Gaussian Process Decentralized Data Fusion and Active Sensing for Spatiotemporal Traffic Modeling and Prediction in Mobility-on-Demand Systems

Gaussian Process Decentralized Data Fusion and Active Sensing for Spatiotemporal Traffic Modeling and Prediction in Mobility-on-Demand Systems

***Abstract*—*Mobility-on-demand* (MoD) systems have recently emerged as a promising paradigm of one-way vehicle sharing for sustainable personal urban mobility in densely populated cities. We assume the capability of a MoD system to be enhanced by deploying robotic shared vehicles that can autonomously cruise the streets to be hailed by users. A key challenge of the MoD system is that of real-time, fine-grained mobility demand and traffic flow sensing and prediction. This paper presents novel *Gaussian process* (GP) decentralized data fusion and active sensing algorithms for real-time, fine-grained traffic modeling and prediction with a fleet of MoD vehicles. The predictive performance of our decentralized data fusion algorithms are theoretically guaranteed to be equivalent to that of sophisticated centralized sparse GP approximations. We derive consensus filtering variants requiring only local communication between neighboring vehicles. We theoretically guarantee the performance of our decentralized active sensing algorithms. When they are used to gather informative data for mobility demand prediction, they can achieve a dual effect of fleet rebalancing to service mobility demands. Empirical evaluation on real-world datasets shows that our algorithms are significantly more time-efficient and scalable in the size of data and fleet while achieving predictive performance comparable to that of state-of-the-art algorithms.**

***Note to Practitioners*—Knowing, understanding, and predicting spatiotemporally varying traffic phenomena in real time has become increasingly important to the goal of achieving smooth-flowing, congestion-free traffic in densely-populated urban cities, which motivates our work here. This paper addresses the following fundamental problem of data fusion and active sensing: How can a fleet of autonomous robotic vehicles or mobile probes actively cruise a road network to gather and assimilate the most informative data for predicting a spatiotemporally varying traffic phenomenon like a mobility demand pattern or traffic flow? Existing centralized solutions are poorly suited because they suffer from a single point of failure and incur huge communication, space, and time overheads with large data and fleet. This paper proposes novel efficient and scalable decentralized data fusion and active sensing algorithms with theoretical performance guarantees. The practical applicability of our algorithms is not restricted to traffic monitoring; they can be used in other environmental sensing applications such as precision agriculture, monitoring of ocean/freshwater phenomena (e.g., plankton bloom), forest ecosystems, pollution (e.g., oil spill), or contamination. Note that the decentralized data fusion component of our algorithms can also be used for static sensors and passive mobile probes and, interestingly, adapted to parallel implementations to be run on a cluster of machines for achieving efficient and scalable probabilistic prediction (i.e., with predictive uncertainty) with large data. Empirical results show that our algorithms can perform well with two datasets featuring real-world traffic phenomena in the densely-populated urban city of Singapore. A limitation of our algorithms is that the decentralized data fusion components assume independence between multiple traffic phenomena while the decentralized active sensing components only work for a single traffic phenomenon. So, in our future work, we will generalize our algorithms to perform active sensing of multiple traffic phenomena and remove the assumption of independence between them.**

***Index Terms*—Decentralized/distributed data fusion, decentralized active sensing, distributed consensus filtering, Gaussian process, log-Gaussian process, relational Gaussian process, adaptive sampling, active learning, mobility demand prediction, traffic flow forecasting, spatiotemporal modeling, environmental sensing and monitoring, vehicular sensor network**

Methodologies and Applications

8. Sensing and networks, 7. mobility and navigation, 5. transportation

Preliminary results of this research have been published in Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI), 2012 [1] and the Robotics: Science and Systems Conference (RSS), 2013 [2]. This paper is an expanded version of the conference papers with new algorithmic and empirical results.

I. Introduction

PRIVATE automobiles are becoming unsustainable personal mobility solutions in densely populated urban cities because the addition of parking and road spaces cannot keep pace with their escalating numbers due to limited urban land. For example, Hong Kong and Singapore have, respectively, experienced 27.6% and 37% increase in private vehicles from 2003 to 2011 [3]. However, their road networks have only expanded less than 10% in size. Without implementing sustainable measures, traffic congestions and delays will grow more severe and frequent, especially during peak hours.

Mobility-on-demand (MoD) systems [4] (e.g., Vélib system of over 20000 shared bicycles in Paris, experimental car-sharing systems described in [5]) have recently emerged as a promising paradigm of one-way vehicle sharing for sustainable personal urban mobility, specifically, to tackle the problems of low vehicle utilization rate and parking space caused by private automobiles. Conventionally, a MoD system provides stacks and racks of light electric vehicles distributed throughout a city: When a user wants to go somewhere, he simply walks to the nearest rack, swipes a card to pick up a vehicle, drives it to the rack nearest to his destination, and drops it off. In this paper, we assume the capability of a MoD system to be enhanced by deploying robotic shared vehicles (e.g., General Motors Chevrolet EN-V 2.0 prototype [6]) that can autonomously drive and cruise the streets of a densely populated urban city to be hailed by users (like taxis) instead

of just waiting at the racks to be picked up. Compared to the conventional MoD system, the fleet of autonomous robotic vehicles provides greater accessibility to users who can be picked up and dropped off at any location in the road network. As a result, it can service regions of high mobility demand but with poor coverage of stacks and racks due to limited space for their installation.

The key factors in the success of a MoD system are the costs to the users and system latencies, which can be minimized by managing the MoD system effectively. To achieve this, two main technical challenges need to be addressed [7]: (a) Real-time, fine-grained mobility demand sensing and prediction, and (b) real-time active fleet management to balance vehicle supply and demand and satisfy latency requirements at sustainable operating costs. Existing works on load balancing in MoD systems [5], dynamic traffic assignment problems [8], dynamic one-to-one pickup and delivery problems [9], and location recommendation and dispatch for cruising taxis [10]–[14] have tackled variants of the second challenge by assuming the necessary inputs of mobility demand and traffic flow information to be perfectly or accurately known using prior knowledge or offline processing of historic data. Such an assumption does not hold for densely populated urban cities because their mobility demand patterns and traffic flow are often subject to short-term random fluctuations and perturbations due to frequent special events (e.g., storewide sales, exhibitions), unpredictable weather conditions, unforeseen emergencies (e.g., breakdowns in public transport services), or traffic incidents (e.g., accidents, vehicle breakdowns, roadworks). So, in order for the active fleet management strategies to perform well in fleet rebalancing and route planning to service the mobility demands, they require accurate, fine-grained predictive information of the spatiotemporally varying mobility demand patterns and traffic flow in real time, the former of which is the desired outcome of addressing the first challenge. To the best of our knowledge, there is little progress in the algorithm design and development to take on the first challenge, which will be a focus of our work in this paper.

In practice, it is non-trivial to achieve real-time, accurate prediction of spatiotemporally varying traffic phenomena such as mobility demand patterns and traffic flow because the quantity of sensors that can be deployed to observe an entire road network is cost-constrained. For example, static sensors such as loop detectors [15], [16] are traditionally placed at designated locations in a road network to collect data for predicting the traffic flow. However, they provide sparse coverage (i.e., many road segments are not observed, thus leading to data sparsity), incur high installation and maintenance costs, and cannot reposition by themselves in response to changes in the traffic flow. Low-cost GPS technology allows the collection of traffic flow data using passive mobile probes [17] (e.g., taxis/cabs). Unlike static sensors, they can directly measure the travel times along road segments. But, they provide fairly sparse coverage due to low GPS sampling frequency (i.e., often imposed by taxi/cab companies) and no control over their routes, incur high initial implementation cost, pose privacy issues, and produce highly-varying speeds and travel times while traversing the same road segment due to inconsistent driving behaviors. A critical mass of probes is needed on each road segment to ease the severity of the last drawback [18] but is often hard to achieve on non-highway segments due to sparse coverage. In contrast, we propose using the autonomous robotic vehicles as active mobile probes [19] to overcome the limitations of static and passive mobile probes. In particular, they can be directed to explore any segments of a road network to gather real-time mobility demand data (e.g., pickup counts of different regions) and traffic flow data (e.g., speeds and travel times along road segments) at a desired GPS sampling rate while enforcing consistent driving behavior.

