Then, after separately solving the MIP associated with each intersection, all the intersections communicate to each other to improve their local traffic control strategies (guided by heuristics), obtain estimated traffic flows, and update parameters in their MIPs for the next time period. Tajalli et al. (2020) also formulate the traffic signal control problem as an MIP based on CTM for each intersection but consider the coordination of intersections in a certain time period by penalizing the difference between estimated and real traffic flow in the objective function. Furthermore, a series of literature follows the procedure that first fixes traffic signals for each intersection and runs simulations on the overall traffic network, and then updates traffic signals according to the estimated information from the simulation (see, e.g., Hajbabaie & Benekohal, 2015; Liang et al., 2021; Stevanovic et al., 2007). All of the aforementioned papers fail to ensure the optimality of the solution and do not consider traffic uncertainties.

ADMM was initially developed to solve convex programs that can be decomposed into multiple sub-clusters and the objective function is the summation of functions related to each sub-cluster. For convex programs, it is proved that both the objective value and solutions obtained by ADMM converge to the optimum (Boyd et al., 2011). Timotheou et al. (2014) propose a distributed algorithm that decomposes a traffic network into individual intersections for traffic signal control and they apply ADMM to obtain the coordination between intersections as well as solving the MIP for each intersection individually. However, vehicle turning is not considered in their model and the work does not take into account the stochasticity of traffic demand or vehicle turning.

Another class of literature is about the max pressure control, which is originally studied in the communication network for routing and scheduling (Neely, 2010; Tassiulas & Ephremides, 1990). Varaiya (2013) firstly introduces the pressure-based method to solve the distributed traffic signal control problem by assigning traffic signals according to queue lengths of different directions. The max pressure control is derived based on the store-and-forward model and proved to be throughput-optimal, which

means it can stabilize the network queue lengths when the traffic demand is within the network capacity. Many studies have been conducted to improve the max pressure control from different perspectives (Levin et al., 2020; Li & Jabari, 2019; Wang, 2022b; Wu et al., 2017). For example, Zaidi et al. (2016) modify the original max pressure method to conduct traffic signal control taking fixed and adapt routing into consideration in a general traffic network. Levin et al. (2020) propose to utilize a cyclic phase structure which is easier to be implemented in the real world. However, the store-and-forward model has some strong assumptions which essentially ignore the vehicle length and travel time along the link (Aboudolas et al., 2009). Besides, the max pressure control only ensures the stability of the network by showing that the overall network queue lengths are bounded. According to Little's law (Little & Graves, 2008), bounded total queue lengths guarantee a bounded total delay but not the optimal minimum delay. By taking the real-time traffic state (e.g., queue length) as the input, the implementation of the max pressure control has a high requirement for real-time traffic monitoring and detection, which is not available for a large proportion of signalized intersections. As aforementioned, this paper focuses on generating background or fixed-time traffic signal timing for those intersections without real-time monitoring capability.

## 2.2. Traffic signal control under uncertainties

Starting from Heydecker (1987) investigating the impact of uncertain traffic flows and showing that using the average flow results in unsatisfactory performance, a series of studies have been conducted on proposing traffic signal timing plan under flow fluctuations. Yin (2008) and Zhang et al. (2010) consider fixed-time signal timing plans for road corridors under stochastic traffic demand. They ensure that the probability of the travel delay exceeding a given threshold is sufficiently small and apply heuristics to solve the stochastic MIPs. Based on their work, Shirke et al. (2022) propose a metaheuristic approach to design robust traffic signal timing plans efficiently. Another class of existing literature focuses on optimizing the average performance given the distribution of uncertain parameters. For example, Tong et al. (2015) propose a two-stage stochastic linear program (LP) for controlling traffic signals at one intersection to minimize the expected traffic delay under uncertain traffic demand. Li et al. (2018) extend the previous work and formulate a two-stage stochastic LP for coordinated traffic signal control on a corridor. However, the models in both Tong et al. (2015) and Li et al. (2018) relax the integrality constraints. Li et al. (2021) propose a two-stage MIP and apply a heuristic stochastic gradient method to solve the model. Moreover, in all these related work, the authors only consider one intersection or corridors instead of complex road networks. For general traffic networks, Chiou (2019) develop a two-stage mixed-integer nonlinear program based on user equilibrium instead of the CTM and apply a Quasi-Newton method to solve the model in centralization.

## 2.3. Main contributions of our paper

Compared to the prior work, the main contributions of this paper are threefold. First, we consider a fixed-time traffic signal control problem and formulate it on general grid networks as an MIP based on the CTM, which takes into account the queue propagation along links and optimizes traffic delay directly. We consider both vehicle movements along corridors and turning movements of vehicles, to obtain better coordination between intersections within a traffic network. Second, we extend the deterministic MIP to a stochastic one by incorporating various types of input parameter

uncertainties, such as stochastic traffic demand and turning ratios. Third, we speed up the computation of the formulation by integrating Benders decomposition and ADMM approaches. Specifically, we ignore the traffic flow relationship between intersections and obtain a signal timing plan for each intersection. Then, we coordinate the information of neighboring intersections and update the parameters of each intersection. Such a distributed computational framework can best utilize advanced technologies in intelligent transportation systems such as distributed micro-computers – that can be installed at traffic intersections. In Table 1, we compare our work with the most relevant papers, in terms of modeling methods, instance scales, types of decisions, assumptions on parameters, and solution approaches.

## 3. A Stochastic MIP for traffic signal control

We define the notation and introduce CTM in Section 3.1. We then formulate the fixed-time traffic signal control as a two-stage stochastic MIP using CTM in Section 3.2, where we aim to maximize the expected total throughput (i.e., the number of vehicles going through the whole traffic network). The model takes the initial traffic condition, samples and probabilities of source demand and vehicle turning ratios as input parameters and computes optimal signal control plans at all intersections.

### 3.1. Problem description and notation

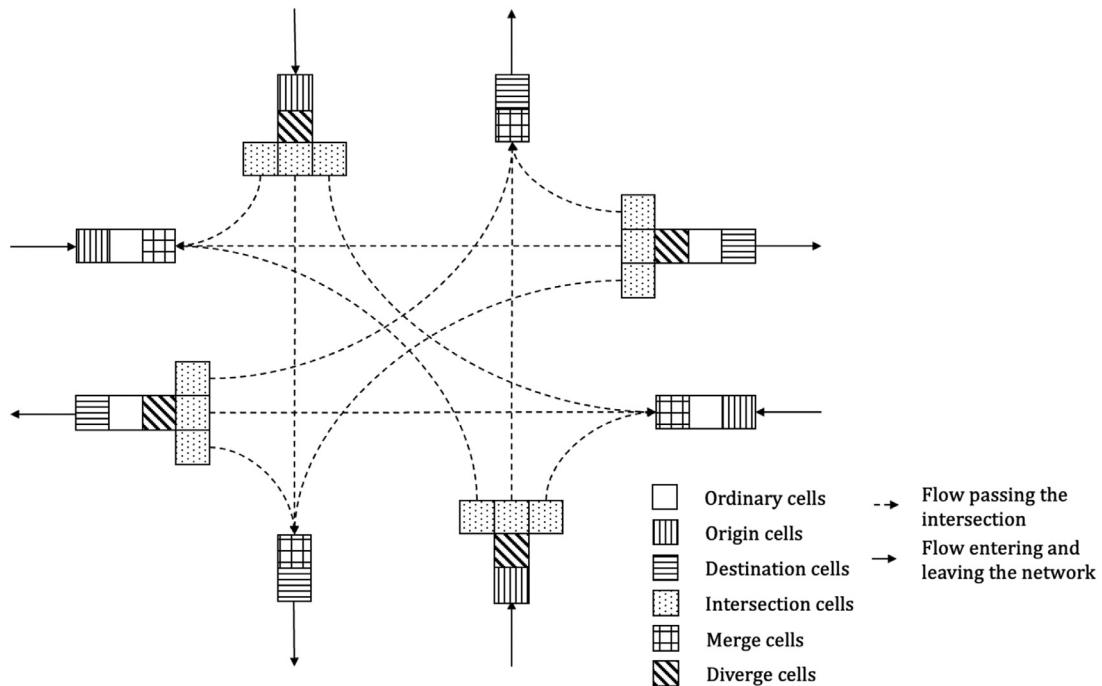
*Structure of the CTM* Consider a road network  $G(V, E)$  where  $V$  is the set of nodes and  $E$  is the set of arcs. Here the nodes are signalized intersections and the arcs refer to road segments connecting pairs of the intersections. In this paper, we focus on road networks in which at most four arcs are connected to the same node, representing North, South, East and West directions and therefore the underlying network is a grid. Each arc is partitioned into homogeneous sections called cells, where the length of each cell is the distance traveled by a vehicle at normal speed without traffic congestion in a unit of time. Let  $\mathcal{C} = \mathcal{E} \cup \mathcal{O} \cup \mathcal{D} \cup \mathcal{I} \cup \mathcal{M} \cup \mathcal{V}$  be the set of cells where  $\mathcal{E}$  is the set of ordinary cells,  $\mathcal{O}$  the set of origin cells,  $\mathcal{D}$  the set of destination cells,  $\mathcal{I}$  the set of intersection cells,  $\mathcal{M}$  the set of merge cells and  $\mathcal{V}$  the set of diverge cells. Ordinary cells have both inflow and outflow of vehicles from other cells. Origin cells receive exogenous inflow and destination cells send outflow traffic outside the network. Intersection cells are cells where vehicles can choose to turn left, go straight or turn right. Merge cells receive inflow traffic from more than one cell, and diverge cells send outflow traffic to more than one cell. Let  $d(c)$  be the set of cells sending their outflow to a cell  $c \in \mathcal{C}$  and  $p(c)$  be the set of cells receiving their inflow from a cell  $c \in \mathcal{C}$ . Let  $\mathcal{R} = \{1, 2, \dots, N_I\}$  be the set of intersections where  $N_I$  denotes the total number of intersections. A signal phase of a signalized intersection is a representation of moving directions that can allow vehicles to pass the intersection at the same time. Let  $\mathcal{F}_i = \{1, 2, \dots, |\mathcal{F}_i|\}$  be the set of indices of all the phases of a signalized intersection  $i$  for all  $i \in \mathcal{R}$ . Let  $I_{ij}$  be the set of the phases of the intersection cells of intersection  $i$  and phase  $j$  for all  $i \in \mathcal{R}$  and  $j \in \mathcal{F}_i$ . We discretize the whole time horizon into  $T$  time steps. Figure 1 depicts the way of transforming one intersection and its related arcs into cells in CTM. We refer the interested readers to the legend in Fig. 1 for different types of the aforementioned cells in the CTM and the traffic flow.

*Input parameters* Let  $n_c^{\text{init}}$  be the number of initial vehicles inside each cell  $c \in \mathcal{C}$ , and let  $\beta_{cc'}$  denote the turning ratio of a diverging cell  $c \in \mathcal{V}$  moving towards the direction of an intersection cell  $c' \in d(c)$ . Notice that  $\sum_{c' \in d(c)} \beta_{cc'} = 1$  for each  $c \in \mathcal{V}$ . Let  $Q_{ct}$  and  $N_{ct}$  respectively represent the maximum number of vehicles that can flow through and reside in cell  $c \in \mathcal{C}$ , and  $W_{ct}$  be

**Table 1**

Comparison between our paper and other traffic signal control papers.

| Paper   | Model                                     | Scale                 | Decision   | Input         | Approach      |
|---|---|-----------------------|------------|---------------|---------------|
| Our work  | Two-stage stochastic MIP based on the CTM | network               | fixed-time | stochastic    | decentralized |
| Lo (1999)   | Deterministic MIP                         | corridor              | real-time  | deterministic | centralized   |
| Shirke et al. (2022); Yin (2008); Zhang et al. (2010) | Stochastic MIP                            | corridor              | fixed-time | stochastic    | centralized   |
| Tong et al. (2015)                                    | Two-stage stochastic LP                   | isolated intersection | real-time  | stochastic    | centralized   |
| Li et al. (2018)                                      | Two-stage stochastic LP                   | corridor              | real-time  | stochastic    | centralized   |
| Li et al. (2021)                                      | Two-stage stochastic MIP                  | corridor              | real-time  | stochastic    | centralized   |
| Chiou (2019)  | Two-stage stochastic MIP                  | network               | fixed-time | stochastic    | centralized   |
| Tajalli et al. (2020); Timotheou et al. (2014)        | Deterministic MIP                         | network               | real-time  | deterministic | decentralized |
| Varaiya (2013); Zaidi et al. (2016)                   | Pressure-based method                     | network               | real-time  | deterministic | decentralized |

**Fig. 1.** A CTM-based transformation of an intersection.

the ratio between the shock-wave propagation speed and the flow-free speed of a cell  $c$  during each time interval  $[t, t+1]$  for  $t \in \{1, \dots, T\}$ . Moreover,  $D_{ct}$  denotes the number of vehicles entering an origin cell  $c \in \mathcal{O}$  during time  $[t, t+1]$  (i.e., the source demand considered in this paper). Parameter  $G_{\min}$  and  $G_{\max}$  indicate the minimum and maximum green time, meaning that any green signal set at time  $t$  cannot change during time  $[t, t+G_{\min}]$  and the green signal of the same direction cannot last for more than  $G_{\max}$  time (Zhang & Wang, 2010), respectively. Let  $N_{cy}$  be the maximum number of cycles in the whole time horizon where traffic signal plans are the same in each cycle. We use  $n^{\text{init}}$ ,  $\beta$ ,  $Q$ ,  $N$ ,  $W$ ,  $D$  to represent the vector forms of all the above parameters, respectively. In addition, we introduce a sufficiently large parameter  $U$ , a sufficiently small parameter  $\epsilon$ , and a weight parameter  $\alpha$  to formulate constraints and the objective function in our stochastic MIP described later.

We denote  $\xi = [D, \beta]$  as the overall uncertain parameter and  $P$  as the probability distribution of  $\xi$ , which is assumed known and can be derived from historical data. Without loss of generality, we assume a discrete distribution  $P$  and a finite set  $\Xi$  of realizations such that  $\Xi = \{\xi^1, \dots, \xi^K\}$  and each realization  $\xi^k$  is associated with probability  $p^k$  such that  $\sum_{k=1}^K p^k = 1$ .

**Decision variables** We determine the traffic signal control plan by finding the beginning time and ending time when the traffic signal is green for each phase and each cycle. Correspondingly, we

define a continuous variable  $l_i$  as the cycle length and a continuous variable  $o_i$  as the offset of each intersection  $i \in \mathcal{R}$ . Variable  $g_{ij}$  indicates the interval length when the traffic signal of an intersection  $i$  and a phase  $j$  is green, and continuous variables  $b_{ijm}$  and  $e_{ijm}$  indicate the beginning and ending green time at intersection  $i$ , phase  $j$  during cycle  $m$ , for each  $i \in \mathcal{R}$ ,  $j \in \mathcal{F}_i$ , and  $m \in \{1, \dots, N_{cy}\}$ . For  $t \in \{1, \dots, T\}$ , we define binary variables  $z_{1ijmt}$  and  $z_{2ijmt}$  to describe the relationship between the time step  $t$ , the beginning green time  $b_{ijm}$  and the ending green time  $e_{ijm}$  of the intersection  $i$ , phase  $j$  and cycle  $m$ , of which the details are provided in constraints (1a)–(1c). For each cell  $c \in \mathcal{C}$  and time  $t \in \{1, \dots, T\}$ , we define continuous variables  $y_{ct}$  and  $n_{ct}$  as the number of vehicles leaving and inside cell  $c$  during time  $[t, t+1]$ , respectively. We use  $l, o, b, e, g, z_1, z_2, y, n$  to represent the vector forms of all the above continuous and binary decision variables.

### 3.2. Two-stage stochastic optimization model

We consider a two-stage stochastic optimization approach for controlling traffic signals, where in the first stage, we determine traffic-signal related decisions, such as the cycle length  $l$ , offset  $o$ , start and end of green time intervals  $b, e$  and auxiliary binary variables  $z_1, z_2$ ; in the second stage, for each realized sample  $\xi^k = [D^k, \beta^k]$ , we can use network flow constraints to build a linear program to determine the actual number of vehicles leaving and

**Table 2**

A summary of notation of the stochastic traffic signal control problem.

| Parameters  |   |
|---|---|
| $\mathcal{C}$   | Set of all cells (where $\mathcal{C} = \mathcal{O} \cup \mathcal{D} \cup \mathcal{I} \cup \mathcal{M} \cup \mathcal{V}$ )                       |
| $\mathcal{O}, \mathcal{D}, \mathcal{I}, \mathcal{M}, \mathcal{V}$ | Set of origins, destinations, intersection, merge, diverge cells  |
| $N_I$   | The total number of intersections   |
| $\mathcal{R}$   | Set of intersections (where $ \mathcal{R}  = N_I$ )   |
| $\mathcal{F}_i$   | Set of signal phases at intersection $i$  |
| $T$   | The total number of time steps  |
| $N_{cy}$  | The total number of cycles in the traffic signal timing plan  |
| $G_{min}, G_{max}$  | Minimum and maximum green time  |
| $d(c)$  | Cells receiving inflow from $c \in \mathcal{C}$   |
| $p(c)$  | Cells sending outflow to $c \in \mathcal{C}$  |
| $n_c^{init}$  | Number of initial vehicles inside cell $c \in \mathcal{C}$  |
| $Q_{ct}, N_{ct}$  | Maximum number of vehicles flowing through and reside in cell $c \in \mathcal{C}$ during time $[t, t+1]$ for $t \in \{1, \dots, T\}$            |
| $W_{ct}$  | Ratio between shock-wave and free-flow speed of cell $c \in \mathcal{C}$ during time $[t, t+1]$ for $t \in \{1, \dots, T\}$                     |
| $D_{ct}^k$  | Source demand realization at origin $c \in \mathcal{O}$ during time $[t, t+1]$ in scenario $k = 1, \dots, K$                                    |
| $\beta_{cc'}^k$   | Turing ratio realization from cell $c \in \mathcal{V}$ to cell $c' \in d(c)$ in scenario $k = 1, \dots, K$                                      |
| Decision variables  |   |
| $l_i, o_i$  | Cycle length and offset of intersection $i \in \mathcal{R}$   |
| $g_{ij}$  | Green length of intersection $i \in \mathcal{R}$ at phase $j \in \mathcal{F}_i$   |
| $b_{ijm}, e_{ijm}$  | Beginning and ending time of the green phase $j \in \mathcal{F}_i$ in cycle $m = 1, \dots, N_{cy}$ of intersection $i \in \mathcal{R}$          |
| $z_{1ijmt}, z_{2ijmt}$  | Binary variables describing the relationship between $b_{ijm}, e_{ijm}$ and time $[t, t+1]$ for $t \in \{1, \dots, T\}$                         |
| $y_{ct}^k, n_{ct}^k$  | Number of vehicles leaving and inside cell $c \in \mathcal{C}$ during time $[t, t+1]$ for $t \in \{1, \dots, T\}$ in scenario $k = 1, \dots, K$ |

inside cells, given by values of variables  $y^k$  and  $n^k$ , respectively. **Table 2** provides a summary of the definitions of all the sets, parameters and variables used in this paper.

