National-scale Traffic Model Calibration in Real Time with Multi-source Incomplete Data

DESHENG ZHANG, Rutgers University
TIAN HE, University of Minnesota
FAN ZHANG, Shenzhen Institutes of Advanced Technology, China

Real-time traffic modeling at national scale is essential to many applications, but its calibration is extremely challenging due to its large spatial and fine temporal coverage. The existing work is focused on urban-scale calibration with complete field data from *single data sources* (e.g., loop sensors or taxis), which cannot be generalized to national scale because complete single-source field data at national scale are almost impossible to obtain. To address this challenge, in this article, we design MultiCalib, a model calibration framework to optimize traffic models based on *multiple incomplete data sources* at national scale in real time. Instead of simply combining multi-source data, we theoretically formulate a multi-source model calibration problem based on real-world contexts and multi-view learning. In particular, we design (i) convex multi-view learning to integrate multi-source data by quantifying biases of data sources, and (ii) context-aware tensor decomposition to infer incomplete multi-source data by extracting real-world contexts. More importantly, we implement and evaluate MultiCalib with two heterogeneous nationwide vehicle networks with 340,000 vehicles to infer traffic conditions on 36 expressways and 119 highways, along with four cities across China. The results show that MultiCalib outperforms baseline calibration by 25% on average with the same input data. Based on the proposed national-scale traffic model calibration, we design a novel dispatching framework integrated with our speed calibration model where we guide a vehicular fleet among national-scale highways with a routing strategy to reduce general traveling time. The results show that a routing strategy based on MultiCalib outperforms a routing strategy based on a state-of-the-art traffic model by 45% on average.

Categories and Subject Descriptors: H.4 [**Information System Application**]: Miscellaneous

General Terms: Algorithms, Model, Experimentation, Application

Additional Key Words and Phrases: Cyber-physical system, incomplete data, model calibration

ACM Reference format:
Desheng Zhang, Tian He, and Fan Zhang. 2019. National-scale Traffic Model Calibration in Real Time with Multi-source Incomplete Data. *ACM Trans. Cyber-Phys. Syst.* 3, 2, Article 21 (February 2019), 26 pages.
https://doi.org/10.1145/3300186

This work was supported by US NSF Grants CNS-1446640 and CNS-1544887, and China 973 Program 2015CB352400, and Research Program Grants of Shenzhen. A preliminary result of this work was published at (Zhang et al. 2016).
Authors' addresses: D. Zhang, 110 Frelinghuysen Rd, Piscataway Township, NJ 08854; email: desheng.zhang@cs.rutgers.edu; T. He, 200 Union Street SE Minneapolis, MN 55455; email: tianhe@umn.edu; F. Zhang, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, P. R. China, 518055; email: zhangfan@siat.ac.cn.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2378-962X/2019/02-ART21 $15.00
https://doi.org/10.1145/3300186

1 INTRODUCTION

Traffic demand and supply modeling are essential for transportation management and planning (Mitsch et al. 2012). A demand model assigns estimated traffic volumes to specific routes; whereas a supply model employs assigned traffic volumes to infer traffic speeds on routes based on a density-speed function (Le Ny et al. 2014). These two kinds of models tie demand and supply together to determine temporal propagations of traffic flows. The resulting traffic conditions, such as travel times, delay, volumes, and speeds, are used for various planning and management applications, such as navigation and dispatching (Rahman et al. 2013), to improve transportation efficiency.

To increase accuracy of these models, a model calibration process is typically required where various parameters in traffic models are optimized based on real-world field data (Balakrishna et al. 2011). Many techniques (Antoniou et al. 2007; Ciuffo and Lima Azevedo 2014; Nouir et al. 2008) have been proposed for demand/supply model calibration at *urban scale* based on field data, e.g., loop sensor data, camera data, or taxi GPS data. However, to our knowledge, there is little work, if any, on calibration on supply/demand models at *national scale*, which is essential for commercial logistics. Moreover, many applications require fine-grained traffic information, e.g., real-time navigation on road segment levels (Ueda et al. 2015), which calls for model calibration at a spatiotemporal level of road segments and minutes. A national-scale coverage with fine-grained resolutions poses a new challenge for us, which cannot be addressed by existing urban-scale model calibration. This is because they typically employ data from *single sources*, e.g., loop sensors (Aslam et al. 2012) or commercial vehicles such as taxis (Ge et al. 2010), with satisfactory urban-scale coverages and resolutions. But for the national-scale fine-grained calibration, these single-source data are often incomplete, which lead to a significant bias as shown in our motivation section.

In this work, we argue that with the recent advance of intelligent transportation systems, many vehicle networks at national scale are equipped with GPS and cellular devices, which enable national-scale traffic data collection without dedicated infrastructures, e.g., loop sensors or traffic cameras. It provides us an unprecedented opportunity to capture traffic dynamics from *national-scale multiple sources*. Therefore, in this article, we combine two national-scale vehicle networks (e.g., both commercial and private vehicles) as a nationwide system, and employ their real-time data to calibrate models for national-scale logistics of our data providers. This approach is technically challenging because (i) isolated national-scale vehicle networks (either commercial or private) are still limited in spatiotemporal coverage due to their operational characteristics, and thus lead to *incomplete data*; (ii) naively combining data sources from multiple networks together has a bias against general traffic flows, which leads to *overfitting* of the calibration process, as shown by our motivation section.

To address these challenges, we propose a calibration framework called MultiCalib for national-scale real-time traffic with two novel techniques: (i) we address overfitting of simple multi-source model calibration based on *convex multi-view learning* with an iterative online process, and (ii) we improve data completeness of data sources at national scale by *context-aware tensor decomposition* with three extracted contexts. A novel combination of these two techniques makes our work different from previous model calibration where traffic models are often calibrated by complete urban-scale data from single sources. Further, MultiCalib is a generic technique independent from specific models, and can be applied to a wide range of models for our data providers to improve their data completeness and model accuracy. In particular, in previous calibration methods (Antoniou et al. 2007; Ciuffo and Lima Azevedo 2014; Nouir et al. 2008), only one data source (e.g., taxis GPS data) is utilized, which leads to data bias (e.g., models are not effective for private vehicles) and data sparsity (e.g., models are not effective for the spatiotemporal combination without data). For the data bias issue, we design a model calibration technique based on data fusion to integrate multi-source data; for the data sparsity, we design a context-aware tensor-decomposition technique to infer missing data. Our key contributions are as follows:

— We design the first generic framework MultiCalib for traffic model calibration based on real-time multi-source data at national scale. To our knowledge, the proposed framework has two unique features: (i) it captures two kinds of heterogenous vehicle systems at national scale, i.e., commercial and private, which provide valuable diversity; (ii) it has one of the largest quantitative and spatiotemporal coverage, i.e., 340,000 vehicles spanning over an area of 6 million km^2 for 12 months.
— We formulate convex multi-view learning based on multi-source data to address biases of traffic model calibration with single-source data. In particular, we optimize parameters for both demand and supply models by minimizing an objective function that measures overall weighted deviation from model outcomes to the ground truth. To efficiently solve this optimization, we design an iterative learning process to alternatively update model parameters and weights assigned to different data sources until no further improvement can be made for the objective function. We formally prove the convexity and convergence of our optimization.
— We design context-aware tensor decomposition based on real-time contexts and historical data to improve completeness of field data at fine-spatiotemporal granularity for the calibration. In particular, we extract three novel contexts, i.e., a temporal pattern and a spatial pattern for historical traffic info, along with vehicle density, for joint tensor decomposition. It is a generic data inference technique, and can be used to improve various national-scale modeling.
— We implement a state-of-the-art model and calibrate it based on MultiCalib with 3TB data from two national-scale vehicle systems with 340,000 vehicles. We evaluate MultiCalib with 36 expressways and 119 highways, along with four cities, i.e., Beijing, Shanghai, Guangzhou, and Shenzhen. The results show that compared to state-of-the-art calibration, MultiCalib increases model accuracy by 25% on average.
— Based on the proposed national-scale traffic model, we propose a novel application where we guide a fleet among several national-scale highways with a dispatching strategy based on traffic models to reduce general traveling time. The accurate traffic prediction based on MultiCalib provides better routing information for dispatching centers to navigate their national-scale fleets. The results show that a dispatching strategy based on MultiCalib outperforms a routing strategy based on a state-of-the-art traffic model by 39% on average. We further improve the performance of the dispatching strategy by 42% on average by considering online calibration and route updates based on real-time vehicle data.

2 MOTIVATION

To show our motivation, we first investigate spatiotemporal coverages of single-source data at national scale, and then explore the potential of multi-source data.

2.1 Single-Source Data Coverage

We select one of the major expressways in China called JingGang'Ao coded G4. It has a length of 2,283KM with 2,384 road segments, crossing five provinces in China, and connecting Beijing to several major cities in the southern part of China, e.g., Wuhan, Guangzhou, and Shenzhen. With two national-scale commercial and private vehicle networks with 340,000 vehicles (details in Section 3), we investigate spatiotemporal coverage of them on G4 along with its origin, Beijing, and destination, Shenzhen. Specifically, we divide 24 hours into a total of 288 5-min slots, and for a particular slot, we calculate the 1-week average percentage of road segments with at least one

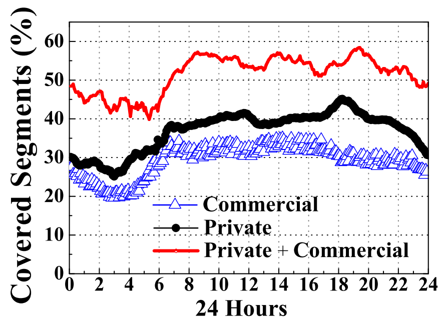

Fig. 1. Daily coverage.

GPS point from a particular vehicle network among all road segments on G4 and these two cities as shown in Figure 1. We have the following observations.

(i) During the daytime, both networks' coverage percentages are better than those of the early morning and late night. For the morning rush hour, the coverage percentages for private and commercial vehicles increase significantly; whereas for the evening rush hours, they decrease gradually. One possible explanation is the morning commutes are more regular than the evening commutes.

(ii) The private vehicles have higher coverage percentages than the commercial vehicles, due to their large volume (295,000 vs. 45,000). But the difference is not significant due to the long travel duration of commercial vehicles. The commercial vehicle network has a more stable coverage percentage through the daytime because commercial vehicles such as logistic trucks have a stable travel time schedule compared to private vehicles even during the late night.

(iii) In general, both vehicle networks have incomplete data issues shown by the percentages of covered road segments. For example, among a total of 288 5-min slots, the private vehicle network can only cover 45% of all road segments at most; whereas the commercial vehicle network can only cover 34% of all road segments at most. On average, the commercial vehicle network has a coverage percentage of 29.3% and the private vehicle network has a coverage percentage of 36.8%, both of which lead to incomplete data issues for model calibration (Balakrishna et al. 2011).

2.2 Multi-Source Data Coverage

In this work, we argue that as a promising solution to incomplete data issues, combining these two heterogenous vehicle networks together can provide a better coverage from complementary perspectives due to their own operational features. As shown by the red curve in Figure 1, we show the coverage percentage of a virtual vehicle network consisting of both private and commercial vehicles. We found that combining them provides valuable diversity in terms of coverages, e.g., the maximum coverage is close to 60% and the average coverage is 51.4%, which is an encouraging coverage percentage due to the large spatial area at fine temporal intervals. Nevertheless, because of different purposes of private and commercial vehicles, simply combining them together may lead to overfitting of model calibration (Zhang et al. 2015). Therefore, we propose a framework in this article to *intelligently* integrate multiple national-scale systems together to increase their coverage percentages for national-scale model calibration as introduced in the next section.

