Observability quantification of public transportation systems with heterogeneous data sources: An information-space projection approach based on discretized space-time network flow models

Jiangtao Liu
School of Sustainable Engineering and the Built Environment,
Arizona State University, Tempe, AZ, 85281, USA
Email: jliu215@asu.edu

Xuesong Zhou
School of Sustainable Engineering and the Built Environment,
Arizona State University, Tempe, AZ, 85281, USA
Email: xzhou74@asu.edu
Tel.: +1 480 9655827
(Corresponding Author)

Submitted for publication in Transportation Research Part B

Abstract

Focusing on how to quantify system observability in terms of different interested states, this paper proposes a modeling framework to systemically account for the multi-source sensor information in public transportation systems. By developing a system of linear equations and inequalities, an information space is generated based on the available data from heterogeneous sensor sources. Then, a number of projection functions are introduced to match the relation between the unique information space and different system states of interest, such as, the passenger flow/density on the platform or in the vehicle at specific time intervals, the path flow of each origin-destination pair, the earning collected from the tickets to different operation companies etc., in urban rail transit systems as our study object. Their corresponding observability represented by state estimate uncertainties is further quantified by calculating its maximum feasible state range in proposed space-time network flow models. All of proposed models are solved as linear programming models by Dantzig-Wolfe decomposition, which allows us to use parallel computing techniques for large-scale networks. Finally, numerical experiments are conducted to demonstrate our proposed methodology and algorithms.

Keywords

System observability quantification; Information space; Heterogeneous data sources; Public transportation system; Dantzig-Wolfe decomposition;

1. Introduction

The currently rapid innovations and developments of transportation system intelligence in multi-source sensing and information sharing continuously generate huge volumes of various data and information for planners and managers to better observe time-varying traffic conditions and accordingly propose adaptive travel demand management and supply (capacity) control strategies. However, the data sparsity problem still exists, because it is impossible to install fixed sensors on each link or to cover all links by point-to-point moving sensors. As a result, it surely requires new methodologies to recover the system-wide transportation conditions based on the limited observations (Zheng et al., 2014). It leads to a fundamental question; that is, how well the time-dependent transportation system states can be estimated or observed based on currently available heterogeneous source data. In other word, the research motivation is how to quantify the transportation system observability for further optimal control and policy making.

Specifically, as shown in Fig.1, the continuous multi-source sensor data that include fixed sensor data (loop detector, video imaging processing data), mobile sensor data (GPS trajectory data, mobile phone data, AVI, AVL, AFC) and social media with useful travel information, provide a heterogeneous information flow to transportation systems for observing different level of system states (flow, density, travel time), covering from link level, route level, OD pair level to the whole network level, which are usually evaluated by transportation system managers under different goals for further operation, planning and policy making as the fundamental inputs. To reach the desired transportation system performances, it requires to actively manage the demand and supply sides in corresponding level of requests as follows. Travel cost (pricing/incentive), route choice (route guidance, travel reservation), departure time choice (transit station entry flow control, travel reservation), and total demand control (rationing, license plate lottery) are usually considered in the demand management side. On the other hand, the available transportation supply resources are optimized or increased, for example, adaptive signal control, dynamic lane use control, dynamic speed limit control, transit vehicle update and rescheduling, new infrastructure constructions. Meanwhile, the feedback loop between system states and optimal control keeps moving forward along the time horizon with any external new disturbances, such as, weather, incidence, special events, new land use, population changes, etc.

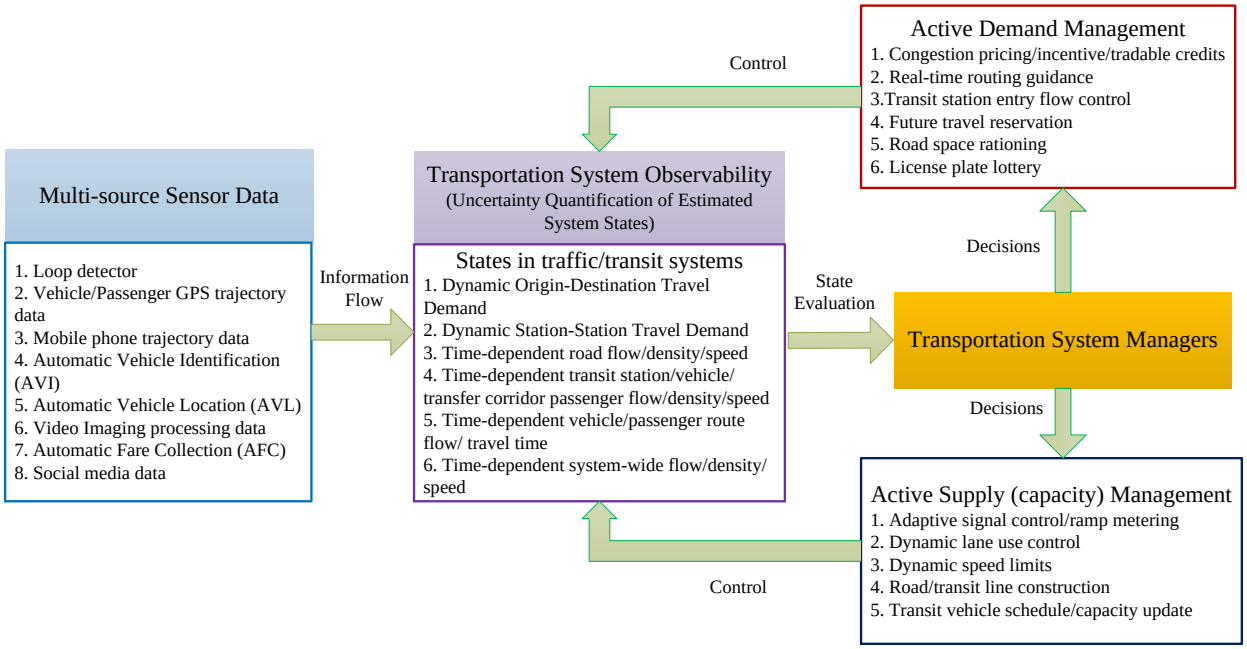


Fig.1 Transportation system management under multi-source sensor data

As pointed out by Daganzo (2007), the most recent computer transportation models can theoretically predict almost anything but not ready in practice, because it is difficult to obtain accurate dynamic origin-destination matrices, capture travelers' unpredictable gaming behaviors, and catch the hypersensitive response of congested networks to any mini inputs. However, the increasingly available multi-source big data and powerful computation capability are creating great opportunities to continuously provide accurate model inputs, capture individually complicated travel behaviors, and efficiently calculate the interested outputs. For example, the AFC data, mobile phone data and travel request mobile phone apps greatly improve the accuracy of observed OD travel demand, and the GPS-enable devices enable us to record each passenger/vehicle's high-resolution trajectory in real time ways. In the meantime, it should

not be neglected that the overwhelming volume of data are also incurring new challenges on data use and transportation modeling.

In the data use side: (i) Is the big data useful enough and what is the value of the data? (ii) Under what goals, one kind of data is more useful than others? (iii) How to fuse multi-source data to keep observation result consistency? Meanwhile, in the transportation modeling side: (i) how to mathematically represent the available multi-source information so that different system states can be estimated in a unified modeling framework; (ii) how to model the exact inner relation between the information and some interested system states, (iii) how to quantify the system observability (uncertainty of estimated states) for further optimal control and management, and (iv) how to design efficient and scalable algorithms for solving those models. Motivated by those general challenges stated above, this paper aims to explore the theoretical relation among sensor data, system states, and system observability in public transportation systems by proposing insightful analysis and theoretically sound linear programming models, especially taking the urban rail transit system as a starting point.

1.1 State estimation and sensor network design in traffic systems

Observability is a concept introduced by Kalman (1959) for linear dynamic systems in control theory. It is a measure for how well internal states of a system can be inferred by knowledge of its external outputs. In other words, it aims to quantify or measure the uncertainty of estimated internal states based on the available external observations under a given sensor environment with sampling errors, sensor error and model errors. A comprehensive literature review can be found at the paper by Castillo et al. (2015). As for evaluating the estimation uncertainty or accuracy, origin-destination (OD) trip matrix estimation is a widely studied classical problem due to its under-determination attribute, which means that there is an infinite number of OD trips that can generate link flows consistent with the observations. Yang et al. (1991) first introduced the concept of Maximum Possible Relative Error (MPRE) to theoretically investigate the estimation uncertainty and reliability of the OD estimated trips obtained by the entropy model. Bianco et al. (2001) further explored the accuracy of estimated OD matrix bound under different sensor location strategies. In addition, Bierlaire (2002) proposed the novel concept of total demand scale as a new measure to examine the quality of estimated OD trip tables from link counts, by maximizing/minimizing the total travel demand satisfying all observations. In the general transportation observability problems, a number of studies (Castillo et al., 2007; Castillo et al., 2008; Gentili and Mirchandani, 2012) modeled the problems as a system of linear equations and/or inequalities and then determine whether the system or one unknown variable is observable or not by analyzing the properties of its coefficient matrix. Meanwhile, in the system of linear inequalities, a general bound of unknown variables can be derived through the dual cone approach. In general, the observability problem more cares about the list of variables to be observed rather than the specific system states uncertainty ranges.

In order to increase the estimation quality, the integration of ubiquitous sensor network design and state estimation has received growing attentions in the past two decades. Yang and Zhou (1998) proposed integer linear programming models and four sensor location rules to determine the optimal number and location of point sensors for origin-destination matrix estimation. Based on the trace of the a posteriori covariance matrix produced in a Kalman filtering model, Zhou and List (2006) offered an information-theoretic framework for locating fixed sensors in the traffic OD demand estimation problem. In addition, based on the observability problem definition, the optimal count number and location of active sensors (Gentili and Mirchandani, 2005) and counting and scanning sensors (Castillo et al., 2012) for estimating path/link flows are studied by analyzing a set of linear equations. Hu et al. (2009) proposed one "basis link" method to find the smallest subset of links in a network to locate sensors so that the traffic flow of all links can be accurately estimated under steady-state conditions. Further, Ng (2012) introduced a new solution approach (''synergistic sensor location'') to avoid possible path enumeration under the assumed steady-state conditions. Xu et al. (2016) proposed a robust network sensor design to completely observe link flows while accounting for the accumulation of observation errors.

For other traffic states, link travel time estimation errors are commonly selected as the optimization criterion for point sensor location problems (Ban et al., 2009; Danczyk and Liu, 2011), and a reliable sensor location method is proposed to consider probabilistic sensor failures (Li and Ouyang, 2011). Herrera et al. (2010) developed a real-time traffic monitoring system by using vehicles with GPS-enable mobile phones and suggested that 2-3% penetration rate of mobile phones can provide accurate traffic speed estimations. Based on a Kalman filtering structure, Xing et al. (2013) developed measurement and uncertainty quantification models to explicitly consider several important sources of errors in path travel time estimation/prediction. In the real-time traffic situations, Eisenman et al. (2006) conducted a sensitivity analysis of estimation and prediction accuracy under different sensor locations and coverage scenarios based on a real-time dynamic simulation system, DYNASMART-X. Boyles and Waller (2011) studied the optimal location selection for providing the real-time traffic information to drivers with the adaptive travel behavior by proposing heuristic algorithms, and Ban et al. (2011) studied the real-time queue length estimation at signalized

intersections by focusing on queuing delay patterns and queue length changes based on travel times from mobile traffic sensors.

Under the framework of urban computing, Zheng et al (2014) reviewed the related date-mining and machine-learning approaches used in transportation system state estimation and prediction. Thiagarajan et al. (2009) applied the WiFi signals to estimate route travel time and identify delay-prone segments by using a hidden Markov model (HMM)-based map matching scheme. Focusing on the GPS trajectory data, the normal traffic patterns are mined for real-time city-wide travel time estimation and prediction by building landmark graphs (Yuan et al., 2011) and for estimating/detecting the traffic anomalies based on the representative terms from twitter-like social media data (Pan et al., 2013). Zhang et al. (2015) proposed a context-aware tensor factorization model to estimate the time spent of each GPS-equipped taxicab on gas stations with consideration of gas price, brand, and weather conditions. Tang et al. (2016) proposed a novel time-dependent graph model to estimate the most likely space-time paths of vehicles with point-to-point data and then implemented a dynamic programming algorithm for the offline and online map-matching applications. Using crowdsourced data from location-based service apps, Zhao and Zhang (2017) examined the individual dynamic choices of activity chains by proposing a data-driven Markov chain approach in activity-travel space-time-state network. Recently, Zhu et al (2018) developed a data-driven link-based network sensor location model to maximize the travel time information gain with accounting for the uncertainty in the prior travel time distribution. Wu et al. (2018) proposed a novel deep-learning-based framework to simultaneously estimate static OD demand and road traffic conditions based on multi-source sensor information by developing a transportation computational graph tool.

1.2 State estimation and smart card use in public transportation systems

In public transportation systems, the automatic fare collection system (AFC) or smart card usually records both the time and station for entry and exit for each passenger in rail transit systems, but only the boarding time and stop and route number normally can be reported in bus transit systems. A comprehensive literature review about smart card data use can be found in the papers (Pelletier et al., 2011; Ma et al., 2013). Obviously, the unknown destination information greatly increase the state uncertainty of bus transit systems. Trépanier et al. (2007) estimate the alighting point for each passenger based on the smallest distance to the boarding stop of his/her next route from individually continuous riding records in smart card. Seaborn et al. (2009) proposed maximum elapsed time thresholds to identify transfers for bus-to-underground, underground-to-bus, and bus-to-bus to identify and assess multi-modal trips in London. Meanwhile, Munizaga and Palma (2012) estimated a multimodal transport OD matrix from smartcard and GPS data whiling consider unobserved trips by expansion factors in Santiago, Chile. Yuan et al. (2013) proposed a space alignment approach by aligning the monetary space and geospatial space with the temporal space to infer each passenger's trajectory and the results improve the detection of uses' home and work places. Nassir et al. (2015) applied the smart card data to detect activity and identify transfers to estimate the true origins and destinations. Nunes et al. (2016) further proposed four endogenous spatial validation rules to enhance the accuracy of estimated passenger destination choice. Alsger et al. (2016) evaluated and improved existing OD estimation method according to available OD information and assessed the previous last destination assumptions in bus transit systems. Under the situation that passenger's boarding stop information is not record in smart cards, Ma et al. (2012) developed a Markov chain based Bayesian decision tree algorithm to estimate the sequential stops on the bus route and then match those stops with the recorded boarding time to infer passengers' origin. Further, Ma et al. (2015) improved their previous algorithms to increase the estimation accuracy and computation efficiency. In addition, in case that buses don't have Automatic Vehicle Location data, Zimmerman et al. (2011) developed a system named Tiramisu that can estimate and predict the real-time bus arrival time by applying the crowd-sourcing data from commuters sharing their GPS-enabled mobile phones.