We first formulate all the constraints in the first stage for configuring a feasible traffic signal timing plan, and denote the feasible region as:

$$X = \left\{ -U \cdot z_{1ijmt} \leq b_{ijm} - t \leq U(1 - z_{1ijmt}) - \epsilon, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, m = 1, \dots, N_{cy}, t = 1, \dots, T \right\} \quad (1a)$$

$$-U \cdot z_{2ijmt} + \epsilon \leq t - e_{ijm} \leq U(1 - z_{2ijmt}), \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, m = 1, \dots, N_{cy}, t = 1, \dots, T \quad (1b)$$

$$\sum_{j \in \mathcal{F}_i} (z_{1ijmt} + z_{2ijmt}) \leq |\mathcal{F}_i| + 1, \forall i \in \mathcal{R}, m = 1, \dots, N_{cy}, t = 1, \dots, T \quad (1c)$$

$$o_i \leq l_i, \forall i \in \mathcal{R} \quad (1d)$$

$$b_{i1m} = l_i \cdot (m - 1) - o_i, \forall i \in \mathcal{R}, m = 1, \dots, N_{cy} \quad (1e)$$

$$e_{ijm} = b_{ijm} + g_{ij}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, m = 1, \dots, N_{cy} \quad (1f)$$

$$b_{ijm} = e_{ij-1m}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i / \{1\}, m = 1, \dots, N_{cy} \quad (1g)$$

$$l_i = \sum_{j \in \mathcal{F}_i} g_{ij}, \forall i \in \mathcal{R} \quad (1h)$$

$$G_{min} \leq g_{ij} \leq G_{max}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i \quad (1i)$$

$$z_{1ijmt}, z_{2ijmt} \in \{0, 1\}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, m = 1, \dots, N_{cy}, t = 1, \dots, T \quad (1j)$$

Constraints (1a) and (1b) describe the relationship between each time step  $t$ , the start time  $b$  and the end time  $e$  of the time

interval when the traffic signal of each phase is green. For each intersection  $i \in \mathcal{R}$ , each phase  $j \in \mathcal{F}_i$  and each cycle  $m \in \{1, \dots, N_{cy}\}$  at each time step  $t$ ,  $z_{1ijmt} = 1$  if  $t \geq b_{ijm}$  and  $z_{1ijmt} = 0$  otherwise. Similarly,  $z_{2ijmt} = 1$  if  $t \leq e_{ijm}$  and  $z_{2ijmt} = 0$  otherwise. Constraints (1c) indicate that at each time step, there is only one phase with green light. Constraints (1d) indicate that the offset should be less than the cycle length. Constraints (1e)–(1g) detail the steps for computing the start and the end of the time interval at each cycle, when the traffic signal of each phase is green based on the cycle length and the length of the interval. Constraints (1h) indicate that the sum of the green time over all phases should be equal to the cycle length. Constraints (1i) bound the green time from below and above. Constraints (1j) require  $z_1$  and  $z_2$  being binary valued.

The overall two-stage stochastic MIP is formulated as:

$$\min_{l, o, g, b, e, z_1, z_2, y, n} \sum_{k=1}^K p^k \left( - \sum_{c \in \mathcal{C}} \sum_{t=1}^T n_{ct}^k + \alpha \sum_{c \in \mathcal{C}} \sum_{t=1}^T (T-t)y_{ct}^k \right) \quad (2a)$$

$$\text{s.t. } (l, o, g, b, e, z_1, z_2) \in X \\ y_{ct}^k \leq n_{ct}^k, \forall c \in \mathcal{C}, t = 1, \dots, T, k = 1, \dots, K \quad (2b)$$

$$y_{ct}^k \leq Q_{ct}, \forall c \in \mathcal{C} \cup \mathcal{O} \cup \mathcal{M} \cup \mathcal{V}, t = 1, \dots, T, k = 1, \dots, K \quad (2c)$$

$$y_{ct}^k \leq \sum_{m=1}^{N_{cy}} (z_{1ijmt} + z_{2ijmt} - 1) Q_{ct}, \forall c \in \mathcal{I}_{ij}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, t = 1, \dots, T, k = 1, \dots, K \quad (2d)$$

$$y_{ct}^k \leq Q_{ct}, \forall c \in \mathcal{C} \setminus \mathcal{V}, \forall c' \in d(c), t = 1, \dots, T, k = 1, \dots, K \quad (2e)$$

$$\beta_{cc'}^k y_{ct}^k \leq Q_{c't}, \forall c \in \mathcal{V}, \forall i \in \mathcal{R}, \forall j \in \mathcal{F}_i, \forall c' \in d(c) \cap I_{ij}, t = 1, \dots, T, k = 1, \dots, K \quad (2f)$$

$$y_{ct}^k \leq W_{ct}(N_{ct} - n_{ct}^k), \forall c \in \mathcal{C} \setminus \mathcal{V}, \forall c' \in d(c), t = 1, \dots, T, k = 1, \dots, K \quad (2g)$$

$$\beta_{cc'}^k y_{ct}^k \leq W_{c't}(N_{c't} - n_{c't}^k), \forall c \in \mathcal{V}, \forall c' \in d(c), t = 1, \dots, T, k = 1, \dots, K \quad (2h)$$

$$n_{ct+1}^k = n_{ct}^k + \sum_{c' \in p(c)} y_{c't}^k - y_{ct}^k, \quad \forall c \in \mathcal{C}/\mathcal{O}/\mathcal{I}, \quad t = 1, \dots, T, \quad k = 1, \dots, K \quad (2i)$$

$$n_{ct+1}^k = n_{ct}^k + D_{ct}^k - y_{ct}^k, \quad \forall c \in \mathcal{O}, \quad t = 1, \dots, T, \quad k = 1, \dots, K \quad (2j)$$

$$n_{ct+1}^k = n_{ct}^k + \sum_{c' \in p(c)} \beta_{c'c}^k y_{c't}^k - y_{ct}^k, \quad \forall c \in \mathcal{I}, \quad t = 1, \dots, T, \quad k = 1, \dots, K \quad (2k)$$

$$n_{c1}^k = n_c^{\text{init}}, \quad \forall c \in \mathcal{C}, \quad k = 1, \dots, K \quad (2l)$$

$$y_{ct}^k \geq 0, \quad n_{ct}^k \geq 0, \quad \forall c \in \mathcal{C}, \quad t = 1, \dots, T, \quad k = 1, \dots, K, \quad (2m)$$

where in the objective function (2a) we minimize the negative value of the expected vehicle throughput of the network over all time steps plus a so-called CTM objective term  $\sum_{c \in \mathcal{C}} \sum_{t=1}^T (T-t)y_{ct}^k$ ,  $\forall k = 1, \dots, K$ . This will force  $y_{ct}^k$  for each cell  $c \in \mathcal{C}$  at each time step  $t \in \{1, \dots, T\}$  to obtain the minimum of the right-hand sides of constraints (2b)–(2h), and then to ensure all vehicles traveling forward as much as possible. (Each  $y_{ct}^k$  is weighted by the remaining  $T-t$  time steps.) The parameter  $\alpha$  is set to balance between these two objectives, where a larger  $\alpha$  can eliminate the vehicle-holding problem to a certain extent. Constraints (2b)–(2m) involve both first-stage variables  $z_1, z_2$  and second-stage recourse variables  $y^k, n^k$ , and establish the fundamental relationships in CTM. Specifically, constraints (2b) indicate that the number of vehicles leaving a cell  $c$  is limited by the number of vehicles inside cell  $c$ . Constraints (2c) and (2d) imply that the number of vehicles leaving a cell  $c$  is limited by the flow capacity of cell  $c$ . Notice that for an intersection cell, its capacity is determined by the related traffic signal, such that if the signal is red, then the capacity should be zero. Constraints (2e) and (2f) indicate that the number of vehicles leaving a cell  $c$  is also limited by the flow capacity of its processing cell  $c'$ . Notice that a diverging cell has more than one processing cell and the number of vehicles entering each processing cell is estimated by the turning ratio. Constraints (2g) and (2h) indicate that the number of vehicles leaving a cell  $c$  should be limited by the number of vehicles that can enter its processing cells  $d(c)$ . Constraints (2i)–(2k) are flow conservation equations for cells, enforcing that the difference between the number of vehicles in a cell  $c$  between two consecutive time steps  $t$  and  $t+1$  equals to the number of vehicles coming from preceding cells minus the number of vehicles leaving cell  $c$  during time  $[t, t+1]$ . Notice that the number of vehicles entering each origin cell is the source demand. Constraints (2l) present the initial number of vehicles inside each cell. Constraints (2m) indicate that the number of vehicles leaving and inside each cell should be non-negative.

#### 4. Decentralized and decomposition algorithms

The scalability issue of solving the two-stage stochastic MIP mainly comes from the number of intersections and time steps. In this section, we propose algorithms for solving Model (2) using Benders decomposition combined with spatial and temporal decomposition. We propose an exact ADMM-based spatially decentralized Benders algorithm and prove the optimality in Section 4.1. We develop a heuristic temporal decomposition technique to further reduce the computational time in Section 4.2.

#### 4.1. ADMM-based spatially decentralized benders algorithm

At each iteration of the standard Benders algorithm, we solve the relaxed master problem (first-stage problem), the subproblem (second-stage problem), and add derived constraints to the relaxed master problem (Shapiro et al., 2014). We present the details of applying the Benders algorithm to solve our problem directly in Appendix A. However, the computational complexity of solving first-stage problems and second-stage problems increases significantly as network sizes and time steps increase. Therefore, we propose a distributed algorithm that solves first-stage problems as well as second-stage problems separately for each intersection based on the partition of the network. In Section 4.1.1 we propose a distributed formulation of first-stage and second-stage problems. In Section 4.1.2 we solve the second-stage problems for each intersection individually by applying ADMM (Boyd et al., 2011). In Section 4.1.3 we generate optimality cuts based on the optimal value and solutions obtained by ADMM and prove that the objective value obtained by our proposed algorithm converges to the optimal objective value.

##### 4.1.1. Distributed formulation

We partition the network into  $N_I$  areas where each area contains only one intersection. For each area with an intersection  $i$  for  $i \in \mathcal{R}$ , let  $\mathcal{C}_i, \mathcal{E}_i, \mathcal{O}_i, \mathcal{D}_i, \mathcal{M}_i, \mathcal{V}_i$  be the corresponding sets of all cells, ordinary cells, origin cells, destination cells, merge cells, and diverge cells, respectively. Because signal constraints (1a)–(1j) can be written separately for each intersection, by defining variables  $\theta_i^k$  for each scenario  $\xi^k$ ,  $k = 1, \dots, K$  and each intersection  $i \in \mathcal{R}$ , we formulate an intersection-based relaxed master problem as follows.

$$(\mathbf{RMP}_i) \quad \min \quad \sum_{k=1}^K p^k \theta_i^k \quad (3a)$$

$$\text{s.t.} \quad \text{Constraints (1a)–(1j) corresponding to the intersection } i \\ (z_{1i}, z_{2i}, \theta_i) \in \Sigma_i(z_{1i}, z_{2i}, \theta_i), \quad (3b)$$

where  $\Sigma_i(z_{1i}, z_{2i}, \theta_i)$  is the set of cuts as linear functions of  $z_{1i}, z_{2i}$  generated up to the current iteration. The problems  $(\mathbf{RMP}_i)$  allow us to solve the relaxed master problem for each area separately which only contains one intersection  $i \in \mathcal{R}$ .

Next, we consider the distributed formulation of second-stage subproblems. For each intersection  $i \in \mathcal{R}$ , we partition the set of cells  $\mathcal{C}_i$  into two parts – one consists of all the cells where the constraints of these cells are related to cells in other intersections, called boundary cells and the other consists of all the remaining cells, called internal cells (Timothéou et al., 2014). For each area containing intersection  $i \in \mathcal{R}$ , let  $\mathcal{B}_i^l$  be the set of input boundary cells receiving inflow traffic from a cell of a neighboring area, and  $\mathcal{B}_i^o$  be the set of output boundary cells sending outflow traffic to a cell of a neighboring area. In a centralized stochastic programming model, constraints (2g) and (2i) are related to boundary cells. We rewrite these constraints separately for boundary cells and internal cells. Notice that in a grid network setting, for each boundary cell  $c \in \mathcal{B}_i^o \cup \mathcal{B}_i^l$ , there is only one cell receiving the inflow traffic from  $c$  and one cell sending the outflow traffic to  $c$ , denoted by  $d_c$  and  $p_c$  such that  $d(c) = \{d_c\}$  and  $p(c) = \{p_c\}$ .

For each intersection  $i \in \mathcal{R}$ , the constraints related to boundary cells include variables corresponding to other intersections. For each  $k = 1, \dots, K$  and  $i \in \mathcal{R}$ , we introduce new decision variable vectors  $\tilde{y}_i^k \in \mathbb{R}^{|\mathcal{B}_i^l| \times T}$  and  $\tilde{n}_i^k \in \mathbb{R}^{|\mathcal{B}_i^o| \times T}$  to estimate the corresponding  $y^k$  and  $n^k$  of cells of neighboring intersections. Constraints (2g) and (2h) are rewritten as equality constraints for boundary cells by defining auxiliary variables  $s_{ct}^k \geq 0$  for each cell  $c \in \mathcal{B}_i^l \cup \mathcal{B}_i^o$ ,  $\forall i \in \mathcal{R}$ ,

at each time step  $t = 1, \dots, T$  and for each scenario  $k = 1, \dots, K$ . Let  $s^k = [s_{ct}^k, c \in \cup_{i \in \mathcal{R}} (\mathcal{B}_i^l \cup \mathcal{B}_i^0), t = 1, \dots, T]^T, \forall k = 1, \dots, K$ . Given the solutions  $\hat{z}_{1i}, \hat{z}_{2i}$  of the first-stage problems  $(\text{RMP}_i)$ ,  $i \in \mathcal{R}$ , we substitute  $z_1 = \hat{z}_{1i}$ ,  $z_2 = \hat{z}_{2i}$  into constraints (2d) and formulate the second-stage problem as a distributed formulation:

$$(\text{SP-D}^k) \min - \sum_{i \in \mathcal{R}} \sum_{c \in \mathcal{D}_i} \sum_{t=1}^T n_{ct}^k - \alpha \sum_{i \in \mathcal{R}} \sum_{c \in \mathcal{C}_i} \sum_{t=1}^T (T-t)y_{ct}^k \quad (4a)$$

$$\text{s.t. Constraints (2b)–(2m) for the scenario } \xi^k, \forall i \in \mathcal{R}, \forall c \in \mathcal{C}_i / \mathcal{B}_i^l / \mathcal{B}_i^0 \quad (4b)$$

$$s_{ct}^k \geq 0, \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^l \cup \mathcal{B}_i^0, \forall c' \in d(c), t = 1, \dots, T \quad (4c)$$

$$y_{ct}^k + s_{ct}^k = W_{dct}(N_{dct} - \tilde{n}_{ict}^k), \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^0, t = 1, \dots, T \quad (4d)$$

$$n_{ct+1}^k = n_{ct}^k + \tilde{y}_{ict}^k - y_{ct}^k, \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^l, t = 1, \dots, T \quad (4e)$$

$$\tilde{y}_{ict}^k = y_{pc}^k, \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^l, t = 1, \dots, T \quad (4f)$$

$$\tilde{n}_{ict}^k = n_{dct}^k, \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^0, t = 1, \dots, T. \quad (4g)$$

Constraints (4b) correspond to constraints (2b)–(2m) for the internal cells under scenario  $\xi^k$ . Constraints (4d) and (4e) refer to the equality form of constraints (2g) and (2i) related to boundary cells given  $\xi^k$ . Constraints (4f) and (4g) indicate that for each intersection  $i \in \mathcal{R}$ ,  $\tilde{y}_i^k$  and  $\tilde{n}_i^k$  should be equal to the value of  $y_i^k$  and  $n_i^k$  of cells of neighboring intersections receiving flow from or sending flow to boundary cells.

The objective function and constraints (4b)–(4e) are separable for each intersection  $i \in \mathcal{R}$ . Because for each cell  $c \in \mathcal{B}_i^l \cup \mathcal{B}_i^0$ , the cell sending or receiving flow  $p_c$  or  $d_c$  belong to other intersections, constraints (4f) and (4g) are not separable. We relax the linear constraints corresponding to all the original variables  $y^k, n^k, s^k$  and estimate variables  $\tilde{y}^k, \tilde{n}^k$  using a Lagrangian penalty function and propose an ADMM algorithm to solve the second-stage problem  $(\text{SP-D}^k)$  in the next section.

#### 4.1.2. ADMM for second-stage problems

In the distributed formulation, each intersection is considered as a block and the model contains several additional coupled (linear) constraints between pairs of intersection blocks. We apply ADMM to solve each second-stage problem  $(\text{SP-D}^k)$ ,  $\forall k = 1, \dots, K$ . Denote  $\mathbf{Y}_i$  as the convex feasible region of  $[y_i^k, n_i^k, s_i^k]$  defined by constraints (4b) and (4c) for each intersection  $i \in \mathcal{R}$ . Introducing dual variables  $\kappa_{ict}^k, \lambda_{ict}^k, \mu_{ict}^k, v_{ict}^k$  for constraints (4d)–(4g), respectively, and using  $\kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k$  to represent vector forms of these dual variables, we define an augmented dual Lagrangian function with a feasible set  $\mathbf{Y}_i$  for each intersection  $i \in \mathcal{R}$  as follows. (Here  $L$  represents the Lagrangian penalty parameter.)

$$\begin{aligned} \mathcal{L}_i^k(y_i^k, n_i^k, s_i^k, \tilde{y}_i^k, \tilde{n}_i^k, \kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k) \\ = - \sum_{c \in \mathcal{D}_i} \sum_{t=1}^T n_{ct}^k - \alpha \sum_{c \in \mathcal{C}_i} \sum_{t=1}^T (T-t)y_{ct}^k \\ + \sum_{c \in \mathcal{B}_i^0} \sum_{t=1}^T \kappa_{ict}^k (y_{ct}^k + s_{ct}^k - W_{dct}(N_{dct} - \tilde{n}_{ict}^k)) \\ + \sum_{c \in \mathcal{B}_i^l} \sum_{t=1}^T \lambda_{ict}^k (n_{ct+1}^k - n_{ct}^k - \tilde{y}_{ict}^k + y_{ct}^k) \end{aligned}$$

$$\begin{aligned} & + \sum_{c \in \mathcal{B}_i^l} \sum_{t=1}^T \mu_{ict}^k (\tilde{y}_{ict}^k - y_{pc}^k) + \sum_{c \in \mathcal{B}_i^0} \sum_{t=1}^T v_{ict}^k (\tilde{n}_{ict}^k - n_{dct}^k) \\ & + \frac{L}{2} \sum_{c \in \mathcal{B}_i^0} \sum_{t=1}^T \|y_{ct}^k + s_{ct}^k - W_{dct}(N_{dct} - \tilde{n}_{ict}^k)\|^2 \\ & + \frac{L}{2} \sum_{c \in \mathcal{B}_i^l} \sum_{t=1}^T \|n_{ct+1}^k - n_{ct}^k - \tilde{y}_{ict}^k + y_{ct}^k\|^2 \\ & + \frac{L}{2} \sum_{c \in \mathcal{B}_i^l} \sum_{t=1}^T \|\tilde{y}_{ict}^k - y_{pc}^k\|^2 + \frac{L}{2} \sum_{c \in \mathcal{B}_i^0} \sum_{t=1}^T \|\tilde{n}_{ict}^k - n_{dct}^k\|^2. \end{aligned} \quad (5)$$

The subproblem  $(\text{SP-D}^k)$  can be rewritten as the minimization problem of the augmented Lagrangian function with a feasible set composed by constraints (4b)–(4e). Following the definitions of dual variables, the augmented Lagrangian function is the summation of functions  $\mathcal{L}_i^k$  for all intersections  $i \in \mathcal{R}$ , i.e.,

$$\begin{aligned} \mathcal{L}^k(y^k, n^k, s^k, \tilde{y}^k, \tilde{n}^k, \kappa^k, \lambda^k, \mu^k, v^k) \\ = \sum_{i \in \mathcal{R}} \mathcal{L}_i^k(y_i^k, n_i^k, s_i^k, \tilde{y}_i^k, \tilde{n}_i^k, \kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k). \end{aligned} \quad (6)$$

The minimization problem of  $\mathcal{L}^k$  is equivalent to a series of problems that minimize  $\mathcal{L}_i^k$  for each intersection  $i \in \mathcal{R}$ .