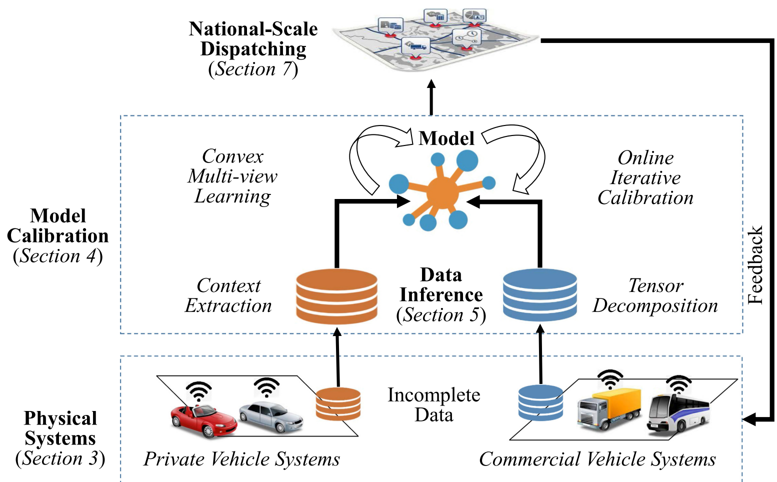

Fig. 2. MultiCalib architecture.

3 NATIONAL-SCALE MODEL CALIBRATION

We first present the architecture of MultiCalib, and then introduce its front-end physical systems, and finally show its back-end traffic modeling and calibration.

3.1 MultiCalib Architecture

In MultiCalib, we use a set of national-scale systems for national-scale traffic demand/supply modeling and their calibration. From a broad perspective, we treat individual components, i.e., vehicles, in these systems as probing sensors in MultiCalib to sense traffic conditions at national scale in real time. Built upon an integration of two national-scale systems, MultiCalib provides unseen national dynamics under extremely fine spatiotemporal resolutions to support real-world applications, which cannot be achieved by any systems with single data sources in isolation, e.g., a monolithic urban-scale infrastructure such as a taxi network. These three components span the whole data-processing chain for national-scale model calibration in MultiCalib.

In Figure 2, we outline MultiCalib with three components, i.e., physical systems, model calibration, and data inference, providing a road map for the article. (i) In Section 3.2, we introduce two national-scale vehicle networks as concrete front-end systems along with their data; and then in Section 3.3, we show our back-end data-driven calibration. (ii) In Section 4, we explore multi-source data from these systems to calibrate traffic models based on iterative convex multi-view learning. (iii) In Section 5, for better calibration, we improve incomplete data of individual systems by context-aware tensor decomposition where we extract three contexts for joint decomposition. (iv) In Section 7, to validate usefulness of our traffic model, we utilize the obtained model in a real-world application where we dispatch fleets among national-scale road networks to reduce their travel time.

3.2 Front-end Physical Systems

We have been collaborating with several logistics companies for both commercial and private networks and accessing their data feeds to obtain status of vehicles. In this version of MultiCalib implementation, we consider two vehicle networks, which detect national-scale traffic dynamics from complementary perspectives.

Beginning Date	2015/1/1	Total Data Size	3 TB
Private Vehicle Network		Commercial Vehicle Network	
# of Vehicles	294,849	# of Vehicles	45,237
# of Daily Records	240 million	# of Daily Records	85 million
Format		Format	
Device ID	Date&Time	Plate ID	Date&Time
Direction	GPS&Speed	Odometer	GPS&Speed

Fig. 3. National-scale datasets.

—For the private vehicle network, it has 295,000 vehicles, which are used to detect real-time traffic at national scale. Their data are collected through onboard devices installed inside vehicles, which are mainly used for navigation purposes. We access these data through a navigation service provider to which every involved vehicle uploads their real-time status to a cloud server by a cellular network. The private vehicle users can choose to opt out of this optional data uploading service, but most users still upload data in order to access navigation services with real-time traffic info. One-day data collected from all private vehicles in the network are about 9GB with an average uploading interval of 10 seconds when devices are turned on.

—The commercial vehicle network has 45,000 vehicles, which are used to detect real-time traffic by commercial vehicles. Their data are also collected through onboard devices and a cellular network, which are mainly used for monitoring and fleet management. We access these data through a logistic management company, which operates a fleet of commercial freight vehicles traveling on major national highways and expressways in China. Every vehicle uploads its real-time GPS locations and speeds back to a company's server as long as its engine is on, and then data are routed to our server. One-day data collected from all commercial vehicles in the network are about 7GB with an average uploading interval of 15 seconds.

The above two systems are extremely valuable for national traffic modeling and calibration. More importantly, they are spatiotemporally complementary to each other due to their own purposes. In particular, the private vehicles cover almost all major cities in China, along with 119 national highways and 36 national expressways in China; but the average accumulated uploading duration for these private vehicles is 100 mins on average for one day due to the purpose of personal usage, e.g., daily commutes and occasional long-distance travels, instead of frequent long-distance travels. In contrast, the commercial vehicles only cover major cities and highways in China, but due to their commercial purposes, their accumulated uploading time is 8 hours on average for one day. As a result, the private vehicles have better spatial coverages; whereas the commercial vehicles have better temporal coverages, which leads to complementary features for national-scale modeling and calibration.

3.3 Back-end Modeling and Calibration

Based on the data from front-end physical systems, we generate and calibrate traffic models to capture national-scale traffic conditions in our back-end server. Given the existence of a plethora of traffic models (Aslam et al. 2012; Ben-Akiva et al. 2012; Yuan et al. 2011a), we decide to focus on calibration of existing models, instead of designing new models. Next, we briefly give some preliminaries of traffic demand and supply modeling, along with existing calibration based on single-source data. Note that although we use the following models as an example, our MultiCalib is a generic technique to other demand and supply models as well.

3.3.1 Existing Demand and Supply Models. The demand model is also known as route choice models. The input of a demand model is an origin-destination (OD) matrix $\mathcal{X}_t$ for a particular time interval t based on OD estimation; its output is vehicle density $D_{t \cdot s}$ for a road segment s during a time slot t. In particular, we use one of the most common demand models as an example (Ben-Akiva et al. 2012) to show the details.

- Simulating behaviors of individual drivers with a fixed OD matrix by selecting a particular route r (consisting of multiple segments) from a route candidate set $\mathcal{R}$ generated by a road network with a probability P_r, e.g.,

$$P_r = \frac{\exp(U_r + \beta^1_{t\cdot r} \cdot CF_r)}{\sum_{l\in\mathcal{R}} \exp(U_l + \beta^1_{t\cdot r} \cdot CF_l)}, \tag{1}$$

  where U is a utility function of a route, i.e., the benefit of selecting this particular route, e.g., reduced travel time; CF is the commonality factor and indicates similarity between a particular route and other routes, e.g.,

$$CF_r = \ln \sum_{l\in\mathcal{R}} \left(\frac{L_{rl}}{\sqrt{L_r \cdot L_l}}\right)^{\beta^2_{t\cdot r}}, \tag{2}$$

  where L_r and L_l are the length of routes r and l, and L_{rl} are the common length between routes r and l. $\beta^1_{t\cdot r}$ and $\beta^2_{t\cdot r}$ are parameters to be estimated. We ignore temporal subscripts for U, CF, and R for simplicity.
- Aggregating trips at road segment level based on their specific routes to output a vehicle density $D_{t\cdot s}$ for a road segment s during a time slot t.

The supply model is to take output of a demand model, i.e., $D_{t\cdot s}$, as input, and then output traffic speeds at road segment level based on speed-density functions. One example of speed-density functions is

$$v_{t\cdot s} = v^*_{t\cdot s} \cdot \left[1 - \left(\frac{\max(0, D_{t\cdot s} - \underline{D}_{t\cdot s})}{D^{\text{jam}}_{t\cdot s}}\right)^{\beta^3_{t\cdot s}}\right]^{\beta^4_{t\cdot s}}, \tag{3}$$

where $v_{t\cdot s}$ is the unknown speed on a road segment s during t; $v^*_{t\cdot s}$ is a given free flow speed; $\underline{D}_{t\cdot s}$ is a given minimum density; $D^{\text{jam}}_{t\cdot s}$ is a given jam density, i.e., the extreme density associated with completely stopped traffic on s during t. $\beta^3_{t\cdot s}$ and $\beta^4_{t\cdot s}$ are parameters to be estimated. Based on a supply model, we have an expected speed $v_{t\cdot s}$ to infer congestion.

3.3.2 Model Calibration by Single-Source Data. The existing model calibration is to adjust parameters of models based on urban-scale field data from single data sources such as loop sensors or taxis. Typically, calibration is presented in the form of an optimization problem, where an objective function based on goodness-of-fit measures is formulated. For example, one technique that calibrates demand and supply simultaneously is given as follows (Chiu et al. 2011).

$$\begin{aligned} \min_{\mathcal{X}_t, \mathcal{P}_{t\cdot s}} \mathbf{F} = \mathbf{D_1}(\bar{v}_{t\cdot s}, v_{t\cdot s}) + \mathbf{D_2}(\mathcal{X}^a_t, \mathcal{X}_t) + \mathbf{D_3}(\mathcal{P}^a_{t\cdot s}, \mathcal{P}_{t\cdot s}), \\ \text{s.t.}, v_{t\cdot s} = \mathbf{M}(\mathcal{X}_t, \mathcal{P}_{t\cdot s}), \end{aligned} \tag{4}$$

where $\mathcal{X}_t$ is passenger demand; $\mathbf{M}$ is a demand/supply model; e.g., given in Equation (3); $\mathcal{P}_{t\cdot s}$ is a set of parameters in $\mathbf{M}$ (e.g., $\beta^1_{t\cdot s}$ to $\beta^4_{t\cdot s}$ in previous sections); $\mathcal{P}^a_{t\cdot s}$ and $\mathcal{X}^a_t$ are *a priori* estimates for $\mathcal{P}_{t\cdot s}$ and $\mathcal{X}_t$; $\mathbf{D_1}$, $\mathbf{D_2}$, and $\mathbf{D_3}$ are goodness-of-fit functions to measure a distance between two values; $v_{t\cdot s}$ is the speed outputted by $\mathbf{M}$; $\bar{v}_{t\cdot s}$ is the ground truth of the speed based on a single data source. The $\mathcal{X}_t$ and $\mathcal{P}_{t\cdot s}$ that minimize $\mathbf{F}$ are selected as the optimal parameters.

3.3.3 Motivation for Calibration with Data Fusion and Tensor Decomposition. In this article, we argue that the above model calibration works fine at urban scale where field data from single sources are complete for urban scale. However, as for our national-scale calibration, field data from single sources are often incomplete to obtain $\bar{v}_{t\cdot s}$ as shown by our motivation section. As follows, we design a calibration technique by multi-view learning driven by multi-source data to address overfitting in Section 4, which is built upon and a direct extension of our previous work (Zhang et al. 2015).