How then can the vacant autonomous robotic vehicles in a MoD system actively cruise a road network to gather and assimilate the *most informative* data for predicting a spatiotemporally varying traffic phenomenon like a mobility demand pattern or traffic flow[1]? To solve this problem, a centralized approach to data fusion and active sensing [15], [20]–[22] is poorly suited because it suffers from a single point of failure and incurs huge communication, space, and time overheads with large data and fleet. Hence, we propose novel decentralized data fusion and active sensing algorithms for real-time, fine-grained traffic sensing, modeling, and prediction with a fleet of autonomous robotic vehicles in a MoD system. Note that the decentralized data fusion component of our proposed algorithms can also be used for static sensors and passive mobile probes. The specific contributions of our work here include:

- Modeling and predicting a mobility demand pattern and traffic flow using, respectively, rich classes of Bayesian nonparametric models called a *log-Gaussian process* (ℓGP) model[2] (Section II-C) and a relational Gaussian process model, the latter of whose spatiotemporal correlation structure can exploit both the road segment features and road network topology information (Section II-B);
- Developing novel *Gaussian process decentralized data fusion* algorithms for cooperative perception of traffic phenomena called GP-DDF and GP-DDF$^+$ (Section III) whose predictive performance are theoretically guaranteed to be equivalent to that of sophisticated centralized sparse approximations of the *full-rank Gaussian process* (full GP for short) model: The computation of such sparse approximate GP models can thus be distributed among the MoD vehicles, thereby achieving efficient and scalable probabilistic prediction;
- Deriving consensus filtering variants of GP-DDF and GP-DDF$^+$ (Section III-A) that require only local communication between neighboring MoD vehicles instead of assuming all-to-all communication between MoD vehicles;
- Devising decentralized active sensing algorithms (Section IV) (a) whose performance, when coupled with GP-DDF, can be theoretically guaranteed to realize the effect

[1]When a vacant autonomous robotic vehicle is tasked as an active mobile probe to gather the most informative mobility demand (traffic flow) data, it can simultaneously serve as a passive mobile probe to collect traffic flow (mobility demand) data, albeit not most informative.

[2]The ℓGP model is widely used in geostatistics to model natural environmental phenomena, but not urban traffic. Hence, its application to modeling a traffic phenomenon (e.g., mobility demand pattern) is novel.

of the spatiotemporal correlation structure of the traffic phenomenon and various parameter settings of the MoD system, and (b) that, when used for sampling a mobility demand pattern, can be analytically shown to exhibit, interestingly, a cruising behavior of simultaneously exploring demand hotspots and sparsely sampled regions that have higher likelihood of picking up users, hence achieving a dual effect of fleet rebalancing to service the mobility demands;
- Analyzing the time and communication overheads of our proposed algorithms (Section V): We prove that our algorithms can scale better than existing state-of-the-art centralized algorithms in the size of the data and fleet;
- Empirically evaluating the predictive accuracy, time efficiency, scalability, and performance of servicing mobility demands (i.e., average cruising length of vehicles, average waiting time of users, total number of pickups) of our proposed algorithms on two datasets featuring real-world traffic phenomena such as a mobility demand pattern over the central business district of Singapore and speeds of road segments over an urban road network in Singapore (Section VI).

II. Modeling Traffic Phenomena with Gaussian Processes (GPs)

A *Gaussian process* (GP) can be used to model a spatiotemporally varying traffic phenomenon as follows: A traffic phenomenon is defined to vary as a realization of a GP. Let V be a set of sampling units representing the domain of the phenomenon such that each sampling unit $s \in V$ is associated with a p-dimensional feature vector x_s and a realized (random) measurement z_s (Z_s) if s is sampled/observed (unobserved). Let $\{Z_s\}_{s\in V}$ denote a GP, that is, every finite subset of $\{Z_s\}_{s\in V}$ has a multivariate Gaussian distribution [23]. The GP is fully specified by its *prior* mean $\mu_s \triangleq \mathbb{E}[Z_s]$ and covariance $\sigma_{ss'} \triangleq \mathrm{cov}[Z_s, Z_{s'}]$ for all $s, s' \in V$, the latter of which characterizes the spatiotemporal correlation structure of the phenomenon and can be defined by the widely-used squared exponential covariance function:

$$\sigma_{ss'} \triangleq \sigma_s^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{p}\left(\frac{[x_s]_i - [x_{s'}]_i}{\ell_i}\right)^2\right) + \sigma_n^2\delta_{ss'} \quad (1)$$

where $[x_s]_i$ ($[x_{s'}]_i$) is the i-th component of the feature vector x_s ($x_{s'}$), the hyperparameters $\sigma_n^2, \sigma_s^2, \ell_1, \ldots, \ell_p$ are, respectively, noise and signal variances and length-scales that can be learned using maximum likelihood estimation, and $\delta_{ss'}$ is a Kronecker delta that is 1 if $s = s'$ and 0 otherwise.

Given that a column vector z_D of realized measurements is observed for some set $D \subset V$ of sampling units, a full GP model can exploit these data (D, z_D) to predict the measurements for any set $S \subset V$ of unobserved sampling units as well as provide their corresponding predictive uncertainties using the Gaussian predictive distribution $\mathcal{N}(\mu_{S|D}, \Sigma_{SS|D})$ with the following *posterior* mean vector and covariance matrix, respectively:

$$\mu_{S|D} \triangleq \mu_S + \Sigma_{SD}\Sigma_{DD}^{-1}(z_D - \mu_D) \quad (2)$$

$$\Sigma_{SS|D} \triangleq \Sigma_{SS} - \Sigma_{SD}\Sigma_{DD}^{-1}\Sigma_{DS} \quad (3)$$

where μ_S (μ_D) is a column vector with mean components μ_s for all $s \in S$ ($s \in D$), Σ_{SD} (Σ_{DD}) is a covariance matrix with covariance components $\sigma_{ss'}$ for all $s \in S, s' \in D$ ($s, s' \in D$), and Σ_{SD} is the transpose of Σ_{DS}. The posterior mean vector $\mu_{S|D}$ (2) is used to predict the measurements for any set S of unobserved sampling units. The posterior covariance matrix $\Sigma_{SS|D}$ (3), which is independent of the measurements z_D, can be processed in two ways to quantify the uncertainty of these predictions: The trace of $\Sigma_{SS|D}$ yields the sum of posterior variances $\Sigma_{ss|D}$ over all $s \in S$ while the determinant of $\Sigma_{SS|D}$ is used in calculating the Gaussian posterior joint entropy

$$\mathbb{H}[Z_S|Z_D] \triangleq \frac{1}{2}\log(2\pi e)^{|S|}\left|\Sigma_{SS|D}\right|. \quad (4)$$

In contrast to the first measure of uncertainty that assumes conditional independence between measurements in the set S of unobserved sampling units, the entropy-based measure (4) accounts for their correlation, thereby not overestimating their uncertainty. Hence, we will focus on using the entropy-based measure of uncertainty in this paper.

A. Subset of Data Approximation

The expressive power of a full GP model comes at a cost of poor scalability (i.e., cubic time) in the size $|D|$ of data; this can be observed from computing its Gaussian predictive distribution (i.e., (2) and (3)), which requires inverting covariance matrix Σ_{DD} that incurs $\mathcal{O}(|D|^3)$ time. If $|D|$ is expected to be large, full GP prediction cannot be performed in real time. For practical usage, we have to resort to computationally cheaper approximate GP prediction.

A simple method of approximation is to select only a subset U of the entire set D of observed sampling units (i.e., $U \subset D$) to compute the posterior distribution of the measurements for any set $S \subseteq V \setminus D$ of unobserved sampling units. Such a sparse subset of data approximation method produces the following Gaussian predictive distribution, which closely resembles that of the full GP model (i.e., by simply replacing D in (2) and (3) with U):

$$\mu_{S|U} = \mu_S + \Sigma_{SU}\Sigma_{UU}^{-1}(z_U - \mu_U) \quad (5)$$

$$\Sigma_{SS|U} = \Sigma_{SS} - \Sigma_{SU}\Sigma_{UU}^{-1}\Sigma_{US}\ . \quad (6)$$

Notice that the covariance matrix Σ_{UU} to be inverted only incurs $\mathcal{O}(|U|^3)$ time, which is independent of $|D|$.