For each intersection  $i \in \mathcal{R}$ , ADMM consists of three main steps: (i) updating original variables  $y_i^k, n_i^k, s_i^k$  by solving the minimization problem of the Lagrangian function  $\mathcal{L}^k$  with fixed estimates variables  $\tilde{y}_i^k, \tilde{n}_i^k$  and dual variables, (ii) updating estimates variables  $\tilde{y}_i^k, \tilde{n}_i^k$  by solving the minimization problem of the Lagrangian function  $\mathcal{L}^k$  with fixed original variables  $y_i^k, n_i^k, s_i^k$  and dual variables, and (iii) updating dual variables  $\kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k$  by gradient ascent. Specifically, in iteration  $l+1$ , we update variables based on current values  $y^{kl}, n^{kl}, s^{kl}, \tilde{y}^{kl}, \tilde{n}^{kl}, \kappa^{kl}, \lambda^{kl}, \mu^{kl}, v^{kl}$  as follows.

$$[y_i^{kl+1}, n_i^{kl+1}, s_i^{kl+1}] = \arg \min_{[y_i^k, n_i^k, s_i^k] \in \mathbf{Y}_i} \mathcal{L}_i^k(y_i^k, n_i^k, s_i^k, \tilde{y}_i^k, \tilde{n}_i^k, \kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k), \forall i \in \mathcal{R}, \quad (7a)$$

$$[\tilde{y}_i^{kl+1}, \tilde{n}_i^{kl+1}] = \arg \min_{\tilde{y}_i^k, \tilde{n}_i^k} \mathcal{L}_i^k(y_i^{kl+1}, n_i^{kl+1}, s_i^{kl+1}, \tilde{y}_i^k, \tilde{n}_i^k, \kappa_i^{kl}, \lambda_i^{kl}, \mu_i^{kl}, v_i^{kl}), \forall i \in \mathcal{R}, \quad (7b)$$

$$\kappa_{ict}^{kl+1} = \kappa_{ict}^{kl} + L(y_{ct}^{kl+1} + s_{ct}^{kl+1} - W_{dct}(N_{dct} - \tilde{n}_{ict}^{kl+1})), \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^0, t = 1, \dots, T, \quad (7c)$$

$$\lambda_{ict}^{kl+1} = \lambda_{ict}^{kl} + L(n_{ct+1}^{kl+1} - n_{ct}^{kl+1} - \tilde{y}_{ict}^{kl+1} + y_{ct}^{kl+1}), \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^l, t = 1, \dots, T, \quad (7d)$$

$$\mu_{ict}^{kl+1} = \mu_{ict}^{kl} + L(\tilde{y}_{ict}^{kl+1} - y_{pc}^{kl+1}), \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^l, t = 1, \dots, T, \quad (7e)$$

$$v_{ict}^{kl+1} = v_{ict}^{kl} + L(\tilde{n}_{ict}^{kl+1} - n_{dct}^{kl+1}), \forall i \in \mathcal{R}, \forall c \in \mathcal{B}_i^0, t = 1, \dots, T. \quad (7f)$$

Because the procedure of updating variables is separable for each intersection, the computation can be conducted in parallel for different intersections in each iteration, which can speed up the computation drastically. In practice, each subproblem can be solved by local computers installed at each intersection and then communicate with each other to update the duals, to fully utilize computing technologies in connected transportation systems.

#### 4.1.3. ADMM-based spatially decentralized benders algorithm

We use the same formulation of  $(\text{RMP}_i)$  as first-stage problems for all the intersections  $i \in \mathcal{R}$  and solve them directly to obtain solutions  $\hat{z}_{1i}, \hat{z}_{2i}, \theta_i^k$ . Given first-stage solutions, for each  $k = 1, \dots, K$ , we solve the second-stage problem by ADMM and obtain the optimal values  $\hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i})$  and the optimal dual solutions  $\hat{\sigma}_{ct}^k, c \in \mathcal{I}_{ij}, j \in \mathcal{F}_i, t = 1, \dots, T$  (associated with constraints (2b)) for each intersection  $i \in \mathcal{R}$ . We generate a linear optimality cut for  $\theta_i^k$  as:

$$\begin{aligned} \theta_i^k \geq \hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i}) - \sum_{j \in \mathcal{F}_i} \sum_{c \in \mathcal{I}_{ij}} \sum_{m=1}^{N_{cy}} \sum_{t=1}^T (\hat{z}_{1ijmt} + \hat{z}_{2ijmt} - 1) Q_{ct} \hat{\sigma}_{ct}^k \\ + \sum_{j \in \mathcal{F}_i} \sum_{c \in \mathcal{I}_{ij}} \sum_{m=1}^{N_{cy}} \sum_{t=1}^T (z_{1ijmt} + z_{2ijmt} - 1) Q_{ct} \hat{\sigma}_{ct}^k. \end{aligned} \quad (8)$$

**Theorem 1.** The optimality cut (8) is a valid cut.

**Proof.** For each  $k = 1, \dots, K$  and  $i \in \mathcal{R}$ , let  $\hat{y}_i^k, \hat{n}_i^k, \hat{\kappa}_i^k, \hat{\lambda}_i^k, \hat{\mu}_i^k, \hat{v}_i^k$  be the optimal solutions of variables  $\hat{y}_i^k, \hat{n}_i^k, \kappa_i^k, \lambda_i^k, \mu_i^k, v_i^k$  in ADMM. From the convergence property of ADMM on convex problems, we have

$$\hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i}) = \min_{[y_i^k, n_i^k, s_i^k] \in \mathbb{Y}_i} \mathcal{L}_i^k(y_i^k, n_i^k, s_i^k, \hat{y}_i^k, \hat{n}_i^k, \hat{\kappa}_i^k, \hat{\lambda}_i^k, \hat{\mu}_i^k, \hat{v}_i^k). \quad (9)$$

Consider the minimization problem on the right-hand side of (9). Associate dual variables  $\rho_{ct}^k, \sigma_{ct}^k, \pi_{cc't}^k, \gamma_{cc't}^k, \delta_{ct}^k, \tau_c^k$  to constraints (2b)–(2l), respectively. The dual Lagrangian function of this minimization problem is in the form of

$$\begin{aligned} & \hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i}, y_i^k, n_i^k, s_i^k, \rho^k, \sigma^k, \pi^k, \gamma^k, \delta^k, \tau^k) \\ &= \mathcal{L}_i^k(y_i^k, n_i^k, s_i^k, \hat{y}_i^k, \hat{n}_i^k, \hat{\kappa}_i^k, \hat{\lambda}_i^k, \hat{\mu}_i^k, \hat{v}_i^k) + \sum_{c \in \mathcal{C}_i / \mathcal{I}_i} \sum_{t=1}^T (Q_{ct} - y_{ct}^k) \sigma_{ct}^k \\ &+ \sum_{j \in \mathcal{F}_i} \sum_{c \in \mathcal{I}_{ij}} \sum_{m=1}^{N_{cy}} \sum_{t=1}^T (\hat{z}_{1ijmt} + \hat{z}_{2ijmt} - 1) Q_{ct} - y_{ct}^k) \sigma_{ct}^k \\ &+ \sum_{c \in \mathcal{C}_i} \sum_{c' \in d(c)} \sum_{t=1}^T (Q_{c't} - y_{ct}^k) \pi_{cc't}^k \\ &+ \sum_{c \in \mathcal{C}_i / \mathcal{B}_i^0} \sum_{c' \in d(c)} \sum_{t=1}^T (W_{c't} N_{c't} - y_{ct}^k - W_{c't} n_{ct}^k) \gamma_{cc't}^k \\ &+ \sum_{c \in \mathcal{C}_i} \sum_{t=1}^T (D_{ct}^k - y_{ct}^k + n_{ct}^k - n_{ct+1}^k) \delta_{ct}^k + \sum_{c \in \mathcal{C}_i} (n_c^{\text{init}} - n_{c1}^k) \tau_c^k. \end{aligned} \quad (10)$$

Because the minimization problem of  $\mathcal{L}_i^k$  is a semi-definite quadratic program, the strong duality holds. Therefore,

$$\hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i}) = \max_{\rho^k \geq 0, \sigma^k \geq 0, \pi^k \geq 0, \gamma^k \geq 0, \delta^k, \tau^k} \min_{y_i^k, n_i^k, s_i^k} \hat{\mathcal{L}}_i^k(\hat{z}_{1i}, \hat{z}_{2i}, y_i^k, n_i^k, s_i^k, \rho^k, \sigma^k, \pi^k, \gamma^k, \delta^k, \tau^k), \quad (11)$$

where  $\Gamma^k$  is the feasible set of  $\gamma^k$  such that  $\gamma_{cc't} \geq 0, \forall c \in \mathcal{C}_i / \mathcal{B}_i^0, c' \in d(c), t = 1, \dots, T$ . Furthermore, the feasibility set of  $\rho^k, \sigma^k, \pi^k, \gamma^k, \delta^k, \tau^k$  generated by  $\min_{y_i^k, n_i^k, s_i^k} \hat{\mathcal{L}}_i^k$  does not depend on first-stage decision variables. Substituting (11) into the cut (8), we can verify that the cut is valid.  $\square$

The procedure of ADMM-based spatially decentralized Benders algorithm is described in [Algorithm 1](#).

**Remark 1.** Because the added cuts are all valid, the summation of the objective values of  $(\text{RMP}_i)$  for all  $i \in \mathcal{R}$  provides a lower bound of the original stochastic programming model. The following theorem guarantees the optimality of the ADMM-based decentralized Benders algorithm. Furthermore, the summation of  $\hat{\mathcal{L}}_i^k$  provides an upper bound of the original stochastic programming model since

**Algorithm 1:** An ADMM-based spatially decentralized Benders algorithm for solving model (2).

---

```

1: for  $i \in \mathcal{R}$  do
2:   Initialize  $(\text{RMP}_i)$  with  $\Sigma_i = \emptyset$ .
3: end for
4: while the termination criteria is not satisfied do
5:   for  $i \in \mathcal{R}$  do
6:     Solve  $(\text{RMP}_i)$  to obtain optimal solution  $(\hat{z}_{1i}, \hat{z}_{2i}, \hat{\theta}_i)$ .
7:   end for
8:   for  $k = 1, \dots, K$  do
9:     Solve  $(\text{SP-D}^k)$  by ADMM in parallel according to Section 4.1.2 and obtain the optimal value  $\mathcal{L}_i^{k*}$  and optimal dual variables  $\hat{\sigma}$ .
10:    for  $i \in \mathcal{R}$  do
11:      if  $\hat{\theta}_i^k < \mathcal{L}_i^{k*}$  then
12:        Add optimality cut-(8) to  $\Sigma_i$ .
13:      end if
14:    end for
15:   end for
16: end while
17: Return the objective value as  $\sum_{k=1}^K p^k \sum_{i \in \mathcal{R}} \hat{\theta}_i^k$  and the solutions of  $(\text{RMP}_i), i \in \mathcal{R}$ .

```

---

it is the objective value of a feasible solution. We terminate the algorithm when the gap between the upper and lower bounds is sufficiently small.

**Theorem 2.** With the assumption that the objective function of the relaxed master problem is separable for the intersections, when the algorithm terminates, the centralized relaxed master problem  $(\text{RMP})$  is equivalent to the union of distributed relaxed master problems  $(\text{RMP}_i), i \in \mathcal{R}$ . Furthermore, Benders-ADMM converges to the optimal solutions and the optimal objective value.

**Proof.** For each particular first-stage solution  $\hat{z}_1, \hat{z}_2$ , we denote  $\theta^k \geq F_D^k(\hat{z}_1, \hat{z}_2)$  as the optimality cut generated by Benders, and  $\theta_i^k \geq F_{D_i}^k(\hat{z}_{1i}, \hat{z}_{2i})$  as the cut generated by Benders-ADMM. It holds that

$$F_D^k(\hat{z}_1, \hat{z}_2) = \sum_{i \in \mathcal{R}} F_{D_i}^k(\hat{z}_{1i}, \hat{z}_{2i}). \quad (12)$$

When the algorithm terminates, for any solution  $(z_1^*, z_2^*, \theta^{k*})$  of  $(\text{RMP})$ , we solve the second-stage subproblems by ADMM given  $(z_1^*, z_2^*)$  and let  $\theta_i^{k*} = \mathcal{L}_i^*$ . From the procedure of Benders-ADMM and [Eq. \(12\)](#),  $(z_1^*, z_2^*, \theta_i^{k*})$  is a solution to  $(\text{RMP}_i)$  for each intersection  $i \in \mathcal{R}$ . On the other hand, for any solution  $(z_1^*, z_2^*, \theta_i^{k*})$  of  $(\text{RMP}_i)$ , we can construct a solution  $\prod_{i \in \mathcal{R}} z_{1i}^*, \prod_{i \in \mathcal{R}} z_{2i}^*, \sum_{i \in \mathcal{R}} \theta_i^{k*}$  for  $(\text{RMP})$  because of the Benders procedures and [Eq. \(12\)](#). It is well known that Benders decomposition algorithm converges to the optimal solutions ([Shapiro et al., 2014](#)) and therefore, Benders-ADMM also converges to the optimal solutions and the optimal objective value.  $\square$

**Remark 2.** ADMM-based spatially decentralized Benders algorithm ([Algorithm 1](#)) can be also employed when the partitioned areas contain multiple intersections.

#### 4.2. Heuristic temporal decomposition

Our ADMM-based spatially decentralized Benders algorithm solves the relaxed master problems and subproblems in a distributed manner, which reduces the computational time and guarantees optimality. To further accelerate the convergence of our Benders algorithm, we introduce a heuristic temporal decomposition technique to pre-determine the cycle length in this section.

The heuristic technique sacrifices the global optimality but still guarantees the optimality under a certain cycle length constraint. For the first-stage problem of each intersection  $i \in \mathcal{R}$ , the computational complexity mainly depends on the number of time steps,  $T$ .

**Lemma 1.** For each  $i \in \mathcal{R}$ , the number of variables and constraints in  $(\text{RMP}_i)$  grows as  $O(T^2)$  with the number  $T$  of time steps.

**Proof.** The number of variables in  $(\text{RMP}_i)$  is  $2|\mathcal{F}_i|N_{\text{cy}}T + 2|\mathcal{F}_i|N_{\text{cy}}$ . As  $|\mathcal{F}_i|$  does not depend on  $T$  and  $N_{\text{cy}}$  grows linearly with  $T$ , the number of variables grows as  $O(T^2)$ . Similarly, the number of constraints is  $O(|\mathcal{F}_i|N_{\text{cy}}T)$ , which also grows as  $O(T^2)$ . This completes the proof.  $\square$

To reduce the size of the first-stage problem, we pre-determine the cycle lengths  $l_i$  as  $\hat{l}_i$  for each  $i \in \mathcal{R}$  based on the source demand volume according to Koonce & Rodegerdts (2008). Since the traffic signal control plan is the same for each cycle, the signal status of time step  $(m-1)\hat{l}_i + t'$  is exactly the same as  $t'$  for all the cycles  $m = 1, \dots, N_{\text{cy}}$ . Therefore, we redefine binary variables  $z_{1ijmt}, z_{2ijmt}$  for each intersection  $i \in \mathcal{R}$ , each phase  $j \in \mathcal{F}_i$ , only for the first two cycles  $m = 1, 2$  and time steps  $t = 1, \dots, \hat{l}_i$ . The first-stage problems  $(\text{RMP}_i)$  and  $(\text{RMP})$  are reformulated by setting  $N_{\text{cy}} = 2$  and  $T = \hat{l}_i$ . In the following lemma, we show that it is sufficient to describe the signal status at each time step  $t = 1, \dots, T$  with the re-defined decision vectors  $z_{1i}, z_{2i}$  for the first two cycles and the first  $\hat{l}_i$  time steps.

**Lemma 2.** For each intersection  $i \in \mathcal{R}$ , with given cycle length  $\hat{l}_i$ , the flow upper bound constraints for cells controlled by signals (2d) are equivalent to

$$\begin{aligned} y_{ct}^k &\leq \sum_{m=1}^2 (z_{1ijmt} + z_{2ijmt} - 1)Q_{ct}, \quad \forall c \in \mathcal{I}_{ij}, \\ \forall i \in \mathcal{R}, \quad \forall j \in \mathcal{F}_i, \quad t = 1, \dots, T, \quad k = 1, \dots, K, \end{aligned} \quad (13)$$

where  $t' = (t-1) \bmod \hat{l}_i + 1$  is the index of time step corresponding to time step  $t$ .