Further, the multi-source data-driven calibration will work fine if we have enough data for implementation sites. However, even with multi-source data, we still face a data sparsity issue due to the fine spatiotemporal modeling. To address this issue, we design a tensor-decomposition technique to infer the missing data based on their correlations with existing data. This data inference based on tensor decomposition aims to improve data completeness for multi-source data-driven calibration. The details of this component are given in Section 5.

4 MULTI-VIEW MODEL CALIBRATION

In this section, we first formulate an optimization problem for multi-view calibration in Section 4.1, develop an iterative process to solve this optimization in Section 4.2, and analyze calibration performance in terms of convexity and convergence in Section 4.3. Note that although our private and commercial vehicle data can only form two views to calibrate traffic models, i.e., private vehicle view and commercial vehicle view, we aim to tackle a generic problem, i.e., multi-view calibration, and thus, our double-view calibration is a concrete implementation of multi-view calibration.

4.1 Calibration Optimization

The objective of our multi-view calibration is to obtain optimized parameters for given demand/supply models for a time period based on multiple streaming datasets from multiple systems. Thus, we design a multi-view calibration framework for given demand and supply models to analyze congestion levels of particular road segments. The calibration is given as follows.

— A demand/supply model $\mathbf{M}$ is given with a parameter set $\mathcal{P}_{t\cdot s}$ to output the estimated speed $v_{t\cdot s}$ for each road segment s during a time slot t;
— An input demand set $\mathcal{X}_t$ is given by dynamic origin-destination (OD) matrices;
— *A priori* estimates for the model parameters in $\mathcal{P}_{t\cdot s}$ and $\mathcal{X}_t$ are given as $\mathcal{P}^a_{t\cdot s}$ and $\mathcal{X}^a_t$;
— K individual systems generate K field datasets from K different views in terms of traffic speed $\bar{v}^k_{t\cdot s}, \forall k \in [1, K]$, which are improved by our tensor-decomposition-based data inference introduced next. Note that in our implementation, we set $K = 2$ because we have only two data sources, i.e., commercial and private vehicle networks.

Based on Equation (4), we formulate an optimization problem to calibrate a given demand/supply model by optimizing its parameter set $\mathcal{P}_{t\cdot s}$ for a particular road segment s.

$$\min_{\mathcal{X}_t, \mathcal{P}_{t\cdot s}, \mathcal{W}_{t\cdot s}} \mathbf{F} = \sum_{k=1}^{K} \left[w^k_{t\cdot s} \cdot \left(\mathbf{D}_1\left(\bar{v}^k_{t\cdot s}, v_{t\cdot s}\right) + \mathbf{D}_2(\mathcal{X}^a_t, \mathcal{X}_t) + \mathbf{D}_3(\mathcal{P}^a_{t\cdot s}, \mathcal{P}_{t\cdot s}) \right) \right],$$
$$\text{s.t., } v_{t\cdot s} = \mathbf{M}(\mathcal{X}_t, \mathcal{P}_{t\cdot s}); \mathbf{R}(\mathcal{W}_{t\cdot s}) = 1,$$

where $\mathcal{W}_{t\cdot s} = \{w^1_{t\cdot s}, \dots, w^K_{t\cdot s}\}$ is a set of parameters to decide weights of particular data-driven views. $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are goodness-of-fit functions to measure a distance between two values; $\mathbf{R}(\mathcal{W}_{t\cdot s})$ is a constraint function, which gives the distribution of view weights. Without this constraint, the optimization problem is essentially unbounded. For the sake of simplicity, we set $\mathbf{R}(\mathcal{W}_{t\cdot s}) = 1$. Other constraint functions can also be used here since we can divide $\mathbf{R}(\mathcal{W}_{t\cdot s})$ by

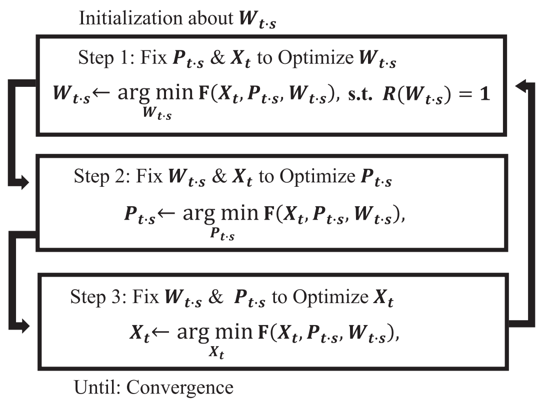

Fig. 4. Iterative multi-view calibration.

a constant. Therefore, **F** indicates the overall weighted distance between multiple observed traffic speeds to the model output, under consideration of *a priori* estimates of $\mathcal{P}_{t\cdot s}$ and $\mathcal{X}_t$. The rationale behind **F** is that for a view with a higher weight, we have a high penalty if the model output has a longer distance to the speed observed from this view; whereas, for a view with a lower weight, we have a low penalty if the output has a longer distance to the speed observed from this view. Thus, to minimize the objective function, the model output relies on the views with higher weights. To solve this, we aim to find the $\mathcal{X}_t$, $\mathcal{P}_{t\cdot s}$, and $\mathcal{W}_{t\cdot s}$ that minimize this overall weighted distance by an iterative learning process as follows.

4.2 Iterative Calibration

We develop an iterative learning technique based on the block coordinate descent. Since, in our objective function, we have three sets of parameters, i.e., a demand matrix $\mathcal{X}_t$, a model parameter set $\mathcal{P}_{t\cdot s}$, and a view-weight parameter set $\mathcal{W}_{t\cdot s}$, we aim to iteratively yet alternatively optimize these three sets of parameters until the final result converges. Specifically, we optimize the values of one parameter set to minimize the objective function while keeping the values of the other two parameter sets fixed, and then we continue this process by swapping the fixed parameter sets and the optimized parameter set until the result converges. We give a description of our iterative technique in Figure 4. We first initialize $\mathcal{W}_{t\cdot s}$ based on the estimation, because it does not affect the final result given the property of the block coordinate descent. In Step 1, we find the optimal $\mathcal{W}_{t\cdot s}$ that minimizes **F** by fixing initial $\mathcal{X}_t$ and $\mathcal{P}_{t\cdot s}$ (given by $\mathcal{X}_t^a$ and $\mathcal{P}_{t\cdot s}^a$). In Step 2, we find the optimal $\mathcal{P}_{t\cdot s}$ that minimizes **F** by fixing the optimized $\mathcal{W}_{t\cdot s}$ and initial $\mathcal{X}_t$. In Step 3, we find the optimal $\mathcal{X}_t$ that minimizes **F** by fixing the optimized $\mathcal{W}_{t\cdot s}$ and $\mathcal{P}_{t\cdot s}$. Next, we go back to Step 1 to optimize $\mathcal{W}_{t\cdot s}$ again based on updated $\mathcal{P}_{t\cdot s}$ and $\mathcal{X}_t$, and so on. Thus, this process is iterative by alternatively optimizing $\mathcal{P}_{t\cdot s}$, $W_{t\cdot s}$, and $\mathcal{X}_t$ until the result converges. The convergence is based on the distance and constraint functions in **F** according to the property of the block coordinate descent (Bertsekas 1999). As follows, we theoretically analyze the performance of this multi-view learning with respect to convergence.

4.3 Calibration Convergence

We use negative log function $\mathbf{R}(\mathcal{W}_{t\cdot s}) = \sum \mathrm{e}^{-w_{t\cdot s}^k}$ as our constraint function of view weights. This is because it maps a number between 0 and 1 to a number from 0 to ∞, which enlarges the difference between different view weights for better calibration. As for distance functions $\mathbf{D_1}$, $\mathbf{D_2}$, and $\mathbf{D_3}$, we use Normalized Squared Loss function given as $\mathbf{D}(\bar{v}^k_{t\cdot s}, v_{t\cdot s}) = \frac{(\bar{v}^k_{t\cdot s} - v_{t\cdot s})^2}{\mathrm{STD}(\bar{v}^1_{t\cdot s},\ldots,\bar{v}^k_{t\cdot s},\ldots,\bar{v}^K_{t\cdot s})}$. This function is an effective method to measure a distance between two variables and at the same time consider the distribution of $\bar{v}^k_{t\cdot s}$. As follows, we show the convexity and convergence of this iterative process given the constraint and distance functions.

Theorem. *If the above two functions are used, then convergence of the iterative calibration in Figure 4 is guaranteed.*

Proof. Based on the convergence proposition on the block coordinate descent (Bertsekas 1999), if the optimizations in Steps 1, 2, and 3 are convex, then the iterative process will converge to a stationary point. For Step 1, we aim to prove that if $\mathcal{X}_t$ and $\mathcal{P}_{t\cdot s}$ are fixed, the optimization for $\mathcal{W}_{t\cdot s}$ is convex. We introduce a variable $z^k_{t\cdot s} = \mathrm{e}^{-w^k_{t\cdot s}}$. Thus, the optimization becomes a new function with only one variable $z^k_{t\cdot s}$. $\min_{z^1_{t\cdot s},\ldots,z^K_{t\cdot s}} \mathbf{F} = \sum_{k=1}^{K}[-\log(z^k_{t\cdot s}) \cdot (\mathbf{D}(\bar{v}^k_{t\cdot s}, v_{t\cdot s}) + \mathbf{D}(\mathcal{X}^a_t, \mathcal{X}_t) + \mathbf{D}(\mathcal{P}^a_{t\cdot s}, \mathcal{P}_{t\cdot s}))]$ under a constraint $\sum_{k=1}^{K} z^k_{t\cdot s} = 1$. Thus, we have a linear constraint function with this variable $z^k_{t\cdot s}$, and a linear combination of negative log functions as the objective function. Both the constraint and objective functions are convex, so any local optimal solution is also the global optimal solution for Step 1. For the convexity of Steps 2 and 3, we can formulate the objective function with one variable as an unconstrained optimization. Since the distance function is convex in Steps 2 and 3, the objective function is also convex as they are a linear combination of convex functions. □

This theorem indicates our iterative calibration process can quickly converge based on these two functions, which makes it suitable for real-time modeling and calibration.

5 CONTEXT-AWARE DATA INFERENCE

We address a key practical challenge for multi-view model calibration introduced above, i.e., data incompleteness. Given national-scale and real-time requirements, data from neither private vehicle networks nor commercial vehicle networks are complete to serve as field data for multi-view calibration. Thus, we focus on improving single-source data by constructing a 3D tensor in Section 5.1, extracting three contexts in Section 5.2, and performing our context-aware decomposition in Section 5.3.

5.1 Tensor Construction

We infer traffic speeds on particular road segments for specific time slots by a 3D tensor $\mathcal{A} \in \mathbb{R}^{N\times K\times M}$.

- A traffic speed dimension indicates traffic speed categories (e.g., 0–5 mph, 5–10 mph): $[v_1, \ldots, v_N]$.
- A temporal dimension indicates a specific time slot (e.g., a 5-min slot from 0:00 AM to 0:05 AM): $[t_1, \ldots, t_K]$.
- A spatial dimension indicates specific spatial units (e.g., a road segment in a nation highway): $[s_1, \ldots, s_M]$.
- An entry $\mathcal{A}(n, k, m)$ indicates the number of GPS records given by a particular data source for a traffic speed category n in a spatial unit m during a slot k.