Depending on the available OD travel information from smart card in urban rail transit systems, a number of studies focus on the route choices and transfer patterns, which can be viewed as different system states required for estimation. Kusakabe et al. (2010) focused on the passengers' train choice behavior by assuming that each passenger aims to minimize the total waiting time at the departure station, loss time at the arrival station, and the transfer frequency. Zhao et al. (2007) chose the logit discrete choice model, but the tight side constraints (e.g. strict vehicle capacity constraint) are still hard to include. Ceapa et al. (2012) mined the regular spatial-temporal trip relations from AFC data to estimate and predict the crowding level for providing more accurate personalized trip planning services. Sun and Xu (2012) estimated the path choice based on the observed overall probability density of journey time and the derived distribution of individual path travel time from the rail transit smart card. In addition, Zhou and Xu (2012) used a matching degree function value to assign the trip to the most likely path based on derived boarding plan of path. Remarked that the verification of those assigned path flows above has not been soundly performed due to the limit of observed data and complicated route choice behaviors, such as passenger's particular travel preference. Kusakabe and

Asakura (2014) proposed a data fusion methodology to consider both the smart card data and person trip survey data by Bayes probabilistic model to estimate behavioral attributes of trips in the smart card data. Based on passenger OD matrix information and vehicle stop time and location data, Zhu et al. (2017a; 2017b) proposed probabilistic models to estimate the individual train loads, left behind probabilities, time-dependent crowding levels at stations under tight vehicle capacity considerations. In addition, Nair et al. (2013) focused on a large-scale bicycle sharing system and analyze the connection between bicycle usage and public transit systems.

1.3 Potential contributions and structure of this paper

USDOT (2015) listed optimizing traffic flow on congested freeways and arterial streets as one of the fundamental urban mobility challenges for Smart City, and pointed out that outdated traffic signal timing causes more than 10 percent of all traffic delay on major routes in urban areas. In the subway system of Beijing, 96 stations have implemented the passenger flow control policy to relieve the system congestion during peak hours in 2018. As explained before, the basic question is how well the system can be observed, and then we can provide the best control to reach our goals, whatever it is in supply side to optimally control signal timing, lane use, speed limit, vehicle rescheduling or in the demand side to influence travelers' departure time choice, route choice and trip generation. Note that most previous studies mainly focused on most likely system state estimation rather than system observability quantification. Hence, this paper aims to develop a modeling framework capable of incorporating multi-source sensor data to address the observability quantification problem in public transportation systems. The contributions of our work are specifically listed as follows.

(i) The information space is generated by a system of linear equations and inequality constraints based on the multi-source sensor data and physical transportation system representation in a unified framework; (ii) Different projection functions are proposed to map the unique information space with our focused different system states (e.g., passenger density on the station platform, in the vehicle, or in the transfer corridor in urban rail transit system) for further system observability quantifications; (iii) Our proposed space-time network flow models are finally solved by Dantzig-Wolfe decomposition, which allows us to use parallel computing techniques for large-scale networks; (iv) The observation errors are also considered by a least square model to correct the directly observed measurements, such as, trip time of grouped passengers from transit smart card, aggregated passenger count from video systems, path choice information from cell phone trajectory data in public transportation systems.

The remainder of this paper is organized in the following manner. The following section conceptually illustrates the general relation between information and system states. Section 3 shows how to construct a space-time network to model the transportation systems, how to generate the information space based on available multi-source information and why Dantzig-Wolfe decomposition is selected to solve our space-time network flow models. The state definitions and observability quantification are provided in Section 4, which also describes the general process of Dantzig-Wolfe algorithms for our specific models. Finally, numerical experiments are performed to demonstrate our proposed methodology in Section 5.

2. Conceptual illustration

Through applying some concepts from game theory and control theory into our problem (LaValle, 2012), a state space is defined as all possible internal system state based on the external physical transportation world, and an **information space** is a place where the internal states live when available information is involved. A **state** is specifically defined and can be associated with the available information. As shown in Fig. 2, the information space is formed by the available information, and the states are tightly connected by different projection functions, which mathematically define the states according to the managers' needs. The bound among all possible states represents the state uncertainty to reflect system observability under current available information.

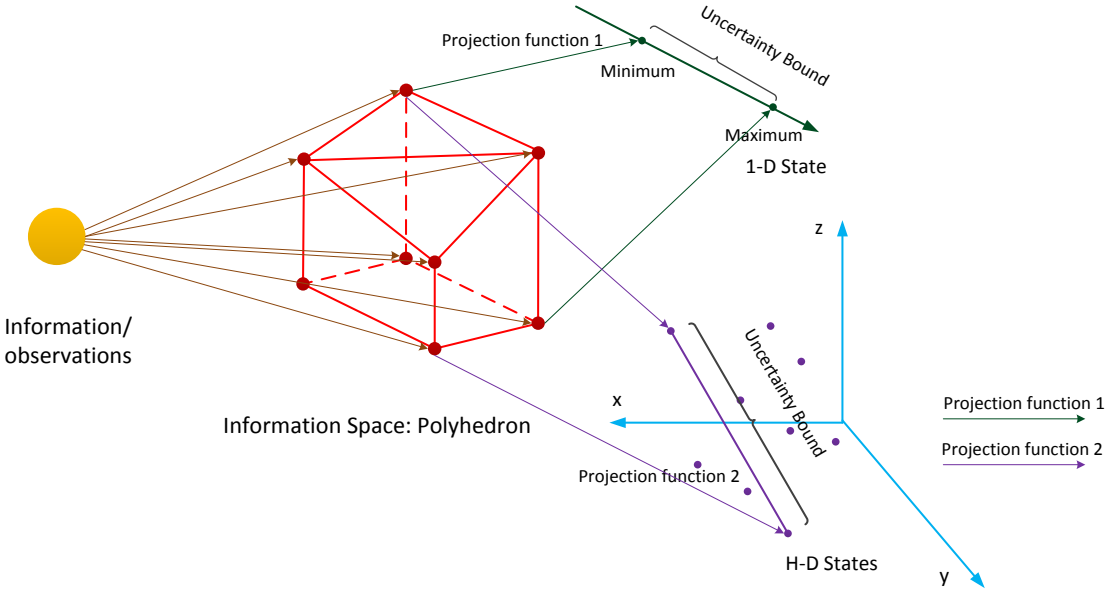
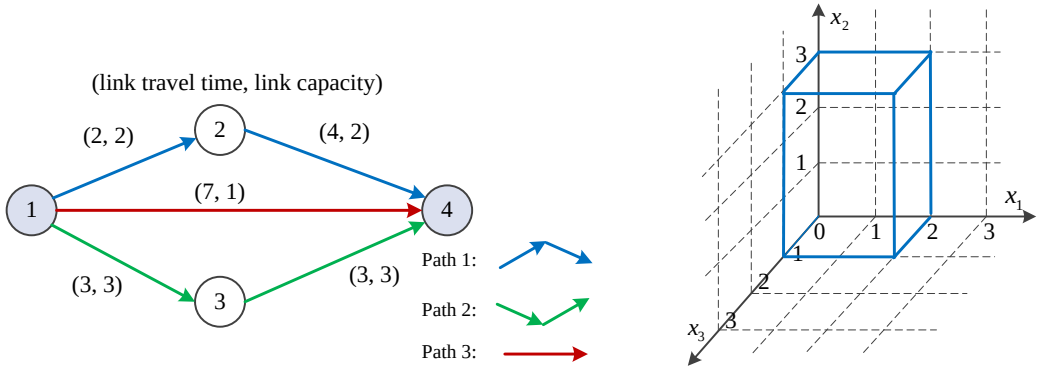


Fig. 2 Relation among information, information space, and flexible states

When the information space is generated as one single point, it can state that the system is observable; otherwise, a non-empty set or space leads to unobservable or partially observable system. In this section, we aim to build the connection between the internal states and the external states (observation or information) by information space as a bridge or communication channel, and further quantify the corresponding uncertainty of states defined by users. How to design sensor network configurations to alter information space and further increase the system observability will be addressed in the future research.

For illustrative purposes, Fig. 3(a) depicts a simple transportation network with four nodes and five links. The link travel time and capacity are also provided as physical network attributes. Let x_1, x_2 and x_3 represent the path flow on paths 1, 2 and 3. Based on the tight capacity constraints, the following relation can be obtained: $0 \le x_1 \le 2$, $0 \le x_2 \le 3$, and $0 \le x_3 \le 1$, which defines the system state space shown as a blue cuboid in Fig. 3(b).



(a) Physical transportation network (b) State space of path flow

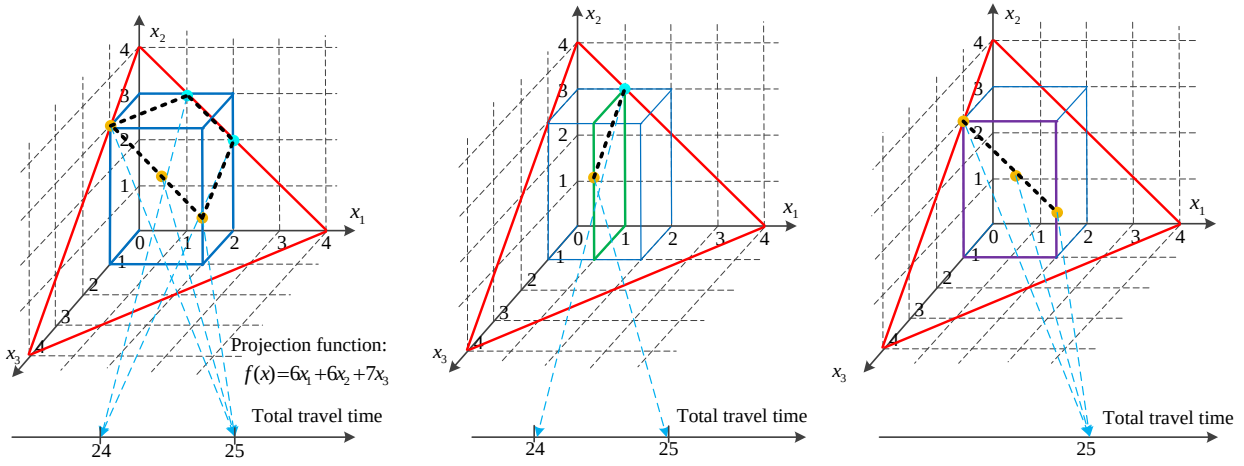
Fig. 3 An illustrative transportation network and its state space

Assume that the OD information is available through survey that there are four vehicles departing from node 1 to node 4. Then, one corresponding constraint will be $x_1 + x_2 + x_3 = 4$. Fig. 4(a) displays the information space as the intersection of the red triangle and the blue cuboid based on the available OD information. Two scenarios are designed as follows to analyze the relation between system states and available information.

Scenario 1: Assume that there is one flow count detector on link $(2, 4)$ and its link count is 1. The relation gets updated as follows: $1 + x_2 + x_3 = 4$, $0 \le x_2 \le 3$, and $0 \le x_3 \le 1$, so the corresponding information space is reduced to be the intersection of the red triangle and the green rectangle shown in Fig. 4(b).

Scenario 2: Suppose that the automatic vehicle identification (AVI) detectors are available at nodes 1 and 4. One vehicle's travel time is observed as 7min. Since only path 3's travel time is 7 and its capacity is just 1, it implies that path flow $x_3 = 1$. As a result, the relation changes as follows: $x_1 + x_2 + 1 = 4$, $0 \le x_1 \le 2$, and $0 \le x_2 \le 3$. The

corresponding information space becomes the intersection of the purple rectangle and the red triangle displayed in Fig. 4(c).



(a) with OD information (b) with OD and loop detector information (c) with OD and bluetooth detector information

Fig. 4 Information spaces and its projected bound under different available information

As shown in Fig. 4, the information spaces are generated as polyhedrons based on different available information. A projection function is defined to map the information space into one dimension state (total travel time). It should be remarked that, the projection functions could be defined based on different goals interested by the system managers, such as, the total system travel time, the number of vehicles in one area, etc. In Fig. 4(a), one projection function is defined as $f(\boldsymbol{x}) = 6x_1 + 6x_2 + 7x_3$, which means that the total system travel time is the analysis goal. Accordingly, different optimization models are solved by maximizing and minimizing $f(\boldsymbol{x})$ subject to different information spaces.

As demonstrated in Fig. 4(a), there are five feasible integer solutions in the information space and the bound of total travel time formed by projection is [24,25]. When the link count data are added in Fig. 4(b), the information space by integer solutions is reduced, but the projected bound is still the same. It indicates that the new information from link count doesn't contribute to reduce the uncertainty of this state estimation (or increase the observability of this state), even though a smaller information space is generated in Fig. 4(b). It reminds us that the value of "big data" is determined by not only "big volume" but also its usefulness of information to the system.

In addition, the point-to-point Bluetooth data (end-to-end passenger id detector data) in Fig. 4(c) makes the projected bound converged to be one unique value, which implies that the point-to-point data is more powerful than the point detector for increasing the observability of system travel time in this case. Therefore, evaluating the values of different information should be based on which states the system manager really cares. One information that seems worthless for one goal may be much useful for other state estimation applications. Moreover, the information space in Fig. 4(c) is larger than that in Fig. 4(b), but the uncertainty of state estimate in Fig. 4(c) is 0 and less than that in Fig. 4(b). Thus, the volume of information space might not be the best criteria to judge the bound of state estimate uncertainty or system observability.

Except for those information from physical sensors above, the previous travel experiences or currently published traffic information from transportation agencies could also take important roles in quantifying observability for managers' further actions. For example, if everyone has perfect information over the network attributes based on his/her experiences and intends to find the best route for his/her trip, which is usually entitled as Wardrop's first principle, the information space will be redefined as, $x_1 + x_2 = 4, 0 \le x_1 \le 2, 0 \le x_2 \le 3$. Compared with the two scenarios above, the information space could be further reduced by this assumed and (potentially questionable) travel behavior. Therefore, one accurate travel behavior can also provide richer information to determine system states, and this could help us understand why a number of studies focus on travel behavior estimation to better understand the system, from the information space perspective.

As a remark, the information space concept in game theory and control theory for storing the amounts of ambiguity in state also may have a connection with Shannon's information theory using entropy-based constructs. For example, the entropy-maximization approach is usually used to represent the most likely traffic state, a kind of stable equilibrium state, for origin-destination travel demand estimation problems. However, in information space theory, the system state is unknown and gradually constructed based on available information. It is possible that the system state derived from the information space is finally the same as the equilibrium state based on maximum entropy.

3 Problem Statement

Table 1 lists the general indices, sets, parameters and variables in our proposed models appeared in Sections 3 and 4.