**Proof.** Recalling the definition of binary variables  $z_1, z_2$ , we have that  $\sum_{m=1}^{N_{\text{cy}}} (z_{1ijmt} + z_{2ijmt} - 1) = 1$  if and only if the signal of phase  $j$  at intersection  $i$  is green at time step  $t$ . Because the signal timing plan is the same for each cycle, we know that

$$\begin{aligned} \sum_{m=1}^{N_{\text{cy}}} (z_{1ijmt} + z_{2ijmt} - 1) \\ = \sum_{m=1}^{N_{\text{cy}}} (z_{1ijmt'} + z_{2ijmt'} - 1), \quad t = 1, \dots, T, \quad t' = (t-1) \bmod \hat{l}_i + 1. \end{aligned} \quad (14)$$

Next, we show that the right-hand-side is equivalent to taking the summation with  $N_{\text{cy}} = 2$ . From constraints (1d)–(1h), given  $l_i = \hat{l}_i$ , we have

$$-\hat{l}_i \leq b_{i11} \leq 0, \quad \forall i \in \mathcal{R}; \quad (15a)$$

$$e_{i|\mathcal{F}_i|2} \geq \hat{l}_i, \quad \forall i \in \mathcal{R}. \quad (15b)$$

The time horizon we consider is  $\{1, \dots, \hat{l}_i\}$ , which is a subset of  $[b_{i11}, e_{i|\mathcal{F}_i|2}]$ . Hence, according to the definition of  $z_1, z_2$ , we have that  $z_{1ijmt'} = z_{2ijmt'} = 0$  for all the cycles  $m \geq 2$ .  $\square$

**Theorem 3.** By pre-determining the cycle length, the problem size of reformulated first-stage problems  $(\text{RMP}_i)$  for each  $i \in \mathcal{R}$  only grows linearly with the cycle length. The statement also holds for the reformulated centralized first-stage problem  $(\text{RMP})$ .

**Proof.** Combining Lemmas 1 and 2, we show that the number of variables reduces to  $4|\mathcal{F}_i|\hat{l}_i + 4|\mathcal{F}_i|$  and number of constraints reduce to  $O(|\mathcal{F}_i|\hat{l}_i)$ . It is clear that the number of variables and the number of constraints in  $(\text{RMP}_i)$  are both  $O(\hat{l}_i)$ . Because the number of variables and constraints in  $(\text{RMP})$  is the summation of the number of variables and constraints in  $(\text{RMP}_i)$  for all  $i \in \mathcal{R}$ , the conclusion holds if we pre-determine the cycle lengths in  $(\text{RMP})$ .  $\square$

With heuristic pre-determined cycle lengths and reformulated first-stage and second-stage problems, Benders-ADMM only ensures convergence to the optimal solutions and optimal objective value of the original centralized model with additional signal cycle length constraints  $l_i = \hat{l}_i, \forall i \in \mathcal{R}$ .

**Remark 3.** The pre-determined cycle length for all the intersections are the same (Koonce & Rodegerdts, 2008). The cycle length  $l^{\text{fix}}$  is set as the mean value of  $l_i, i \in \mathcal{R}$  computed by the following formula (Webster, 1958).

$$l_i = \left\lceil \frac{|\mathcal{F}_i| * 7.5 + 5}{1 - \sum_{j=1}^4 D_j/Q_j} \right\rceil \quad (16)$$

where  $D_j$  and  $Q_j$  are the source demand and maximum flow capacity of the direction corresponding to a phase  $j$ .

## 5. Numerical studies

We apply algorithms proposed in Section 4 to solve the traffic signal control problem on instances of randomly generated grid networks and real-world traffic networks. In Section 5.1, we introduce the experimental design, including the warm-up initialization in Section 5.1.1, the network design of instances in Sections 5.1.2 and 5.1.3 and the out-of-sample evaluation in Section 5.1.4. In Sections 5.2 and 5.3, we present the computational results to demonstrate the efficacy of our approaches.

### 5.1. Experimental design

#### 5.1.1. Warm-up initialization

We introduce a warm start technique according to Webster (1958) to simulate the initialized state of traffic networks and obtain the number  $n_c^{\text{init}}$  of vehicles inside each cell  $c \in \mathcal{C}$ . We define a fixed traffic signal time plan and compute the number of vehicles in each cell  $c \in \mathcal{C}$  by solving the model with an objective function as the second term of (2a) and constraints (2b)–(2m). For each phase  $j \in \mathcal{F}_i$ , the green time of a phase  $j$  is defined as

$$g_j^{\text{fix}} = \frac{D_j}{\sum_{j=1}^4 D_j} l^{\text{fix}}, \quad (17)$$

where  $D_j$  is the source demand of the direction corresponding to a phase  $j$  and  $l^{\text{fix}}$  is the cycle length. For each cell  $c \in \mathcal{C}$ , we set  $n_c^{\text{init}}$  as the number of vehicles inside  $c$  at the last time step.

#### 5.1.2. Randomly generated grid networks

We conduct numerical studies on randomly generated grid networks with the size  $N_{\text{row}} \times N_{\text{col}}$ , where  $N_{\text{row}}$  is the number of rows and  $N_{\text{col}}$  is the number of intersections in each row that has the same structure of intersections and road segments. We define the set of phases  $\mathcal{F}_i$  for all the intersections  $i \in \mathcal{R}$  as  $\mathcal{F}_i = \{1, 2, 3, 4\}$ , where  $j = 1$  means turning left in East-West direction,  $j = 2$  means going straight or turning right in East-West direction,  $j = 3$  means turning left in North-South direction, and  $j = 4$  means going straight or turning right in North-South direction. We set the values of input parameters as follows. Let “veh” denote the number of vehicles. For each intersection cell  $c \in \mathcal{I}$  and time step  $t = 1, \dots, T$ ,  $Q_{ct} = 1.5$  veh and  $N_{ct} = 6$  veh, meaning that at most

**Table 3**

Instances of randomly generated source demand.

| Instance | Mean (veh/h)         | SD/Mean |
|----------|----------------------|---------|
| 1        | 200 (E-W), 50 (S-N)  | 2       |
| 2        | 200 (E-W), 50 (S-N)  | 3       |
| 3        | 200 (E-W), 50 (S-N)  | 4       |
| 4        | 400 (E-W), 100 (S-N) | 2       |
| 5        | 400 (E-W), 100 (S-N) | 3       |
| 6        | 400 (E-W), 100 (S-N) | 4       |

E-W: direction of east and west.

S-N: direction of south and north.

1.5 and 6 vehicles can flow through and reside in an intersection cell at any time, respectively. For the other cells  $c \in \mathcal{C}/\mathcal{I}$  and time step  $t = 1, \dots, T$ ,  $Q_{ct} = 3$  veh and  $N_{ct} = 12$  veh, meaning that at most 3 and 12 vehicles can flow through and reside in other cells at each time. The ratio of shock-wave speed over free-flow speed is the same for all the cells in  $\mathcal{C}$  for each  $t \in \{1, \dots, T\}$ , which is set as  $W = 1/3$ . The initialized number of vehicles  $n^{init}$  is generated by the warm-start technique described in Section 5.1.1. The minimum green time  $G_{min} = 6$  seconds and the maximum green time  $G_{max} = 75$  seconds. The whole time horizon is half an hour and has 600 time steps. We set the weight parameter  $\alpha = 0.001$  in the objective function.

We generate random samples of source demand and turning ratios as follows. We assume uniform arrivals of vehicles during the half-an-hour time horizon, and therefore values of  $D_{ct}$  are the same for all time steps  $t = 1, \dots, T$  for any origin cell  $c \in \mathcal{O}$ . The source demand for each cell  $c \in \mathcal{O}$  follows a truncated Normal distribution defined on  $[0, \infty)$  shown in Table 3, where Column “SD/Mean” represents the ratio between standard deviation and the mean value. The unit of source demand is the number of vehicles per hour (veh/h).

The turning ratios of all the diverge cells  $c \in \mathcal{D}$  follow truncated Normal distribution defined on  $[0, 1]$  with mean values = 0.15, 0.72, 0.13, representing the ratio of turning left, going straight and turning right, respectively. The ratio between the standard deviation and the mean value is set to 0.3. We test our approaches on randomly generated grid networks having sizes  $4 \times 4$ ,  $2 \times 8$ ,  $6 \times 6$  and  $10 \times 10$  and the distribution instance #4 of source demand given in Table 3. We test the other types of distributions in Table 3 only for the  $4 \times 4$  grid network.

We generate 100 in-sample scenarios to formulate the stochastic optimization model. As a benchmark, we firstly use Gurobi to solve the stochastic programming model directly. Then, we employ Benders with/without temporal decomposition described in Section 4.2 to solve the model. Furthermore, based on the temporal decomposition, we apply Benders and the ADMM-based spatially decentralized Benders. For the  $4 \times 4$  grid network, we build a deterministic counterpart by setting the source demand and turning ratios as the expectation and solve the model by applying ADMM-based spatially decentralized Benders with temporal decomposition.

### 5.1.3. Real-world traffic networks

We also test all the algorithms on a real traffic network based on the road network of Downtown Ann Arbor (a city in Michigan, United States). This traffic network contains 14 corridors, 37 signalized intersections, and 27 unsignalized intersections with stop signs and one-way roads. We present the map of our considered network area and the neighboring area in Fig. 2.

Notice that the test instance is a typical downtown area with grid traffic networks with multiple arterials in each direction. We mark the main arterials in the direction of East and West by red lines, including Miller Avenue, Huron Street, Washington Street, and Liberty Street. We mark the main arterials in the direction

of North and South by purple lines, including Ashley Street, Main street, Division Street, and State Street. The downtown area of Ann Arbor is mainly the center area of our network, near the intersections of main arterials, which are marked by blue text. In addition, we mark the area of the University of Michigan as the Southeast area of the traffic network. In the morning, people mostly travel from West to East for studying and working at the university. In the afternoon, people mostly travel from East to West for going back home.

All the parameters, including mean values of source demand and turning ratios, are computed based on collected real-world traffic data. We test two instances with high average source demand during morning and afternoon peak hours where we set the SD/Mean ratio = 2 and an instance with low average source demand during off-peak hours with SD/Mean ratio = 3. The ratio between standard deviation and mean of turning ratios is 0.3. We generate 10 scenarios to formulate the stochastic optimization model and apply ADMM-based spatially decentralized Benders with temporal decomposition to solve the model. To construct a deterministic counterpart, we use the mean values of the source demand and turning ratios as input parameters and apply ADMM-based spatially decentralized Benders with temporal decomposition.

**Remark 4.** We set a heuristic lower bound of the green time that the green length of each phase  $j$  should be no less than 60% of  $g_j^{\text{fix}}$  and apply early termination to make the solutions more practical. Notice that our algorithm still works without these settings.

#### 5.1.4. Out-of-sample evaluation

**Metrics for evaluation** After obtaining a traffic signal control plan, we evaluate its performance by conducting CTM simulation according to Daganzo (1992). Let  $y_{ct}^*$  be the obtained solution of the number of vehicles leaving a cell  $c \in \mathcal{C}$  at a time step  $t = 1, \dots, T$ . We define a link as a set of cells that belong to the same road segment connecting two neighboring intersections. For each link starting from a cell  $c_1$  and ending at a cell  $c_2$ , we define the cumulative number of vehicles at time step  $t = 1, \dots, T$  of the inflow and outflow as  $CN_{\text{in}}(t)$  and  $CN_{\text{out}}(t)$ . The computation of these two metrics follows:

$$CN_{\text{in}}(t) = \sum_{t'=1}^t y_{\text{proc}(c_1)t}^* \quad (18a)$$

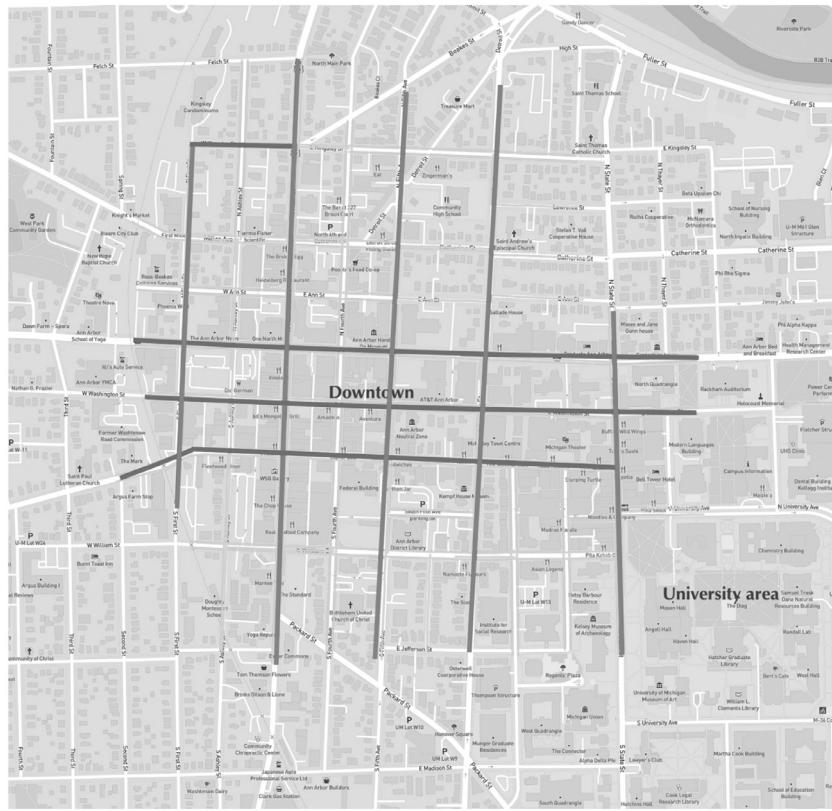
$$CN_{\text{out}}(t) = \sum_{t'=1}^t y_{c_2t}^* \quad (18b)$$

We define the cumulative number of vehicles outflow at time step  $t$  as  $CN^*(t)$  by assuming free-flow speed, and compute the value as

$$CN^*(t) = CN_{\text{in}}(t - N_c), \quad (19)$$

where  $N_c$  is the number of cells contained in the link. The performance of a traffic signal control plan is evaluated for every link during the  $T$  time steps by the total travel time  $\sum_{t=1}^T (CN_{\text{in}}(t) - CN_{\text{out}}(t))$  and the total delay  $\sum_{t=1}^T (CN^*(t) - CN_{\text{out}}(t))$ . We also compute the average travel time and the average delay of each vehicle. In addition, the total number of vehicles traveling through a traffic network during the time horizon is the sum of  $CN_{\text{out}}(T)$  for all the links ending at destination cells.

**Out-of-sample evaluation procedures** For both randomly generated grid networks and the real-world traffic network, we generate 5 replications of instances with the same parameter settings, each having 100 independently identically distributed scenarios with the same distribution as the one used in-sample computation. We conduct the out-of-sample tests based on CTM simulation



**Fig. 2.** Map of Ann Arbor downtown.

for different traffic signal control plans on these replications. We present the averages of the in-sample and out-of-sample objective values, the average travel delay, and the total throughput across the total 500 scenarios. For the real-world traffic network, we also compare the performance with our baseline solution that sets the green time of each phase as  $g_j^{\text{fix}}$ .

All the numerical experiments of randomly generated grid networks are conducted on Windows Server 2012 R2 Standard with 128 gigabytes RAM and 2.20 gigahertz processor. All the numerical experiments of real-world traffic networks are conducted on a Windows computer with 32 gigabytes RAM and 3.60 gigahertz processor.

## 5.2. Results of randomly generated grid networks

### 5.2.1. Computational time comparison

We set the time limit for Gurobi to 7200 seconds. If we do not apply any decomposition schemes and directly solve the problem, Gurobi is not able to provide a feasible solution or even an upper bound of the objective value within the time limit. When using the generic Benders without temporal decomposition, we are not able to solve the first-stage problem within the time limit, and thus it cannot provide a feasible solution to the second-stage problem. Based on these results, the temporal decomposition is necessary.

We present the computational time results of the three algorithms with the temporal decomposition in Table 4, where we vary network sizes. In the table,  $N_{\text{row}}$  and  $N_{\text{col}}$  represent the number of intersections of each row and column in the grid networks while "Benders" and "Benders-ADMM" represent Benders decomposition algorithm and ADMM-based spatially decentralized Benders algorithm, respectively. In the table, "MP-min", "MP-max", "MP-A", "SP-min", "SP-max", "SP-A" stand for the minimum, maximum and average computational time in seconds of solving first-stage master

problems and second-stage subproblems during all the iterations, respectively. If an algorithm is not able to return a feasible solution due to the time limit, we mark the related results as “–” in the table.

The results show that the CPU time varies significantly during the iterations of different algorithms. With added cuts, the computational time of the first-stage problem increases drastically. The computational time of second-stage problems varies depending on first-stage solutions. The results also indicate that for every instance, Benders-ADMM performs significantly better than Benders when solving first-stage problems since the size of first-stage problems is reduced by the spatial decomposition. For second-stage problems, Benders-ADMM outperforms Benders since the parallel computing is able to be applied so that each intersection can solve the second-stage problem at the same time. Moreover, for all the algorithms, the computational time increases when the network size is larger. For networks with the same number of intersections, the algorithms take less time to solve the model of an asymmetric network than a symmetric one.