With either our private or commercial vehicle data, we fill this tensor $\mathcal{A}$, and then obtain a traffic speed distribution under a specific spatiotemporal partition. But a key challenge is that the tensor $\mathcal{A}$ is incomplete because of our fine-grained spatial-temporal coverage. For example, for a road

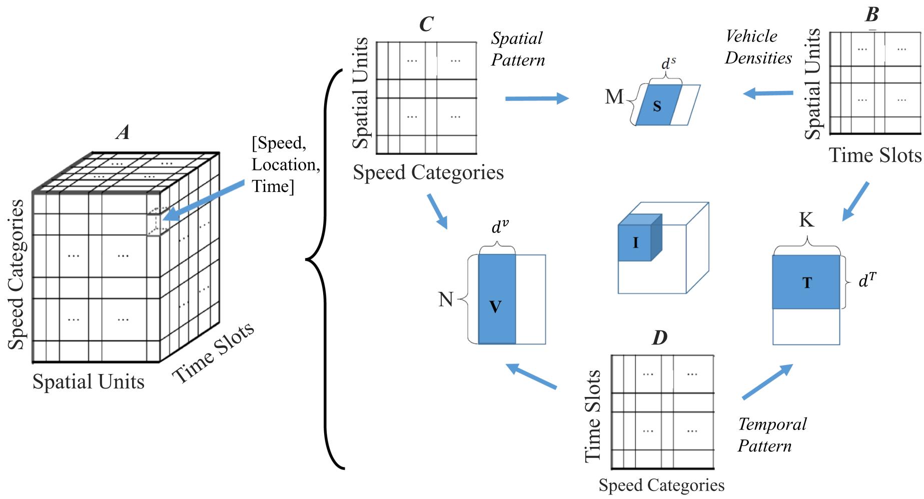

Fig. 5. Context-aware tensor decomposition.

segment without any vehicle in particular time slots, its corresponding entries are empty due to lacking GPS data.

A common approach by the machine learning community to address this data incomplete issue is to use tensor decomposition. For example, as in Figure 5, we have a tensor $\mathcal{A}$ with these three dimensions. An entry denotes a tuple [speed, location, time]. But $\mathcal{A}$ is sparse due to fine spatiotemporal granularity, e.g., we may have millions of road segments under 1 minute slots. Thus, we can decompose $\mathcal{A}$ into a core tensor $\mathcal{I}$ along with the other three matrices, $\mathcal{V} \in \mathbb{R}^{N \times d^v}$, $\mathcal{S} \in \mathbb{R}^{M \times d^s}$, and $\mathcal{T} \in \mathbb{R}^{K \times d^t}$, based on the Tucker decomposition model (Kolda and Bader 2009). $\mathcal{V}$, $\mathcal{S}$, and $\mathcal{T}$ infer correlations between different speed categories, different spatial units, and different time slots, respectively. d^v, d^s, and d^t are the numbers of latent factors and very small. The following objective function is often used to optimize the above decomposition by minimizing the decomposition error.

$$\min_{\mathcal{I},\mathcal{V},\mathcal{S},\mathcal{T}} \mathbf{L}(\mathcal{I},\mathcal{V},\mathcal{S},\mathcal{T}) = \mathbf{D}_4(\mathcal{A}, \mathcal{I} \times \mathcal{V} \times \mathcal{S} \times \mathcal{T}),$$

where $\mathbf{D}_4$ usually is a l_2-norm-based measurement, e.g., $||\mathcal{A} - \mathcal{I} \times \mathcal{V} \times \mathcal{S} \times \mathcal{T}||^2$. By minimizing this objective function, we obtain optimized $\mathcal{I}$, $\mathcal{V}$, $\mathcal{S}$, and $\mathcal{T}$ by the sparse tensor $\mathcal{A}$, which is given by real-time GPS data from either private vehicle networks or commercial vehicle networks. As a result, we use $\mathcal{I} \times \mathcal{V} \times \mathcal{S} \times \mathcal{T} = \mathcal{A}'$ to approximate $\mathcal{A}$ where $\times$ is the tensor-matrix multiplication.

However, in this work, we have a new challenge, i.e., $\mathcal{A}$ is over sparse at national scale with either private or commercial vehicles as shown in our motivation section, which leads to poor performance of traditional decomposition. To address this issue, we design a technique to use historical GPS data to establish several correlated contexts for context-aware tensor decomposition as follows.

5.2 Real-world Contexts

We use historical data to establish three contexts to provide additional info for our decomposition, i.e., temporal/spatial patterns for historical traffic, along with vehicle density. To formulate three contexts, we design three matrices as in Figure 5.

— Vehicle Densities, as in a matrix $\mathcal{B}$ where a row denotes a spatial unit, a column denotes a slot, and an entry denotes the average vehicle count in this spatial unit for this slot over a period of time.
— Spatial Patterns for Speed, as in a matrix $\mathcal{C}$ where a row denotes a spatial unit, a column denotes a speed category, and an entry denotes the number of GPS records in this speed category and this spatial unit over a period of time.
— Temporal Patterns for Speed, as in a matrix $\mathcal{D}$ where a row denotes a slot, a column denotes a speed category, and an entry denotes the number of GPS records in this speed category and this slot over a period of time.

All $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are obtained with historical GPS data.

5.3 Context-aware Tensor Decomposition

With the extracted contexts, we design joint context-aware decomposition with an objective function as follows.

$$\min_{\mathcal{I},\mathcal{V},\mathcal{S},\mathcal{T}} \mathbf{L}(\mathcal{I},\mathcal{V},\mathcal{S},\mathcal{T}) = \mathbf{D}_4(\mathcal{A}, \mathcal{I}\times\mathcal{V}\times\mathcal{S}\times\mathcal{T}) \\ +\mathbf{D}_4(\mathcal{B}, \mathcal{S}\times\mathcal{T}) + \mathbf{D}_4(\mathcal{C}, \mathcal{S}\times\mathcal{V}) + \mathbf{D}_4(\mathcal{D}, \mathcal{T}^T\times\mathcal{V}), \tag{5}$$

where the $\mathbf{D}_4$ usually is the l_2 norm; the first term measures decomposition errors about $\mathcal{A}$; the next three terms measure the error of factorizing matrices $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$. In this function, $\mathcal{A}$ and $\mathcal{B}$ share $\mathcal{S}$ and $\mathcal{T}$; $\mathcal{A}$ and $\mathcal{C}$ share $\mathcal{S}$ and $\mathcal{V}$; $\mathcal{A}$ and $\mathcal{D}$ share $\mathcal{V}$ and $\mathcal{T}$. Since $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are obtained by historical data, they lead to accurate $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{V}$, which improve decomposition of $\mathcal{A}$ itself. Thus, the historical traffic speed patterns are transferred into the decomposition of $\mathcal{A}$, which leads to better tensor decomposition for sparse tensors.

For the decomposition purpose, we first normalize all values to [0, 1] and set $d^v = d^s = d^t$. Next, we use an element-wise optimization algorithm as a numeric method (Karatzoglou et al. 2010) to obtain a local optimal solution for $\mathcal{I}$, $\mathcal{V}$, $\mathcal{S}$, and $\mathcal{T}$, since this objective function does not have a closed-form solution to find the global optimal solution. Finally, after we obtain $\mathcal{I}$, $\mathcal{V}$, $\mathcal{S}$, and $\mathcal{T}$, we use $\mathcal{I}\times\mathcal{V}\times\mathcal{S}\times\mathcal{T} = \mathcal{A}'$ to obtain complete tensors for both private and commercial vehicle networks as complete field data for multi-view model calibration.

6 IMPLEMENTATION AND EVALUATION

We test MultiCalib based on our data in Figure 3. Along with service providers from whom we have data access, we implement MultiCalib on 36 expressways and 119 highways in China, together with four major cities in China, i.e., Beijing, Shanghai, Guangzhou, and Shenzhen, to gain insights for their national-scale dispatching. As in Figure 6, we create a visualization of our implementation for three cities where we found that the commercial vehicles are mostly focused on major expressways, and the private vehicles are more evenly distributed around landmarks in the cities. In particular, we investigate MultiCalib on these cities' roads classified as motorways in OpenStreetMap (OSM 2016), which are highest-performance roads, e.g., in Shenzhen, we have 564 segments classified as motorways.

In this article, we focused on calibration to optimize model parameters with real-time traffic speeds, and it is independent from underlaying traffic models, and it has the capability to be applied to a wide range of traffic models. As a result, we implement a classic underlying traffic model based on C-logit as given in Equation (1) and then aim to compare two cases: (i) calibrating this model based on an existing traffic calibration model using a combination of two data sources without sophisticated data fusion and missing data inference; (ii) calibrating this model based on MultiCalib using multi data sources with data fusion and.

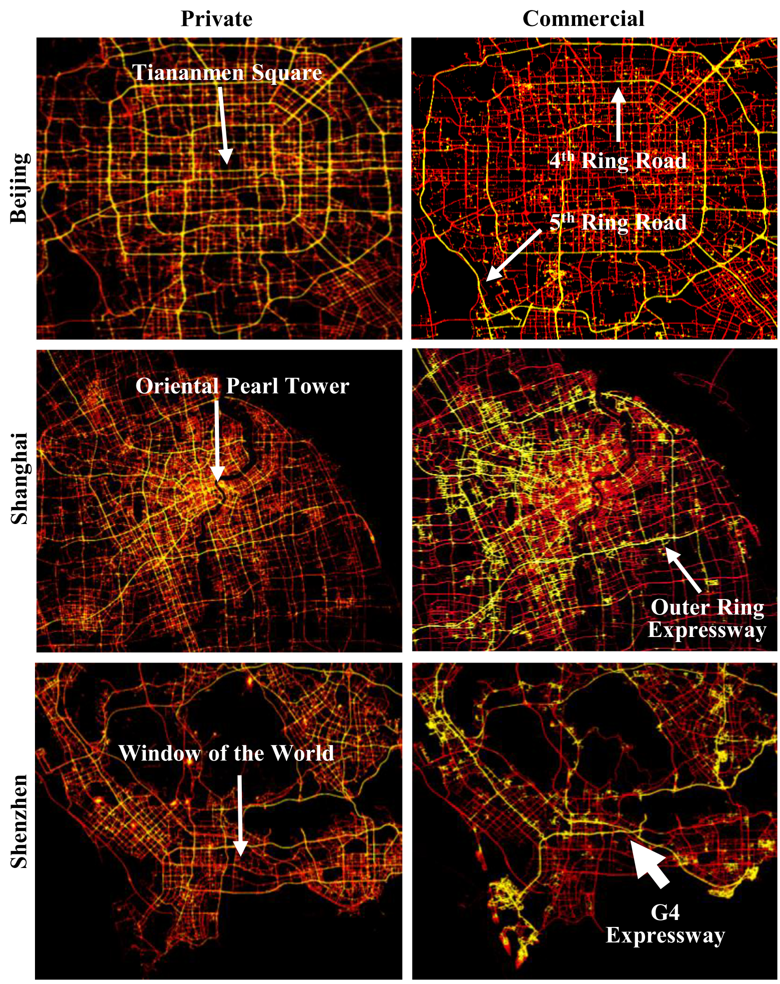

Fig. 6. Three cities for implementation and evaluation.