Table 1 Indices, sets, parameters and variables

Indices	Definition
i, j	Index of nodes, $i, j \in N$
(i, j)	Index of physical link between two adjacent nodes, $(i, j) \in L$
a	Index of passenger group, $a \in A$
$o(a)$	Index of origin node of group a
$d(a)$	Index of destination node of group a
t, s	Index of time intervals in the space-time network
τ	Index of time period for the observed passenger flow
p	Index of paths, $p \in P$
r	Index of transit companies
Sets	
N	Set of nodes in the physical transit network
L	Set of links in the physical transit network
A	Set of passenger groups
V	Set of vertices in the space-time network
E	Set of edges/arcs in the space-time network
G	Set of time period for the observed passenger flows
$S_{p,a}$	Set of paths p of group a
$G(i, j, \tau)$	Set of arcs on observed link (i, j) at time period τ
Parameters	
β_1, β_2	The weights on target passengers' trip time and observed link/arc flows, respectively
μ_a	The observed aggregated trip time of group a from smart card data
$\mu_{i,j,\tau}$	The observed aggregated passenger count on link (i, j) during time period τ
w_p	The travel time of path p
c_p^r	The earning collected on path p of transit company r
$Cap_{i,j,t,s}$	Capacity of traveling arc (i, j, t, s) in the space-time network
DT^a	The departure time of group a
AT^a	The assumed arrival time of group a
D_a	The number of passengers in group a
$c_{i,j,t,s}$	Travel cost of traveling arc (i, j, t, s) in the space-time network
T	The time horizon in the space-time network
$\delta_{(i,j,t,s)}^{p,a}$	Path-link incidence index of route p of group a on arc (i, j, t, s)
w^p	The path travel time of path p
Variables	
$x_{i,j,t,s}^a$	The number of passengers in group a is assigned on traveling/waiting arc (i, j, t, s) in the space-time network
$\theta_a, \theta_{i,j,\tau}$	Continuous positive deviation variables for group a's trip time and link (i, j) during time period τ, respectively
x_a^p	The number of passengers of group a choosing their feasible path p
μ_a^*	The preprocessed aggregated trip time of group a from smart card data
$\mu_{i,j,\tau}^*$	The preprocessed aggregated passenger count on link (i, j) during time period τ

3.1 Space-time Network Construction in Public Transit Systems

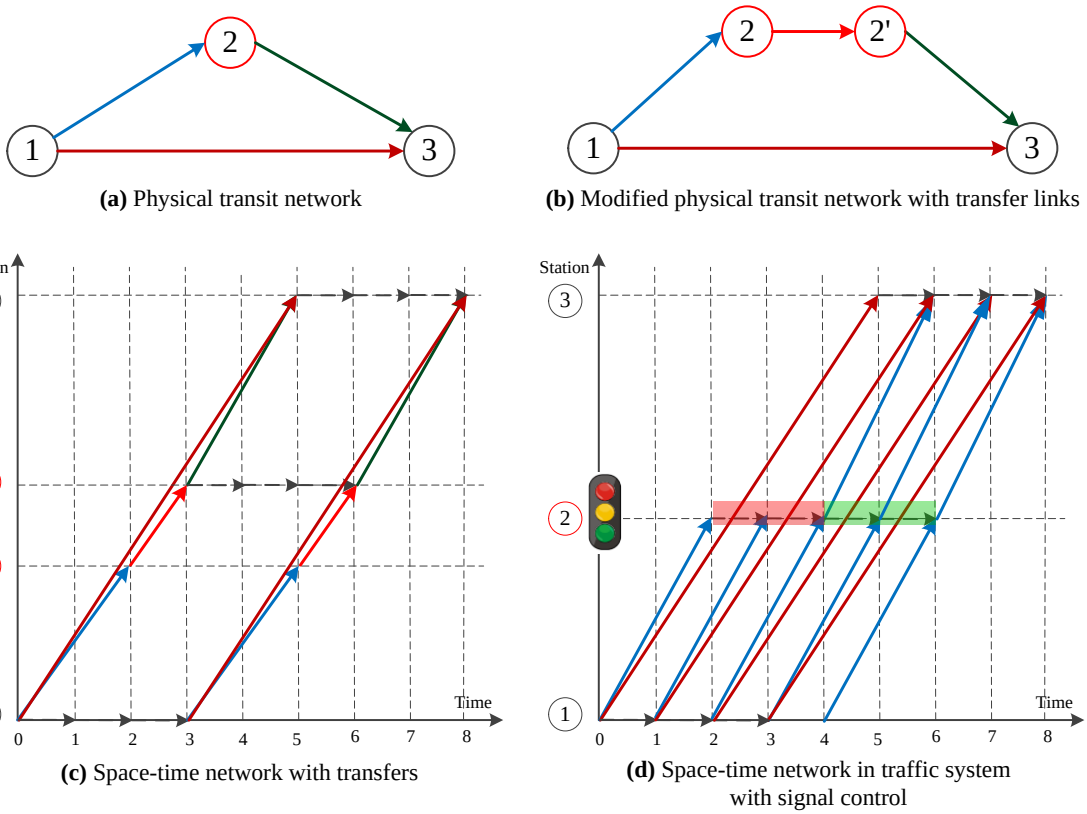
To properly account for the evolution of system dynamics over time, Ford and Fulkerson (1958) first introduced dynamic network flow models to solve the dynamic maximum flow problem in time extended networks. The space-time network flow models are then widely used in dynamic transportation systems, such as, dynamic system optimal with a point queue model (Zawack and Thompson, 1987), dynamic user equilibrium with a spatial queue model (Drissi-Kaitouni and Hameda-Benchekroun, 1992), dynamic system optimal with departure time, route choice and

congestion toll (Yang and Meng, 1997), dynamic user equilibrium with link travel time functions (Chen and Hsueh, 1998), and activity-based dynamic user equilibrium (Lam and Yin, 2001). Recently, in order to maximize network accessibility, Tong et al. (2015) proposed a space-time network flow model with binary decision variables, which actually derives a number of agent-based models in space-time networks later.

There are a number of studies providing how to construct specific time-expanded networks for different transportation systems, such as, freeway network (Lu et al., 2016), road network with signal settings (Li et al., 2016), urban transit network (Liu and Zhou, 2016), bike-sharing network (Lu, 2016), road network with activity requests (Liu et al., 2017), and vehicle trajectory network (Wei et al., 2017). In this section, we consider a physical urban rail transit network with a set of nodes (stops/stations) N and a set of links L as a starting point. Each link can be denoted as a directed link (i, j) from upstream node i to downstream node j. A deterministic transit schedule is supposed to be obtained from Automatic Vehicle Location (AVL) data from vehicle tracking systems. Then, we construct a space-time network, where V is the set of vertices and E is the set of arcs. Node i is extended to a set of vertices (i, t) at each time interval t in the study horizon, $t = 1,2, \dots, T$, where T is the length of the optimization horizon. The transit schedule from node i to node j from time t to time s can be represented by a travelling arc (i, j, t, s) where $(s - t)$ is the exact scheduled/running link travel time and should be integer multipliers of one time interval. The capacity of travelling arcs can be viewed as the transit vehicle's carrying capacity. In addition, a waiting arc is built from (i, t) to $(i, t + 1)$ at node i with waiting time of 1 unit of time and its capacity is defined as the station/platform storage capacity.

In urban rail transit systems, individual passengers should have a trip record with origin, departure time, destination and arrival time from the smart card. However, transit agencies may just provide aggregated trip data for groups of passengers. Each group a with D_a passengers has a departure time DT^a at origin node $o(a)$ to its destination node $d(a)$. At each destination node, there is one assumed large arrival time T for all groups so that the following proposed model will be one-origin-one-destination problem in the space-time network. It should be noted that the travel cost of waiting arcs on the destination node is 0, which means that once the passengers in a group arrive at the destination, the waiting cost to the super-destination (at larger arrival time T) is 0. Finally, the estimated trip time in the model should be equal to the observed trip time, which will be presented in the following sections.

In addition, one transfer node can be divided as multiple nodes, depending on how many transit lines intersect at this node. One illustrative example is shown in Fig. 5(a) where two lines intersect at node 2 and make it as a transfer station. Then node 2 is split to node $2'$ and node 2" and the modified physical network is shown in Fig. 5(b). The travel time of transfer links could be the actual walking time, and its capacity is the maximum passenger throughput at transfer corridors. As a remark, based on the maximum transfer distance accepted by passengers, it is possible to connect different stops by transfer links or extended to multimodal networks. Fig. 5(c) shows the transfer process where all transfer time is assumed to be 1 unit of time. In addition, it is feasible to consider the uncertainty of walking time on transfer links or from station entry to the platform through constructing more service/travelling arcs with different arc travel times. Furthermore, in traffic networks, the road can provide its service at each time interval with a specific arc capacity and the signal timing rules whether those service arcs are open or closed, which is represented in a space-time network in Fig. 5(d) to show our unified modeling framework.



(a) Physical transit network

(b) Modified physical transit network with transfer links

(c) Space-time network with transfers

(d) Space-time network in traffic system with signal control

Fig. 5 space-time network construction

3.2 Information space generation based on multi-source sensor data

Information space (LaValle, 2012) works as a communication channel to connect the external physical world and the internal system states. The external physical world is sensed by heterogeneous sensors in terms of different observations or information, which finally forms a corresponding information space. Meanwhile, the internal system states are reflected through the information space based on the specific state definitions.

In addition to the physical transit lines and schedules as useful information, the possible observations in the urban transit systems are summarized in Table 2.

Table 2 Available trip information in urban transit systems

Bus transit system	Rail transit system
(1) origin, boarding time, destination, alighting time for each person in individual bus	(1) origin, entry time, destination, exit time for each person or aggregated for each passenger group in transit network
(2) origin and boarding time for each person in individual bus	(2) origin and boarding time for each person or aggregated for each passenger group in transit network
(3) historical time-dependent OD information for transit networks	

In bus transit systems, (i) when the origin, boarding time, destination, alighting time for each person in individual buses are available from smart card data and the transit schedule or transit vehicle trajectory is available from AVL data, the state in each vehicle can be completely observed . Those accurate trip information can provide great values for the operational transit planning. (ii) If only the origin, and boarding time can be recorded, algorithms are needed to estimate the individual destination. Usually it is estimated as the nearest stop to the boarding stop of traveler's next route based on the continuous riding records, so the corresponding alighting time will be also available. However, if travelers just have a single transit trip, it is still difficult to estimate the destination, which causes a large uncertainty for state estimation. (iii) If the market penetration rate of smart card is very low or the goal is for operational transit planning, the historical time-dependent OD information has to be used to perform transit network assignment with assumed travel behaviors (Szeto and Jiang, 2014; Jiang and Szeto, 2016; Cats et al., 2016; Liu and Zhou, 2016; Codina

et al., 2017), which could create a larger uncertainty in the system and needs to be carefully calibrated and validated by real-world survey and observations.

In rail transit systems, (i) the origin, entry time, destination, and exit time usually are available for each passenger or group, but the path/vehicle/transfer selection in the network level still has a large uncertainty. (ii) When the destination and exit time to stations are not recorded, the system uncertainty will be increased more.

In addition, with the development of sensing technologies, more available sensor information from big data applications can be used in the public transportation systems.

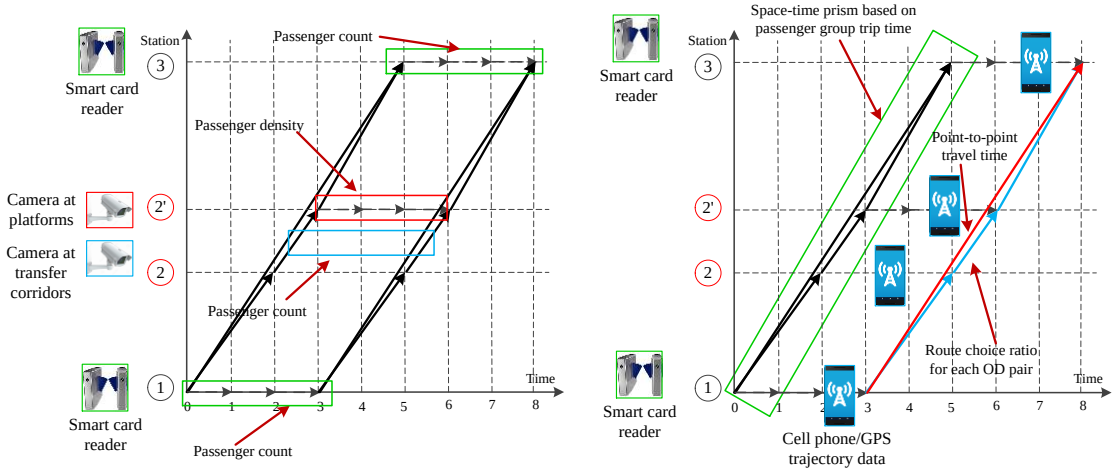
(i) Video data processed to gain the aggregated passenger flow at key points during specific time periods, such as, transfer corridors, the entry and exit of stations, or the stop/platforms.

(ii) Cellphone/GPS based point-to-point trajectory data. A general path choice ratio bound is available when the penetration/sample rate of cell phone used as sensors is large enough. The granularity of the trajectory points is highly dependent on the cell tower locations. Also, Bluetooth data can provide a point-to-point travel time and general path choice ratio.

(iii) General travel behavior data (e.g. preference) through survey. It can provide the path choice of some specific passengers, so the path choice uncertainty of all passengers can be reduced, to some extent.

(iv) Social media: it can share the location information, users' interests, and events happening around them, which could be used to mine the human mobility pattern and predict the travel demand related to those events.

Taking the transit network in subsection 3.1 as an example, the possible observations are illustrated in Fig. 6(a) and 6(b) for fixed sensors and mobile sensors, respectively.



(a) Observations from fixed sensors

(b) Observations from mobile sensors

Fig. 6 Observations from multi-source sensors

The specific modeling on generating information space based on those available sensor data is developed as follows. It should be remarked that, the first-in-first-out (FIFO) rule is not incorporated in our proposed space-time network, because it can be violated in transit networks and the multi-source information can better present travelers' decision by reducing the feasible information space. Also, readers who are interested in FIFO in space-time networks can refer to the details in the paper (Shang et al., 2018).

Taking the rail transit system as our modeling example, we have the following formulation.

(i) According to the physical network, transit schedule, and dynamic OD information from smart card data, the standard flow balance constraint can be given as follows:

$$\sum_{i,t:(i,j,t,s)\in E} x^a_{i,j,t,s} - \sum_{i,t:(j,i,s,t)\in E} x^a_{j,i,s,t} = \begin{cases} -D_a & \forall a, j = o(a), s = DT^a \\ D_a & \forall a, j = d(a), s = T \\ 0 & otherwise \end{cases} \quad (1)$$

(ii) Strict vehicle and station platform capacity constraint

$$\sum_a x^a_{i,j,t,s} \le Cap_{i,j,t,s}, \ \forall (i,j,t,s) \in A \quad (2)$$

(iii) As stated in subsection 3.1, the estimated trip time of each group in the model should be consistent with the observation (average trip time of each group) from smart card.

$$\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) = D_a \times \mu_a, \forall\, a \quad (3)$$

(iv) Estimated aggregated passenger flow count on link (i,j) during time period τ is expected to be the observation from video data or counting by person.

$$\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} = \mu_{i,j,\tau}, \forall(i,j,\tau) \quad (4)$$

(v) Non-negative arc flow variables

$$x^a_{i,j,t,s} \geq 0 \quad (5)$$

Note that if a passenger is viewed as a group of passengers and $x^a_{i,j,t,s}$ will be a binary variable, the above modeling is still available and we will have an agent-based model. Usually, agent-based trajectory data can provide more point-to-point travel time information rather than just path choice. The above formulation presents an arc-based formulation for constructing the information space. In comparison, a path-based formulation based on the feasible path generation is offered as follows:

(i) Flow balance constraint:

$$\sum_p x^p_a = D_a, \forall a \quad (6)$$

(ii) Capacity constraint:

$$\sum_{(p,a)\in S_{(p,a)}} (\delta^{p,a}_{(i,j,t,s)} \times x^p_a) \leq Cap_{i,j,t,s}, \forall(i,j,t,s) \in A \quad (7)$$

(iii) Trip time constraint:

$$\sum_p x^p_a * w^p = D_a \times \mu_a, \forall a \quad (8)$$

(iv) Aggregated passenger flow count constraint:

$$\sum_a \sum_{t\in\tau} (\delta^{p,a}_{(i,j,t,s)} \times x^p_a) = \mu_{i,j,\tau}, \ \forall(i,j,\tau) \quad (9)$$

(v) Non-negative path flow:

$$x^p_a \geq 0 \quad (10)$$

Given the time-expanded space-time network constructed in subsection 3.1, the general feasible path set for a passenger group with specific OD pair and a departure time can be generated by a forward label correcting algorithm from the vertex (origin and departure time) to its destination node based on the observed trip time of that group as a prism.