### 5.2.2. Evaluation results

**Overall objective values** We use Benders-ADMM (Algorithm 1) to solve the deterministic (Deter) and stochastic (SP) MIP models and present the objective values of both models in Table 5. Column "Mean" presents mean values of source demand and Column "SD/Mean" presents the ratios between standard deviations and demand mean values of the solutions to different types of instances. Columns "In-sample Obj" and "Out-of-sample Obj" present the in-sample and out-of-sample objective values. Column "Gap" presents the gaps between in-sample and out-of-sample objectives. In all cases, gaps of the stochastic MIP model are smaller than gaps of the deterministic one, indicating that the stochastic approach better describes the real traffic conditions. When we increase demand mean values, all the objective values increase since there

**Table 4**

CPU time of different algorithms for solving grid networks with various sizes.

| $N_{\text{row}}$ | $N_{\text{col}}$ | Benders      |            |         |            |           |          |
|------------------|------------------|--------------|------------|---------|------------|-----------|----------|
|                  |                  | MP-min (s)   | MP-max (s) | MP-A(s) | SP-min (s) | SP-max(s) | SP-A(s)  |
| 4                | 4                | 1.18         | 123.99     | 54.98   | 2005.10    | 6905.05   | 4632.75  |
| 2                | 8                | 0.96         | 76.20      | 33.78   | 1687.63    | 4834.43   | 3252.43  |
| 6                | 6                | 1.78         | 685.31     | 132.60  | 17810.42   | 43607.42  | 34603.78 |
| 10               | 10               | —            | —          | —       | —          | —         | —        |
| $N_{\text{row}}$ | $N_{\text{col}}$ | Benders-ADMM |            |         |            |           |          |
|                  |                  | MP-min (s)   | MP-max (s) | MP-A(s) | SP-min (s) | SP-max(s) | SP-A(s)  |
| 4                | 4                | 1.73         | 46.22      | 18.81   | 539.44     | 594.26    | 571.98   |
| 2                | 8                | 1.68         | 33.96      | 14.71   | 295.12     | 519.81    | 426.21   |
| 6                | 6                | 4.26         | 101.51     | 42.43   | 1060.78    | 1685.95   | 1404.67  |
| 10               | 10               | 13.73        | 77.55      | 45.64   | 2780.25    | 3682.51   | 3231.38  |

**Table 5**

In-sample and out-of-sample objective values of randomly generated grid networks.

| Mean (veh/h)         | SD/Mean | In-sample Obj (veh·s) |             | Out-of-sample Obj (veh·s) |             | Gap    |        |
|----------------------|---------|-----------------------|-------------|---------------------------|-------------|--------|--------|
|                      |         | Deter                 | SP          | Deter                     | SP          | Deter  | SP     |
| 200 (E-W), 50 (S-N)  | 2       | -420597.37            | -556225.39  | -597997.23                | -596944.23  | 29.65% | 6.81%  |
| 200 (E-W), 50 (S-N)  | 3       | -420597.37            | -649147.73  | -715975.68                | -727529.79  | 41.25% | 10.76% |
| 200 (E-W), 50 (S-N)  | 4       | -420597.37            | -751408.39  | -794377.55                | -818953.97  | 47.05% | 8.23%  |
| 400 (E-W), 100 (S-N) | 2       | -687354.94            | -865289.83  | -981940.02                | -994091.40  | 30.00% | 12.96% |
| 400 (E-W), 100 (S-N) | 3       | -687354.94            | -986676.06  | -978864.21                | -1092679.91 | 29.77% | 9.69%  |
| 400 (E-W), 100 (S-N) | 4       | -687354.94            | -1044571.92 | -1122114.76               | -1179086.03 | 38.74% | 9.87%  |

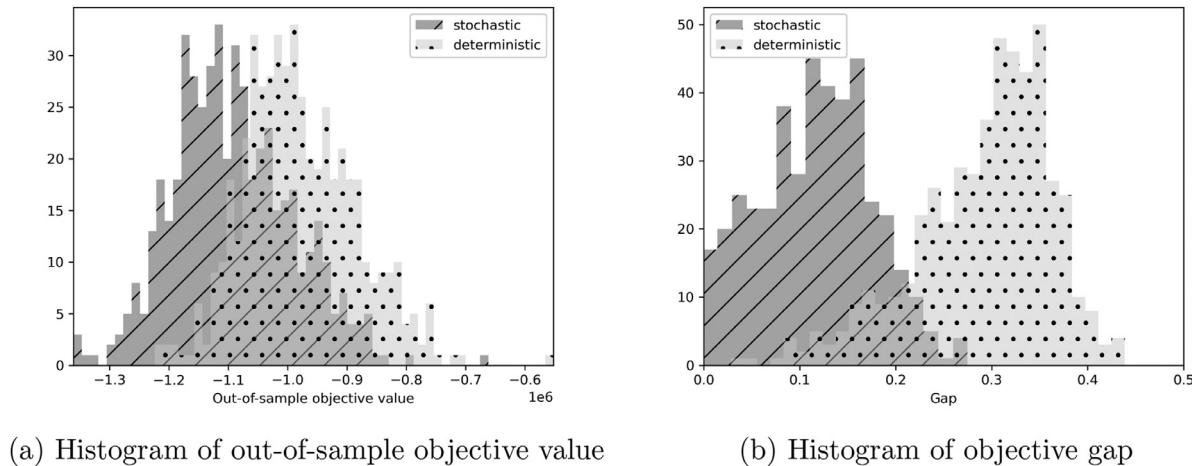


Fig. 3. Out-of-sample performance and gap results of the instance with demand mean as 400 (E-W), 100 (S-N) and SD/Mean ratio as 3.

are more vehicles entering the network. When the deviations increase, objective gaps of the deterministic model increase while the stochastic model can still maintain relatively low gaps, showing its solution robustness against parameter uncertainty.

Figure 3 shows the histograms of out-of-sample objective values and the gaps for the instance with 400 vehicles per hour in East-West direction and SD/Mean ratio = 3. Here the rectangles filled with slashes are associated with the stochastic model and the ones with dots are associated with the deterministic model. In most cases, the objective value and the gap of the stochastic model are smaller than the ones of the deterministic model. The deviation of the gap of the stochastic model is also less than the deterministic model, showing the solution robustness of the former.

**Other performance metrics** We use travel delay and throughput as the metrics to evaluate signal timing plans. In Table 6, we present their values in out-of-sample tests of signal timing plans obtained from the stochastic and deterministic models. Column “Gap” presents the gaps of the corresponding metrics between the two types of models. In most cases, the travel delay of the stochastic model is less than the one of the deterministic counterpart

and the throughput of the stochastic one is larger. Therefore, the signal timing plans obtained by the stochastic model outperform the ones of the deterministic model, demonstrating the importance and benefits of considering data uncertainties in traffic signal control. When the demand mean value increases, more improvements are brought by the stochastic model in most cases. For the same mean value, the largest improvement of the stochastic model is often attained when the traffic network is not too idle or too congested (i.e., SD/Mean = 3).

We show the distribution of out-of-sample results in Table 7, where we present the standard deviation (SD) and ratio between standard deviation and mean (SD/Mean) of travel delay (D) and throughput (T), indicated by “SD-D”, “SD-T”, “SD/Mean-D”, “SD/Mean-T”, respectively. The ratio between standard deviation and mean of travel delay is less than the ratio of the throughput. In most cases, when the mean value and deviation of the demand increase, there exist more samples where the traffic network is fully congested, leading to less deviation of traffic delay and throughput. Figure 4 shows the histograms of out-of-sample tests on the instance with 400 vehicles per hour in the East-West direction and

**Table 6**

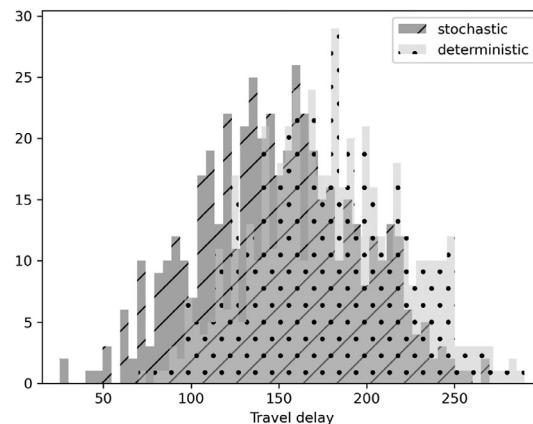
Out-of-sample evaluation of solutions for randomly generated grid networks with varying demand mean values and SD/Mean ratios.

| Mean (veh/h)         | SD/Mean | Average travel delay (s) |        |        | Total arrival (veh) | Total throughput (veh) |         |       |
|----------------------|---------|--------------------------|--------|--------|---------------------|------------------------|---------|-------|
|                      |         | Deter                    | SP     | Gap    |                     | Deter                  | SP      | Gap   |
| 200 (E-W), 50 (S-N)  | 2       | 29.36                    | 30.00  | -2.14% | 2105.61             | 1869.71                | 1880.02 | 0.55% |
| 200 (E-W), 50 (S-N)  | 3       | 60.20                    | 54.76  | 9.04%  | 2872.96             | 2280.55                | 2346.54 | 2.89% |
| 200 (E-W), 50 (S-N)  | 4       | 96.81                    | 87.24  | 8.24%  | 3658.03             | 2536.44                | 2583.79 | 1.87% |
| 400 (E-W), 100 (S-N) | 2       | 96.67                    | 92.31  | 4.70%  | 4195.30             | 2966.80                | 3030.21 | 1.80% |
| 400 (E-W), 100 (S-N) | 3       | 179.12                   | 149.07 | 16.57% | 5825.33             | 3191.74                | 3351.85 | 4.68% |
| 400 (E-W), 100 (S-N) | 4       | 207.95                   | 197.14 | 5.20%  | 7300.13             | 3364.67                | 3535.66 | 5.08% |

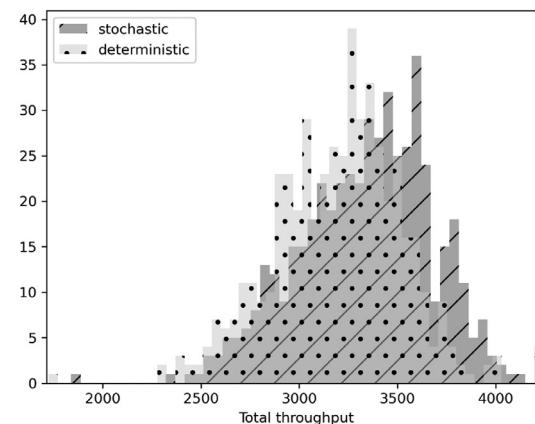
**Table 7**

Standard deviation of out-of-sample evaluation of solutions for randomly generated grid networks with varying demand mean values and SD/Mean ratios .

| Mean of demand (veh/h) | SD/Mean of demand | SD-D (s) |       | SD-T (veh) |        | SD/Mean-D |      | SD/Mean-T |      |
|------------------------|-------------------|----------|-------|------------|--------|-----------|------|-----------|------|
|                        |                   | Deter    | SP    | Deter      | SP     | Deter     | SP   | Deter     | SP   |
| 200 (E-W), 50 (S-N)    | 2                 | 15.03    | 14.28 | 312.95     | 325.95 | 0.51      | 0.48 | 0.17      | 0.17 |
| 200 (E-W), 50 (S-N)    | 3                 | 29.75    | 27.52 | 334.76     | 366.29 | 0.49      | 0.50 | 0.15      | 0.16 |
| 200 (E-W), 50 (S-N)    | 4                 | 37.64    | 36.83 | 328.85     | 308.30 | 0.39      | 0.42 | 0.13      | 0.12 |
| 400 (E-W), 100 (S-N)   | 2                 | 35.09    | 34.17 | 336.78     | 362.72 | 0.36      | 0.37 | 0.11      | 0.12 |
| 400 (E-W), 100 (S-N)   | 3                 | 42.18    | 44.82 | 318.49     | 346.43 | 0.24      | 0.30 | 0.10      | 0.10 |
| 400 (E-W), 100 (S-N)   | 4                 | 48.63    | 48.11 | 277.71     | 325.89 | 0.23      | 0.24 | 0.08      | 0.09 |



(a) Histogram of traffic delay



(b) Histogram of throughput

**Fig. 4.** Out-of-sample delay and throughput results of the instance with demand mean as 400 (E-W), 100 (S-N) and SD/Mean ratio as 3.

SD/Mean ratio being 3. In most cases, the stochastic model outperforms the deterministic model in the out-of-sample tests.

### 5.3. Results of real-world traffic networks

In this section, we present the results of the ADMM-based spatially decentralized Benders on real-world instances. Because solving the model without temporal decomposition is computationally impractical, we pre-determine the cycle length heuristically. In our numerical experiments, we show that the performance of signal timing plans is not sensitive to the cycle length and set the cycle length as the average value of each intersection's cycle length, computed by the rule in Remark 3. Our numerical results show that the signal timing plan with an in-sample scenario size  $K = 10$  obtains the best performance and 500 scenarios are enough for simulating the uncertain traffic reality in the out-of-sample test. The gap between in-sample and out-of-sample tests is around 20%, mainly because we terminate early in solving second-stage problems by ADMM allowing at most 20% gap between the obtained objective value and the optimal one. The details of sensitivity analysis for cycle length, the in-sample and out-of-sample scenario sizes are presented in Appendix B.1.

We select both instances of morning peak and off-peak hours of the real-world network, to present the computational time and solution performance evaluation in Sections 5.3.1 and 5.3.2, respectively. Similar results of the instance of afternoon peak hours are presented in Appendix B.3.

#### 5.3.1. Results of morning peak hours

We present the computational time of the deterministic and stochastic models in seconds in Table 8. The results in Table 8 show that our ADMM-based spatially decentralized Benders algorithm is able to solve both models within an acceptable time limit.

We present the in-sample objective values, out-of-sample objective values, and gaps between the deterministic and stochastic models in Table 9. To evaluate the traffic timing plans, we also present the average travel delay, total arrival and total throughput in the second half of the table. In Columns "Average Delay" and "Total Throughput", the percentages in Row "Deterministic" show the improvements of relative metrics compared to the baseline, and the percentages on the row "Stochastic" show the improvements of related metrics shown in each column as compared to the deterministic model, respectively.

**Table 8**

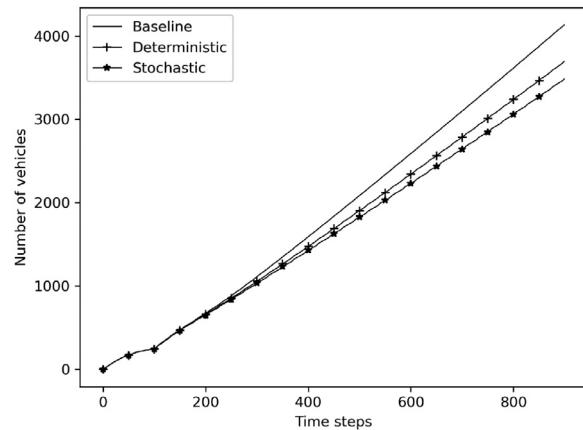
CPU time results of the traffic network of Downtown Ann Arbor during morning peak hours.

|               | MP-min (s) | MP-max (s) | MP-A (s) | SP-min(s) | SP-max (s) | SP-A (s) |
|---------------|------------|------------|----------|-----------|------------|----------|
| Deterministic | 0.24       | 0.79       | 0.48     | 41.35     | 41.66      | 41.54    |
| Stochastic    | 1.55       | 2.67       | 1.94     | 447.30    | 474.69     | 464.51   |

**Table 9**

Out-of-sample evaluation results of Downtown Ann Arbor during morning peak hours.

|               | In-sample Obj (veh-s) | Out-of-sample Obj (veh-s) | Gap                    |
|---------------|-----------------------|---------------------------|------------------------|
| Deterministic | −1498923.97           | −1939263.92               | 22.70%                 |
| Stochastic    | −1540021.95           | −2023592.97               | 23.88%                 |
| Baseline      | Average delay (s)     | Total arrival (veh)       | Total throughput (veh) |
| Deterministic | 352.76                | 7956.79                   | 3295.30                |
| Stochastic    | 297.40 (15.69%)       | 7956.79                   | 3789.91 (15.01%)       |
|               | 276.33 (7.62%)        | 7956.79                   | 4009.33 (5.47%)        |



**Fig. 5.** Average number of vehicles in the traffic network during morning peak hours for all the scenarios.

The results in [Table 9](#) show that although the gaps between in-sample and out-of-sample objective values of deterministic and stochastic models are similar, the stochastic model outperforms the deterministic counterpart in terms of the average delay and total throughput. Therefore, it is valuable to take into account uncertainties in real-world traffic signal control. We show that both deterministic and stochastic models significantly outperform the baseline in delays and throughputs, with at most 15.69% and 21.67% improvement, respectively, illustrating the advantages of considering the coordination between different intersections.

We visualize the average number of vehicles in the network across all the scenarios over time in [Fig. 5](#). The line represents the baseline, the line marked by plus signs represents the deterministic model and the line marked by stars represents the stochastic model. We find that the number of vehicles in the traffic network increases the fastest in the baseline setting, while it increases the slowest in the stochastic model, which also suggests the benefits of stochastic models in preventing congestion. We also provide the visualizations of the number of vehicles under the best and worst scenarios with respect to arrival and delay in the [Appendix B.2](#).

[Figure 6](#) provides the snapshots of the spatial distribution of the number of vehicles of deterministic and stochastic models under scenarios with the minimum and maximum delay (i.e., the best scenario and worst scenario) at time step  $t = 800$ . In the figure, we visualize the occupancy ratio, which stands for the ratio between the number of vehicles and the maximum allowed number of vehicles in the cell. The road segments are darker if the occupancy ratio is higher, meaning worse congestion. [Figure 6\(a\)](#) and

(b) show less congestion given by the stochastic model under the best scenario. Comparing [Fig. 6\(c\)](#) and (d), although both models have congestion, there is fewer number of congested roads given by the signal timing plan produced by the stochastic model. The figures also show that for the deterministic model, the congestion in the East–West direction is worse while for the stochastic model, the congestion in the North–South direction is worse.

We present the traffic delay of solutions using the three different methods (i.e., Baseline, Deterministic, and Stochastic) for each intersection in [Fig. 7\(a\)](#). In addition, we present the histograms of the traffic delay for all the intersections in [Fig. 7\(b\)](#).

In [Fig. 7](#), not all the delays at intersections are reduced. Compared to the baseline, the delays at 60% of the intersections are reduced or unchanged by the deterministic model, and 69% are reduced or unchanged by the stochastic model. The histograms show that the maximum delays at all intersections by the deterministic and stochastic models are significantly lower compared to the baseline, showing that considering coordination leads to avoiding over-saturation. The stochastic model shows the best performance by obtaining the highest counts of intersections for the minimum delay group, which is less than 11 seconds. The spatial distribution of intersections with delay reduced or unchanged is presented in [Appendix B.2](#).

### 5.3.2. Results of off-peak hours

The computational time of the deterministic and stochastic models under off-peak hours are presented in [Table 10](#). Both models can be solved within acceptable computational time. Compared to the results under peak hours, it takes less time to solve the models with lower demand.

We compare the performance of the baseline, the deterministic model, and the stochastic model in [Table 11](#). We also show the solution improvements of the deterministic model compared to the baseline solution, and the solution improvements of the stochastic model compared to the deterministic one, both by percentage. For all the signal timing plans, the average delay of off-peak hours is less than the one of peak hours.