The predictive performance of subset of data approximation is sensitive to the selection of subset U. In practice, random subset selection often yields poor performance. This issue can be resolved by actively selecting an informative subset U in an iterative greedy manner: Firstly, U is initialized to be an empty set. Then, all sampling units in $D \setminus U$ are scored based on a criterion that can be chosen from, for example, the works of [20], [24], [25]. The highest-scored sampling unit is selected for inclusion into U and removed from D. This greedy selection procedure is iterated until U reaches a predefined size. Among the various criteria introduced earlier, the differential entropy score [24] is reported to perform well [26]; it is a monotonic function of the posterior variance $\Sigma_{ss|U}$ (6), thus resulting in the greedy selection of a sampling unit $s \in D \setminus U$ with the largest variance in each iteration.

B. Modeling Traffic Flow with Relational Gaussian Process

To model traffic flow (e.g., speeds of road segments) over an urban road network with a relational GP, let each sampling unit $s \in V$ denote a road segment and its associated measurement z_s quantify the traffic flow condition such as road speed. Similar to the standard GP, a relational GP uses the GP prior covariance $\sigma_{ss'}$ to measure the pairwise "similarity" of road segments. For traffic flow (e.g., road speeds), the correlation of measurements between pairs of road segments depends not only on their features (e.g., length, number of lanes, speed limit, direction) but also on the relational information residing within the road network topology. So, different from the standard GP, a relational GP defines the GP prior covariance not just in terms of the features, but also by exploiting the topology information, which will be described next.

Definition 1 (Road Network): Let the road network be represented as a weighted directed graph $G \triangleq (V, E, m)$ that consists of

- a set V of vertices denoting the domain of all possible road segments,
- a set $E \subseteq V \times V$ of edges where there is a edge (s, s') from $s \in V$ to $s' \in V$ iff the end of segment s connects to the start of segment s' in the road network, and
- a weight function $m : E \rightarrow \mathbb{R}^+$ measuring the standardized Manhattan distance [27] $m((s, s')) \triangleq \sum_{i=1}^{p} |[s]_i - [s']_i|/r_i$ of each edge (s, s') where $[s]_i$ ($[s']_i$) is the i-th component of the feature vector specifying road segment s (s'), and r_i is the range of the i-th feature. The weight function m serves as a dissimilarity measure between adjacent road segments.

The next step is to compute the shortest path distance $d(s, s')$ between all pairs of road segments $s, s' \in V$ (i.e., using Floyd-Warshall or Johnson's algorithm) with respect to the topology of the weighted directed graph G. Such a distance function is again a measure of dissimilarity, rather than one of similarity, as required by the GP prior covariance. Furthermore, a valid GP prior covariance needs to be positive semidefinite and symmetric [28], which are clearly violated by d.

To construct a valid GP prior covariance from d, multi-dimensional scaling [27] is applied to embed the domain of road segments into the p'-dimensional Euclidean space $\mathbb{R}^{p'}$. Specifically, a mapping $g : V \rightarrow \mathbb{R}^{p'}$ is determined by minimizing the squared loss $g^* = \arg\min_g \sum_{s,s' \in V} (d(s, s') - \|g(s) - g(s')\|)^2$. With a small squared loss, the Euclidean distance $\|g^*(s) - g^*(s')\|$ between $g^*(s)$ and $g^*(s')$ is expected to closely approximate the shortest path distance $d(s, s')$ between any pair of road segments s and s'. After embedding into Euclidean space, a conventional covariance function such as the squared exponential one (1) that is modified to reflect our notations here can be used:

$$\sigma_{ss'} \triangleq \sigma_s^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{p'}\left(\frac{[g^*(s)]_i - [g^*(s')]_i}{\ell_i}\right)^2\right) + \sigma_n^2 \delta_{ss'}$$

where $[g^*(s)]_i$ ($[g^*(s')]_i$) is the i-th component of the p'-dimensional vector $g^*(s)$ ($g^*(s')$). The resulting GP prior covariance $\sigma_{ss'}$[3] is guaranteed to be valid. By plugging this newly constructed GP prior covariance into (2), (3) and (4), the relational GP model can predict the measurements of the unobserved sampling units and quantify their corresponding predictive uncertainties in the same manner as the standard GP model.

C. Modeling Mobility Demand Pattern with Log-Gaussian Process (ℓGP)

To model a mobility demand pattern over an urban city with a ℓGP, its service area can be represented as a directed graph $G \triangleq (V, E)$ where V denotes a set of all regions generated by gridding the service area, and $E \subseteq V \times V$ denotes a set of edges such that there is an edge (s, s') from $s \in V$ to $s' \in V$ iff at least one road segment in the road network starts in s and ends in s'. Then, each sampling unit $s \in V$ denotes a region such that its associated p-dimensional feature vector x_s represents its context information (e.g., location, time, precipitation) and its corresponding measurement y_s quantifies its mobility demand[4]. Since it is often impractical in terms of sensing resource cost to determine the actual mobility demand of a region, a common practice is to use the pickup count of the region as a surrogate measure. To elaborate, the user pickups made by vacant MoD vehicles cruising in a region contribute to its pickup count. Since we do not assume a data center to be available to keep track of the pickup count, a fully distributed gossip-based protocol [29] is utilized to aggregate these pickup information from the vehicles in the region that are connected via an ad hoc wireless communication network. Consequently, any vehicle entering the region can access its pickup count simply by joining its ad hoc network.

As observed in [11], [13] and our real-world dataset (Fig. 1a), a mobility demand pattern over a large service area in an urban city is typically characterized by spatiotemporally correlated demand measurements and contains a few small-scale hotspots exhibiting extreme measurements and much higher spatiotemporal variability than the rest of the demand pattern. That is, if the measurements are put together into a 1D sample frequency distribution, a positive skew results. But, the GP covariance structure is sensitive to strong positive skewness and easily destabilized by a few extreme measurements [30]. In practice, this can cause reconstructed patterns to display large hotspots centered about a few extreme measurements and predictive variances to be unrealistically small in hotspots [31], which are undesirable. So, if the GP is used to model a demand pattern directly, it may not predict well. To resolve this, a standard statistical practice is to take the log of the measurements (i.e., $z_s = \log y_s$) to remove skewness and extremity, and use the GP to model the demand pattern in the *log-scale* instead.

[3]For spatiotemporal traffic modeling, the GP prior covariance $\sigma_{ss'}$ can be extended to account for the temporal dimension.

[4]The service area is represented as a grid of regions instead of a network of road segments like in Section II-B because we observe less smoothly-varying, noisier demand measurements (hence, lower spatial correlation) for the latter in our dataset featuring a real-world mobility demand pattern (Section VI-A) since many road segments do not permit stopping of vehicles.

Demand measurements may not be observed in some regions because vacant MoD vehicles did not cruise into them. Since our ultimate interest is to predict them in the *original scale*, GP's predicted log-measurements for these unobserved regions must be transformed back *unbiasedly*. To achieve this, we utilize a widely-used variant of GP in geostatistics called the ℓGP that can model the demand pattern in the original scale. Let $\{Y_s\}_{s\in V}$ denote a ℓGP: If $Z_s \triangleq \log Y_s$, then $\{Z_s\}_{s\in V}$ is a GP. So, $Y_s = \exp\{Z_s\}$ denotes the original random demand measurement of unobserved region s and is predicted using the log-Gaussian posterior mean (i.e., best unbiased predictor):

$$\widehat{\mu}_{s|D} \triangleq \exp(\mu_{s|D} + \Sigma_{ss|D}/2) \tag{7}$$

where $\mu_{s|D}$ and $\Sigma_{ss|D}$ are simply the Gaussian posterior mean (2) and variance (3) of GP, respectively. The uncertainty of predicting the measurements for any set $S \subset V$ of unobserved regions can be quantified by the following log-Gaussian posterior joint entropy, which will be exploited by our decentralized active sensing algorithm (Section IV):

$$\mathbb{H}[Y_S|Y_D] \triangleq \frac{1}{2}\log(2\pi e)^{|S|}\left|\Sigma_{SS|D}\right| + \mu_{S|D}\cdot\mathbf{1} \tag{8}$$

where $\mu_{S|D}$ and $\Sigma_{SS|D}$ are the Gaussian posterior mean vector (2) and covariance matrix (3) of GP, respectively.