It should be remarked that when considering the bus transit systems, the smart card data usually only have the origin and departure time without passengers' destination and arrival time information. To model this condition that dynamic OD trips are unknown, D_a will be a variable in equation (1) and the summation of D_a with same origin and departure time should be equal to the recorded total trip generation at this origin and departure time from smart card data. If the structure of each OD pair with departure time is given based on the historical OD information, the number of unknown OD variables will be greatly reduced. In addition, from the data mining perspective, interesting readers can be referred to the paper (Ma et al., 2013) to find studies related to estimating destination probabilities.

3.3 Dantzig-Wolfe decomposition for special flow-balance blocks

As shown in the arc-based and path-based formulation in section 3.2, the flow balance constraint is a special block that can be solved by classical shortest path algorithms and further be incorporated by Dantzig–Wolfe decomposition. Actually, this kind of methods had been adopted for static traffic assignment (Larsson and Patriksson, 1992), side constrained traffic equilibrium (Larrson et al., 2004), time constrained shortest path problem (Desrosiers and Lubbecke, 2005), etc. The advantage of this decomposition allows us to solve the special blocks in parallel via independent computation threads to address large-scale networks, especially when the computer hardware has a rapid development in current days. It also has the re-optimization capability if the travel demand, arc performance function or network topology has any changes in future (Larsson and Patriksson, 1992).

Specifically, Dantzig–Wolfe decomposition is originally proposed by Dantzig and Wolfe (1960) for solving linear programming problems with special structure. A general primal linear program can be represented as: min $c^T\boldsymbol{x}$, subject to, $A\boldsymbol{x} \leq \boldsymbol{b}$, $D\boldsymbol{x} \leq \boldsymbol{d}$, and $\boldsymbol{x} \geq 0$. According to Minkowski-Weyl's Theorem, given the convex set $X = \{x \in \mathbb{R}^n | Ax \leq b\}$ where $Ax \leq b$ is a **special block**, X can be represented by the extreme points and extreme rays of X: $X = \{x = \sum_i \lambda_i x^i + \sum_j \mu_i y^j | \sum_i \lambda_i = 1, \lambda_i \geq 0, \mu_j \geq 0\}$. When X is a bounded polyhedron, X can be represented by the extreme points, $X = \{x = \sum_i \lambda_i x^i | \sum_i \lambda_i = 1\,, \lambda_i \geq 0\}$.

Substituting the expression above to the original model leads to the following Master Problem:

$$\min \sum_i c^T \lambda_i \boldsymbol{x}^i \quad (11)$$

$$\text{Subject to, } \sum_i D\lambda_i \boldsymbol{x}^i \leq \boldsymbol{d}, \sum_i \lambda_i = 1 \text{ and } \lambda_i \geq 0 \quad (12)$$

Suppose that a subset of extreme points P is available. The Restricted Master Problem (**RMP**) can be obtained by min $\sum_{i\in P} c^T \lambda_i \boldsymbol{x}^i$, subject to, $\sum_{i\in P} D\lambda_i \boldsymbol{x}^i \leq \boldsymbol{d}$, $\sum_{i\in P} \lambda_i = 1$ and $\lambda_{i\in P} \geq 0$. Assume that λ^* and (π, ω) is the optimal and dual solutions to the RMP, respectively. The reduced cost is defined as $\gamma(\boldsymbol{x}) = c^T\boldsymbol{x} - \pi^T A\boldsymbol{x} - \omega$. Then, we solve the subproblem: min $c^T\boldsymbol{x} - \pi^T A\boldsymbol{x} - \omega$, subject to $A\boldsymbol{x} \leq b$ and $\boldsymbol{x} \geq 0$. If the reduced cost is non-negative, the solution is optimal; otherwise, the solution can be viewed as a new extreme point and added to the RMP until the reduced cost is non-negative.

With different objective functions related to different estimated states, our proposed models in section 4.1 based on the generated information space in Section 3.2, will be solved under the framework of Dantzig–Wolfe decomposition in section 4.2. Specifically, based on the flow-balance constraint, the flow on a particular path (or path flow for a passenger group a) can represent one extreme point. A path flow uniquely corresponds to its path, so a particular path implicitly indicates a specific extreme point. This enables us to express the arc flow of group a on arc (i,j,t,s) as $x^a_{i,j,t,s} = \sum_h (\delta^{p(h),a}_{(i,j,t,s)} \times x^{p(h)}_a \times \lambda_{(a,h)})$, where $x^{p(h)}_a = D_a$ for each generated extreme point h, and $\sum_{h \epsilon H(a)} \lambda_{(a,h)} = 1$. Since variable $x^a_{i,j,t,s}$ is continuous rather than discrete, it should be a continuous combination of extreme points and $\lambda_{(h,a)} \geq 0$ according to Minkowski-Weyl's Theorem. On the other hand, the link flow vector of each group $\boldsymbol{x}^a_{i,j,t,s}$ can also be seen as one extreme point, as it is the result of one specific path flow vector.

4. Observability quantification of different states under heterogeneous data sources

The system observability reflected by state uncertainty mainly arises from two sources: one is lack of useful information, which results in the many-to-one mapping between the many possible system states and one partial observation, and the other is the possible measurement error due to the noise and disturbance in sensing systems. This section will focus on quantifying the uncertainty of state estimates based on available limited useful observations. How to address the measurement error issues will also be discussed later.

4.1 Projection-function-based observability quantification

As illustrated in Section 2, the generated information space can work as a channel to connect available observations with different states. This section will propose different projection functions as the mapping between the feasible information space and specific transit system states. Table 3 introduces our focused states, which are also displayed in Fig. 7.

Table 3 Focused states and motivations

Focused states	Motivations
(1) **Arc flow/density state**: passenger density on station platforms, in vehicles, and transfer corridors	(i) identify possible dangerous spots for safety; (ii) make decisions on vehicle updates, line/timetable changes and stop location adjustment
(2) **Path flow state**: the number of passengers taking one specific line segment	(i) clear the total ticket fare to each company based on the service they provide; (ii) evaluate the current liquidation policy and quantify the unreasonable income bound for each company
(3) **Path flow state**: the path flow range of each time-dependent OD pair	(i) compare or verify the traditional logit route choice model for better understanding travel behavior
(4) **Network-level arc flow/density state**: network-level time-dependent passenger flow/density states on serval key stations/vehicles	(i) distribute the network-level transit condition and intelligent passenger trip guidance (ii) evaluate network-level control and policy

The focused states and its uncertainty quantification are listed as follows.

Projection function 1 for arc flow state: the number of passengers on one specific arc (i,j,t,s) (station platform, vehicle, transfer corridor) in the space-time network is represented as $\sum_a x^a_{i,j,t,s}$, so the arc flow uncertainty can be quantified by maximizing and minimizing $\sum_a x^a_{i,j,t,s}$, subject to constraints (1) to (5).

Projection function 2 for path flow state 1: the earnings that one transit company r can obtain is represented as $\sum_a \sum_p (x^p_a \times c^r_p)$, where c^r_p is the income of using the segment in company r's operation area of path p. It can be calculated as a parameter in advance based on the ticket price and segment and path distance. Therefore, the earning bound is estimated by maximizing and minimizing$\sum_a \sum_p (x^p_a \times c^r_p)$ subject to constraints (6) to (10).

Projection function 3 for path flow state 2: the flow rate on path p is $\sum_a x^p_a$, so the uncertainty bound of path flow is measured by maximizing and minimizing $\sum_a x^p_a$, subject to constraints (6) to (10).

Projection function 4 for network-level arc flow state: the passenger flow (density) states on key station platforms at one time index (e.g., at 7:30am) is a high-dimensional vector $\{\boldsymbol{q}(\boldsymbol{i},\boldsymbol{t})\}$ where i is one of the key stations. For one specific station i, $q(i,t) = \sum_a x^a_{i,j,t,s}$ is the number of passengers at station i at time t. Since the state is not one dimension anymore, the concept of the Maximal Possible Relative Error (MPRE) first introduced by Yang et al. (1991) is adopted to quantify the state uncertainty of high-dimensional variables. As shown in Fig. 8(a), the state solution (vector $\{\boldsymbol{x}^{\boldsymbol{a}}_{\boldsymbol{i,j,t,s}}\}$) based on different projection functions for one-dimensional state above is one feasible solution in the information space, so each solution (vector $\{\boldsymbol{x}^{\boldsymbol{a}}_{\boldsymbol{i,j,t,s}}\}$) can be mapped to high-dimensional states to

generate new state points (vector $\boldsymbol{q}(\boldsymbol{i},\boldsymbol{t})$) illustrated in Fig. 8(b), which are used as sample points to approximately obtain the MPRE. Specifically, we need to calculate the average relative error between any two points, and find the maximal one as the MPRE.

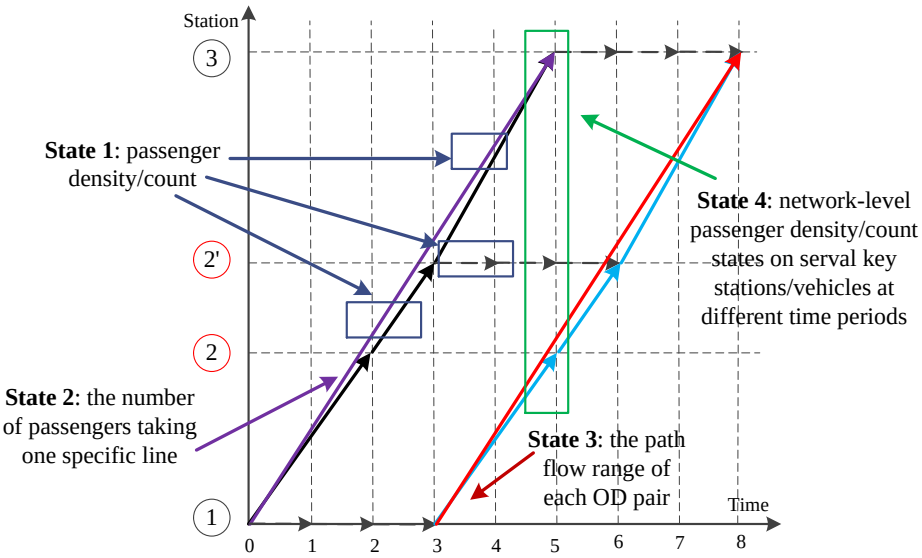
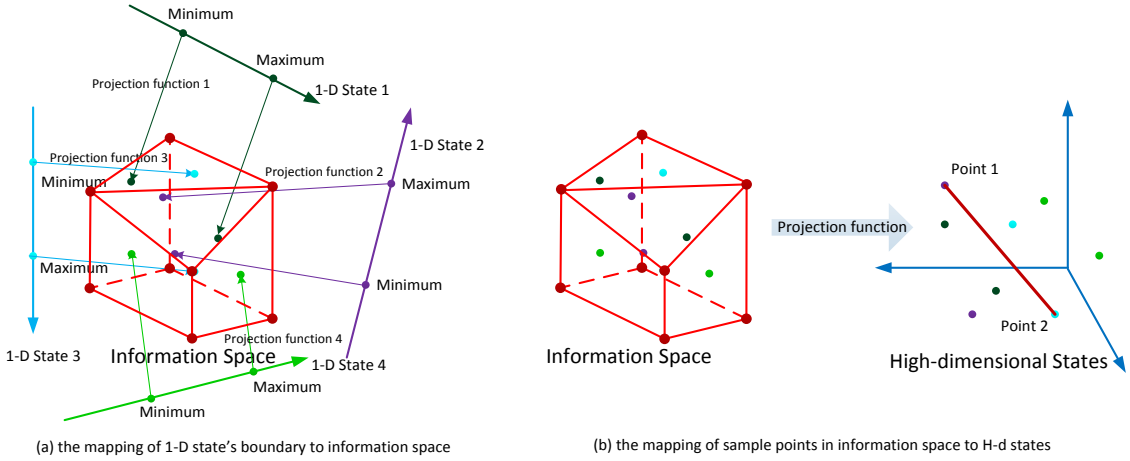


Fig. 7 States illustration in a space-time network



(a) the mapping of 1-D state's boundary to information space

(b) the mapping of sample points in information space to H-d states

Fig. 8 Relation of information space and different types of states

For example, the average relative error between point 1 and point 2 is calculated as follows (Yang et al., 1991), where $\boldsymbol{q_1}(\boldsymbol{i},\boldsymbol{t})$ and $\boldsymbol{q_2}(\boldsymbol{i},\boldsymbol{t})$ are a m-dimensional vector recording m stations' passenger flow at time t. The relative deviation is $\lambda_{(1,2,i,t)} = \frac{q_1(i,t)-q_2(i,t)}{q_1(i,t)}$ and the average relative deviation $AV(\lambda_{(1,2,t)}) = \sqrt{\frac{\phi(\lambda_{(1,2,t)})}{m}}$, where$\phi(\lambda_{(1,2,t)}) = \sum_{i=1}^{m} \lambda_{(1,2,i,t)}{}^2$and $\lambda_{(1,2,t)} = \{\lambda_{(1,2,1,t)}, \lambda_{(1,2,2,t)}, \dots, \lambda_{(1,2,m,t)}\}$. In addition, Yang et al. (1991) defined the concept of Estimation Reliability as a measure about the state uncertainty; that is, $Re = \frac{1}{1+AV(\lambda)}$, which shows that when the $AV(\lambda)$ is 0, the reliability of the estimated state is 1. In contrast, when $AV(\lambda)$ tends to infinity, there is almost no reliability guarantee. This result is just based on some sample points, so it is still an approximation approach.

4.2 The solution procedure of Dantzig-Wolfe decomposition

From the perspective of Dantzig-Wolfe decomposition, based on the master problem in formulas (11) and (12), $\boldsymbol{x}^i$ as one extreme point can be replaced by variable vector $\boldsymbol{x}^{\boldsymbol{a}}_{\boldsymbol{i},\boldsymbol{j},\boldsymbol{t},\boldsymbol{s}}$ and variable $\boldsymbol{x}^{\boldsymbol{p}}_{\boldsymbol{a}}$ for arc-based and path-based models above, respectively, and c^T is the corresponding cost on each variable.

Taking minimizing the flow on arc (i,j,t,s) as an example, the general procedure of the algorithm is shown in Fig. 9 and described as follows.