Comparing [Table 11](#) with [Table 9](#), the gaps between in-sample and out-of-sample objective values of the two models are larger than the gaps during peak hours since the deviation of the source demand is higher. The improvements of the deterministic and stochastic models are both more significant in terms of average delay while the improvement of throughput are similar compared to the results of peak hours. We visualize the average number of vehicles in the network across all the scenarios over time in [Fig. 8](#). The increase rates of the number of vehicles of all models are slower than the results of peak hours since the mean source demand of off-peak hours is smaller. In [Fig. 9](#), we visualize the num-

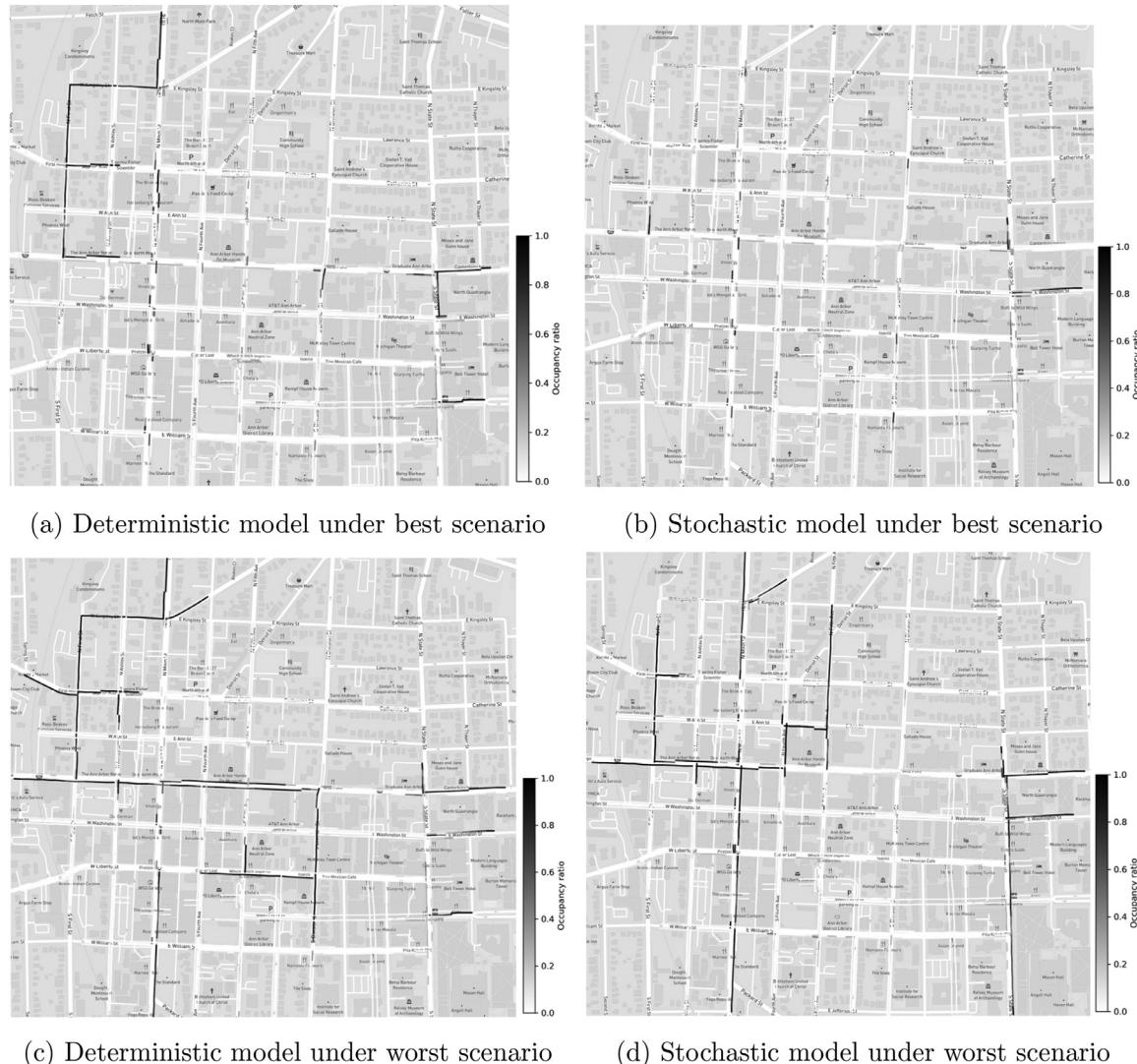


Fig. 6. Spatial distribution of vehicles at time step 800 during morning peak hours.

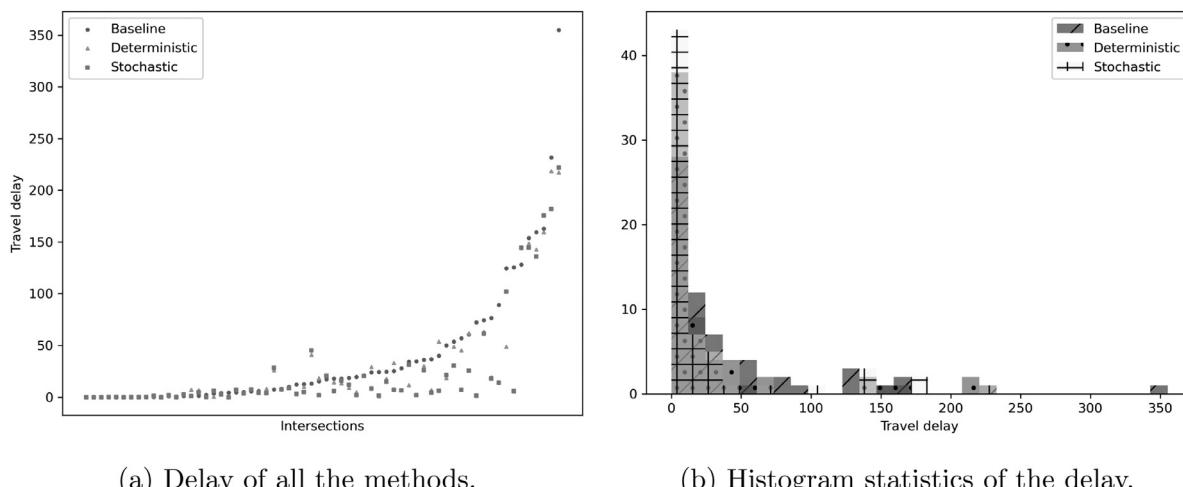


Fig. 7. Average travel delay for all the intersections during morning peak hours.

**Table 10**

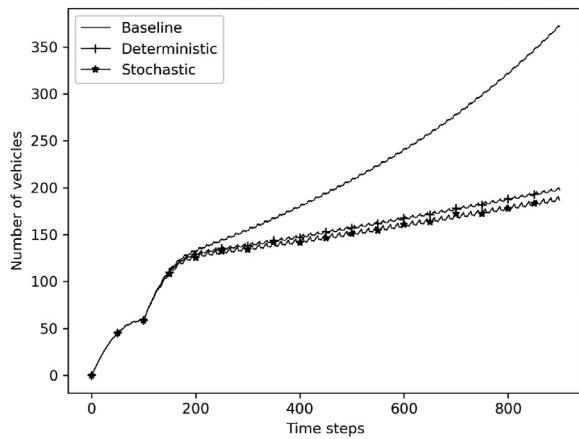
CPU time results of the traffic network of Downtown Ann Arbor during off-peak hours.

|               | MP-min (s) | MP-max (s) | MP-A (s) | SP-min(s) | SP-max (s) | SP-A (s) |
|---------------|------------|------------|----------|-----------|------------|----------|
| Deterministic | 0.13       | 0.40       | 0.23     | 40.70     | 42.36      | 41.28    |
| Stochastic    | 0.89       | 2.72       | 1.79     | 424.50    | 450.26     | 439.44   |

**Table 11**

Out-of-sample evaluation results of Downtown Ann Arbor during off-peak hours.

|               | In-sample Obj (veh-s) | Out-of-sample Obj (veh-s) | Gap                    |
|---------------|-----------------------|---------------------------|------------------------|
| Deterministic | −580971.48            | −1121107.27               | 48.17%                 |
| Stochastic    | −592697.54            | −1140783.78               | 48.04%                 |
| Baseline      | Average delay (s)     | Total arrival (veh)       | Total throughput (veh) |
| Deterministic | 117.99                | 2905.69                   | 2040.82                |
| Stochastic    | 59.36 (49.69%)        | 2905.69                   | 2419.07 (18.53%)       |
|               | 51.19 (13.77%)        | 2905.69                   | 2466.60 (1.96%)        |

**Fig. 8.** Average number of vehicles in the traffic network during afternoon peak hours for all the scenarios.

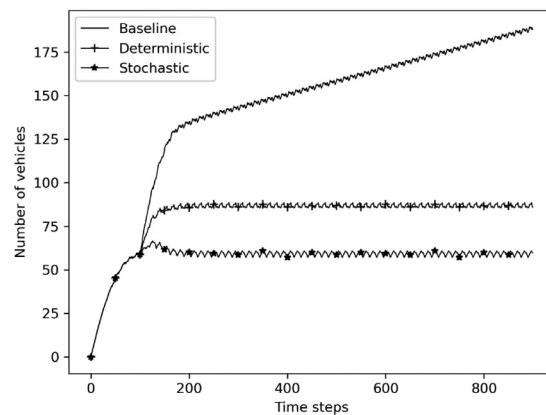
ber of vehicles in the network of the scenarios with the minimum delay and maximum delay of the deterministic model, stochastic model and baseline setting. **Figure 9(a)** shows that under the scenario with minimum delay, the number of vehicles in the traffic network keep stable for both deterministic and stochastic models, and there are fewer vehicles of the stochastic model. **Figure 9(b)** shows that under the scenario with maximum delay, the number of vehicles of all the settings increases and the increasing rate of

the stochastic model is slower than the other two settings. We also provide the visualizations of the number of vehicles under the best and worst scenarios with respect to arrival in [Appendix B.2](#).

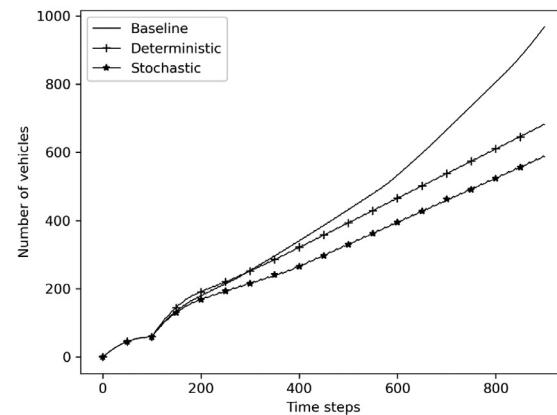
**Figure 10** provides the snapshots of the spatial distribution of the number of vehicles of the deterministic and stochastic models under the best and worst scenario at time step  $t = 800$ . There are fewer congested intersections in the results of the stochastic model under both scenarios, compared to the peak hours. While the solutions returned by stochastic model can lead to almost fully empty roads, deterministic counterpart still leaves several roads congested.

We present the average traffic delay of solutions given by the three methods during off-peak hours for each intersection in **Fig. 11(a)**, and the traffic delay for all intersections in **Fig. 11(b)**. All the legends are the same as in **Fig. 7** for the morning peak hours instance.

Compared to baseline, the delays at 64% of the intersections are reduced or unchanged by the deterministic model, and at 70% intersections are reduced or unchanged by the stochastic model. These percentages are higher compared to the ones during morning peak hours. The histograms show that both centralized models improve the maximum delays at all intersections compared to the baseline case. The stochastic model obtains the most number of intersections in the minimum delay group, which is less than 2.5 seconds during off-peak hours, demonstrating the benefits of considering coordination and uncertainties. We present the spatial distribution of intersections with delay reduced or unchanged in [Appendix B.2](#).



(a) Scenario with minimum delay



(b) Scenario with maximum delay

**Fig. 9.** Number of vehicles in the Downtown Ann Arbor traffic network during off-peak hours.

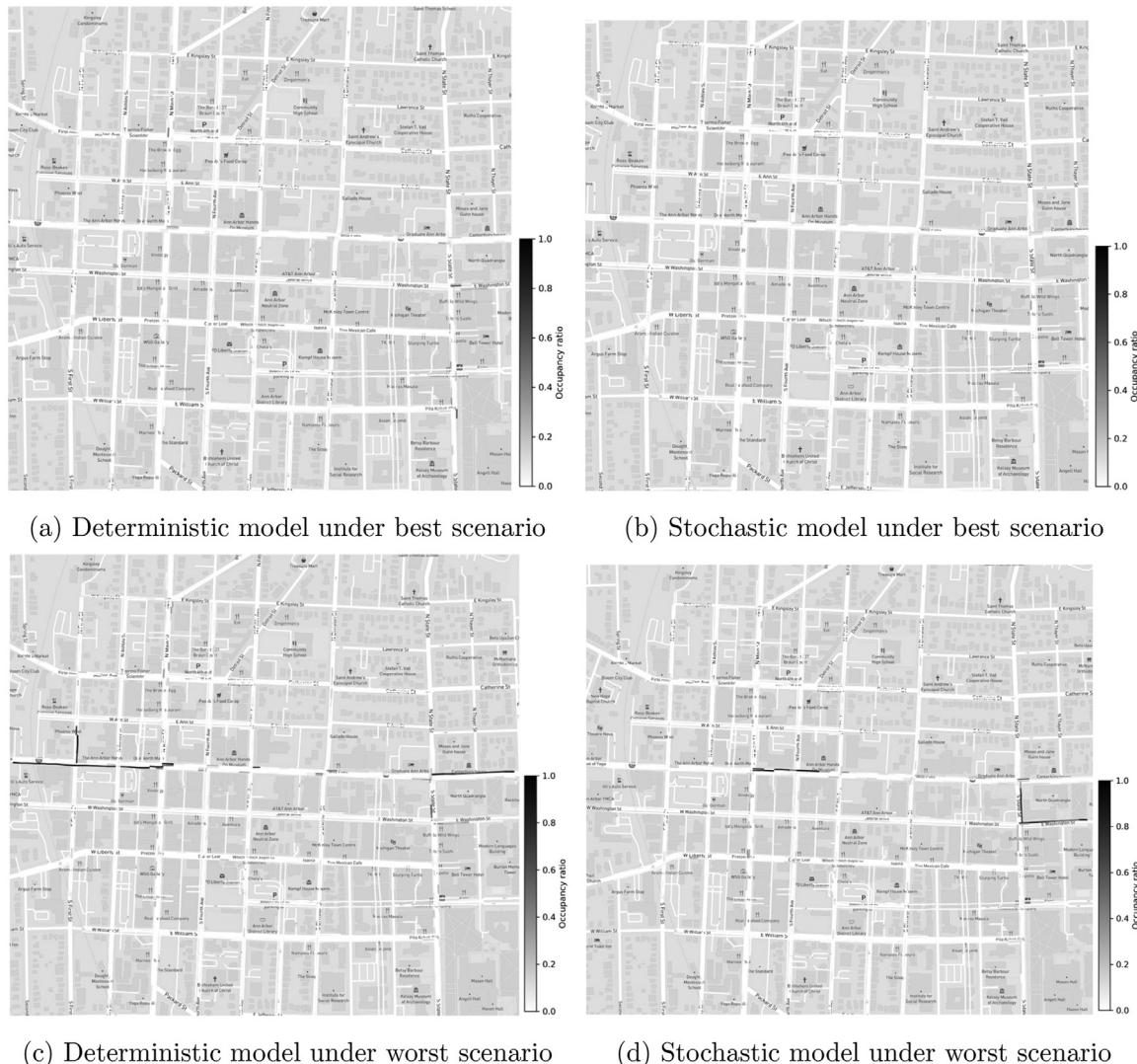


Fig. 10. Spatial distribution of vehicles at time step 800 during off-peak hours.

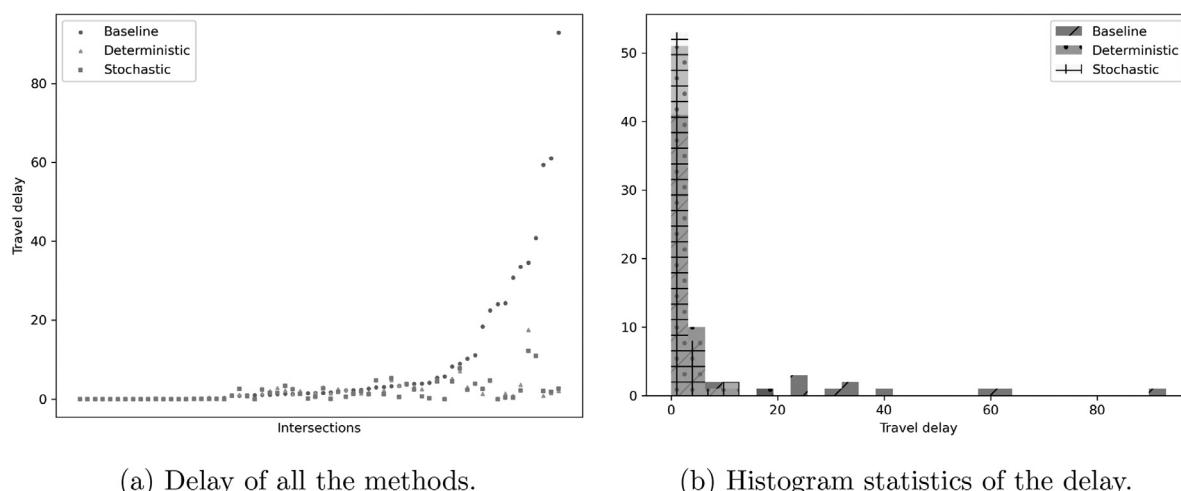


Fig. 11. Average travel delay for all the intersections during off-peak hours.

Based on all the above results, we show the benefits of considering travel demand and turning ratio uncertainties as well as the coordination of intersections in traffic signal control with real-world data and road networks. Our models work well for real-world instances during both peak hours and off-peak hours and our stochastic model is more appropriate for instances with high standard deviations of travel demand.

## 6. Conclusion

In this paper, we built a CTM-based MIP for traffic signal control and extended the deterministic model to a two-stage stochastic optimization model by considering random source demand and turning ratios. We proposed efficient algorithms for solving the models and overcoming the scalability difficulties. Our algorithm not only reduced the computational time but also ensured the optimality for the non-convex model with mixed-integer variables.

With the numerical results obtained from randomly generated grid networks and real-world traffic networks, we firstly showed the reduction of the computational time of our algorithm. Then we demonstrated the benefits of considering stochastic traffic demand. Furthermore, we illustrated the advantages of our model to consider the coordination of all the intersections in a real-world traffic network. We noted that the parameters in the model need fine-tuning in practice, and this can be achieved easily since our decision models are solved offline. The proposed method can be used by traffic engineers to generate the fixed-time or background signal timing plan by taking stochastic traffic demand as input in forms of distribution or different scenarios, which is more robust than only considering the average traffic volume.

There are several limitations and corresponding possible directions for future research. Firstly, our proposed formulation only optimizes traffic signal timing parameters based on a pre-determined phase sequence and structure. It will be more flexible if the phase structure (i.e., lead and lag phase) can also be considered as a decision variable. Secondly, like most existing traffic signal control literature (Lo, 1999; Varaiya, 2013), vehicle re-routing and induced demand due to the change of the traffic signal timing plan is not explicitly included in this work. The overall traffic control problem can also be formulated as a more complicated bi-level programming problem by taking vehicle re-routing into consideration. We leave this for future study. In addition, with the development of connected and autonomous vehicles (CAV), the distribution of traffic demand can be directly estimated from real-world data. The combination of data collected from CAV and our proposed approach provides a complete framework for traffic signal optimization in practice. Furthermore, demographics and socioeconomics can be used to predict long-term traffic demand, providing the potential to estimate traffic flow distribution better, which is helpful to guide the planning and operation of urban traffic optimization.