In our implemented version of this model, OD flows for multiple time intervals in different departure times of day are sequentially estimated with filed data (e.g., GPS data and loop detector data) and a route assignment model given in Equations (1) to (3). In particular, since we only consider cars, we do not have a mode choice model, i.e., all demand inferred is associated with cars. The survey paper (GiacomoPrato 2009) presented a detailed discussion of the route choice modeling.

We compare MultiCalib with one state-of-the-art calibration technique called Online calibration for Dynamic Traffic Assignment (OC-DTA) (Chiu et al. 2011), which is an efficient model calibration on the classic supply/demand model called Dyna Massachusetts Institute of Technology (MIT) (Ben-Akiva et al. 2012) based on single-source data. We aggregate GPS data from both vehicle networks to feed OC-DTA for a fair comparison. OC-DTA serves as a simple multi-source approach for model calibration where multiple data sources are combined straightforwardly. In contrast, MultiCalib uses multi-view-learning-based calibration with tensor decomposition for multi-source calibration.

We use cross-validation to evaluate the performance of MultiCalib with 13 weeks of data. We divide the 13-week dataset in Figure 3 into two subsets: one as a testing dataset containing data about a particular day, e.g., day d_1, and it is used to serve as the real-time streaming data; the other as a historical dataset containing data about the rest of days, serving as the historical training data for demand and supply models. If we use 5-min slots for modeling, for the first slot t_1, i.e., from 00:00 to 00:05, with the vehicle data in the historical training dataset, we use MultiCalib to calibrate a traffic model to find the optimal model parameters for all road segments in this slot. Then, we use the testing dataset (which gives us traffic speeds of a particular day) as the ground truth to verify the calibrated model. For some cities and highways, e.g., Shenzhen and G4, we have loop detector data to serve as the ground truth, which are inductive loops installed in selected major road segments and can detect metal and, thus, accurately detect vehicle speeds. We let the testing dataset rotate among all data, leading to multiple sets of experiments. The average results are reported.

We test the MultiCalib with Mean Average Percent Error (MAPE) as

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|\bar{\mathbf{v}}_i - \mathbf{v}_i|}{\bar{\mathbf{v}}_i},$$

where n is the total number of temporal-spatial combinations we tested, e.g., in 10 min slots, we have $24 \times \frac{60}{10} \times (36 + 119) \times 1{,}250 = 2{,}790{,}000$ combinations for a one-day evaluation given we have (36+119) routes and 1,250 segments per route. Under a temporal-spatial combination i, $\mathbf{v}_i$ is the traffic speed inferred by a calibrated model, while $\bar{\mathbf{v}}_i$ is ground truth. An accurate calibration yields a small MAPE, and vice versa.

In addition to the end-to-end results on the speed calibration accuracy, we also provide intermediate results on how well we infer the missing data based on our tensor-decomposition technique to quantitatively evaluate this technical contribution. This evaluation is based on cross-validation where we randomly delete x% of data we already have from our tensor and assume these data were missing data. Then, we use our techniques to infer these data along with other truly missing data based on the remaining known data. Since we have the ground truth of these x% deleted data, we compare our inferred results with the ground truth and then use the percentage of inference accuracy as a metric to evaluate the performance of our missing data inference. We let this deleted data rotate among all known data, leading to multiple sets of experiments. The average results are reported.

We first show results in four cities and four particular expressways from G1 to G4, along with the average result on all highways. Then, we study impacts of slot lengths. Further, we investigate the impact of historical data sizes on the running time and the accuracy of MultiCalib to show its feasibility and robustness for real-world calibration. Then, we report the performance of our data inference based on tensor decomposition. Finally, we present an evaluation summary.

6.1 Impact of Cities

Figures 7–10 plot the average MAPE under 5-min slots for all motorways in Beijing, Shanghai, Guangzhou, and Shenzhen, respectively. In general, we found that MultiCalib has better performance than OC-DTA, e.g., in Beijing, MultiCalib outperforms OC-DTA by 28%. This is because OC-DTA only straightforwardly uses multi-source GPS data to calibrate traffic models without data inference and multi-view perspectives; whereas MultiCalib uses a data inference technique to improve the completeness of data and then uses a multi-view-learning-based calibration to optimize the model. For all cities, we found the performance of calibration becomes better for both techniques in the morning and evening rush hour because there are abundant field data to calibrate the model during that time with low speeds. But for the early morning or late night, their performance becomes worse due to incomplete field data. As for Beijing, the variation of calibration performance is larger than Shanghai and Guangzhou, e.g., during the evening rush hour, the

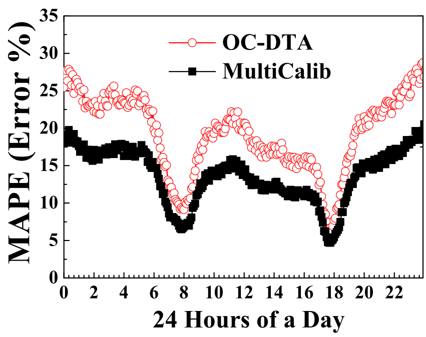

Fig. 7. Beijing.

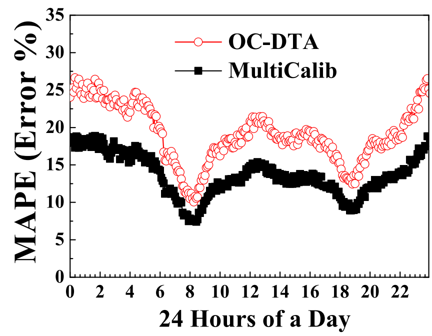

Fig. 8. Shanghai.

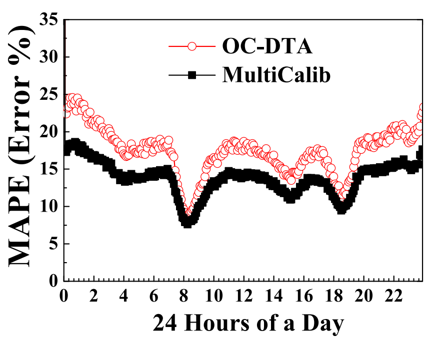

Fig. 9. Guangzhou.

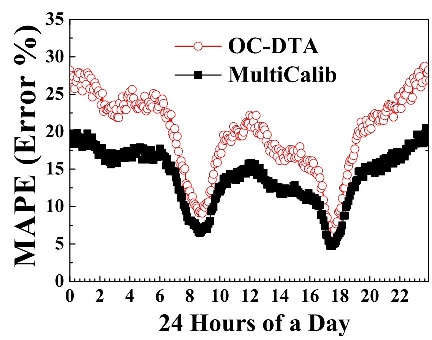

Fig. 10. Shenzhen.

MAPE in Beijing is close to 5%, but for the early morning its MAPE is up to 29%. It may be because of the heavy traffic flow in Beijing compared to other cities. For Shenzhen, the MAPE becomes lower starting at the early afternoon compared to others, which indicates a different pattern for the evening rush hour in Shenzhen. For the morning rush hour, all cities' performance becomes the best around 8AM, which is the most crowded hour in many cities.

6.2 Impact of Expressways

Figures 11–14 plot the average MAPE under 5-min slots for all major road segments in China's national expressways G1, G2, G3, and G4, respectively. We also found that MultiCalib outperforms OC-DTA in general. But the main difference between the calibration on national expressways and cities is that there are no clear morning and even rush hours in expressways. Typically, for all expressways, the performance of calibration for both techniques becomes better from 6AM and worse from 10PM. One of the possible explanations is that during these time periods, the field data are more complete due to a large number of vehicles on the expressways, which leads to better calibration. For G1 connecting Beijing to Harbin, MAPE is smooth during the daytime, which may be because it has a short distance and is the only expressway to the northeastern part of China, thus leading to heavy traffic. For G2 connecting Beijing to Shanghai, MAPE is also smooth during the daytime but with some effects of the evening rush hour because it connects two of the biggest cities in China. For G3 connecting Beijing to Fuzhou (with future planning to Taiwan), MAPE is higher during the noon or afternoon, which may be because of limited numbers of vehicles on this expressway due to its current construction and no major cities in the middle of G3. For G4 connecting Beijing and HongKong, MAPE is lower in the morning rush hour, the late afternoon,

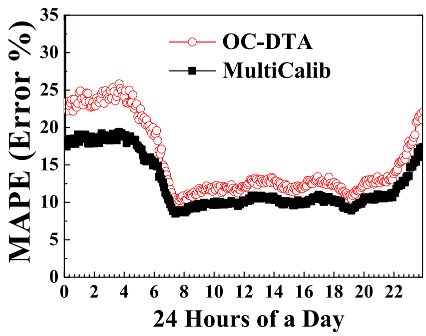

Fig. 11. G1 expressway.

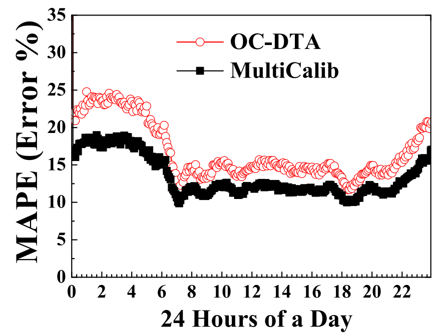

Fig. 12. G2 expressway.

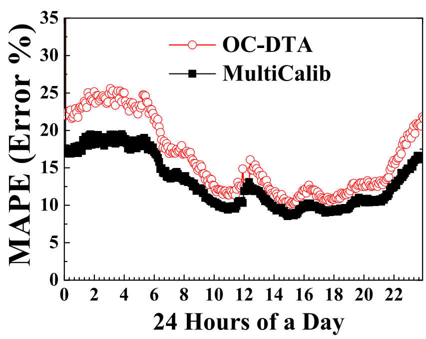

Fig. 13. G3 expressway.

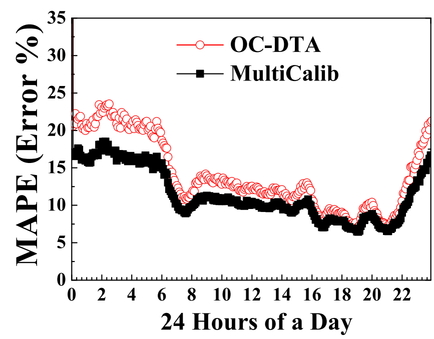

Fig. 14. G4 expressway.

and the early evening due to heavy traffic on this major expressway with several big cities with tens of millions of people.

Figure 15 gives the average MAPE for all 119 national highways and 36 expressways under 10-min slots in 24 hours. The MAPE of both calibration techniques are typically higher than the MAPE we found from Figures 11 to 14. This is because the data coverage percentage may change dramatically between different highways and some highways with few vehicles uploading GPS data lead to higher MAPE. But the relative performance between the two calibration techniques is similar. MultiCalib has an average MAPE of 17.2%, while OC-DAT has an average MAPE of 23.2%, which leads to a 25% performance gain for MultiCalib. The performance gains are more obvious in the early morning, which may be the result from our inference technique.

In addition to these end-to-end results, we further investigate the impact of our tensor-decomposition-based data inference on the performance of MultiCalib. In particular, we show that removing the data inference part on MultiCalib, and use a model calibration without tensor decomposition in Figure 15. In this variation of MultiCalib, we utilize a simple interpolation to infer the spatial temporal unites without data, whereas MultiCalib utilizes the tensor-decomposition-based technique to infer the missing data. We found that the performance of MultiCalib without Decomposition is between the full version of MultiCalib and OC-DTA. This is because without context-aware tensor decomposition, MultiCalib based on the interpolation cannot have high-quality data for calibration. However, we found that the MultiCalib without Decomposition still outperforms OC-DAT, which indicates the data fusion component still works fine even with a simple data inference technique.