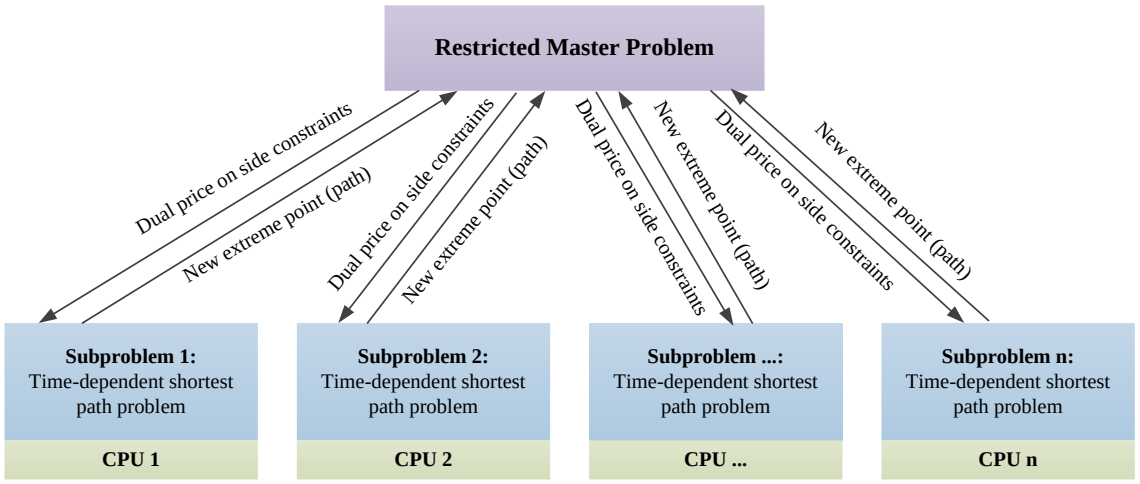


Fig. 9 The procedure of Dantzig-Wolfe decomposition algorithm

Step 1: Initialization. Find one feasible passenger arc flow vector $\{\boldsymbol{x}_{i,j,t,s}^{a,k}\}$ on the shortest path as the k^{th} extreme point for each passenger group a. It indicates that $(k-1)$ extreme points have been generated before finding one feasible solution, which will be explained after Step 3 as a remark.

Step 2: Solve the restricted master problem to obtain the duals of side constraints.

$$\text{Min } \sum_k \sum_a (x_{i,j,t,s}^{a,k} \times \lambda_{a,k}) \tag{13}$$

Subject to,

$$\sum_k \sum_a (x_{i,j,t,s}^{a,k} \times \lambda_{a,1}) \le Cap_{i,j,t,s} \quad \forall (i,j,t,s) \in A \tag{14}$$

$$\sum_k \sum_{(i,j,t,s)} [(x_{i,j,t,s}^{a,k} \times \lambda_{a,1}) \times c_{i,j,t,s}] = D_a \times \mu_a, \forall\, a \tag{15}$$

$$\sum_k \sum_a \sum_{t \in \tau} (x_{i,j,t,s}^{a,k} \times \lambda_{a,1}) = \mu_{i,j,\tau}, \forall (i,j,\tau) \tag{16}$$

$$\sum_k \lambda_{a,k} = 1, \forall\, a \tag{17}$$

$$\lambda_{a,k} \ge 0 \tag{18}$$

$\pi_{i,j,t,s}$, π_a, $\pi_{i,j,\tau}$ and ω_a are the duals of side constraints (14)-(17), respectively.

Step 3: Solve each sub-problem as a time-dependent shortest path problem to calculate its reduced cost for each passenger group, which can be implemented by parallel computing techniques, such as, Multi Process Interface(MPI) (Liu and Zhou, 2016).

Sub-problem for each passenger group a:

$$\text{Min } \left(c_{i,j,t,s}^{a} \times x_{i,j,t,s}^{a,k+1}\right) - \sum_{(i,j,t,s)} (\pi_{i,j,t,s} \times x_{i,j,t,s}^{a,k+1}) - \pi_a \times \sum_{(i,j,t,s)} \left(c_{i,j,t,s} \times x_{i,j,t,s}^{a,k+1}\right) - \sum_{(i,j,\tau)} (\pi_{i,j,\tau} \times \sum_{t \in \tau} x_{i,j,t,s}^{a,k+1}) - \omega_a \tag{19}$$

Subject to

$$\sum_{i,t:(i,j,t,s) \in E} x_{i,j,t,s}^{a,k+1} - \sum_{i,t:(j,i,s,t) \in E} x_{j,i,s,t}^{a,k+1} = \begin{cases} -D_a & \forall a, j = o(a), s = DT^a \\ D_a & \forall a, j = d(a), s = T \\ 0 & otherwise \end{cases} \tag{20}$$

Actually, the reduced cost is $\left(c_{i,j,t,s}^{a} \times x_{i,j,t,s}^{a,k+1}\right) - \sum_{(i,j,t,s)} (\pi_{i,j,t,s} \times x_{i,j,t,s}^{a,k+1}) - \pi_a \times \sum_{(i,j,t,s)} \left(c_{i,j,t,s} \times x_{i,j,t,s}^{a,k+1}\right) - \sum_{(i,j,\tau)} (\pi_{i,j,\tau} \times \sum_{t \in \tau} x_{i,j,t,s}^{a,k+1}) - \omega_a$. If it is negative, we add the solution of the sub-problem to the restricted master problem at step 2 and begin next iteration. When the reduced costs of all sub-problems are non-negative, the optimal solution is achieved.

In the initialization step, in order to find one feasible solution from the shortest path problem as initial extreme points, we can introduce artificial variables for those coupling constraints and solve the problem by the Dantzig-Wolfe decomposition again (Kalvelagen, 2003). For example, corresponding the example in section 3.3, the coupling side constraint is $\sum_j D_{i,j} x_i \le d_i$, so we can add artificial variable $y_i \ge 0$ to have $\sum_j D_{i,j} x_i - y_i \le d_i$, and minimize $\sum_i y_i$ as a master problem. Based on the Dantzig-Wolfe decomposition algorithm, when the $\sum_i y_i$ is equal to 0, we can conclude that one feasible solution for our primal problem is obtained and can be used for step 2.

In addition, for the maximum problem, we can transform it as a minimum problem by changing the positive arc costs to be negative. Since there is no circle in the space-time network, the label correcting algorithm can always be used to find the shortest path.

4.3 Discussions

4.3.1 Preprocessing on available measurements

The analyses on small card data (Trépanier et al., 2007; Barry et al., 2009) show that the data must be thoroughly validated and corrected prior to the practical use. Therefore, it might happen that no feasible solution exists when the observed data are directly used in built models. Even though each observation is tested in the model and can provide feasible solutions, it is still possible to have infeasible solutions when different measurement/observations are considered simultaneously, because the inconsistency among different kinds of sensors may still exist. Hence, we need to preprocess the data to obtain estimated measurements which are as close as possible to the corresponding measurements under real-world physical constraints. There are different approaches to clean and verify those measurements. The approach adopted here is to minimize the generalized least squares between the observed and corrected measurements, subject to constraints (1), (2) and (5). The model is explained at Appendix A in detail. Therefore, the measurement values μ_a and $\mu_{i,j,\tau}$ in information space generation and uncertainty quantification need to be replaced by μ_a^* and $\mu_{i,j,\tau}^*$ from our proposed prepressing model.

The model is a linearly constrained quadratic programming model. The Frank-Wolfe algorithm is usually used to solve the optimization problem where the objective function is convex differentiable real-valued function and the feasible region of side constraints is compact convex (Frank and Wofle, 1956). Therefore, the model at Appendix A can be well solved under the framework of the Frank-Wolfe algorithm, which is explained at Appendix B in detail.

4.3.2 Real-time state uncertainty quantification

The uncertainty of real-time system state increases the difficulty of real-time state prediction and optimal control. Compared with the offline state observability in this paper, the challenges in the real-time condition include (i) the real-time rail transit OD travel information is not available and (ii) the state transition along the time is highly required.

(i) Real-time OD demand estimation: Based on day-to-day historical and accurate dynamic OD demands in urban rail transit systems, we can classify k representatives $OD_{o,d,\tau}^k$ for each OD pair at different time periods, so the estimated real-time OD demand is $OD_{o,d,\tau} = \sum_k (w_k \times OD_{o,d,\tau}^k)$, where w_k is a binary variable, which indicates that only one OD candidate k will be chosen. As a result, the dynamic OD travel demand's spatial structure can be well captured, compared with those OD estimation models which mainly optimize one departure time profile for all or one-class total static OD trips. In addition, the real-time trip generation at each station/origin with departure time is available from the smart card data, so $\sum_d OD_{o,d,\tau} = OD_{o,\tau}^{obs}$ provides more information to generate the real-time information space.

(ii) Real-time state transition: the rolling horizon approach has been widely chosen for real-time transportation operations and control (Peeta and Mahmassani, 1995; Zhou and Mahmassani, 2007; Meng and Zhou, 2011). Under this mechanism, when focusing on one time period, it needs a look-back period and a look-ahead period, because the generated passengers from the look-back period could still be in the network during our focused time period, and in the look-ahead period all passengers can arrive at their destination for our network modeling. Along the planning time horizon, once some trips are finished at our focused time period, their true OD information can be obtained in real time, so the corresponding estimated OD trips can be replaced by the real ones, which can also reduce the information space for our state observability quantification.

Similar to the offline modelling, the states can be flexibly defined based on the managers' analysis goals. The min/max models on one-dimensional state and the MPRE for multi-dimensional states are also available for quantifying the real-time state observability, which provides a fundamental input for the measure of future real-time prediction and optimal control.

5. Experiments

5.1 Tests in a hypothetic network

This section will demonstrate the proposed models and algorithms in Sections 4 and implement them in a general purpose optimization package GAMS. All source codes can be downloaded at the website: https://www.researchgate.net/publication/326020738_Observability_Scenarios_1-4.The experiments are performed

in the following transit network shown in Fig. 10(a), where seven urban rail lines exist in the transit systems. In order to model the passenger count observation at transfer corridors, specific transfer links are built as shown in Fig. 10(b).

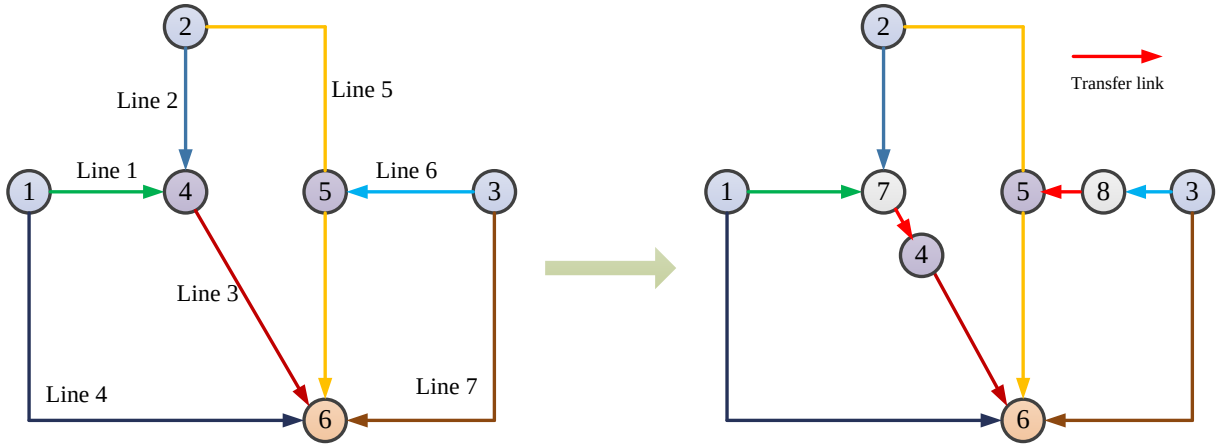


Fig. 10 Hypothetic urban rail transit network

5.1.1 Given multi-source sensor data

(1) Table 4 lists the existing transit service arcs based on the given timetable of the seven transit lines; the corresponding space-time network is constructed in Fig. 11.

(2) The origin, destination, departure time and aggregated trip time of each passenger group are listed in Table 5, and each group represents 100 passengers in this test.

(3) The vehicle capacity of each line is assumed in

Group No	OD Pair	Departure Time	Average Trip Time	Group No	OD Pair	Departure Time	Average Trip Time
1	1 → 6	0	6	15	1 → 6	3	7.5
2	1 → 6	0	7	16	1 → 6	3	7
3	1 → 6	0	8	17	1 → 6	3	8
4	1 → 6	0	6.5	18	2 → 6	3	6
5	2 → 6	0	7	19	2 → 6	3	7
6	2 → 6	0	7.5	20	2 → 6	3	6.5
7	2 → 6	0	6.5	21	2 → 6	3	7.5
8	2 → 6	0	6	22	2 → 6	3	8
9	3 → 6	0	7	23	2 → 6	3	6.8
10	3 → 6	0	7.5	24	3 → 6	3	7
11	3 → 6	0	8	25	3 → 6	3	7.5
12	1 → 6	3	6	26	3 → 6	3	7.4
13	1 → 6	3	7	27	3 → 6	3	7.8
14	1 → 6	3	6.5	28	3 → 6	3	8

Table 6, where it can be observed that the capacity of rail transit vehicles could have its adjustment at different time periods by increasing or decreasing the number of train units.

(4) The passenger count data from video processed data at transfer corridor $(7, 4)$ is available; that is, 450 and 810 passengers are observed at time points 3 and 6.

Table 4 Hypothetic transit service arcs lists

Service Arc	Start Time	End Time	Service Arc	Start Time	End Time
(1,7)	0	3	(1,7)	3	6
(7,4)	3	4	(7,4)	6	7
(4,6)	4	6	(4,6)	7	9
(1,6)	0	8	(1,6)	3	11

(2,7)	0	3	(2,7)	3	6
(2,5)	0	4	(2,5)	3	7
(5,6)	4	7	(5,6)	7	10
(3,8)	0	3	(3,8)	3	6
(8,5)	3	4	(8,5)	6	7
(3,6)	0	8	(3,6)	3	11

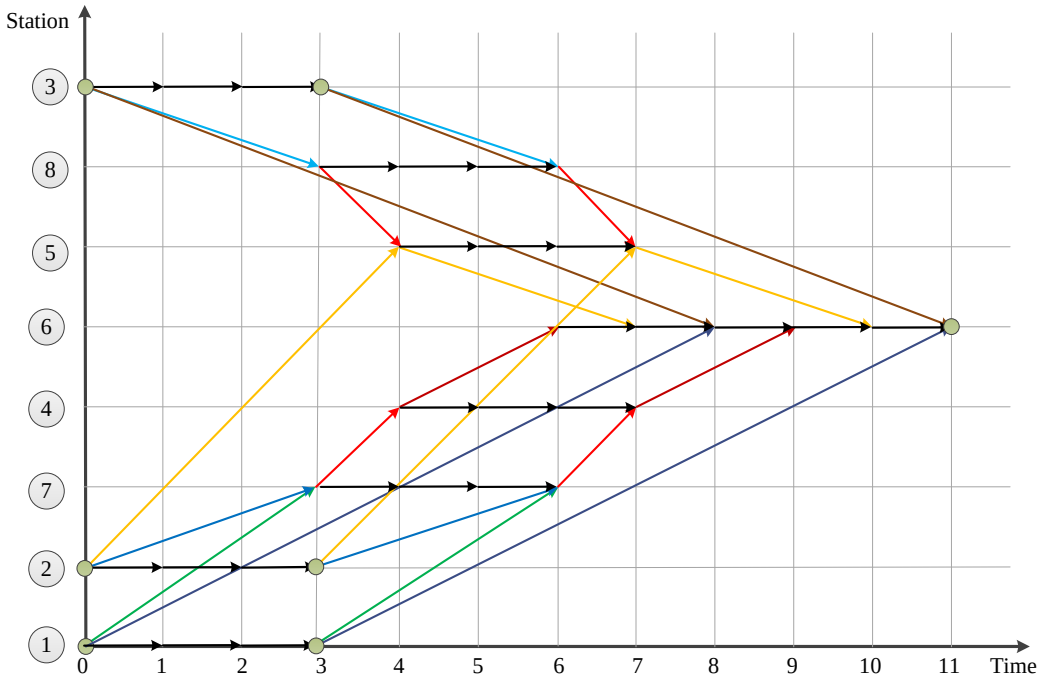


Fig. 11 The corresponding space-time transit service network

Table 5 Trip attributes of each passenger group

Group No	OD Pair	Departure Time	Average Trip Time	Group No	OD Pair	Departure Time	Average Trip Time
1	1 → 6	0	6	15	1 → 6	3	7.5
2	1 → 6	0	7	16	1 → 6	3	7
3	1 → 6	0	8	17	1 → 6	3	8
4	1 → 6	0	6.5	18	2 → 6	3	6
5	2 → 6	0	7	19	2 → 6	3	7
6	2 → 6	0	7.5	20	2 → 6	3	6.5
7	2 → 6	0	6.5	21	2 → 6	3	7.5
8	2 → 6	0	6	22	2 → 6	3	8
9	3 → 6	0	7	23	2 → 6	3	6.8
10	3 → 6	0	7.5	24	3 → 6	3	7
11	3 → 6	0	8	25	3 → 6	3	7.5
12	1 → 6	3	6	26	3 → 6	3	7.4
13	1 → 6	3	7	27	3 → 6	3	7.8
14	1 → 6	3	6.5	28	3 → 6	3	8

Table 6 Vehicle capacity of transit lines

Line No	L1	L2	L3	L4	L5	L6	L7
Capacity of vehicles departing at time 0	300	300	600	200	400	300	200
Capacity of vehicles departing at time 3	400	400	800	300	600	400	300

5.1.2 Focused states for observability quantification

The states focused in this experiment are listed as follows.