## Acknowledgment and disclaimer

Funding for this research was provided by the Center for Connected and Automated Transportation under Grant no. 69A3551747105 of the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (OST-R), University Transportation Centers Program. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

## Appendix A. Benders decomposition algorithm

We first create new variables  $\theta^k$ ,  $k = 1, \dots, K$  and define a relaxed master problem as follows.

$$\begin{aligned} (\text{RMP}) \quad \min \quad & \sum_{k=1}^K p^k \theta^k \\ \text{s.t.} \quad & (l, o, g, b, e, z_1, z_2) \in X, \\ & (z_1, z_2, \theta) \in \Sigma(z_1, z_2, \theta). \end{aligned}$$

Here,  $\Sigma(z_1, z_2, \theta)$  is the set of Benders cuts as linear functions of  $z_1, z_2$  generated up to the current iteration from sample-based subproblems:

$$\begin{aligned} (\text{SP}^k) \quad \mathcal{Q}^k(z_1, z_2) = \min \quad & - \sum_{c \in \mathcal{D}} \sum_{t=1}^T n_{ct}^k + \alpha \sum_{c \in \mathcal{C}} \sum_{t=1}^T (T-t)y_{ct}^k \\ \text{s.t.} \quad & (y^k, n^k) \in \text{Constraints (2b)–(2m) for each } k. \end{aligned}$$

Given first-stage integer solutions  $\hat{z}_1, \hat{z}_2$ , one can solve  $(\text{SP}^k)$  with  $z_1 = \hat{z}_1$ ,  $z_2 = \hat{z}_2$  and obtain  $\mathcal{Q}^k(\hat{z}_1, \hat{z}_2)$ , for each  $k = 1, \dots, K$ . The signal constraints (1a)–(1j) in the first stage ensure that for each  $i \in \mathcal{R}$ ,  $j \in \mathcal{F}_i$ ,  $m = 1, \dots, N_{cy}$  and  $t = 1, \dots, T$ , we have  $\hat{z}_{1ijmt} + \hat{z}_{2ijmt} \geq 1$ . Given non-negative parameters  $Q, N, D, n^{\text{init}}, W$  and positive parameter  $\epsilon$ , each subproblem  $(\text{SP}^k)$  always has at least one feasible solution  $y^k = n^k = \mathbf{0}$  for any given  $\hat{z}_1, \hat{z}_2$ . Therefore, we only need to add optimality cuts to the set  $\Sigma$ .

We next derive optimality cuts from the dual formulations of the second-stage linear programs  $(\text{SP}^k)$ . Associate dual variables  $\rho_{ct}^k$ ,  $\sigma_{ct}^k$ ,  $\pi_{cc't}^k$ ,  $\gamma_{cc't}^k$ ,  $\delta_{ct}^k$ ,  $\tau_c^k$  to constraints (2b)–(2l), respectively. Following weak duality, for each  $k = 1, \dots, K$ , the optimality cut is of the form:

$$\begin{aligned} \theta^k \geq & \sum_{c \in \mathcal{C}/\mathcal{T}} \sum_{t=1}^T Q_{ct} \sigma_{ct}^k + \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{F}_i} \sum_{c \in \mathcal{I}_{ij}} \sum_{m=1}^{N_{cy}} \sum_{t=1}^T (z_{1ijmt} + z_{2ijmt} - 1) Q_{ct} \sigma_{ct}^k \\ & + \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{D}(c)} \sum_{t=1}^T Q_{c't} \pi_{cc't}^k + \sum_{c \in \mathcal{C}} \sum_{c' \in \mathcal{D}(c)} \sum_{t=1}^T W_{c't} N_{c't} \gamma_{cc't}^k \\ & + \sum_{c \in \mathcal{C}} \sum_{t=1}^T D_{ct}^k \delta_{ct}^k + \sum_{c \in \mathcal{C}} n_c^{\text{init}} \tau_c^k. \end{aligned} \quad (\text{A.2})$$

The generic Benders procedures are presented in [Algorithm 2](#). In every iteration, we solve  $(\text{RMP})$  with current Benders' cuts and

---

**Algorithm 2:** A generic Benders decomposition approach for solving model (2).

---

- 1: Initialize  $(\text{RMP})$  with  $\Sigma = \emptyset$ .
  - 2: **while** the termination criteria is not satisfied **do**
  - 3:     Solve  $(\text{RMP})$  to obtain optimal solutions  $(\hat{z}_1, \hat{z}_2, \hat{\theta})$ .
  - 4:     **for**  $k = 1, \dots, K$  **do**
  - 5:         Solve  $(\text{SP}^k)$  to obtain optimal dual solutions.
  - 6:         **if**  $\hat{\theta}^k < \mathcal{Q}^k(\hat{z}_1, \hat{z}_2)$  **then**
  - 7:             Add an optimality cut~(21) to the set  $\Sigma(z_1, z_2, \theta)$ .
  - 8:         **end if**
  - 9:     **end for**
  - 10: **end while**
  - 11: Return the objective value as  $\sum_{k=1}^K p^k \hat{\theta}^k$  and the solutions of  $(\text{RMP})$ .
- 

obtain feasible solutions  $\hat{z}_1, \hat{z}_2$ . The optimal value of  $(\text{RMP})$  provides a lower bound to the original stochastic MIP (2). We then compute optimal dual solutions to the second-stage subproblems  $(\text{SP}^k)$  for each  $k = 1, \dots, K$ , and the expectation of their optimal values provides an upper bound of the overall optimal objective

**Table 12**  
Delay and total throughput of different cycle lengths.

| Cycle length ( $\hat{l}$ ) | 19          | 24          | 29          | 34          | 39          |
|----------------------------|-------------|-------------|-------------|-------------|-------------|
| In-sample Obj              | −1507598.54 | −1498923.98 | −1490859.43 | −1479932.66 | −1427223.55 |
| Average Delay              | 286.34      | 297.40      | 293.04      | 303.72      | 302.76      |
| Total Throughput           | 3881.38     | 3789.91     | 3848.80     | 3773.03     | 3820.57     |

**Table 13**  
Objective value and evaluation metrics of different in-sample scenarios  $k$ .

| In-sample scenarios | 5           | 10          | 15          | 20          |
|---------------------|-------------|-------------|-------------|-------------|
| In-sample Obj       | −1493833.50 | −1540021.95 | −1462882.47 | −1535876.09 |
| Out-of-sample Obj   | −1925990.42 | −2023592.97 | −1992383.23 | −1929821.75 |
| Gap                 | 22.44%      | 23.90%      | 26.58%      | 20.41%      |
| Average delay       | 298.96      | 276.33      | 283.97      | 297.95      |
| Total throughput    | 3811.87     | 4009.33     | 3931.91     | 3817.81     |

value. We terminate the algorithm when the gap between the upper bound and the lower bound is sufficiently small.

## Appendix B. Additional numerical results

### B1. Sensitivity analysis

**Cycle length** We solve the deterministic model with various cycle lengths  $\hat{l} = 19, 24, 29, 34, 39$ , where for all the intersections  $i \in \mathcal{R}$ ,  $\hat{l}_i = \hat{l}$ . We present the average delay and throughput under the out-of-sample test with 500 scenarios in Table 12.

We show that for all the metrics, the gap among different cycle lengths is no larger than 5%, which demonstrates that the performance of signal timing plans is not sensitive to the cycle length. In our following numerical experiments, we follow the rule in Remark 3 and choose cycle length as  $\hat{l}_i = 24$ ,  $\forall i \in \mathcal{R}$ . In real-world applications, one can test the model with multiple selected cycle lengths and choose the one with the best performance. **In-sample scenarios** We choose in-sample scenarios  $K = 5, 10, 15, 20$  and solve the stochastic MIP. We present the in-sample and out-of-sample objective values as well as the gaps between them. We test the signal timing plans on the same out-of-sample test set with 500 scenarios and present the results in Table 13.

We show that when the total number of scenarios is 20, the gaps between in-sample and out-of-sample tests decrease slightly. However, the signal timing plan given by the sample with  $K = 10$  scenarios has the best performance in out-of-sample tests. The main reason is that we terminate early in solving second-stage problems by ADMM allowing at most 20% gap between the obtained objective value and the optimal one. Also, adding in-sample scenarios increases the difficulty of solving the problem, making solutions of  $K = 15, 20$  not as good as  $K = 10$  with the same number of Benders iterations. To obtain the balance between computation and results, we use  $K = 10$  as our in-sample scenario size in our later numerical studies.

**Out-of-sample scenarios** We vary the number of out-of-sample test scenarios among 500, 1000, 1500, 2000 and evaluate the performance of the signal timing plan obtained by the solution of  $K = 10$  instance. We present the out-of-sample objective values, gaps between in-sample and out-of-sample objectives, average delay, total arrival, and the total throughput results in Table 14.

Note that all the results are similar regardless of the choice of out-of-sample scenario size. We use 500 scenarios for our later computation in all out-of-sample tests.

### B2. Number of vehicles in real-world traffic networks

**Results of morning peak hours** To evaluate the traffic signal plans, in Fig. 12, we visualize the number of vehicles in the network

of the scenarios with the lowest and highest arrivals. The figure shows that in both scenarios, the stochastic model outperforms the deterministic model and the baseline, and it is more significant when the arrival is lower.

In Fig. 13, we visualize the number of vehicles in the traffic network under the scenarios with minimum delay and maximum delay. For both scenarios, the number of vehicles keeps increasing. The increasing rate of the stochastic model is lower under the scenario with minimum delay while the increase rate of the deterministic model is lower under the scenario with maximum delay. The increase rate of the stochastic model is higher than the deterministic model because the total arrival of the stochastic model is larger.

In Fig. 14, we present the spatial distribution of intersections compared between the stochastic model and the baseline in the morning peak hours, where the delay of intersections marked by green color is reduced or unchanged and the delay of others is increased. We show that for every corridor, the delay of the majority of intersections is reduced or unchanged. The intersections with increased delay are most in the upper-left corner and around a parking lot.

**Results of off-peak hours** In Fig. 15, we visualize the number of vehicles in the network of the scenarios with the lowest and highest arrivals. Figure 15(a) shows that when the arrival is low, the number of vehicles in the traffic work keeps stable for both deterministic and stochastic models. Figure 15(b) shows that when the arrival is high, the increase rates of the number of vehicles of the stochastic model is slower than the other two settings.

In Fig. 16, we present the spatial distribution of intersections compared between the stochastic model and the baseline in the traffic network of off-peak hours. The intersections with reduced delay are marked by green color. We show that for every corridor, the delay of the majority of intersections is reduced or unchanged. The intersections with increased delay are most in the lower-left corner and around a parking lot.

### B3. Detailed results of afternoon peak hours

We present the computational time of the deterministic and stochastic models in seconds in Table 15. We show that our distributed algorithm solves the problem within an acceptable time limit. The computational time of master problems is slightly higher but the time of subproblems is lower compared to the morning peak hours instances.

We present the in-sample and out-of-sample objective values as well as their gaps of the deterministic and stochastic model in Table 16. We show that the gap of the stochastic model is slightly smaller than the deterministic model because it has a better real-

**Table 14**

Objective value and evaluation metrics of different out-of-sample scenarios.

| Out-of-sample scenarios | 500         | 1000        | 1500        | 2000        |
|-------------------------|-------------|-------------|-------------|-------------|
| Out-of-sample Obj       | −2023592.97 | −2021604.03 | −2024107.83 | −2024080.06 |
| Gap                     | 23.90%      | 23.82%      | 23.92%      | 23.91%      |
| Average delay           | 276.33      | 274.40      | 275.09      | 275.61      |
| Total arrival           | 7956.79     | 7922.89     | 7932.95     | 7942.05     |
| Total throughput        | 4009.33     | 4016.07     | 4020.70     | 4020.16     |

**Table 15**

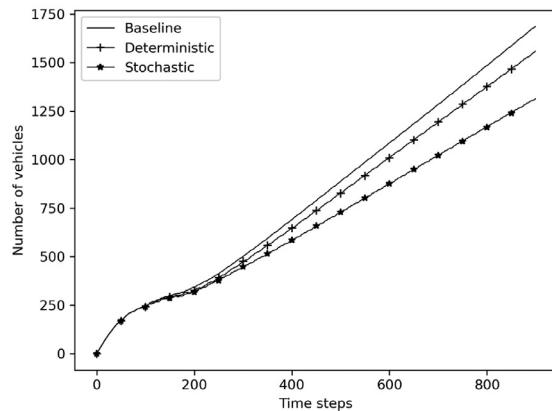
CPU time results of the traffic network of Downtown Ann Arbor during afternoon peak hours .

|               | MP-min (s) | MP-max (s) | MP-A (s) | SP-min(s) | SP-max (s) | SP-A (s) |
|---------------|------------|------------|----------|-----------|------------|----------|
| Deterministic | 0.23       | 0.94       | 0.50     | 37.50     | 37.81      | 37.67    |
| Stochastic    | 1.18       | 3.17       | 2.08     | 378.51    | 393.16     | 386.89   |

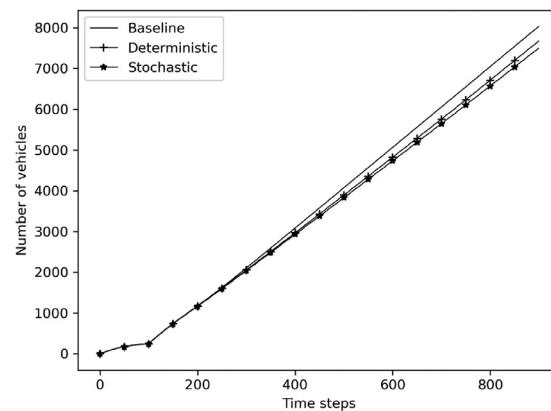
**Table 16**

Out-of-sample evaluation results of Downtown Ann Arbor during afternoon peak hours .

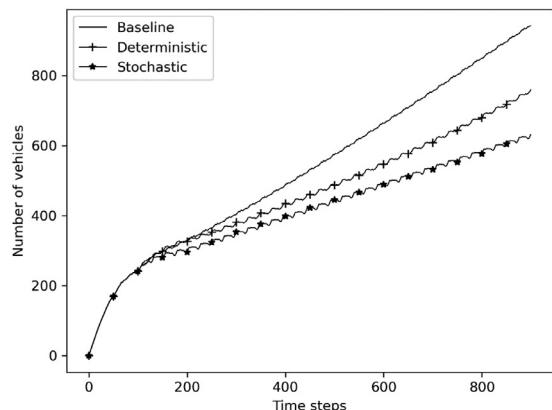
|                   | In-sample Obj (veh·s) | Out-of-sample Obj (veh·s) | Gap                    |
|-------------------|-----------------------|---------------------------|------------------------|
| Deterministic     | −1510461.85           | −2082966.89               | 27.49%                 |
| Stochastic        | −1602873.23           | −2177611.47               | 26.39%                 |
| Average delay (s) |                       | Total arrival (veh)       | Total throughput (veh) |
| Baseline          | 425.93                | 9447.04                   | 3029.47                |
| Deterministic     | 344.58 (23.61%)       | 9447.04                   | 3961.59 (30.77%)       |
| Stochastic        | 323.25 (6.60%)        | 9447.04                   | 4179.38 (5.50%)        |



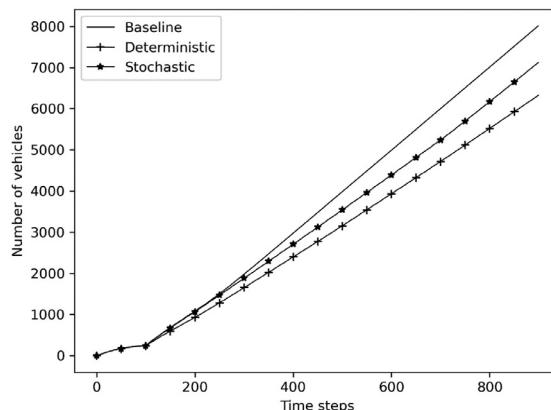
(a) Scenario with lowest arrival



(b) Scenario with highest arrival

**Fig. 12.** Number of vehicles in the traffic network during morning peak hours.

(a) Scenario with minimum delay

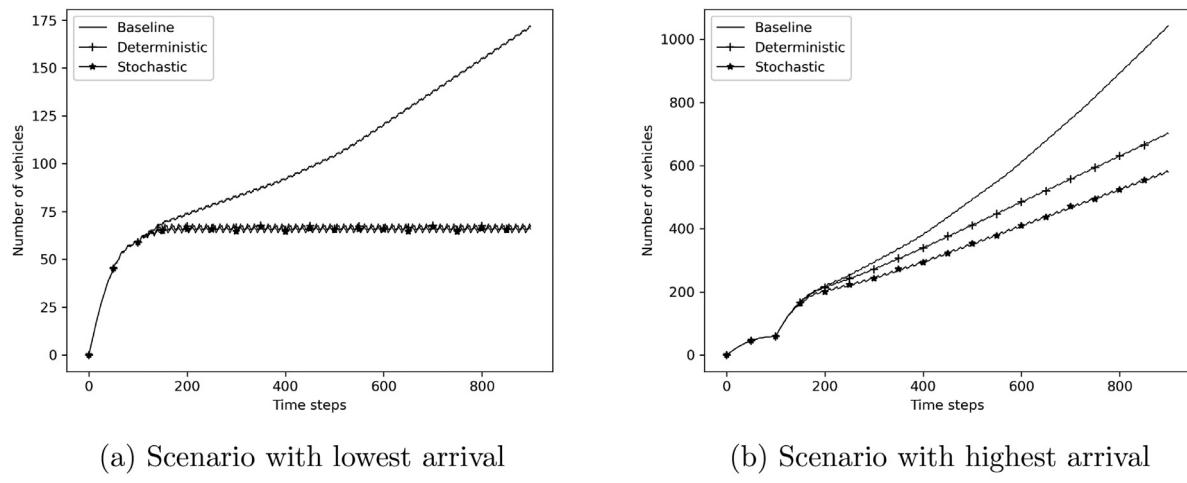


(b) Scenario with maximum delay

**Fig. 13.** Number of vehicles in the traffic network during morning peak hours.



**Fig. 14.** Spatial distribution of intersections under the stochastic model solution with delay improved or unchanged compared to the baseline in the traffic network during morning peak hours, marked by green color in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 15.** Number of vehicles in the traffic network during off-peak hours.

ization of the uncertainties. Both gaps are similar to the results of morning peak hours because they have the same deviation.

Furthermore, we present the average delay and total throughput to evaluate the signal controls. In Row “Deterministic”, we also present the improvement of the deterministic model compared to the baseline. Similarly, in Row “Stochastic”, we present the improvement of the stochastic model compared to the deterministic model. We demonstrate that our deterministic and stochastic models both outperform dramatically compared to the baseline, indicating that our centralized model obtains better network coordination of intersections. The stochastic model obtains a smaller average travel delay and larger total throughput, showing the benefits of considering uncertainties.

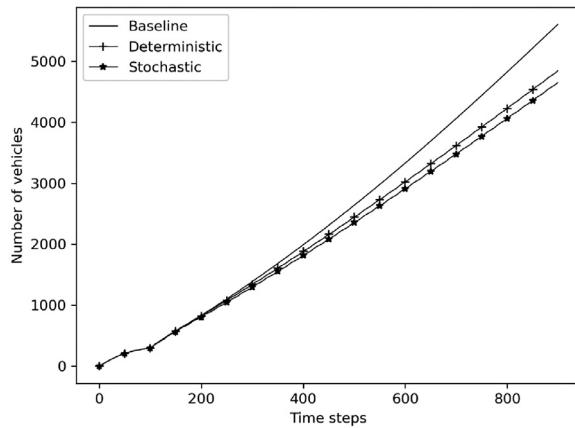
Compared to the results of morning peak hours, the impact of network coordination is more significant, demonstrated by the

higher improvement when being compared with the baseline case. In particular, the improvement of total throughput is larger than morning peak hours and off-peak hours, mainly because the traffic flow is highest during afternoon peak hours. On the other hand, the improvement of the stochastic model compared to the deterministic model is smaller. The main reason is that under extremely high traffic flow, the optimization space for traffic signals is small so the difference between the results of the two models decreases.