III. Gaussian Process Decentralized Data Fusion

The mobility demand data (e.g., pickup counts of different regions) and/or traffic flow data (e.g., speeds of road segments) are gathered distributedly by the vacant MoD vehicles cruising the road network in the service area and have to be assimilated in order to predict the traffic phenomena (i.e., mobility demand pattern and/or traffic flow). A straightforward approach to data fusion is to fully communicate all the data to every vehicle, each of which then performs the same full GP prediction (2) of traffic flow or ℓGP prediction (7) of the mobility demand pattern separately. Such an approach unfortunately cannot scale well and be performed in real time due to its cubic time complexity in the size of the data. In this section, we will discuss two novel Gaussian process decentralized data fusion algorithms for cooperative perception of the traffic phenomena called GP-DDF and GP-DDF$^+$ that can distribute the computational load among the MoD vehicles to achieve efficient and scalable approximate GP and ℓGP prediction.

The intuition of our GP-DDF algorithm is as follows: Each of the K MoD vehicles constructs a local summary of the data taken along its own path in the road network and communicates its local summary to every other vehicles. Then, it assimilates the local summaries received from the other vehicles into a globally consistent summary, which is exploited for predicting the traffic phenomena as well as active sensing. This intuition will be formally realized and described in the paragraphs below.

While exploring the service area, each MoD vehicle summarizes its local data taken along its path based on a common support set $U \subset V$ known to all the other vehicles. Its local summary is defined as follows:

Definition 2 (Local Summary): Given a common support set $U \subset V$ known to all K vehicles, a set $D_k \subset V$ of observed regions or road segments and a column vector z_{D_k} of corresponding measurements local to vehicle k, its local summary is defined as a tuple $(\dot{z}_U^k, \dot{\Sigma}_{UU}^k)$ where

$$\dot{z}_B^k \triangleq \Sigma_{BD_k}\Sigma_{D_kD_k|U}^{-1}(z_{D_k} - \mu_{D_k}) \tag{9}$$

$$\dot{\Sigma}_{BB'}^k \triangleq \Sigma_{BD_k}\Sigma_{D_kD_k|U}^{-1}\Sigma_{D_kB'} \tag{10}$$

such that $\Sigma_{D_kD_k|U}$ is defined in a similar manner to (6).

Remark. Unlike subset of data (Section II-A), the support set U of regions or road segments does not have to be observed since the local summary (i.e., (9) and (10)) is independent of the corresponding measurements z_U. So, U does not need to be a subset of $D = \bigcup_{k=1}^K D_k$. To select an informative support set U from the set V of all possible regions or road segments, an offline active selection procedure similar to that in the last paragraph of Section II-A can be performed just once prior to observing data to determine U. In contrast, subset of data has to perform online active selection every time new regions or road segments are being observed.

By communicating its local summary to every other vehicles, each vehicle can then construct a globally consistent summary from the received local summaries:

Definition 3 (Global Summary): Given a common support set $U \subset V$ known to all K vehicles and the local summary $(\dot{z}_U^k, \dot{\Sigma}_{UU}^k)$ of every vehicle $k = 1, \ldots, K$, the global summary is defined as a tuple $(\ddot{z}_U, \ddot{\Sigma}_{UU})$ where

$$\ddot{z}_U \triangleq \sum_{k=1}^K \dot{z}_U^k \tag{11}$$

$$\ddot{\Sigma}_{UU} \triangleq \Sigma_{UU} + \sum_{k=1}^K \dot{\Sigma}_{UU}^k \ . \tag{12}$$

Finally, the global summary is exploited by each vehicle to compute a globally consistent Gaussian predictive distribution, as detailed in Theorem 1A below, as well as to perform decentralized active sensing (Section IV):

Theorem 1 (GP-DDF): Let a common support set $U \subset V$ be known to all K vehicles.

A. Given the global summary $(\ddot{z}_U, \ddot{\Sigma}_{UU})$, each vehicle computes a globally consistent Gaussian predictive distribution $\mathcal{N}(\overline{\mu}_S, \overline{\Sigma}_{SS})$ of the measurements for any set S of unobserved regions or road segments where

$$\overline{\mu}_S \triangleq \mu_S + \Sigma_{SU}\ddot{\Sigma}_{UU}^{-1}\ddot{z}_U \tag{13}$$

$$\overline{\Sigma}_{SS} \triangleq \Sigma_{SS} - \Sigma_{SU}(\Sigma_{UU}^{-1} - \ddot{\Sigma}_{UU}^{-1})\Sigma_{US} \ . \tag{14}$$

B. Let $\mathcal{N}(\mu_{S|D}^{\text{PITC}}, \Sigma_{SS|D}^{\text{PITC}})$ be the Gaussian predictive distribution computed by the centralized sparse *partially independent training conditional* (PITC) approximation of the full GP model [32] where

$$\mu_{S|D}^{\text{PITC}} \triangleq \mu_S + \Gamma_{SD}\left(\Gamma_{DD} + \Lambda\right)^{-1}(z_D - \mu_D) \tag{15}$$

$$\Sigma_{SS|D}^{\text{PITC}} \triangleq \Sigma_{SS} - \Gamma_{SD}\left(\Gamma_{DD} + \Lambda\right)^{-1}\Gamma_{DS} \tag{16}$$

such that

$$\Gamma_{BB'} \triangleq \Sigma_{BU}\Sigma_{UU}^{-1}\Sigma_{UB'} \tag{17}$$

and Λ is a block-diagonal matrix constructed from the K diagonal blocks of $\Sigma_{DD|U}$, each of which is a matrix $\Sigma_{D_kD_k|U}$ for $k = 1,\ldots,K$ where $D = \bigcup_{k=1}^{K} D_k$. Then, $\overline{\mu}_S = \mu_{S|D}^{\text{PITC}}$ and $\overline{\Sigma}_{SS} = \Sigma_{SS|D}^{\text{PITC}}$.

The proof of Theorem 1B is given in Appendix A.

Remark 1. The computation of the centralized sparse PITC approximation of the full GP model [32] can be distributed among the K vehicles, thereby improving the time efficiency of prediction. Supposing $|S| \leq |U|$ for simplicity, the $\mathcal{O}\big(|D|((|D|/K)^2 + |U|^2)\big)$ time incurred by PITC can be reduced to $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K\big)$ time of running our GP-DDF algorithm on each of the K vehicles, the latter of which scales better with increasing size $|D|$ of data.

Remark 2. We can draw insights from PITC to elucidate an underlying property of our GP-DDF algorithm: It is assumed that $Z_{D_1},\ldots,Z_{D_K},Z_S$ are conditionally independent given Z_U. To potentially reduce the degree of violation of this assumption, an informative support set U is actively selected, as described earlier in this section. Furthermore, the experimental results on our real-world dataset featuring speeds of road segments over an urban road network[5] (Section VI-B) show that GP-DDF can achieve predictive performance comparable to that of the full GP model while enjoying significantly lower computational cost over it, thus demonstrating the practicality of such an assumption for predicting traffic phenomena. The predictive performance of GP-DDF can be improved by increasing the size of U at the expense of greater time and communication overhead.

Remark 3. The Gaussian predictive mean $\overline{\mu}_s$ (13) and variance $\overline{\Sigma}_{ss}$ (14) of GP-DDF can be plugged into (7) to obtain the log-Gaussian posterior mean for predicting the demand measurement of any unobserved region in the original scale.

Though GP-DDF scales very well with large data, it can predict poorly due to (a) loss of information caused by summarizing the measurements and correlation structure of the original data; and (b) sparse coverage of the hotspots (i.e., with higher spatiotemporal variability) by the support set. To alleviate this drawback suffered by GP-DDF, we will now describe our GP-DDF$^+$ algorithm that combines the best of both worlds, that is, the predictive power of full GP and efficiency of GP-DDF. GP-DDF$^+$ is based on the intuition that a vehicle can exploit its local data to improve the predictions for unobserved regions or road segments "close" to its data (in the correlation sense). At the same time, GP-DDF$^+$ can preserve the efficiency of GP-DDF by exploiting its idea of summarizing information, specifically, into the local and global summaries (Definitions 2 and 3). To improve the predictive power of GP-DDF, we develop the following novel GP-DDF$^+$ algorithm that is further augmented by local information.

[5]The work of [32] only illustrated the predictive performance of PITC on a simulated toy example.