(1) Arc flow state: passenger count (congestion) in transfer corridor $(8, 5)$ at time points 3 and 6, respectively.

(2) Path flow state 1: the passenger flow departing at node 2 and time 0 to use line 1.

(3) Path flow state 2: the earning collected in the ticket for company line 1 on its first vehicle.

(4) Network-level arc flow state: the system-wide passenger count (congestion) on the running vehicles at time point 5.

5.1.3 Scenario design

As a short summary, based on the available supply and demand data, we aim to (i) preprocess the measurements if there is no feasible solution due to the possible existence of measurement errors in step 1, and (ii) quantify the uncertainty of our focused states in step 2. Five scenarios are designed to demonstrate the value of information based on our proposed models.

Scenario 1 (S1: base case): it is assumed that the origin, destination, and departure time of each passenger group is given, and no other information is available.

Scenario 2 (S2: base case + count): based on scenario 1, the passenger count data from video processed data at transfer corridor (7, 4) is available.

Scenario 3 (S3: base case + end-to-end travel time): based on scenario 1, the averaged group trip time from smart card is available.

Scenario 4 (S4: base case + end-to-end travel time + count): based on scenario 1, both the passenger count data and average group trip time data are available.

Scenario 5 (S5: ground truth): since the observed data may have its measurement errors, we assume that a ground truth can be obtained and will be compared with other scenarios. The ground truth is assumed as the system conditions based on maximizing the arc flow at time point 3 in scenario 3.

5.1.4 Result analysis

In step 1, the measurement is preprocessed by the proposed model at Appendix A. In step 2, we compute the uncertainty range of states (1)-(3) by maximizing and minimizing the state goals, and state (4) is addressed based on the solutions from the previous three states as a sample-based approximation. Before analyzing different state results in different scenarios, it is important to clearly illustrate the conditions under which those results are obtained from our proposed models.

(1) In scenario 1, there is no available sensor data, so the measurement doesn't need to be preprocessed.

(2) In scenario 2, the measurement is preprocessed for the passenger count data at transfer corridor (7, 4). The total squared errors in objective function (A.1) in step 1 is not equal to 0, which indicates that there will be no feasible solution if the observed measurement is directly used to step 2. The estimated passenger counts at transfer corridor (7, 4) at time points 3 and 6 is 450 and 800, respectively, compared with the observed values of 450 and 810. The total absolute error for the observed passenger count is 10.

(3) In scenario 3, when step 1 is conducted using the observed average trip time, the total error is not equal to 0 as well. The estimated average group trip time for each group is shown in Table 7. The total absolute error for the average group trip time is 2.58.

(4) In scenario 4, in step 1, there are two different sensor data, so it will require weights on different measurements. As discussed by Lu et al. (2013), the weights should reflect the degrees of confidence on different observed data and can be represented by the inverses of the variances of the distinct sources of measurements. Therefore, the weights on aggregated average trip time and passenger count are calculated as 2.36 and 0.31, respectively. Finally, the total absolute errors for observed average group trip time and passenger count are 3.83 and 273, which are greater than the absolute errors in scenario 1 and scenario 2, respectively. It shows that the inconsistency among multi-source data forces the model to find a balance among those observation.

(5) In scenario 5, the preprocessed group trip time in step 1 is used as the input to maximize the passenger count in transfer corridor $(8, 5)$ at time points 3, and the corresponding system condition is assumed as the ground truth in this dynamic transit system.

Table 7 The observed and preprocessed average group trip time for each passenger group

Passenger group No	Observed values	Preprocessed values in scenario 3	Passenger group No	Observed values	Preprocessed values in scenario 3
1	6	6	15	7.5	7.5

2	7	7	16	7	7
3	8	8	17	8	8
4	6.5	6.5	18	6	6
5	7	7	19	7	6.9
6	7.5	7	20	6.5	6.4
7	6.5	6.5	21	7.5	7
8	6	6	22	8	7
9	7	7	23	6.8	6.7
10	7.5	7.5	24	7	7.08
11	8	8	25	7.5	7.57
12	6	6	26	7.4	7.47
13	7	7	27	7.8	7.87
14	6.5	6.5	28	8	8

Fig. 12 shows that the estimated maximum and minimal flow rates on each focused arc under different scenarios. As available information is increased, the range of passenger flow uncertainty on transfer corridor (8, 5) is reduced. Meanwhile, both scenarios 3 and 4 can assert that their estimated state uncertainty is 0 and the state is completely observable. However, the different estimated unique states on arc (8,5,6,7) seem conflicted.

Specifically, in scenario 3, the observed trip time is preprocessed due to its measurement error, and finally the estimated states on transfer corridor (8, 5) is consistent with the states in the ground truth in scenario 5. Note that the estimated states may not be totally consistent with the ground truth, even though the observed data is same as the corresponding data in ground truth, because the observation is only a partial reflection of the whole system condition. It is also possible that the corrected measurement is not consistent with that in this ground truth if other measurement correction approaches rather than the least square method are used in reality in step 1.

In addition, in scenario 4, the inconsistency of observed link count data and observed trip time data makes the corrected measurement different from the corresponding data in the ground truth, so the final estimated unique state in step 2 cannot be the real-world condition anymore. Therefore, in reality, when the transportation system state is estimated by different sensor data, the data quality and assigned weight on each data source in step 1 is important and should be clearly stated. Note that how to balance each observation is beyond the scope of this paper. For more details on knowledge fusion, readers can be referred to the paper (Zheng et al., 2014).

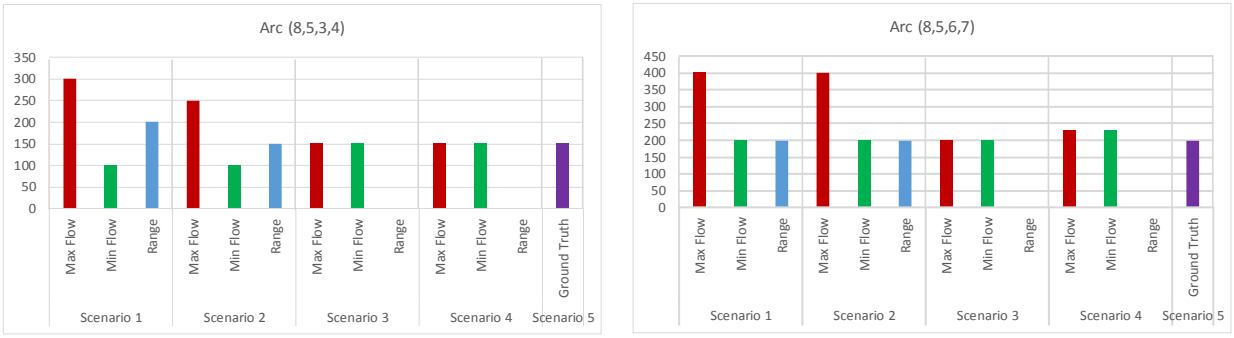


Fig. 12 The estimated flow uncertainty range on each focused arc

Focusing on the passenger flow departing at node 2 and time 0 to use line 1, it is actually the path flow of path (2,0) →(5,4)→ (6,7). The path flow uncertainty is shown in Fig. 13. The uncertainty range is similar to the arc flow above. The estimated unique state in scenario 4 is not consistent with the state value in ground truth. In addition, if line 1 is managed by one company and other lines are managed by other companies, it needs to assign the fare to each company based on their service. However, the number of passengers using one specific line is uncertain in the transit system, so based on our proposed method, we can quantify the uncertainty and estimate the general fare earning for each company rather than just using some simple rules for fare clearing (Gao et al., 2011; Zhou, 2014). For example, one simple rule is to calculate the shortest path and then assume that passengers will choose the shortest path as their selected lines.

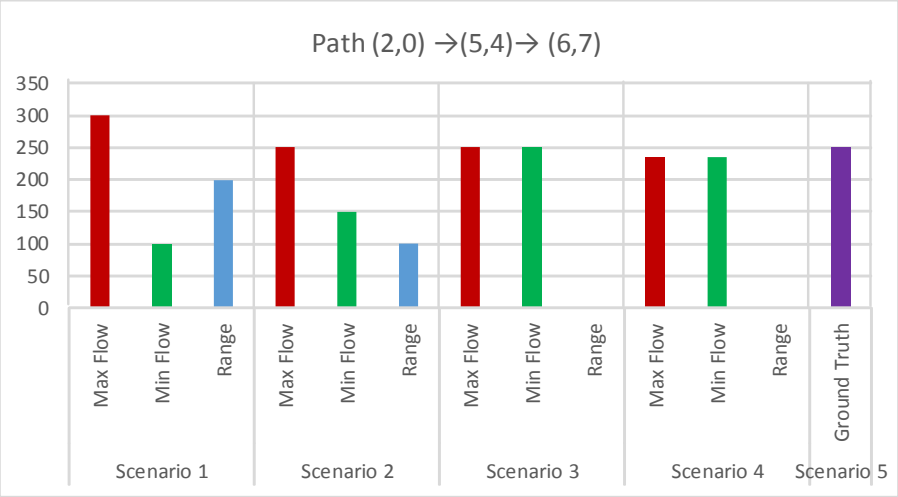


Fig. 13 Estimated flow uncertainty range on the focused path

In each scenario, we maximize and minimize the passenger flow on arc (8,5,3,4), arc (8,5,6,7), and path of line 1 as 6 cases, respectively, so six feasible solutions of $x^a_{i,j,t,s}$ can be obtained as sample points to quantify our defined system-level state observability. For state 4, the system-wide passenger count (congestion) on the running vehicles at time point 5 is be represented by the passenger flow on arcs (1,6,3,11), (1,6,0,8), (1,7,3,6), (2,5,3,7), (2,7,3,6), (3,6,3,11), (3,6,0,8), (3,8,3,6), (4,6,4,6), and (5,6,4,7). The results under six objectives in scenario 1 are listed in Table 8.

Table 8 Estimated passenger flows on arcs under six objectives in scenario 1

Arc(i,j,t,s)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
(1,6,3,11)	300	300	300	300	200	200
(1,6,0,8)	200	200	200	200	100	100
(1,7,3,6)	300	300	300	300	400	400
(2,5,3,7)	400	400	200	400	200	200
(2,7,3,6)	200	200	400	200	400	400
(3,6,3,11)	300	300	100	300	100	100
(3,6,0,8)	0	200	200	200	0	200
(3,8,3,6)	200	200	400	200	400	400
(4,6,4,6)	500	300	300	300	600	400
(5,6,4,7)	400	400	400	400	400	400

Based on the definitions of Maximal Possible Relative Error (MPRE) and Estimation Reliability (Re), the values of MPRE and Re are 16.38 and 5.753%, respectively. As shown in Table 8, the possible flow on arc (3,6,0,8) is a number between 0 and 200, resulting in a huge uncertainty and causing an extremely low estimation reliability. One possible reason is that our proposed models do not assume any travel behavior, so all solutions are based on the physical constraints and available sensor observations.

In scenario 2, with passenger count information, the estimated results of 6 cases for the system-level state are listed in Table 9. The corresponding values of MPRE and Re are 0.267 and 78.93%, respectively. It shows that the estimation reliability gets significantly improved when passenger counts from one key location (transfer corridor) are available, which could avoid a large uncertainty range occuring in scenario 1. In Scenarios 3 and 4, the values of MPRE and Re are 0 and 100%, respectively, but it is still emphasized that the MRPR and Re should be clearly explained with its correspondingly different measurement preprocessing errors (assigned weights) and adopted approach.

Table 9 Estimated passenger flows on arcs under six objectives in scenario 2

Arc(i,j,t,s)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
(1,6,3,11)	200	200	300	300	200	200
(1,6,0,8)	200	100	200	200	200	100

(1,7,3,6)	400	400	300	300	400	400
(2,5,3,7)	200	200	200	400	200	200
(2,7,3,6)	400	400	400	200	400	400
(3,6,3,11)	300	300	100	300	100	100
(3,6,0,8)	50	200	200	200	50	150
(3,8,3,6)	200	200	400	200	400	400
(4,6,4,6)	450	450	300	300	450	450
(5,6,4,7)	400	350	400	400	400	400

5.1.5 Results from Frank-Wolfe algorithm and Dantzig-Wolfe decomposition

In this section, we implement Frank-Wolfe algorithm in step 1 and Dantzig-Wolfe decomposition algorithm in step 2 in GAMS. The case of minimizing the passenger flow on arc (8,5,3,4) in scenario 4 is treated as an example to analyze the performance of those algorithms. The source code can be downloaded at this link (https://www.researchgate.net/publication/324809217_F-W_and_D-W_Observability_Quantification).

In section 5.1.4, the case is solved by the solver MINOS in GAMS directly. In step 1, the solved model is a non-linear programming model, and the minimal total generalized least square error in the objective function is 5.968. When the model is solved by Frank-Wolfe algorithm as a linear programming model, the result shown in Fig.14 finally converge to 7.069 after 20 iterations. The gap is probably caused by the optimal step size, which is found as a constant value at each iteration rather than a constant value vector for each variable. Hence, it could make the final solution converge to a local optimal solution.