We visualize the average number of vehicles in the traffic network for all the scenarios in Fig. 17. We show that although the number of vehicles increases with time for all the methods because of high traffic arrival, the stochastic model has the slowest increase while the baseline has the fastest increase. The increase rate of afternoon peak hours is higher than the two other TODs because the traffic arrival is higher.



**Fig. 16.** Spatial distribution of intersections under the stochastic model solution with delay improved or unchanged compared to the baseline in the traffic network during off-peak hours, marked by green color in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



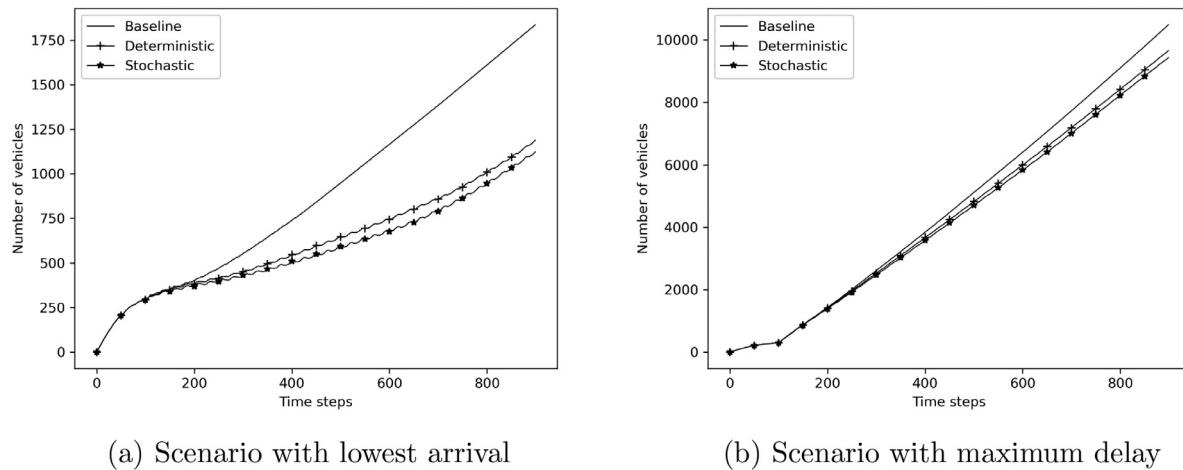
**Fig. 17.** Average number of vehicles in the traffic network during afternoon peak hours for all the scenarios.

Furthermore, we visualize the number of vehicles in the network under the scenario with the lowest and highest traffic arrival in Fig. 18. We show that for both scenarios, the stochastic model outperforms the deterministic model and the improvement is more significant under the scenario with the lowest arrival because, with higher traffic arrival, the intersections are more possible to be saturated under all the signal timing plans, causing an increase of the number of vehicles in the network.

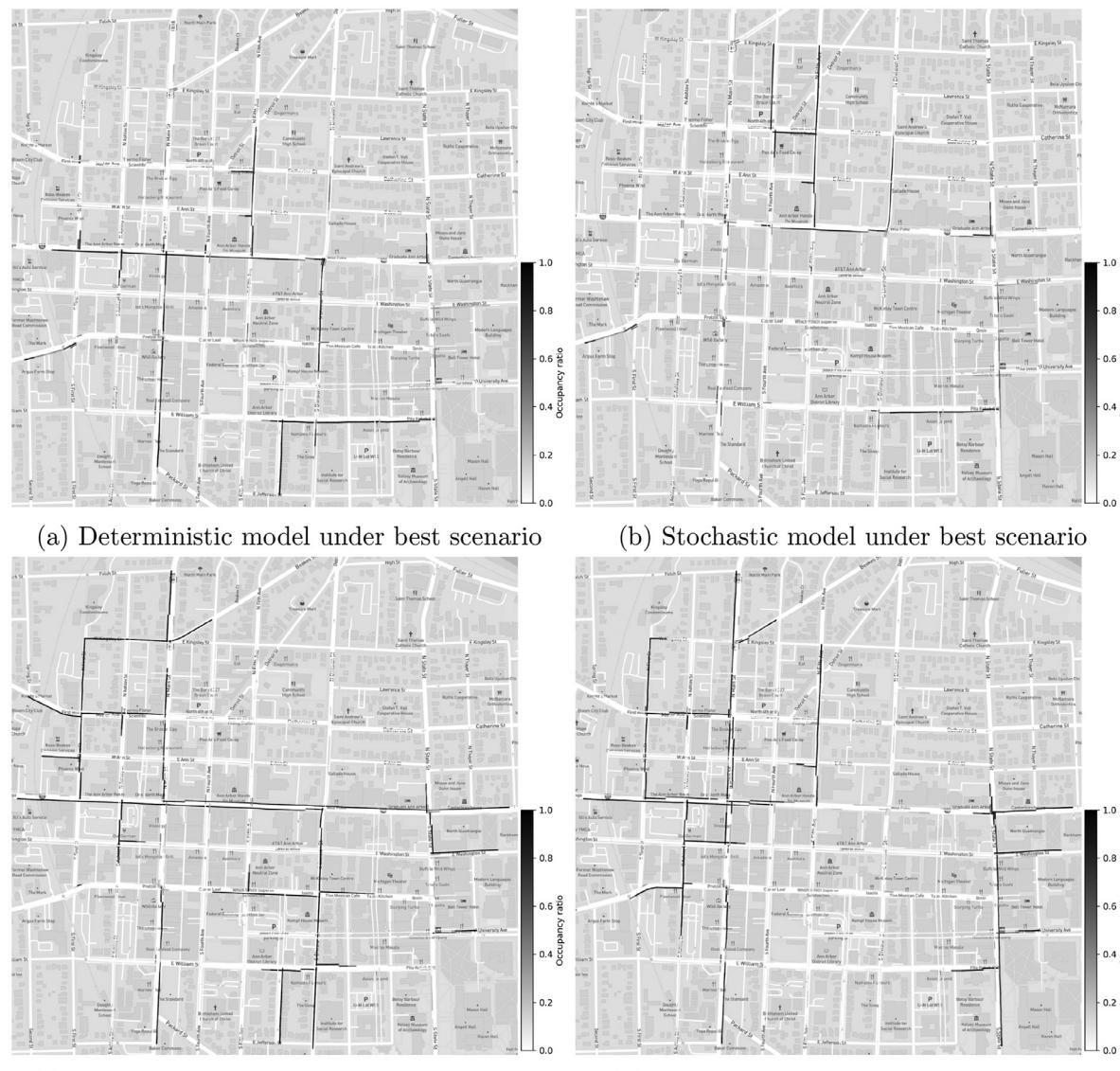
We provide the snapshots of the spatial distribution of the number of vehicles of the deterministic and stochastic models un-

der the best and worst scenario at time step  $t = 800$  in Fig. 19. We show that even under the best scenario, traffic congestion exists on two main corridors, Main Street and Huron Street. The stochastic model obtains significantly better performance on main corridors but still causes traffic congestion on a few local roads with the direction of north and south. For the worst scenario, traffic congestion in the stochastic model is mostly in the direction of north and south direction while traffic congestion in the deterministic model is in both directions.

The average traffic delay of each intersection is presented in Fig. 20(a) and the histograms of all intersections are presented in Fig. 20(b). Compared to the baseline, the delays at 69% of the intersections are reduced or unchanged by the deterministic model, and at 75% of all the intersections are reduced or unchanged by the stochastic model. All the percentages are higher than the other two instances, showing that our model is able to reduce the congestion in more intersections with higher source demand. Compared to the baseline, our centralized model obtains a lower maximum delay. The stochastic model obtains the highest percentage of intersections in the minimum delay group, which is less than 12 seconds during afternoon off-peak hours, showing the advantages of considering network coordination and traffic uncertainties. In Fig. 21, we present the spatial distribution of intersections compared between the stochastic model and the baseline in the traffic network of afternoon peak hours. The intersections with reduced delay are marked by green color and the others are marked by red color. We show that for every corridor, the delay of the majority of intersections is reduced or unchanged.



**Fig. 18.** Number of vehicles in the Downtown Ann Arbor traffic network during afternoon peak hours.



**Fig. 19.** Spatial distribution of vehicles at time step 800 during afternoon peak hours.

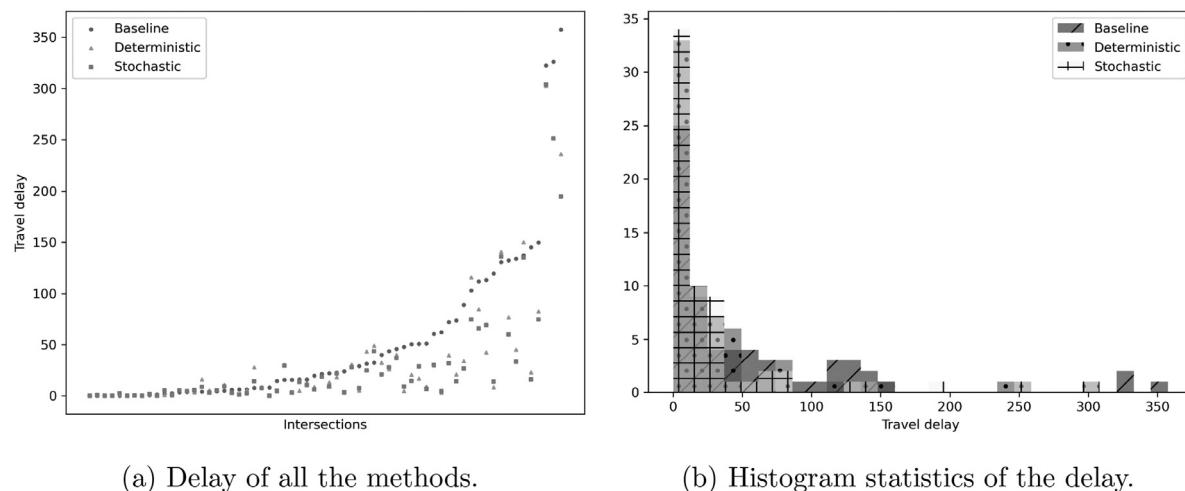


Fig. 20. Average travel delay for all the intersections during off-peak hours.



Fig. 21. Spatial distribution of intersections under the stochastic model solution with delay improved or unchanged compared to the baseline in the traffic network during afternoon peak hours, marked by green color in the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## References

- Aboudolas, K., Papageorgiou, M., & Kosmatopoulos, E. (2009). Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. *Transportation Research Part C: Emerging Technologies*, 17(2), 163–174.
- Adacher, L., & Tirolo, M. (2018). A macroscopic model with the advantages of microscopic model: A review of cell transmission model's extensions for urban traffic networks. *Simulation Modelling Practice and Theory*, 86, 102–119.
- Al Islam, S. B., & Hajbabaie, A. (2017). Distributed coordinated signal timing optimization in connected transportation networks. *Transportation Research Part C: Emerging Technologies*, 80, 272–285.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al., (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1–122.
- Chiou, S. W. (2019). A two-stage model for period-dependent traffic signal control in a road networked system with stochastic travel demand. *Information Sciences*, 476, 256–273.
- Comert, G., & Cetin, M. (2011). Analytical evaluation of the error in queue length estimation at traffic signals from probe vehicle data. *IEEE Transactions on Intelligent Transportation Systems*, 12(2), 563–573.
- Daganzo, C. (1992). The cell transmission model. Part I: A simple dynamic representation of highway traffic.
- Hajbabaie, A., & Benekohal, R. F. (2015). A program for simultaneous network signal timing optimization and traffic assignment. *IEEE Transactions on Intelligent Transportation Systems*, 16(5), 2573–2586.
- Heydecker, B. (1987). Uncertainty and variability in traffic signal calculations. *Transportation Research Part B: Methodological*, 21(1), 79–85.
- Hoogendoorn, S. P., & Bovy, P. H. (2001). State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 215(4), 283–303.
- Isa, N., Yusoff, M., & Mohamed, A. (2014). A review on recent traffic congestion relief approaches. In *2014 4th international conference on artificial intelligence with applications in engineering and technology* (pp. 121–126). IEEE.
- Koonce, P., & Rodegerds, L. (2008). Traffic signal timing manual. *Technical report*. United States. Federal Highway Administration.

- Levin, M. W., Hu, J., & Odell, M. (2020). Max-pressure signal control with cyclical phase structure. *Transportation Research Part C: Emerging Technologies*, 120, 102828.
- Levin, T. (2020). The 31 US cities that had the worst traffic in 2019 according to a study. <https://www.businessinsider.com/us-cities-most-traffic-2019-2020-3>.
- Li, K., & Ioannou, P. (2004). Modeling of traffic flow of automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 5(2), 99–113.
- Li, L., Huang, W., Chow, A. H., & Lo, H. K. (2021). Two-stage stochastic program for dynamic coordinated traffic control under demand uncertainty. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 12966–12976.
- Li, L., Huang, W., & Lo, H. K. (2018). Adaptive coordinated traffic control for stochastic demand. *Transportation Research Part C: Emerging Technologies*, 88, 31–51.
- Li, L., & Jabari, S. E. (2019). Position weighted backpressure intersection control for urban networks. *Transportation Research Part B: Methodological*, 128, 435–461.
- Liang, X., Guler, S. I., & Gayah, V. V. (2021). Decentralized arterial traffic signal optimization with connected vehicle information. *Journal of Intelligent Transportation Systems*, 1–16. <https://doi.org/10.1080/15472450.2021.1990762>.
- Little, J. D., & Graves, S. C. (2008). Little's law. In *Building intuition* (pp. 81–100). Springer.
- Lo, H. K. (1999). A novel traffic signal control formulation. *Transportation Research Part A: Policy and Practice*, 33(6), 433–448.
- Neely, M. J. (2010). Stochastic network optimization with application to communication and queueing systems. *Synthesis Lectures on Communication Networks*, 3(1), 1–211.
- Pipes, L. A. (1953). An operational analysis of traffic dynamics. *Journal of applied physics*, 24(3), 274–281.
- Prashker, J. N., & Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: A review. *Transport reviews*, 24(4), 437–463.
- Schrank, D., Eisele, B., & Lomax, T. (2019). 2019 urban mobility report. *Technical report*. Texas A&M Transportation Institute.
- Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2014). *Lectures on stochastic programming: Modeling and theory*. SIAM.
- Sheffi, Y. (1985). *Urban transportation networks: Equilibrium analysis with mathematical programming methods*. New Jersey: PRENTICE-HALL, INC., Englewood Cliffs.
- Shirke, C., Sabar, N., Chung, E., & Bhaskar, A. (2022). Metaheuristic approach for designing robust traffic signal timings to effectively serve varying traffic demand. *Journal of Intelligent Transportation Systems*, 26(3), 343–355.
- Stevanovic, A., Martin, P. T., & Stevanovic, J. (2007). VisSim-based genetic algorithm optimization of signal timings. *Transportation Research Record*, 2035(1), 59–68.
- Tajalli, M., Mehrabipour, M., & Hajbabaie, A. (2020). Network-level coordinated speed optimization and traffic light control for connected and automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), 6748–6759.
- Tang, K., Boltze, M., Nakamura, H., & Tian, Z. (2019). *Global practices on road traffic signal control: Fixed-time control at isolated intersections*. Elsevier.
- Tassiulas, L., & Ephremides, A. (1990). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. In *29th IEEE conference on decision and control* (pp. 2130–2132). IEEE.
- Timotheou, S., Panayiotou, C. G., & Polycarpou, M. M. (2014). Distributed traffic signal control using the cell transmission model via the alternating direction method of multipliers. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 919–933.
- Tong, Y., Zhao, L., Li, L., & Zhang, Y. (2015). Stochastic programming model for oversaturated intersection signal timing. *Transportation Research Part C: Emerging Technologies*, 58, 474–486.
- Urbanik, T., Tanaka, A., Lozner, B., Lindstrom, E., Lee, K., Quayle, S., ... Gettman, D., et al. (2015). *Signal timing manual*: vol. 1. Transportation Research Board Washington, DC.
- Varaiya, P. (2013). Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, 36, 177–195.
- Wang, X., Jerome, Z., Zhang, C., Shen, S., Kumar, V. V., & Liu, H. X. (2022a). Trajectory data processing and mobility performance evaluation for urban traffic networks. *Transportation Research Record*, 2677(3), 355–370. 0361981221115088
- Wang, X., Yin, Y., Feng, Y., & Liu, H. X. (2022b). Learning the max pressure control for urban traffic networks considering the phase switching loss. *Transportation Research Part C: Emerging Technologies*, 140, 103670.
- Webster, F. V. (1958). Traffic signal settings. *Technical report*.
- Wu, J., Ghosal, D., Zhang, M., & Chuah, C.-N. (2017). Delay-based traffic signal control for throughput optimality and fairness at an isolated intersection. *IEEE Transactions on Vehicular Technology*, 67(2), 896–909.
- Wu, N., & Brilon, W. (1999). Cellular automata for highway traffic flow simulation. In *Proceedings 14th international symposium on transportation and traffic theory (Abbreviated presentations)*.
- Yin, Y. (2008). Robust optimal traffic signal timing. *Transportation Research Part B: Methodological*, 42(10), 911–924.
- Yu, X.-H., & Recker, W. W. (2006). Stochastic adaptive control model for traffic signal systems. *Transportation Research Part C: Emerging Technologies*, 14(4), 263–282.
- Zaidi, A. A., Kulcsár, B., & Wymeersch, H. (2016). Back-pressure traffic signal control with fixed and adaptive routing for urban vehicular networks. *IEEE Transactions on Intelligent Transportation Systems*, 17(8), 2134–2143.
- Zhan, X., Zheng, Y., Yi, X., & Ukkusuri, S. V. (2016). Citywide traffic volume estimation using trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, 29(2), 272–285.
- Zhang, G., & Wang, Y. (2010). Optimizing minimum and maximum green time settings for traffic actuated control at isolated intersections. *IEEE Transactions on Intelligent Transportation Systems*, 12(1), 164–173.
- Zhang, L., Yin, Y., & Lou, Y. (2010). Robust signal timing for arterials under day-to-day demand variations. *Transportation Research Record*, 2192(1), 156–166.
- Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., & Liu, H. X. (2019a). Estimation of queue lengths, probe vehicle penetration rates, and traffic s at signalized intersections using probe vehicle trajectories. *Transportation Research Record*, 2673(11), 660–670.
- Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., & Liu, H. X. (2019b). Various methods for queue length and traffic volume estimation using probe vehicle trajectories. *Transportation Research Part C: Emerging Technologies*, 107, 70–91.
- Zheng, J., & Liu, H. X. (2017). Estimating traffic volumes for signalized intersections using connected vehicle data. *Transportation Research Part C: Emerging Technologies*, 79, 347–362.