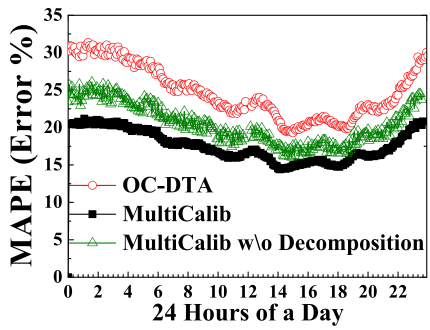

Fig. 15. Average MAPE.

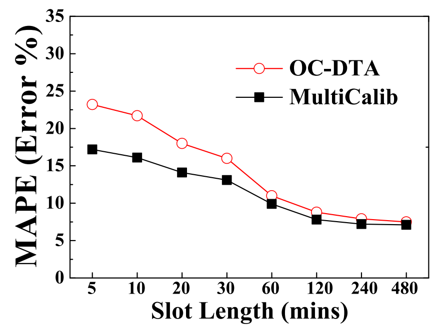

Fig. 16. Effects of lengths.

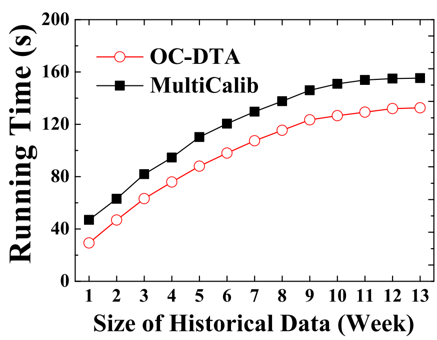

Fig. 17. History vs. Time.

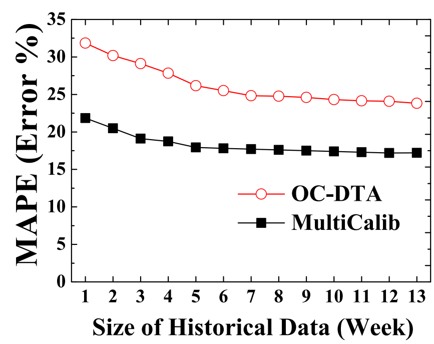

Fig. 18. History vs. MAPE.

6.3 Impact of Slot Lengths

Figure 16 plots both MAPE with different slot lengths with a default value of 5 mins. The MAPE of both calibration techniques decreases with the lengths of the time slots. This is because in a longer slot, more data about vehicles can be accumulated to infer traffic speeds, and also the traffic speed becomes more stable. MultiCalib outperforms OC-DTA significantly if the slot is shorter than 20 mins because MultiCalib uses a data inference technique to improve this data completeness. This advantage is more obvious during short time slots where data are more incomplete. But when the slot becomes longer than 1 hour, both techniques have similar performance. This is because in such a long slot, we have enough data for calibration of relatively stable speeds.

6.4 Impact of Historical Data

We study the impact of historical data on model accuracies and running time by comparing MultiCalib to OC-DTA with a default value of 13 weeks. Figures 17 and 18 plot running time and MAPE on different lengths of historical data in terms of weeks. As expected, the more the historical data, the longer the running time; the more accurate the model calibration, the lower the MAPE. MultiCalib has 15% longer time than OC-DTA, which in turn leads to a 25% lower MAPE. This is because MultiCalib has to perform its tensor decomposition to infer incomplete data along with iterative calibration, which leads to longer running time.

6.5 Performance of Missing Data Inference

In this section, we report our results on missing data inference to show the performance of our data inference techinque based on tensor decomposition. In particular, we investigate the impact

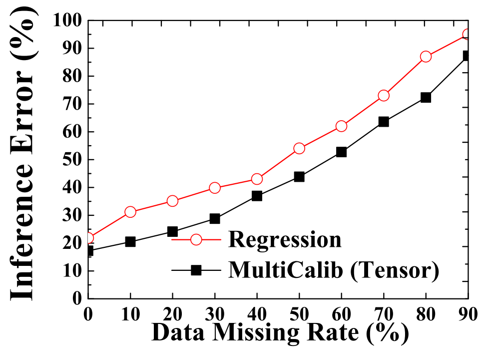

Fig. 19. Missing data inference.

of different data missing rates from 0% missing to the 90% missing on the performance of our technique. In Figure 19, the *X* axis is different data missing rates and the *Y* axis is the percentage of inference errors. In this experiment, we define an inference error if the inferred value is 5% smaller or bigger than the ground truth. We found that with the increase of the data missing rates, the performance of our data inference technique based on tensor decomposition becomes worse. This is because when the data missing rate becomes higher, the remaining data cannot provide enough contexts for a high performance decomposition, which makes the recovered tensors very generic. However, we found that compared to a baseline approach based on regression, our tensor-decomposition-based approach always performs better, which validates the performance of our data inference technique to improve later traffic speed calibration.

6.6 Evaluation Summary

We have the following observations. (i) The calibration performance is highly dependent on both venues and time as shown by Figures 7 and 14. Every city or expressway has its own characteristics, which need to be considered when modeling. On average, all techniques have better performance during the time when more data can be used for calibration as in Figure 15. (ii) The length of slots significantly affects the performance of calibration as in Figure 16. Normally, a longer slot has better performance than a shorter slot, but a longer slot has low usability in real-time applications. (iii) As in Figures 17 and 18, MultiCalib has a longer running time, but it increases its accuracy. (iv) Looking across factors, it seems spatiotemporal contexts have the highest impact, and then slot lengths, and finally historical data length. (v) Our data inference technique performs better than a regression-based baseline approach, but of the approaches, it performs worse when more data were missing, which indirectly validates the importance of speed data for our calibration model.

7 MULTICALIB APPLICATION: NATIONAL-SCALE DISPATCHING

In this section, based on the traffic models obtained by MultiCalib, we design, implement, and evaluate a national-scale dispatching service from commercial fleets to reduce their traveling time for individual vehicles at national-scale road networks. Based on traffic models, we aim to estimate travel time for a few shortest pathes between origins and destinations of individual vehicles in real time. A better model can predict more accurately about travel time, which leads to shorter overall travel time for all vehicles in fleets. In the rest of the section, we first introduce service design, and then present our service evaluation.

7.1 Service Design

Our application is targeted at national-scale fleets with a large number of vehicles traveling on national highways and expressways on a daily basis. Given origin and destination pairs of these vehicles, the current dispatching strategy is very straightforward, mostly based on shortest travel times. However, the national highway and express way networks experience temporal dynamics due to various events, e.g., weather, accidents, or roadwork. These events are hard to predict in advance and thus have significant impacts on travel time on national road networks, leading to inaccurate travel time estimation. An accurate traffic speed model at national scales, e.g., calibrated by MultiCalib, could provide better estimation for travel time on fine-grained road segment levels. Based on an accurate traffic model, a dispatching strategy can potentially reduce the actual travel time for vehicles in a fleet, by finding the route with shorter real-time travel time from all regular routes between an origin-destination pair.

For nationwide trips with shorter distances or on routes with less traffic dynamics, an offline calibrated traffic speed model would suffice for scheduling services. However, for some long-distance nationwide travels on routes with rather dynamic traffic speeds, an offline calibrated traffic speed model, even with a good calibration technique such as MultiCalib, cannot address spatiotemporal dynamics of nationwide traffic. As a result, we design two versions of scheduling based on offline and online traffic speed calibration as follows.

7.2 Offline Calibration-based Dispatching

In our implementation, we utilize a simplified version of a dispatching algorithm (Ichoua et al. 2003) by the following:

- Processing the GPS data at individual vehicle levels to obtain all origin/destination pairs for all vehicles;
- Using a traffic speed model calibrated by a model calibration technique (e.g., MultiCalib or OC-DTA) to estimate travel time for these routes;
- Recommending the route with the shortest expected travel time for the vehicle for this particular origin/destination pair.

We note that a more sophisticated dispatching can improve the total travel time at the application level, but our main objective here is to show, even for a simplified dispatching algorithm, our calibration model MultiCalib can significantly improve its performance compared with a baseline. As a result, we compared the same dispatching algorithm using the same traffic speed model but calibrated by different calibration models, i.e., our calibration model MultiCalib and the baseline model OC-DTA. This is an indirect approach to demonstrate our calibration performance on the application level.

7.3 Online Calibration-based Dispatching

Compared to the offline calibration-based dispatching, our online calibration is representative to a wide range of scheduling algorithms that update the recommended routes based on real-time data. However, our calibration will make the traffic speed model more accurate, which leads to better performance of final dispatching. Our online calibration-based dispatching is given as follows.

- Processing the GPS data at individual vehicle levels to obtain all origin/destination pairs for all vehicles;
- Calibrating a traffic speed model based on a model calibration technique (e.g., MultiCalib or OC-DTA) and the real-time data on these routes;
- Estimating travel time for these routes based on recently-calibrated speed models;

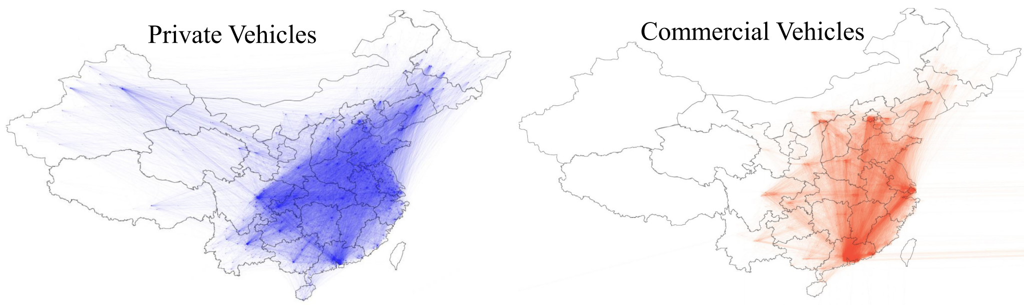

Fig. 20. Nationwide OD pattern.

—Recommending the route with the shortest expected travel time for the vehicles for this particular origin/destination pair based on the latest travel time estimation.
—Periodically going back to the section step to calibrating the traffic speed model again based on recently received real-time speed data.

Compared to the offline calibration-based dispatching, the online calibration-based dispatching considers the constant changes of traffic speeds nationwide and is more responsive to traffic dynamics. Further, a calibration technique based on multi-source data, e.g., MultiCalib, will be more robust to the data uncertainty or data missing issues compared to the calibration based on single-source data. It has the potential to lead to better chosen routes given the same scheduling technique.

In our online calibration-based dispatching, a key issue to address is how often do we calibrate our speed models based on new data and how often do we recalculate the dispatching routes. As for the model calibration frequency, we utilize the update speed model for every 5 mins since travel speeds will not change significantly in such a short time (Zhang et al. 2015); whereas the dispatching results are updated if the existing route is 10% longer that the current route to avoid frequent adjustment of recommended routes.