Definition 4 (GP-DDF$_k^+$): Given a common support set $U \subset V$ known to all K vehicles, the global summary $(\ddot{z}_U, \ddot{\Sigma}_{UU})$, the local summary $(\dot{z}_U^k, \dot{\Sigma}_{UU}^k)$, a set $D_k \subset V$ of observed regions or road segments and a column vector z_{D_k} of corresponding measurements local to vehicle k, its GP-DDF$_k^+$ algorithm computes a Gaussian predictive distribution $\mathcal{N}(\overline{\mu}_S^k, \overline{\Sigma}_{SS}^k)$ of the measurements for any set $S \subset V$ of unobserved regions or road segments where $\overline{\mu}_S^k \triangleq \big(\overline{\mu}_s^k\big)_{s\in S}$ and $\overline{\Sigma}_{SS}^k \triangleq \big(\overline{\sigma}_{ss'}^k\big)_{s,s'\in S}$ such that

$$\overline{\mu}_s^k \triangleq \mu_s + \Big(\gamma_{sU}^k \ddot{\Sigma}_{UU}^{-1}\ddot{z}_U - \Sigma_{sU}\Sigma_{UU}^{-1}\dot{z}_U^k\Big) + \dot{z}_s^k \tag{18}$$

$$\begin{aligned}\overline{\sigma}_{ss'}^k \triangleq \sigma_{ss'} - \Big(&\gamma_{sU}^k\Sigma_{UU}^{-1}\Sigma_{Us'} - \Sigma_{sU}\Sigma_{UU}^{-1}\dot{\Sigma}_{Us'}^k \\ &- \gamma_{sU}^k\ddot{\Sigma}_{UU}^{-1}\gamma_{Us'}^k\Big) - \dot{\Sigma}_{ss'}^k\end{aligned} \tag{19}$$

and

$$\gamma_{sU}^k \triangleq \Sigma_{sU} + \Sigma_{sU}\Sigma_{UU}^{-1}\dot{\Sigma}_{UU}^k - \dot{\Sigma}_{sU}^k \ . \tag{20}$$

Remark 1. Both the Gaussian predictive mean $\overline{\mu}_s^k$ (18) and covariance $\overline{\sigma}_{ss'}^k$(19) of GP-DDF$_k^+$ exploit summary information (i.e., bracketed term) contributed from exchanged summaries among vehicles and local information (i.e., last term) contributed from local data.

Remark 2. The Gaussian predictive mean $\overline{\mu}_s^k$ (18) and variance $\overline{\sigma}_{ss}^k$ (19) of GP-DDF$_k^+$ can be plugged into (7) to obtain the log-Gaussian posterior mean for predicting the demand measurement of any unobserved region in the original scale.

Remark 3. Since different vehicles exploit different local data, their GP-DDF$_k^+$ algorithms provide inconsistent predictions of the traffic phenomena.

It is often desirable to achieve a globally consistent prediction among all vehicles. To do this, each unobserved region or road segment is simply assigned to the vehicle that predicts its measurement best, which can be performed in a decentralized way:

Definition 5 (Assignment Function): An assignment function $\tau : V \mapsto \{1\ldots K\}$ is defined as

$$\tau(s) \triangleq \operatorname*{arg\,min}_{k\in\{1\ldots K\}} \overline{\sigma}_{ss}^k \tag{21}$$

for all $s \in S$ where the predictive variance $\overline{\sigma}_{ss}^k$ is defined in (19). From now on, let $\tau_s \triangleq \tau(s)$ for notational simplicity.

Using the assignment function τ, each vehicle can now compute a globally consistent Gaussian predictive distribution, as detailed in Theorem 2A below:

Theorem 2 (GP-DDF$^+$): Let a common support set $U \subset V$ and a common assignment function τ be known to all K vehicles.

A. The GP-DDF$^+$ algorithm of each vehicle computes a globally consistent Gaussian predictive distribution $\mathcal{N}(\overline{\mu}_S, \overline{\Sigma}_{SS})$ of the measurements for any set $S \subset V$ of unobserved regions or road segments where $\overline{\mu}_S \triangleq$

$(\overline{\mu}_s^{\tau_s})_{s\in S}$ (18) and $\overline{\Sigma}_{SS} \triangleq (\overline{\sigma}_{ss'})_{s,s'\in S}$ such that

$$\overline{\sigma}_{ss'} \triangleq \begin{cases} \overline{\sigma}_{ss'}^{\tau_s} & \text{if } \tau_s = \tau_{s'}, \\ \Sigma_{ss'|U} + \gamma_{sU}^{\tau_s} \ddot{\Sigma}_{UU}^{-1} \gamma_{Us'}^{\tau_{s'}} & \text{otherwise}, \end{cases} \tag{22}$$

and $\gamma_{Us'}^{\tau_{s'}}$ is the transpose of $\gamma_{s'U}^{\tau_{s'}}$.

B. Let $\mathcal{N}(\mu_{S|D}^{\text{PIC}}, \Sigma_{SS|D}^{\text{PIC}})$ be the Gaussian predictive distribution computed by the centralized sparse *partially independent conditional* (PIC) approximation of the full GP model [33] where $\mu_{S|D}^{\text{PIC}} \triangleq \left(\mu_{s|D}^{\text{PIC}}\right)_{s\in S}$ and $\Sigma_{SS|D}^{\text{PIC}} \triangleq \left(\sigma_{ss'|D}^{\text{PIC}}\right)_{s,s'\in S}$ such that

$$\mu_{s|D}^{\text{PIC}} \triangleq \mu_s + \widetilde{\Gamma}_{sD} \left(\Gamma_{DD} + \Lambda\right)^{-1} (z_D - \mu_D) \tag{23}$$

$$\sigma_{ss'|D}^{\text{PIC}} \triangleq \sigma_{ss'} - \widetilde{\Gamma}_{sD} \left(\Gamma_{DD} + \Lambda\right)^{-1} \widetilde{\Gamma}_{Ds'} \tag{24}$$

and $\widetilde{\Gamma}_{Ds'}$ is the transpose of $\widetilde{\Gamma}_{s'D}$ such that

$$\Gamma_{BB'} \triangleq \Sigma_{BU} \Sigma_{UU}^{-1} \Sigma_{UB'} \tag{25}$$

$$\widetilde{\Gamma}_{sD} \triangleq (\widetilde{\Gamma}_{s\bar{s}})_{\bar{s}\in D} \tag{26}$$

$$\widetilde{\Gamma}_{s\bar{s}} \triangleq \begin{cases} \sigma_{s\bar{s}} & \text{if } \tau_s = \tau_{\bar{s}}, \\ \Gamma_{s\bar{s}} & \text{otherwise}, \end{cases} \tag{27}$$

and Λ is a block-diagonal matrix constructed from the K diagonal blocks of $\Sigma_{DD|U}$, each of which is a matrix $\Sigma_{D_kD_k|U}$ for $k = 1, \ldots, K$ where $D = \bigcup_{k=1}^{K} D_k$, and let $\tau_{\bar{s}} \triangleq k$ for all $\bar{s} \in D_k$. Then, $\overline{\mu}_s = \mu_{s|D}^{\text{PIC}}$ and $\overline{\sigma}_{ss'} = \sigma_{ss'|D}^{\text{PIC}}$ for all $s, s' \in S$.

The proof of Theorem 2B is given in Appendix B.

Remark 1. In Theorem 2A, if $\tau_s = \tau_{s'} = k$, then vehicle k can compute $\overline{\mu}_s^{\tau_s}$ (18) and $\overline{\sigma}_{ss'}$ (19) locally and send them to the other vehicles that request them. Otherwise, $\tau_s \neq \tau_{s'}$ and vehicle k has to request $|U|$-sized vectors $\gamma_{sU}^{\tau_s}$ and $\gamma_{s'U}^{\tau_{s'}}$ from the respective vehicles τ_s and $\tau_{s'}$ to compute $\overline{\sigma}_{ss'}$ (22).

Remark 2. The equivalence result of Theorem 2B implies that the computation of the centralized sparse PIC approximation of the full GP model [33] can be distributed among K vehicles, hence improving the time efficiency of prediction. Supposing $|S| \leq |U|$ and $|S| \leq |D|/K$ for simplicity, the $\mathcal{O}\big(|D|((|D|/K)^2 + |U|^2)\big)$ time incurred by PIC can be reduced to $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2 K\big)$ time of running our GP-DDF$^+$ algorithm on each of the K vehicles. Hence, GP-DDF$^+$ scales better with increasing size $|D|$ of data.