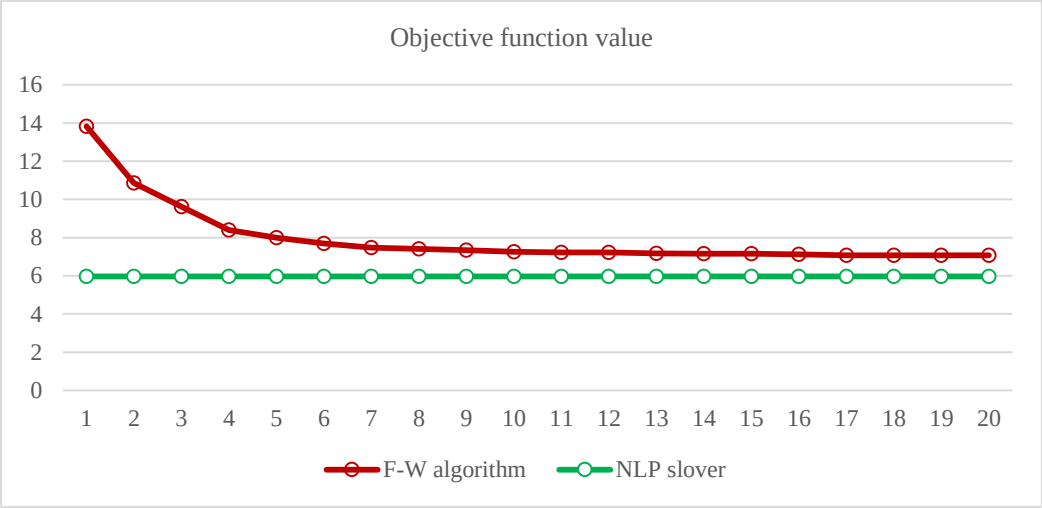


Fig. 14 Objective function values under different solving approaches

In step 2, Dantzig-Wolfe decomposition is applied to generate extreme points for time-dependent OD pairs as subproblems, and the restricted master problem is solved by CPLEX. The solved minimal passenger flow is 142.5 based on the preprocessed measurements by Frank-Wolfe algorithm rather than by the NLP solver. The generated extreme points (feasible paths) and the correspondingly optimal weights are listed in Table 10. However, if the preprocessed measurements in step 1 are directly obtained from the NLP solver, the final minimal passenger count on arc (8,5,3,4) from the Dantzig-Wolfe decomposition is 150 as well.

Table 10 Generated extreme points and optimal weights in Dantzig-Wolfe decomposition

Passenger Group No	Extreme points (path node sequence (ii, tt))	Optimal weights on extreme points	Passenger Group No	Extreme points (path node sequence (ii, tt))	Optimal weights on extreme points
1	(1,0)→(6,8);	0.06	15	(1,3)→(7,6) →(4,7) → (6,9);	0.30
	(1,0)→(7,3) →(4,4) → (6,6);	0.94		(1,3)→(6,11);	0.70

2	(1,0)→(6,8);	0.54	16	(1,3)→(7,6) →(4,7) → (6,9);	0.51
	(1,0)→(7,3) →(4,4) → (6,6);	0.46		(1,3)→(6,11);	0.49
3	(1,0)→(6,8);	0.93	17	(1,3)→(7,6) →(4,7) → (6,9);	0.07
	(1,0)→(7,3) →(4,4) → (6,6);	0.07		(1,3)→(6,11);	0.93
4	(1,0)→(6,8);	0.29	18	(2,3)→(7,6) →(4,7) → (6,9);	1
	(1,0)→(7,3) →(4,4) → (6,6);	0.71		(2,3)→(5,7) →(6,10);	0
5	(2,0)→(5,4) →(6,7);	0.93	19	(2,3)→(7,6) →(4,7) → (6,9);	0.15
	(2,0)→(7,3) →(4,4) → (6,6);	0.07		(2,3)→(5,7) →(6,10);	0.85
6	(2,0)→(5,4) →(6,7);	0.93	20	(2,3)→(7,6) →(4,7) → (6,9);	0.62
	(2,0)→(7,3) →(4,4) → (6,6);	0.07		(2,3)→(5,7) →(6,10);	0.38
7	(2,0)→(5,4) →(6,7);	0.58	21	(2,3)→(7,6) →(4,7) → (6,9);	0
	(2,0)→(7,3) →(4,4) → (6,6);	0.42		(2,3)→(5,7) →(6,10);	1
8	(2,0)→(5,4) →(6,7);	0.13	22	(2,3)→(7,6) →(4,7) → (6,9);	0
	(2,0)→(7,3) →(4,4) → (6,6);	0.87		(2,3)→(5,7) →(6,10);	1
9	(3,0)→(6,8);	0.08	23	(2,3)→(7,6) →(4,7) → (6,9);	0.36
	(3,0)→(8,3) →(5,4) → (6,7);	0.92		(2,3)→(5,7) →(6,10);	0.64
10	(3,0)→(6,8);	0.56	24	(3,3)→(6,11);	0.09
	(3,0)→(8,3) →(5,4) → (6,7);	0.44		(3,3)→(8,6) →(5,7) →(6,10);	0.91
11	(3,0)→(6,8);	0.93	25	(3,3)→(6,11);	0.55
	(3,0)→(8,3) →(5,4) → (6,7);	0.07		(3,3)→(8,6) →(5,7) →(6,10);	0.45
12	(1,3)→(7,6) →(4,7) → (6,9);	0.94	26	(3,3)→(6,11);	0.45
	(1,3)→(6,11);	0.06		(3,3)→(8,6) →(5,7) →(6,10);	0.55
13	(1,3)→(7,6) →(4,7) → (6,9);	0.50	27	(3,3)→(6,11);	0.85
	(1,3)→(6,11);	0.50		(3,3)→(8,6) →(5,7) →(6,10);	0.15
14	(1,3)→(7,6) →(4,7) → (6,9);	0.76	28	(3,3)→(6,11);	0.93
	(1,3)→(6,11);	0.24		(3,3)→(8,6) →(5,7) →(6,10);	0.07

5.2. Tests in a large-scale network

In order to address the computational challenges in large-scale networks, we will propose an approximation-based approach, which provides a k-shortest path set as extreme points for each passenger group (in each OD pair with time-dependent departure time) in advance rather than using Dantzig-Wolfe decomposition to generate extreme point iteration by iteration. In this section, the public Google Transit Feed Specification (GTFS) data from Alexandria Transit Company in 2015 is used as our tested large-scale transit network (https://transitfeeds.com/p/alexandria-transit-company). As shown in Fig.15, it has 12 routes, 1638 trips (866 trips on weekdays, 423 trips on Saturdays, 261 trips on Sundays, and 88 trips on the Christmas day), and 629 stops.

Fig. 15. Alexandria transit network read from GTFS, in Virginia, USA

In this experiment, the trips on weekdays are only considered as the provided schedule. Then, 32,029 vertexes and 713,650 arcs are generated in the corresponding space-time network for one whole weekday. The arcs include vehicle running arcs, passengers' walking arcs from origin to transit stops and from transit stops to destination, transfer arcs, and waiting arcs. The space-time arc generation rules contain (i) the trip (path) travel time is less than 120 minutes; (ii) the maximum number of transfer times is 3; (iii) the maximum transfer/walking time is 30 minutes; (iv) the maximum transfer/walking distance is 0.5 mile.

In order to obtain the time-dependent transit demand, we map the traffic analysis zones in the city of Alexandria to the transit network as the activity locations. As a result, 42 OD pairs are matched. In addition, the time period of 7:00am to 9:00am is divided by 36 time intervals, so the time-dependent OD demand is defined by each 5 minutes. Finally, 1484 time-dependent OD pairs are obtained based on the arc generation rules above.

In addition, as an approximation for those extreme points in Dantzig-Wolfe decomposition for each time-dependent OD pair, we generate 3-shortest paths using our developed k-shortest path algorithm. Finally, 4452 paths are generated with 7,868 arcs in the space-time network. The k-shortest path algorithm for each time-dependent OD pair is shown as follows.

(i) Based on the origin vertex (origin node and departure time) in the space-time network, the label correcting algorithm is used to generate a shortest path tree from origin vertex to all possible vertexes selected on the basis of the space-time arc generation rules.

(ii) According to the physical destination location, we can find a number of candidate vertexes (stop id and stop time in schedule) connected to the physical destination node by walking arcs. Then we can add the label costs of those candidate vertexes and its corresponding walking arc costs to the physical destination, so the destination will have a number of vertexes (destination node and arrival time) with different label cost.

(iii) Sort those label costs of the destination node and select k least-cost destination vertexes and back trace to the origin vertex. As a remark, at each vertex, we also record the transfer state (the number of transfer times), so when back tracing the path to origin vertex, we can obtain different paths which are from one same vertex with same label cost but with different transfer states. Finally, the k-shortest path set can be generated for each time-dependent OD pair.

For simplicity, we assume that all transit vehicle capacity is 35 and the walking, waiting and transfer arc capacity is 9999. Also, the time-dependent demand of each OD pair is assumed to be 1, which means that one passenger will arrive every 5 minutes for each OD pair. The observed passenger trip time is generated as a random value between the minimal and the maximal path costs of the 3-shorest paths. Our focused state is the uncertainties of passenger flow count on transfer links from stop 4,000,644 to stop 4,000,863 and from stop 4,000,745 to stop 4,000,509 based on the 3-hour transit demand. Then, four models are solved by CPLEX optimization engine in GAMS as a linear programming problem implemented on a workstation with Intel(R) Xeon(R) CPU E2680 v2 @ 2.8GHz processors. For each model, there are 10,837 equations and 4,452 variables, and the computation time is around 19 seconds. The results are shown in Fig. 16.

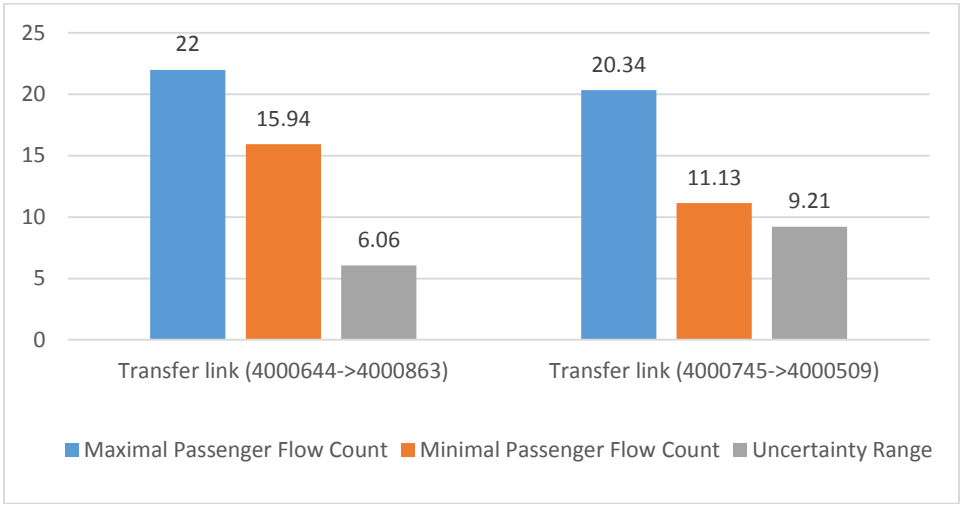


Fig. 16 Uncertainties of passenger flow count at transfer links

7. Conclusion and future research

This research provides some insights about the relationship between multi-source information, information space, state estimation, and system observability quantification by taking the urban transit systems as the analysis object. The information space and information errors are highly respected for state observability. Projection-function-based approaches are presented to quantify the observability of different states under the same information space. The proposed models can explain that the value of information highly relies on its aimed specific system states and sensor location rather than its volume. It provides an analysis to show how to better use available information for different state estimate evaluation and how to design the sensor network for future system state observability improvement.

It should be remarked that, the observability quantification based on different states is just the first step for better observing and controlling the system. The following questions are currently under our considerations for future research: (1) what is the balance among the system observability, the minimally needed information, and the required accuracy of future controls? (2) What is the balance of the sensor data cost, value of information and its computational efficiency in proposed models and algorithms? (3) How to integrate the heterogeneous sensor network design with the real-time system control? (4) How to visualize the real-time uncertainty of different system states in a straightforward way for the public.

Acknowledge

We would like to thank Dr. Yingyan Lou at Arizona State University and Dr. Jianrui Miao at Beijing Jiaotong University for their great comments on our experiments. This paper is supported by National Science Foundation – United States under Grant No. CMMI 1538105 "Collaborative Research: Improving Spatial Observability of Dynamic Traffic Systems through Active Mobile Sensor Networks and Crowdsourced Data". The work presented in this paper remains the sole responsibility of the authors.

Appendix A: preprocessing the measurements

Based on the proposed constraints in subsection 3.2, a nonlinear programming model is presented as follows.

$$\text{Min } \beta_1 \sum_a \left(\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a\right)^2 + \beta_2 \sum_{(i,j,\tau)} \left(\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau}\right)^2 \quad \text{(A.1)}$$

Subject to, flow-balance constrains (1): $AX = B$, capacity constraints(2): $CX \le D$, and variable constraint (5): $x \ge 0$.

The objective function is to minimize the weighted total deviations between the preprocessed and observed data, where β_1 and β_2 are the weights reflecting different degrees of confidence on observations. Those weights can be viewed as the inverses of the variances of the heterogeneous sources of measurements adopted by Lu et al. (2013).

Another technique used to measure the deviation is to quantify the absolute difference as least absolute deviations (LAD). The corresponding objective function will change to be

$$\text{Min } \beta_1 \sum_a \left|\sum_{(i,j,t,s)} (x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a\right| + \beta_2 \sum_{(i,j,\tau)} \left|\sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau}\right| \quad \text{(A.2)}$$

The analysis on the attributes of the methods of least squares and least absolute deviation can be found in the paper by Gorard (2005) and Koenker and Hallock (2001). Specifically, least absolute deviation treats all observation equally, but least squares gives more emphasis to large residuals by squaring the residuals, which could be a better choice when dealing with outliers in which estimated values are far from real-world sensor observations. As a remark, the least absolute deviations can be solved by linear programming through transforming the model. For example, new model based on formula (A.2) could be minimizing $\beta_1 \sum_a \theta_a + \beta_2 \sum_{(i,j,\tau)} \theta_{i,j,\tau}$ with adding new constraints $-\theta_a \leq \sum_{(i,j,t,s)}(x^a_{i,j,t,s} \times c_{i,j,t,s}) - D_a \times \mu_a \leq \theta_a$, $-\theta_{i,j,\tau} \leq \sum_a \sum_{t\in\tau} x^a_{i,j,t,s} - \mu_{i,j,\tau} \leq \theta_{i,j,\tau}$, and $\theta_a \geq 0, \theta_{i,j,\tau} \geq 0$.

Appendix B: Frank-Wolfe algorithm for nonlinear programming models

The algorithm procedure is described as follows.

Step 1: initialization: $k = 0$, and find one feasible solution as $\boldsymbol{x}_0$;

Step 2: Based on the first-order Taylor approximation of $f(x)$ aroundx_k, minimizing the linear approximation: min $\boldsymbol{s}_k^T \nabla f(x_k)$, and $\boldsymbol{s_k}$ is subject to all constraints.

Step 3: Find γ that minimizes $f(x_k + \gamma(s_k - x_k))$, subject to $0 \leq k \leq 1$.

Step 4: Update: $x_{k+1} = x_k + \gamma(s_k - x_k)$. If $|x_{k+1} - x_k| \leq \Delta$ or $k = K$, stop. Otherwise, $k = k + 1$ and go to step 2.

Specifically, at step 2, $\nabla f(x_k) = \beta_1 \sum_a \sum_{(i,j,t,s)}[2 \times c_{i,j,t,s} \times c_{i,j,t,s} \times x^a_{i,j,t,s}(k)] + \beta_2 \sum_{(i,j,\tau)}[\sum_a \sum_{t\in\tau}(2 \times x^a_{i,j,t,s}(k))]$ is constant, so $\boldsymbol{s}_k^T \nabla f(x_k) = \beta_1 \sum_a \sum_{(i,j,t,s)}[2 \times c_{i,j,t,s} \times c_{i,j,t,s} \times x^a_{i,j,t,s}(k) \times s^a_{i,j,t,s}(k)] + \beta_2 \sum_{(i,j,\tau)}[\sum_a \sum_{t\in\tau}(2 \times x^a_{i,j,t,s}(k) \times s^a_{i,j,t,s}(k))]$. Finally, the model proposed in step 2 is a linear programming model with the flow-balance constraint.