7.4 Service Evaluation

In this article, we use 3-month commercial fleet GPS data from 45,237 vehicles to evaluate the performance of our services. For every vehicle in the fleet, we first extract their origin and destination for all their trips. Then we use our dispatching strategy (either online or offline) to recommend routes for all vehicles between their OD pairs. We visualize all major OD pairs for private vehicles and commercial vehicles in Figure 20. We found that commercial vehicles are more concentrated between major cities; whereas private vehicles are spread out around the entire country.

Based on our dispatching service, the total travel time for all vehicles in the fleet is expected to be reduced compared to the current situation without traffic-model-based dispatching. In particular, we divide the 3-month data into two subsets: *Testing Set*, which includes the data for one particular day as the real-time data to obtain origins and destinations for all vehicles in this particular day; *Historical Set*, which includes the data for 90 continuous days preceding the day in the testing set as the historical data for traffic speed modeling.

We compared our two MultiCalib-based dispatching strategies (i.e., the offline and online version) with a dispatching strategy based on OC-DTA (Chiu et al. 2011), which is an efficient model calibration on the classic supply/demand model called DynaMIT (Ben-Akiva et al. 2012) based on single-source data, as introduced in Section 6. Since traffic speed models calibrated by MultiCalib with multi-source data have better performance than OC-DTA based on single-source data, the

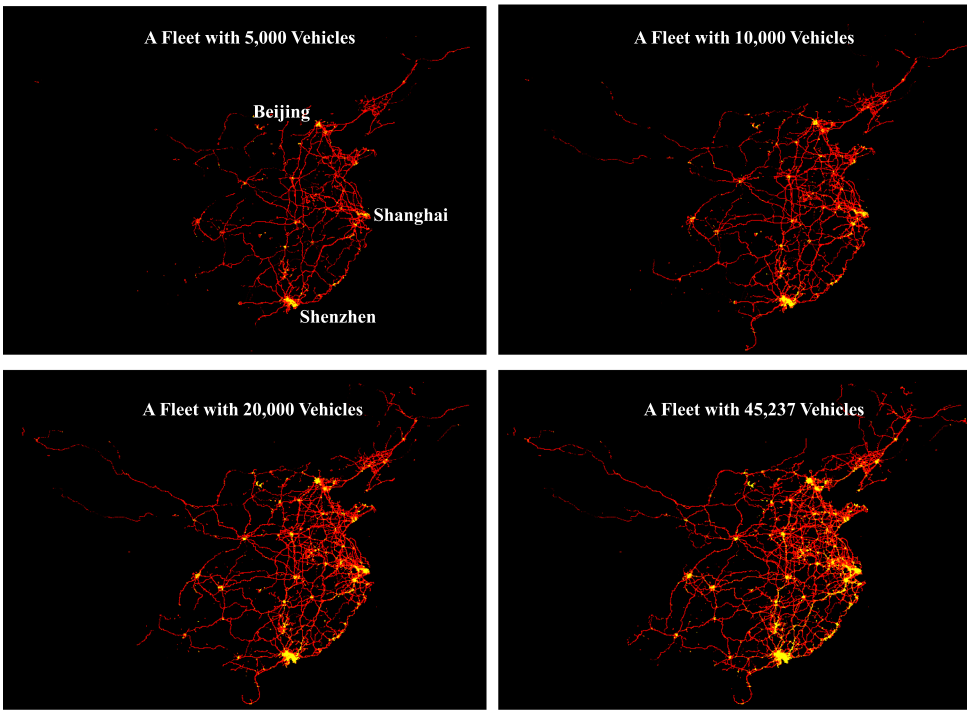

Fig. 21. Visualization of dispatching services at four scales.

dispatching strategy based on MultiCalib is expected to dispatch vehicles to faster routes than the dispatching strategy based on OC-DTA. We use the percentage of reduced travel time as a metric to evaluate the performance of these two dispatching strategies, which is obtained by the original travel time (calculated by GPS traces), and the simulated travel time (calculated by our trace-driven simulation). We also evaluate the impact of fleet scales from 5,000 vehicles to 45,000 vehicles on performance of our dispatching service. Normally, the more vehicles in a fleet, the better modeling for traffic speeds at national-scale road networks, and the shorter expected travel time for all the vehicles in a fleet. A visualization of our dispatching traces at four scales is given in Figure 21.

Figures 22 to 25 plot the percentages of the reduced travel time in 24 hours for four scales of fleets by comparing two dispatching strategies driven by two traffic models, i.e., MultiCalib and OC-DTA. The results are given under 5-min slots in 24 hours. In general, we found that the dispatching strategy based on MultiCalib has better performance than the dispatching strategy based on OC-DTA, e.g., based on a fleet with 5,000 vehicles, the dispatching strategy based on MultiCalib outperforms the one based on OC-DTA by 27% on average. This is because the traffic model used by the dispatching strategy based on OC-DTA only utilizes multi-source GPS data for a straightforward calibration without sophisticated data inference and multi-view perspectives; whereas the dispatching strategy based on MultiCalib uses the introduced data inference technique to improve the completeness of data and then a multi-view-learning-based calibration to optimize the model, which leads to better traffic modeling. For all scales, we found the performance of dispatching becomes better for both techniques in the early morning and late night. One possible reason is that the traffic is more unpredictable during the early morning and late night. The original dispatching strategy without traffic modeling leads to routes with longer travel time. But for the regular daytime, their performance on reduced travel time becomes rather stable.

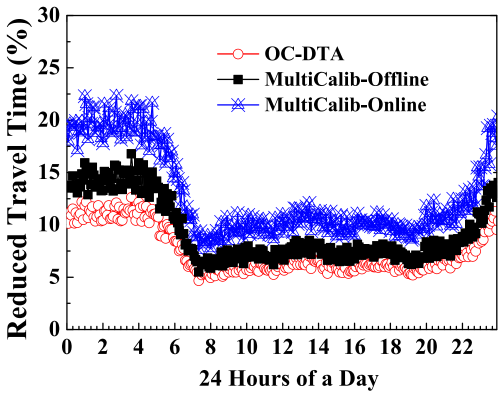

Fig. 22. 5,000 vehicle fleet.

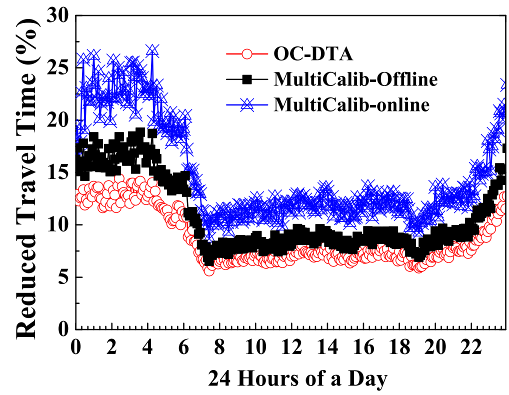

Fig. 23. 10,000 vehicle fleet.

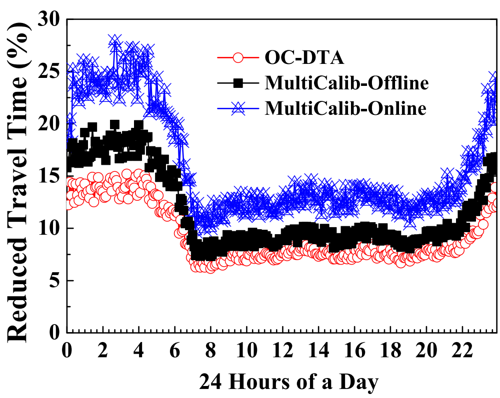

Fig. 24. 20,000 vehicle fleet.

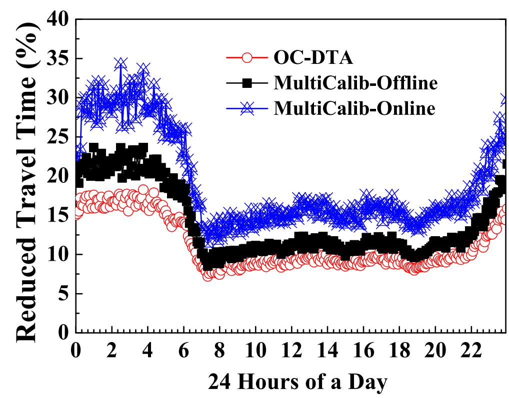

Fig. 25. 45,237 vehicle fleet.

Comparing four different scales, we found that, in general, the larger the fleet scale, the better the performance for all dispatching strategies. For example, with a fleet with 5,000 vehicles, the offline dispatching strategy based on MultiCalib outperforms the one based on OC-DTA by 27% on average; but with a fleet with 10,000 vehicles, this performance gain increases to 33% on average. The largest performance gain is obtained when all 45,237 vehicles are used in the fleet for traffic modeling where the offline dispatching strategy based on MultiCalib outperforms the one based on OC-DTA by 39%. This is because more vehicles lead to high coverage of national road networks, which leads to better traffic modeling and more accurate prediction for expected travel times used in the dispatching strategies.

Further, when we compared the offline and the online dispatching strategies based on MultiCalib, we found the performance of online dispatching is always better than the offline dispatching by 42% on average across all different vehicle densities from 5,000 to 45,237 vehicles. This is because the online dispatching strategy can detect the constant change of traffic speed on the nationwide road segments, and just the recommended routes accordingly if there is a significant travel time improvement. In contrast, the offline dispatching strategy only recommends routes once when the trip started and cannot adjust the routes based on real-time traffic. Interestingly, the performance gain between these two kinds of dispatching increases with densities of vehicles. One possible explanation is that more vehicles provide more timely updates for traffic speeds. As a result, it leads to better performances when more vehicles are contributing data to the speed models and calibration models.

7.5 Potential Insights for Performance Comparsion

In the above section, we provide the experimental discussion as to why MultiCalib performs better in the same scheduling algorithm compared to other baseline model calibration techniques. As follows, we provide some analytical discussion to explain conceptually why MultiCalib is better than the baseline OC-DTA. For the baseline OC-DTA, even though we feed the data of both vehicle networks to OC-DTA for a fair comparison, OC-DTA did not consider the difference between these two vehicle networks (e.g., max speed limitation), and only treats them as a single network with more vehicles. In contrast, our MultiCalib considers the different operational natures of private vehicles and commercial vehicles to weight their data contributions through our iterative learning to the final model calibration process, which helps to update models more accurately for dispatching services. For example, assuming we have a highway route segment with one private vehicle (i.e., cars) and 10 commercial vehicles (i.e., trucks), they will contribute their speed data to the model collaboration process. Since we have many more trucks with lower travel speeds compared to the one car with a higher speed we have on the same road segment, OC-DTA will calibrate the model much more toward the speeds of trucks. In contrast, MultiCalib balances the speed estimations by considering the weights of each type of vehicle, which is irrelevant to their number on road segments. In this case, MultiCalib can provide an unbiased speed estimation to calibrate speed models, which leads to better performance of the resultant scheduling algorithms.

8 RELATED WORK

MultiCalib is related to traffic modeling and calibration as well as multi-source data-driven systems.