Remark 3. The equivalence result also sheds some light on an important property of GP-DDF$^+$ based on the structure of PIC: It is assumed that $Z_{D_1 \bigcup S_1}, \ldots, Z_{D_K \bigcup S_K}$ are conditionally independent given the support set U. As compared to GP-DDF that assumes conditional independence of $Z_{D_1}, \ldots, Z_{D_K}, Z_{S_1}, \ldots, Z_{S_K}$, GP-DDF$^+$ can predict Z_S better since it imposes a weaker conditional independence assumption. Experimental results on our dataset featuring a real-world mobility demand pattern (Section VI-A) also show that GP-DDF$^+$ achieves predictive accuracy comparable to full GP and significantly better than GP-DDF, thus justifying the practicality of such an assumption for predicting a mobility demand pattern.

A. Consensus Filtering-Based GP-DDF and GP-DDF$^+$

GP-DDF and GP-DDF$^+$ have assumed all-to-all communication between the K vehicles, which does not scale well in the size of the fleet. More importantly, such an assumption does not usually hold in practice because each vehicle can only communicate locally with its neighbors situated within its communication range that is often much smaller than the size of the traffic phenomenon of interest. The key idea of refining GP-DDF and GP-DDF$^+$ to meet the requirement of only local communication between neighboring vehicles is to observe that the summation of the local summary $(\dot{z}_U^k, \dot{\Sigma}_{UU}^k)$ of every vehicle $k = 1, \ldots, K$ to form the global summary $(\ddot{z}_U, \ddot{\Sigma}_{UU})$ (i.e., (11) and (12) of Definition 3) makes it amenable to be approximated using consensus filters [34]–[36]. In this subsection, we will discuss how a distributed algorithm called *band-pass consensus filter* (BCF), which is previously utilized in distributed Kalman filters [34], [35], can be used to approximate the global summary.

A consensus filter like BCF is a distributed algorithm that can be used for calculating the average consensus of the local summaries of the K vehicles with limited communication range. Specifically, each vehicle k computes its approximate consensus (υ_k, Υ_k) using the following discrete-time BCF [35] that asymptotically converges to the average consensus $(K^{-1}\ddot{z}_U, K^{-1}(\ddot{\Sigma}_{UU} - \Sigma_{UU}))$:

$$q_k = p_k + \dot{z}_U^k$$

$$p_k \leftarrow p_k + \eta \sum_{j\in N_k} (q_j - q_k)$$

$$\upsilon_k \leftarrow \upsilon_k + \eta \sum_{j\in N_k} (\upsilon_j - \upsilon_k) + \eta \sum_{j\in N_k \bigcup \{k\}} (q_j - \upsilon_k)$$

and

$$Q_k = P_k + \dot{\Sigma}_{UU}^k$$

$$P_k \leftarrow P_k + \eta \sum_{j\in N_k} (Q_j - Q_k)$$

$$\Upsilon_k \leftarrow \Upsilon_k + \eta \sum_{j\in N_k} (\Upsilon_j - \Upsilon_k) + \eta \sum_{j\in N_k \bigcup \{k\}} (Q_j - \Upsilon_k)$$

where p_k, q_k, P_k, and Q_k are necessary internal estimators stored in vehicle k, $N_k \subseteq \{1, \ldots, K\} \setminus \{k\}$ denotes the set of neighboring vehicles situated within the communication range of vehicle k, and step size $\eta < 1/\max_k\{|N_k|+1\}$ guarantees the asymptotic convergence of its approximate consensus (υ_k, Υ_k) to the average consensus $(K^{-1}\ddot{z}_U, K^{-1}(\ddot{\Sigma}_{UU} - \Sigma_{UU}))$. Then, each vehicle k can approximate the global summary $(\ddot{z}_U, \ddot{\Sigma}_{UU})$ using $(K\upsilon_k, K\Upsilon_k + \Sigma_{UU})$. Like a *low-pass consensus filter* (LCF) [34], BCF can filter out high-frequency noise. BCF can also provide filtering of slowly-varying noise like a *high-pass consensus filter* (HCF) [36].

The connectivity of the communication network is important to the consensus filters. HCF guarantees to converge to the average consensus only in a connected network [36]. In contrast, LCF remains a stable filter regardless of the connectivity of the network [35]. In the case that a vehicle k leaves the network (i.e., $N_k = \emptyset$), its BCF to calculate υ_k reduces to

$$\upsilon_k \leftarrow \upsilon_k + \eta \left(p_k + \dot{z}_U^k - \upsilon_k\right)$$

where p_k remains unchanged. Then, BCF reduces to a stable LCF [35] with the input $\dot{z}_U^k$ and constant noise p_k. Its BCF to calculate Υ_k is reduced in a similar manner.

IV. Gaussian Process Decentralized Active Sensing

Suppose that there are K vacant MoD vehicles in the fleet actively cruising the road network G to gather the most informative traffic flow data (e.g., speeds of road segments) for predicting traffic flow being modeled with a relational GP (Section II-B) and each vehicle $k \in \{1 \dots K\}$ has observed a set $D_k \subset V$ of road segments. In an active sensing problem, all vehicles have to jointly select the most informative walks $w_1^*, \dots w_K^*$ of length L along which traffic flow data will be sampled:

$$(w_1^*, \dots, w_K^*) \triangleq \underset{(w_1,\dots,w_K)}{\arg\max}\ \mathbb{H}\left[Z_{\bigcup_{k=1}^K S_{w_k}} \middle| Z_{\bigcup_{k=1}^K D_k}\right] \quad (28)$$

where S_{w_k} denotes the set of unobserved road segments to be visited by the walk w_k. To ease notation, let $w \triangleq (w_1 \dots w_K)$ and $S_w = \bigcup_{k=1}^K S_{w_k}$ (similarly, for w^* and S_{w^*}). Then, each vehicle k executes its walk w_k^* while observing the road segments $S_{w_k^*}$, and updates its residing road segment and stored data.

To derive the most informative joint walk w^*, the posterior entropy (28) of every possible joint walk w has to be evaluated. Such a *centralized active sensing* (CAS) algorithm cannot be performed in real time due to the following two issues: (a) It relies on all the data that are gathered distributedly by the vehicles, thus incurring huge time and communication overheads with large data, and (b) it involves evaluating a prohibitively large number of joint walks (i.e., exponential in the fleet size).

The first issue can be alleviated by approximating the Gaussian posterior joint entropy using the decentralized GP-DDF or GP-DDF$^+$ algorithm, thus distributing its computational load among all vehicles. Then, the active sensing problem (28) is approximated by

$$w^* = \underset{w}{\arg\max}\, \overline{\mathbb{H}}\left[Z_{S_w}\right] \quad (29)$$

$$\overline{\mathbb{H}}\left[Z_{S_w}\right] \triangleq \frac{1}{2}\log(2\pi e)^{|S_w|}\left|\overline{\Sigma}_{S_w S_w}\right|. \quad (30)$$

To obtain $\overline{\mathbb{H}}\left[Z_{S_w}\right]$ (30), $\Sigma_{S_w S_w|D}$ in $\mathbb{H}[Z_{S_w}|Z_D]$ ((4) and (28)) is replaced by $\overline{\Sigma}_{S_w S_w}$ of GP-DDF and GP-DDF$^+$ defined in Theorems 1A and 2A, respectively.

On the other hand, if the vehicles are actively cruising the service area to gather the most informative mobility demand data (e.g., pickup counts of different regions) for predicting a mobility demand pattern being modeled with a ℓGP (Section II-C), then Z_{S_w} and Z_D in (28) and (29) have to be replaced by Y_{S_w} and Y_D, respectively. Similarly, the active sensing problem (28) is approximated by $w^* = \arg\max_w \overline{\mathbb{H}}\left[Y_{S_w}\right]$ where

$$\overline{\mathbb{H}}\left[Y_{S_w}\right] \triangleq \frac{1}{2}\log(2\pi e)^{|S_w|}\left|\overline{\Sigma}_{S_w S_w}\right| + \overline{\mu}_{S_w} \cdot \mathbf{1}\ . \quad (31)$$

To obtain $\overline{\mathbb{H}}\left[Y_{S_w}\right]$ (31), $\mu_{S_w|D}$ and $\Sigma_{S_w S_w|D}$ in $\mathbb{H}[Y_{S_w}|Y_D]$ (8) are replaced by $\overline{\mu}_{S_w}$ and $\overline{\Sigma}_{S_w S_w}$ of GP-DDF and GP-DDF$^+$ defined in Theorems 1A and 2A, respectively.