In addition, at step 1, for finding one feasible solution, we can define a simple linear objective function, so the model will be a linear programming model with the flow balance constraint, which can be solved by the Dantzig-Wolfe algorithm due to the special block structure as well.

References

Alsger, A., Assemi, B., Mesbah, M., Ferreira, L., 2016. Validating and improving public transport origin–destination estimation algorithm using smart card fare data. Transportation Research Part C, 68, 490-506.

Ban, X.J., Hao, P. and Sun, Z., 2011. Real time queue length estimation for signalized intersections using travel times from mobile sensors. Transportation Research Part C: Emerging Technologies, 19(6), 1133-1156.

Ban, X., Herring, R., Margulici, J.D., Bayen, A., 2009. Optimal sensor placement for freeway travel time estimation. Transportation and Traffic Theory, Chapter 34 (W.H.K. Lam, S.C. Wong, H.K. Lo eds.), Springer, 697-721.

Barry, J., Freimer, R. and Slavin, H., 2009. Use of entry-only automatic fare collection data to estimate linked transit trips in New York City. Transportation Research Record: Journal of the Transportation Research Board, (2112), 53-61.

Bianco, L., Confessore, G., Reverberi, P., 2001. A Network Based Model for Traffic Sensor Location with Implications on O/D Matrix Estimates. Transportation Science 35(2), 50-60

Bierlaire, M., 2002. The total demand scale: a new measure of quality for static and dynamic origin–destination trip tables. Transportation Research Part B: Methodological, 36(9), 837-850.

Boyles, S., Waller S. T., 2011. Optimal information location for adaptive routing. Networks and Spatial Economics 11, 233-254.

Castillo, E., Conejo, A.J., Pruneda R., Solares C., 2007. Observability in linear systems of equations and inequalities: Applications. Computer & Operations Research, 34, 1708-1720.

Castillo, E., Conejo, A.J., Menéndez, J.M. and Jiménez, P., 2008. The observability problem in traffic network models. Computer - Aided Civil and Infrastructure Engineering, 23(3), 208-222.

Castillo, E., Rivas, A., Jimenez, P., Menendez, J. M., 2012. Observability in Traffic Networks. Plate Scanning Added by Counting Information. Transportation 39 (6), 1301-1333.

Castillo, E., Grande, Z., Calviño, A., Szeto, W.Y. and Lo, H.K., 2015. A state-of-the-art review of the sensor location, flow observability, estimation, and prediction problems in traffic networks. Journal of Sensors, 2015.

Cats, O., West, J. and Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transportation Research Part B: Methodological, 89, 43-57.

Ceapa, I., Smith, C. and Capra, L., 2012. Avoiding the crowds: understanding tube station congestion patterns from trip data. In Proceedings of the ACM SIGKDD international workshop on urban computing (134-141). ACM.

Chen, H.K. and Hsueh, C.F., 1998. A model and an algorithm for the dynamic user-optimal route choice problem. Transportation Research Part B: Methodological, 32(3), 219-234.

Codina, E. and Rosell, F., 2017. A heuristic method for a congested capacitated transit assignment model with strategies. Transportation Research Part B: Methodological, 106, 293-320.
Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological, 41(1), 49-62.
Danczyk, A., Liu, H., 2011. A mixed-integer linear program for optimizing sensor locations along freeway corridors. Transportation Research Part B, 45 (1) 208–217.
Dantzig, G.B. and Wolfe, P., 1960. Decomposition principle for linear programs. Operations research, 8(1), 101-111.
Desrosiers, J. and Lübbecke, M.E., 2005. A primer in column generation. In Column generation (1-32). Springer US.
Drissi-Kaïtouni, O., Hameda-Benchekroun, A., 1992. A dynamic traffic assignment model and a solution algorithm. Transportation Science 26, 119–128.
Eisenman, S., M., Fei, X., Zhou, X., Mahmassani, H. S., 2006. Number and Location of Sensors for Real-time Network Estimation and Prediction: A Sensitivity Analysis. Transportation Research Record 1964, 253-259.
Frank, M. and Wolfe, P., 1956. An algorithm for quadratic programming. Naval Research Logistics (NRL), 3(1‐2), 95-110.
Ford Jr, L.R. and Fulkerson, D.R., 1958. A suggested computation for maximal multi-commodity network flows. Management Science, 5(1), 97-101.
Gao, L., Wang, B., and Zhang, C., 2011. Comparison of urban rail transit fare clearing model based on travel survey. Modern urban transit, 11, 97-99.
Gentili, M., Mirchandani, P. B., 2005. Locating Active Sensors on Traffic Networks. Annals of Operations Research 136 (1): 229-257.
Gentili, M. and Mirchandani, P.B., 2012. Locating sensors on traffic networks: Models, challenges and research opportunities. Transportation research part C: emerging technologies, 24, 227-255.
Gorard, S., 2005. Revisiting a 90-year-old debate: the advantages of the mean deviation. British Journal of Educational Studies, 53(4), 417-430.
Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q. and Bayen, A.M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. Transportation Research Part C: Emerging Technologies, 18(4), 568-583.
Hu, S. R., Peeta, S., Chu, C., 2009. Identification of Vehicle Sensor Locations for Link-Based Network Traffic Applications. Transportation Research Part B 43 (8), 873-894.
Jiang, Y. and Szeto, W.Y., 2016. Reliability-based stochastic transit assignment: Formulations and capacity paradox. Transportation Research Part B: Methodological, 93, 181-206.
Kalman, R., 1959. On the general theory of control systems. *IRE Transactions on Automatic Control 4*(3), 110-110
Kalvelagen, E., 2003. Dantzig-Wolfe Decomposition with GAMS.
Koenker, R. and Hallock, K.F., 2001. Quantile regression. Journal of economic perspectives, 15(4), 143-156.
Kusakabe, T. and Asakura, Y., 2014. Behavioural data mining of transit smart card data: A data fusion approach. Transportation Research Part C: Emerging Technologies, 46, 179-191.
Kusakabe, T., Iryo, T. and Asakura, Y., 2010. Estimation method for railway passengers' train choice behavior with smart card transaction data. Transportation, 37(5), 731-749.
Lam, W.H. and Yin, Y., 2001. An activity-based time-dependent traffic assignment model. Transportation Research Part B: Methodological, 35(6), 549-574.
Larsson, T. and Patriksson, M., 1992. Simplicial decomposition with disaggregated representation for the traffic assignment problem. Transportation Science, 26(1), 4-17.
Larsson, T., Patriksson, M. and Rydergren, C., 2004. A column generation procedure for the side constrained traffic equilibrium problem. Transportation Research Part B: Methodological, 38(1), 17-38.
LaValle, S.M., 2012. Sensing and filtering: A fresh perspective based on preimages and information spaces. Foundations and Trends® in Robotics, 1(4), 253-372.
Li, P., Mirchandani, P., Zhou, X., 2015. Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork. Transportation Research Part B 81(1), 103–130.
Li, X., Ouyang, Y. 2011. Reliable sensor deployment for network traffic surveillance. Transportation Research Part B 45 (1), 218–231.
Liu, J., Kang, J.E., Zhou, X. and Pendyala, R., 2017. Network-oriented household activity pattern problem for system optimization. Transportation Research Part C: Emerging Technologies.
Liu, J., Zhou, X., 2016. Capacitated transit service network design with boundedly rational agents. Transportation Research Part B: Methodological, 93, 225-250.
Lu, C.C., 2016. Robust multi-period fleet allocation models for bike-sharing systems. Networks and Spatial Economics, 16(1), 61-82.

Lu, C.C., Liu, J., Qu, Y., Peeta, S., Rouphail, N.M., Zhou, X., 2016. Eco-system optimal time-dependent flow assignment in a congested network. Transportation Research Part B 94, 217–239.
Lu, C.C., Zhou, X., Zhang, K., 2013. Dynamic origin–destination demand flow estimation under congested traffic conditions. Transportation Research Part C, 34, 16-37.
Ma, X.L., Liu, C., Liu, J.F., Chen, F. Yu, H., 2015. Boarding stop inference based on transit IC card data. Journal of Transportation Systems Engineering and Information Technology, 15(4), 78-84.
Ma, X.L., Wang, Y.H., Chen, F., Liu, J.F., 2012. Transit smart card data mining for passenger origin information extraction. Journal of Zhejiang University Science C, 13(10), 750-760.
Ma, X.L., Wu, Y.J., Wang, Y., Chen, F. and Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transportation Research Part C: Emerging Technologies, 36, 1-12
Meng, L. and Zhou, X., 2011. Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. Transportation Research Part B: Methodological, 45(7), 1080-1102.
Munizaga, M.A. and Palma, C., 2012. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies, 24, 9-18.
Nair, R., Miller-Hooks, E., Hampshire, R.C. and Bušić, A., 2013. Large-scale vehicle sharing systems: analysis of Vélib'. International Journal of Sustainable Transportation, 7(1), 85-106.
Nassir, N., Hickman, M. and Ma, Z.L., 2015. Activity detection and transfer identification for public transit fare card data. Transportation, 42(4), 683-705.
Ng, M., 2012. Synergistic sensor location for link flow inference without path enumeration: A node-based approach. Transportation Research Part B: Methodological, 46(6), 781-788.
Nunes, A.A., Dias, T.G. and e Cunha, J.F., 2016. Passenger journey destination estimation from automated fare collection system data using spatial validation. IEEE transactions on intelligent transportation systems, 17(1), 133-142.
Pan, B., Zheng, Y., Wilkie, D. and Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (344-353). ACM.
Pelletier, M.P., Trépanier, M. and Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies, 19(4), 557-568.
Peeta, S. and Mahmassani, H.S., 1995. Multiple user classes real-time traffic assignment for online operations: a rolling horizon solution framework. Transportation Research Part C: Emerging Technologies, 3(2), 83-98.
Seaborn, Catherine, John Attanucci, and Nigel H. M. Wilson.2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. Transportation Research Record: Journal of the Transportation Research Board 2121, 55–62.
Shang, P., Li, R., Liu, Z., Yang, L. and Wang, Y., 2018. Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network. Transportation Research Part C: Emerging Technologies, 89, 321-343.
Sun, Y., Xu, R., 2012. Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data. Transportation Research Record (2275), 58-67.
Szeto, W.Y. and Jiang, Y., 2014. Transit assignment: Approach-based formulation, extragradient method, and paradox. Transportation Research Part B: Methodological, 62, 51-76.
Tang, J., Song, Y., Miller, H.J. and Zhou, X., 2016. Estimating the most likely space–time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method. Transportation Research Part C: Emerging Technologies, 66, 176-194.
Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S. and Eriksson, J., 2009. VTrack: accurate, energy-aware road traffic delay estimation using mobile phones. In Proceedings of the 7th ACM conference on embedded networked sensor systems(85-98). ACM.
Tong, L., Zhou, X. and Miller, H.J., 2015. Transportation network design for maximizing space–time accessibility. Transportation Research Part B: Methodological, 81, 555-576.
Trépanier, M., Tranchant, N. and Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, 11(1), 1-14.
USDOT, 2015. Smart City Challenge: Lessons for Building Cities of the Future. Report.
Wei, Y., Avcı, C., Liu, J., Belezamo, B., Aydın, N., Li, P.T. and Zhou, X., 2017. Dynamic programming-based multi-vehicle longitudinal trajectory optimization with simplified car following models. Transportation Research Part B: Methodological, 106, 102-129.

Wu, X., Guo, J., Xian, K., Zhou, X., 2018. Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. Submitted

Xing, T., Zhou, X., Taylor, J., 2013. Designing Heterogeneous Sensor Networks for Estimating and Predicting Path Travel Time Dynamics: An Information-Theoretic Modeling Approach. Transportation Research Part B, 57, 66-90.

Xu, X., Lo, H.K., Chen, A. and Castillo, E., 2016. Robust network sensor location for complete link flow observability under uncertainty. Transportation Research Part B: Methodological, 88, 1-20.

Yang, H., Iida, Y. and Sasaki, T., 1991. An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts. Transportation Research Part B: Methodological, 25(5), 351-363.

Yang, H. and Meng, Q., 1998. Departure time, route choice and congestion toll in a queuing network with elastic demand. Transportation Research Part B: Methodological, 32(4), 247-260.

Yang, H., Zhou, J., 1998. Optimal Traffic Counting Locations for Origin-Destination Matrix Estimation. Transportation Research Part B 32, 109-126.

Yuan, N.J., Wang, Y., Zhang, F., Xie, X. and Sun, G., 2013. Reconstructing individual mobility from smart card transactions: A space alignment approach. In Data Mining (ICDM), 2013 IEEE 13th International Conference, 877-886.

Yuan, J., Zheng, Y., Xie, X. and Sun, G., 2011. Driving with knowledge from the physical world. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (316-324). ACM.

Zawack, D.J. and Thompson, G.L., 1987. A dynamic space-time network flow model for city traffic congestion. Transportation Science, 21(3),153-162.

Zhang, F., Yuan, N.J., Wilkie, D., Zheng, Y. and Xie, X., 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. ACM Transactions on Intelligent Systems and Technology (TIST), 6(3), 37.

Zhao, J., Rahbee, A., Wilson, N.H., 2007. Estimating a Rail Passenger Trip Origin‐Destination Matrix Using Automatic Data Collection Systems. Computer‐Aided Civil and Infrastructure Engineering, 22(5), 376-387.

Zhao, S. and Zhang, K., 2017. Observing individual dynamic choices of activity chains from location-based crowdsourced data. Transportation Research Part C: Emerging Technologies, 85, 1-22.

Zheng, Y., Capra, L., Wolfson, O. and Yang, H., 2014. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 5(3), 38

Zhou, F., Xu, R.H., 2012. Model of passenger flow assignment for urban rail transit based on entry and exit time constraints. Transportation Research Record (2284), 57-61.

Zhou, L., 2014. Verification and optimization of network clearing model. Urban mass transit, 11, 59-66.

Zhou, X., List, G.F., 2010. An information-theoretic sensor location model for traffic origin-destination demand estimation applications. Transportation Science, 44(2), 254-273.

Zhou, X. and Mahmassani, H.S., 2007. A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework. Transportation Research Part B: Methodological, 41(8), 823-840.

Zhu, N., Fu, C., Ma, S., 2018. Data-driven distributionally robust optimization approach for reliable travel-time-information-gain-oriented traffic sensor location model. Transportation Research Part B: Methodological, 113, 91-120.

Zhu, Y., Koutsopoulos, H.N. and Wilson, N.H., 2017a. A probabilistic Passenger-to-Train Assignment Model based on automated data. Transportation Research Part B: Methodological, 104, 522-542.

Zhu, Y., Koutsopoulos, H.N. and Wilson, N.H., 2017b. Inferring left behind passengers in congested metro systems from automated data. Transportation Research Part C: Emerging Technologies.

Zimmerman, J., Tomasic, A., Garrod, C., Yoo, D., Hiruncharoenvate, C., Aziz, R., Thiruvengadam, N.R., Huang, Y. and Steinfeld, A., 2011. Field trial of tiramisu: crowd-sourcing bus arrival times to spur co-design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (1677-1686). ACM.