8.1 Traffic Modeling and Calibration

Traffic modeling and its calibration are very important for urban transportation and planning (Zhang et al. 2015). Recently, with the increasing availability of vehicle GPS data, numerous modeling and calibration techniques have been proposed for transportation services (Rahman et al. 2013), e.g., inferring taxicab passenger demand (Ge et al. 2010; Huang and Powell 2012; Yuan et al. 2011b), assisting regular drivers for route planning (Yuan et al. 2011a); estimating traffic volumes or speeds (Aslam et al. 2012); and assigning dynamic traffic to road segments (Ben-Akiva et al. 2012). To improve the performance of these models, several calibration techniques have also been proposed, e.g., jointly calibrating both demand and supply models (Chiu et al. 2011). These modeling and calibration techniques can be used by many applications, e.g., navigation and dispatching (Rahman et al. 2013), to improve transportation efficiency.

However, all of the above work is for urban-scale modeling and calibration, and their key assumption is that single-source field data for modeling and calibration, e.g., loop sensors, camera data, or taxi data, are complete (Nouir et al. 2008). But for national-scale modeling and calibration, this assumption cannot be held anymore because we do not have single-source infrastructure at national scale to provide compete field data. In this work, from a multi-source data perspective, we investigate and improve incomplete data from two national-scale vehicle networks by a data inference technique based on context-aware tensor decomposition, along with multi-view-learning-based calibration. Under national-scale and multi-source data contexts, these two techniques make our work significantly different from previous traffic model calibration based on single-source data at urban scale.

8.2 Vehicular Data-driven Systems

Our work is also related to vehicular data-driven systems. Recently, numerous novel systems have been proposed based on various vehicular data to improve urban efficiency (Wu et al. 2009). Here, we focus on the work closely related to models based on vehicular GPS data (Zheng et al. 2010). For example, some systems have been proposed to (i) enable taxi drivers to find their passengers more efficiently (Ge et al. 2010; Xie et al. 2018; Zhang and He 2012; Zhang et al. 2013a, 2013b, 2016, 2017; Zheng et al. 2008); (ii) assist transit passengers to query about their travel time, cost, or modalities (Balan et al. 2011; Wu et al. 2012); (iii) help regular drivers based on vehicle GPS data to improve their driving performance (Yuan et al. 2011a), to estimate traffic speeds (Aslam et al. 2012; Zhou et al. 2018; Yang et al. 2018), and to detect driver fraud (Zhang et al. 2011); (vi) model and improve the urban transportation systems (Zheng et al. 2010; Sen and Balan 2013; Shang et al. 2014); and (v) study human mobility based on vehicular GPS data (Zhang et al. 2015, 2014; Bhattacharya et al. 2013; Lathia and Capra 2011; Fang et al. 2018). However, the above systems are mostly based on single-source data.

9 CONCLUSION

In this work, we design MultiCalib to effectively calibrate national-scale traffic models based on multi-source incomplete data in real time. More importantly, we implement and evaluate MultiCalib based on extremely-large-scale infrastructures in China to test its performance. Our efforts lead to a few valuable insights for fellow researchers to design and implement real-time data-driven models at national scale. Specifically, these insights are that (i) heterogeneous physical systems provide complementary data, which can be intelligently integrated together to improve the performance of modeling and calibration; (ii) even though field data are mostly incomplete to model national-scale phenomenon in real time, a data inference technique based on historical data and real-world contexts can improve completeness of data, thus enabling better modeling; (iii) instead of simply combining different data sources together, a weight-based modeling technique from multi-view perspectives leads to better performance by an iterative measuring process.

REFERENCES

Open street map. 2016. Retrieved from http://www.openstreetmap.org/.

C. Antoniou, M. Ben-Akiva, and H. N. Koutsopoulos. 2007. Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. *IEEE Trans. Intell. Transp. Syst.* 8, 4 (2007), 661–670.

Javed Aslam, Sejoon Lim, Xinghao Pan, and Daniela Rus. 2012. City-scale traffic estimation from a roving sensor network. In *Proceedings of 10th ACM Conference on Embedded Network Sensor Systems (SenSys'12).*

Ramachandran Balakrishna, Constantinos Antoniou, Haris N. Koutsopoulos, Yang Wen, and Moshe Ben-Akiva. 2011. Dynamic traffic assignment. In *Transportation Research E-Circular (2011-6).*

Rajesh Krishna Balan, Khoa Xuan Nguyen, and Lingxiao Jiang. 2011. Real-time trip information service for a large taxi fleet. In *MobiSys'11.*

M. Ben-Akiva, M. Bierlaire, H. N. Koutsopoulos, and R. Mishalani. 2012. Real-time simulation of traffic demand-supply interactions within DynaMIT. *Transportation and Network Analysis: Current Trends.* Springer, 19–36.

Dimitri P. Bertsekas. 1999. Non-linear programming. In *Athena Scientific.* Athena Scientific. https://books.google.com/books?id=TwOujgEACAAJ.

Sourav Bhattacharya, Santi Phithakkitnukoon, Petteri Nurmi, Arto Klami, Marco Veloso, and Carlos Bento. 2013. Gaussian process-based predictive modeling for bus ridership. In *UbiComp'13.*

Yichang Chiu, Jon Bottom, Michael Mahut, Alex Paz, Ramachandran Balakrishna, Travis Waller, and Jim Hicks. 2011. Dynamic traffic assignment. In *Transportation Research CIRCULAR E-C153.* Retrieved from http://onlinepubs.trb.org/onlinepubs/circulars/ec153.pdf.

B. Ciuffo and C. Lima Azevedo. 2014. A sensitivity-analysis-based approach for the calibration of traffic simulation models. *IEEE Trans. Intell. Transp. Syst.* 15, 3 (2014).

Zhihan Fang, Fan Zhang, and Desheng Zhang. 2018. MultiCell: Urban population modeling based on multiple cellphone networks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*

Yong Ge, Hui Xiong, Alexander Tuzhilin, Keli Xiao, and Marco Gruteser. 2010. An energy-efficient mobile recommender system. In *KDD'10*.

Carlo GiacomoPrato. 2009. Route choice modeling: Past, present and future research directions. *J. Choice Modelling*.

Yan Huang and Jason W. Powell. 2012. Detecting regions of disequilibrium in taxi services under uncertainty. In *SIGSPATIAL'12*.

Soumia Ichoua, Michel Gendreau, and Jean-Yves Potvin. 2003. Vehicle dispatching with time-dependent travel times. *Eur. J. Oper. Res.*

Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *RecSys'10*.

T. G. Kolda and B. W. Bader. 2009. Tensor decompositions and applications. *SIAM Rev. 51*.

Neal Lathia and Licia Capra. 2011. How smart is your smartcard?: Measuring travel behaviours, perceptions, and incentives. In *UbiComp'11*.

Jerome Le Ny, Ahmed Touati, and George J. Pappas. 2014. Real-time privacy-preserving model-based estimation of traffic flows. In *ICCPS'14*.

Stefan Mitsch, Sarah M. Loos, and Andre Platzer. 2012. Towards formal verification of freeway traffic control. In *ICCPS'12*.

Z. Nouir, B. Sayrac, B. Fourestie, W. Tabbara, and F. Brouaye. 2008. Boosting predictions by calibration of traffic model and learning of indicators' distributions. *IEEE Trans. Veh. Technol.* 57, 4 (2008), 2489–2497.

Mohammad Ashiqur Rahman, Qi Duan, and Ehab Al-Shaer. 2013. Energy efficient navigation management for hybrid electric vehicles on highways. In *ICCPS'13*.

Rijurekha Sen and Rajesh Krishna Balan. 2013. Challenges and opportunities in taxi fleet anomaly detection. In *SENSEMINE'13*.

Jingbo Shang, Yu Zheng, Wenzhu Tong, Eric Chang, and Yong Yu. 2014. Inferring gas consumption and pollution emission of vehicles throughout a city. In *KDD'14*. ACM.

Naonori Ueda, Futoshi Naya, Hitoshi Shimizu, Tomoharu Iwata, Maya Okawa, and Hiroshi Sawada. 2015. Real-time and proactive navigation via spatio-temporal prediction. In *UbiComp'15*.

Eugene Wu, Philippe Cudre-Mauroux, and Samuel Madden. 2009. Demonstration of the TrajStore system. *Proc. VLDB Endow.* (2009).

Wei Wu, Wee Siong Ng, Shonali Krishnaswamy, and Abhijat Sinha. 2012. To taxi or not to taxi?—enabling personalised and real-time transportation decisions for mobile users. In *MDM'12*.

Xiaoyang Xie, Fan Zhang, and Desheng Zhang. 2018. PrivateHunt: Multi-source data-driven dispatching in for-hire vehicle systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 45 (March 2018), 26 pages. DOI: https://doi.org/10.1145/3191777

Yu Yang, Fan Zhang, and Desheng Zhang. 2018. SharedEdge: GPS-free fine-grained travel time estimation in state-level highway systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 48 (March 2018), 26 pages. DOI: https://doi.org/10.1145/3191780

Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. 2011a. Driving with knowledge from the physical world. In *KDD'11*.

Jing Yuan, Yu Zheng, Liuhang Zhang, XIng Xie, and Guangzhong Sun. 2011b. Where to find my next passenger. In *UbiComp'11*.

Desheng Zhang and Tian He. 2012. pCruise: Reducing cruising miles for taxicab networks. In *Proceedings of the IEEE 33rd Real-Time Systems Symposium (RTSS'12)*. IEEE, 85–94.

Desheng Zhang, Tian He, Shan Lin, Sirajum Munir, and John A. Stankovic. 2017. Taxi-passenger-demand modeling from a roving sensor network. *IEEE Trans. Big Data (IEEE TBD)* 3, 3 (2017), 362–374.

Desheng Zhang, Tian He, Yunhuai Liu, and John A. Stankovic. 2013a. A unified recommendation system for carpooling and regular taxicab services. In *Proceedings of the IEEE International Conference on Big Data*. IEEE, 439–447.

Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*. ACM, 201–212.

Daqing Zhang, Nan Li, Zhi-Hua Zhou, Chao Chen, Lin Sun, and Shijian Li. 2011. iBAT: Detecting anomalous taxi trajectories from GPS traces. In *UbiComp'11*.

Desheng Zhang, Ye Li, Fan Zhang, Mingming Lu, Yunhuai Liu, and Tian He. 2013b. coRide: Carpool service with a win-win fare model for taxicab networks. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 9 pages.

Desheng Zhang, Fan Zhang, and Tian He. 2016. MultiCalib: National-scale traffic model calibration in real time with multi-source incomplete data. In *GIS'16*. ACM, New York, NY, Article 19, 10 pages. DOI: https://doi.org/10.1145/2996913.2996918

Desheng Zhang, Fan Zhang, Ming Lu, Yunhuai Liu, and Tian He. 2016. Carpool service for large-scale taxicab networks. *ACM Trans. Sensor Networks (ACM TOSN)*. 12, 3 (2016), 18 pages.

Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015. coMobile: Real-time human mobility modeling at urban scale by multi-view learning. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 40 pages.

Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015. UrbanCPS: A cyber-physical system based on multi-source data with model integration. In *ICCPS'15*.

Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. 2010. Understanding transportation modes based on GPS data for web applications. *ACM Trans. Web* 4, 1 (Jan. 2010), 1.

Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. 2008. Learning transportation mode from raw GPS data for geographic applications on the web. In *WWW'08*.

Qin Zhou, Zhihan Fang, Yunhuai Liu, and Desheng Zhang. 2018. EXIMIUS: A measurement framework for explicit and implicit urban traffic sensing. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. ACM, 1–14.

Received January 2017; revised October 2018; accepted October 2018