To address the second issue, a simple and highly scalable (i.e., in the fleet size) strategy is to adopt a *fully decentralized active sensing* (full DAS) algorithm by assuming that the joint walk $w_1^* \dots w_K^*$ is derived by selecting the locally optimal walk of each vehicle k:

$$w_k^* = \underset{w_k}{\arg\max}\, \overline{\mathbb{H}}\left[Z_{S_{w_k}}\right] \quad (32)$$

where $\overline{\mathbb{H}}\left[Z_{S_{w_k}}\right]$ is defined in the same way as (30). Then, each vehicle can select its locally optimal walk independently of the other vehicles, thus significantly reducing the search space of joint walks. A consequence of such an assumption is that, without coordinating their walks, the vehicles may select suboptimal joint walks (e.g., two vehicles' locally optimal walks are highly correlated). In practice, this assumption becomes less restrictive when the size $|D|$ of data increases to potentially reduce the degree of violation of conditional independence of $Z_{S_{w_1}}, \dots, Z_{S_{w_K}}$.

When our full DAS algorithm is instead used for sampling a mobility demand pattern being modeled with a ℓGP, it becomes $w_k^* = \arg\max_{w_k} \overline{\mathbb{H}}[Y_{S_{w_k}}]$ where $\overline{\mathbb{H}}[Z_{S_{w_k}}]$ is defined in the same way as (31). It can then be observed from (31) that the cruising behavior of full DAS trades off between exploring sparsely sampled regions with high predictive uncertainty (i.e., by maximizing the log-determinant of Gaussian posterior covariance matrix $\overline{\Sigma}_{S_{w_k} S_{w_k}}$ term) and hotspots (i.e., by maximizing the Gaussian posterior mean vector $\overline{\mu}_{S_{w_k}}$ term). As a result, it redistributes vacant MoD vehicles to regions with high likelihood of picking up users. Hence, besides gathering the most informative data for predicting the mobility demand pattern, full DAS is able to achieve a dual effect of fleet rebalancing to service mobility demands.

A. Partially Decentralized Active Sensing (Partial DAS)

Our full DAS algorithm (32) has assumed no coordination between the vehicles in their walks, which may not always seem necessary to gain enough time efficiency to perform real-time active sensing of any traffic phenomenon. To relax this assumption while still preserving scalability, our strategy is to instead partition the vehicles into several small groups such that each group of vehicles selects its joint walk independently, which is the intuition underlying our *partially decentralized active sensing* (partial DAS) algorithm to be described next. Note that, unlike full DAS, our partial DAS algorithm can only be coupled with GP-DDF instead of GP-DDF$^+$ (Section III) because it exploits an inherent structural assumption of GP-DDF for decentralization that does not hold for GP-DDF$^+$.

The key idea of our partial DAS algorithm is to construct a block-diagonal matrix whose log-determinant closely approximates that of $\overline{\Sigma}_{S_w S_w}$ (14) and exploit the property that the log-determinant of such a block-diagonal matrix can be decomposed into a sum of log-determinants of its diagonal blocks, each of which depends only on the walks of a disjoint subset of the K vehicles. Consequently, the active sensing problem can be partially decentralized leading to a reduced space of possible joint walks to be searched, as detailed in the rest of this subsection.

Firstly, we extend an earlier structural assumption of GP-DDF (Section III): $Z_{D_1},\ldots,Z_{D_K},Z_{S_{w_1}},\ldots,Z_{S_{w_K}}$ are conditionally independent given Z_U. Then, it can be shown via the equivalence to PITC (Theorem 1B) that $\overline{\Sigma}_{S_wS_w}$ (14) comprises diagonal blocks of the form $\overline{\Sigma}_{S_{w_k}S_{w_k}}$ for $k=1,\ldots,K$ and off-diagonal blocks of the form $\Sigma_{S_{w_k}U}\ddot{\Sigma}^{-1}_{UU}\Sigma_{US_{w_{k'}}}$ for $k,k'=1,\ldots,K$ and $k\neq k'$. In particular, each off-diagonal block of $\overline{\Sigma}_{S_wS_w}$ represents the correlation of measurements between the unobserved road segments S_{w_k} and $S_{w_{k'}}$ along the respective walks w_k of vehicle k and $w_{k'}$ of vehicle k'. If the correlation between some pair of their possible walks is high enough, then their walks have to be coordinated. This is formally realized by the following coordination graph over the K vehicles:

Definition 6 (Coordination Graph): Define the coordination graph to be an undirected graph $\mathcal{G}\triangleq(\mathcal{V},\mathcal{E})$ that comprises

- a set $\mathcal{V}$ of vertices denoting the K vehicles, and
- a set $\mathcal{E}$ of edges denoting coordination dependencies between vehicles such that there exists an edge $\{k,k'\}$ incident with vehicles $k\in\mathcal{V}$ and $k'\in\mathcal{V}\setminus\{k\}$ iff

$$\max_{s\in S_{W_k},s'\in S_{W_{k'}}}\left|\Sigma_{sU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{Us'}\right|>\varepsilon \tag{33}$$

for a predefined constant $\varepsilon>0$ where W_k denotes the set of possible walks of length L of vehicle k from its residing road segment and $S_{W_k}\triangleq\bigcup_{w_k\in W_k}S_{w_k}$.

Remark. The construction of $\mathcal{G}$ can be decentralized as follows: Since $\ddot{\Sigma}_{UU}$ is symmetric and positive definite, it can be decomposed by Cholesky factorization into $\ddot{\Sigma}_{UU}=\Psi\Psi^\top$ where Ψ is a lower triangular matrix and $\Psi^\top$ is the transpose of Ψ. Then, $\Sigma_{sU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{Us'}=(\Psi\backslash\Sigma_{Us})^\top\Psi\backslash\Sigma_{Us'}$ where $\Psi\backslash B$ denotes the column vector ϕ solving $\Psi\phi=B$. That is, $\Sigma_{sU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{Us'}$ (33) can be expressed as a dot product of two vectors $\Psi\backslash\Sigma_{Us}$ and $\Psi\backslash\Sigma_{Us'}$; this property is exploited to determine adjacency between vehicles in a decentralized manner:

Definition 7 (Adjacency): Let

$$\Phi_k\triangleq\{\Psi\backslash\Sigma_{Us}\}_{s\in S_{W_k}} \tag{34}$$

for $k=1,\ldots,K$. A vehicle $k\in\mathcal{V}$ is adjacent to vehicle $k'\in\mathcal{V}\setminus\{k\}$ in coordination graph $\mathcal{G}$ iff

$$\max_{\phi\in\Phi_k,\phi'\in\Phi_{k'}}\left|\phi^\top\phi'\right|>\varepsilon\ . \tag{35}$$

It follows from the above definition that if each vehicle k constructs Φ_k and exchanges it with every other vehicle, then it can determine its adjacency to all the other vehicles and store this information in a column vector a_k of length K with its k'-th component being defined as follows:

$$[a_k]_{k'}=\begin{cases}1 & \text{if vehicle } k \text{ is adjacent to vehicle } k',\\ 0 & \text{otherwise.}\end{cases} \tag{36}$$

By exchanging its adjacency vector a_k with every other vehicle, each vehicle can construct a globally consistent adjacency matrix $A_{\mathcal{G}}\triangleq(a_1\ldots a_K)$ to represent coordination graph $\mathcal{G}$.

Next, by computing the connected components (say, $\mathcal{K}$ of them) of coordination graph $\mathcal{G}$, their resulting vertex sets partition the set $\mathcal{V}$ of K vehicles into $\mathcal{K}$ disjoint subsets $\mathcal{V}_1,\ldots,\mathcal{V}_{\mathcal{K}}$ such that the vehicles within each subset have to coordinate their walks. Each vehicle can determine its residing connected component in a decentralized way by performing a depth-first search in $\mathcal{G}$ starting from it as root.

Finally, construct a block-diagonal matrix $\widehat{\Sigma}_{S_wS_w}$ to comprise diagonal blocks of the form $\overline{\Sigma}_{S_{w\mathcal{V}_n}S_{w\mathcal{V}_n}}$ for $n=1,\ldots,\mathcal{K}$ where $w_{\mathcal{V}_n}\triangleq(w_k)_{k\in\mathcal{V}_n}$ and $S_{w\mathcal{V}_n}\triangleq\bigcup_{k\in\mathcal{V}_n}S_{w_k}$. The active sensing problem (29) is then approximated by

$$\begin{aligned}&\max_w\ \frac{1}{2}\log(2\pi e)^{|S_w|}\left|\widehat{\Sigma}_{S_wS_w}\right|\\ &\equiv\max_{(w_{\mathcal{V}_1},\ldots,w_{\mathcal{V}_{\mathcal{K}}})}\sum_{n=1}^{\mathcal{K}}\log(2\pi e)^{|S_{w\mathcal{V}_n}|}\left|\overline{\Sigma}_{S_{w\mathcal{V}_n}S_{w\mathcal{V}_n}}\right|\\ &=\sum_{n=1}^{\mathcal{K}}\max_{w\mathcal{V}_n}\ \log(2\pi e)^{|S_{w\mathcal{V}_n}|}\left|\overline{\Sigma}_{S_{w\mathcal{V}_n}S_{w\mathcal{V}_n}}\right|,\end{aligned} \tag{37}$$

which can be solved in a partially decentralized manner by each disjoint subset $\mathcal{V}_n$ of vehicles:

$$\widehat{w}_{\mathcal{V}_n}=\arg\max_{w\mathcal{V}_n}\ \log(2\pi e)^{|S_{w\mathcal{V}_n}|}\left|\overline{\Sigma}_{S_{w\mathcal{V}_n}S_{w\mathcal{V}_n}}\right|. \tag{38}$$

Our partial DAS algorithm becomes fully decentralized (i.e., full DAS) if ε is set to be sufficiently large: More vehicles become isolated in $\mathcal{G}$, consequently decreasing the size $\kappa\triangleq\max_n|\mathcal{V}_n|$ of its largest connected component to 1. As shown in Section V-B1, decreasing κ improves its time efficiency. On the other hand, it tends to a centralized behavior (29) by setting $\varepsilon\rightarrow 0^+$: $\mathcal{G}$ becomes near-complete, thus resulting in $\kappa\rightarrow K$.

Let

$$\xi\triangleq\max_{n,w_{\mathcal{V}_n},i,i'}\left|\left[\left(\overline{\Sigma}_{S_{w\mathcal{V}_n}S_{w\mathcal{V}_n}}\right)^{-1}\right]_{ii'}\right| \tag{39}$$

and $\epsilon\triangleq 0.5\log 1\Big/\Big(1-\big(K^{1.5}L^{2.5}\kappa\xi\varepsilon\big)^2\Big)$. In the result below, we prove that the joint walk $\widehat{w}\triangleq(\widehat{w}_{\mathcal{V}_1},\ldots,\widehat{w}_{\mathcal{V}_{\mathcal{K}}})$ is guaranteed to achieve an entropy $\overline{\mathbb{H}}[Z_{S_{\widehat{w}}}]$ (i.e., by plugging $\widehat{w}$ into (30)) that is not more than ϵ from the maximum entropy $\overline{\mathbb{H}}[Z_{S_{w^*}}]$ achieved by joint walk w^* (29):

Theorem 3 (Performance Guarantee): If $K^{1.5}L^{2.5}\kappa\xi\varepsilon<1$, then $\overline{\mathbb{H}}[Z_{S_{w^*}}]-\overline{\mathbb{H}}[Z_{S_{\widehat{w}}}]\leq\epsilon$.

The proof of Theorem 3 is given in Appendix C. The implication of Theorem 3 is that our partial DAS algorithm can perform comparatively well (i.e., small ϵ) under the following favorable spatiotemporal correlation structure of the phenomenon and parameter settings of the MoD system: (a) The network of K vehicles is not large, (b) the length L of each vehicle's walk to be optimized is not long, (c) the largest subset of κ vehicles being formed to coordinate their walks (i.e., largest connected component in $\mathcal{G}$) is reasonably small, which occurs when the phenomenon exhibits a sufficiently small degree of spatiotemporal correlation given some predefined ε, and (d) the minimum required correlation ε between walks of adjacent vehicles is kept low. Our partial DAS algorithm is not used for sampling a mobility demand pattern being modeled with a ℓGP: It is not time-efficient when κ is large, which is indeed the case because many vehicles tend to cluster within hotspots, as explained earlier.

TABLE I
COMPARISON OF TIME AND COMMUNICATION COMPLEXITY BETWEEN DATA FUSION AND ACTIVE SENSING ALGORITHMS ANALYZED IN SECTION V.

Algorithm	Data Fusion		Active Sensing	
	Computation	Communication	Computation	Communication
GP-DDF$^+$+Full DAS	$\mathcal{O}\Big((\frac{\lvert D\rvert}{K})^3 + \lvert U\rvert^3 + \lvert U\rvert^2 K\Big)$	$\mathcal{O}(\lvert U\rvert^2)$	$\mathcal{O}\Big(\Delta(L^3 + (L\lvert U\rvert)^2 + (\frac{\lvert D\rvert}{K})^2)\Big)$	$\mathcal{O}(\Delta\lvert U\rvert)$
GP-DDF+Full DAS GP-DDF+Partial DAS	$\mathcal{O}\Big((\frac{\lvert D\rvert}{K})^3 + \lvert U\rvert^3 + \lvert U\rvert^2 K\Big)$	$\mathcal{O}(\lvert U\rvert^2)$	$\mathcal{O}\big(\Delta(L^3 + (L\lvert U\rvert)^2)\big)$ $\mathcal{O}\big(\kappa\Delta L\lvert U\rvert^2 + \Delta^2 L^2\lvert U\rvert(K+\kappa^2) + \Delta^\kappa(\kappa L)^3\big)$	– $\mathcal{O}(\Delta L\lvert U\rvert + K)$
Full GP+Full DAS Full GP+CAS	$\mathcal{O}(\lvert D\rvert^3)$	$\mathcal{O}\Big(\frac{\lvert D\rvert}{K}\Big)$	$\mathcal{O}\big(\Delta(L^3 + (L\lvert D\rvert)^2)\big)$ $\mathcal{O}\big(\Delta^K KL(\lvert D\rvert^2 + (KL)^2)\big)$	– –
Subset of Data+CAS	$\mathcal{O}(\lvert U\rvert^3\lvert D\rvert)$	$\mathcal{O}\Big(\frac{\lvert D\rvert}{K}\Big)$	$\mathcal{O}\big(\Delta^K KL(\lvert U\rvert^2 + (KL)^2)\big)$	–

V. TIME AND COMMUNICATION ANALYSIS

A. GP-DDF$^+$ coupled with Full DAS

In this section, the time and communication overheads (TABLE I) of our proposed GP-DDF$^+$ coupled with full DAS algorithm (Algorithm 1) are analyzed and compared to that of both full GP (Section II-C) and GP-DDF (Section III) coupled with full DAS (Section IV) algorithms. As shown in Section VI-A later, Algorithm 1 will be run on each vehicle k to perform active sensing of a real-world mobility demand pattern being modeled with a ℓGP. Note that GP-DDF$^+$ cannot be coupled with partial DAS, as explained earlier in Section IV-A.

Algorithm 1: GP-DDF$^+$+Full DAS$(U, K, L, k, D_k, y_{D_k})$

while *true* **do**
- `/* Data fusion (Section III) */`
- Construct local summary by (9) & (10)
- Exchange local summary with every vehicle $i \neq k$
- Construct global summary by (11) & (12)
- Construct assignment function by (21)
- Predict demand measurements for unobserved regions by (18) & (22)
- `/* Active Sensing (Section IV) */`
- Compute local maximum-entropy walk w_k^* by (32)
- Execute walk w_k^* and observe its demand measurements $Y_{w_k^*}$
- Update local information D_k and y_{D_k}

1) Time Complexity: Firstly, each vehicle k has to evaluate $\Sigma_{D_k D_k|U}$ in $\mathcal{O}\big(|U|^3 + |U|(|D|/K)^2\big)$ time and invert it in $\mathcal{O}\big((|D|/K)^3\big)$ time. After that, GP-DDF$^+$ constructs the local summary in $\mathcal{O}\big(|U|^2|D|/K + |U|(|D|/K)^2\big)$ time by (9) and (10), and subsequently the global summary in $\mathcal{O}\big(|U|^2K\big)$ time by (11) and (12). To construct the assignment function for any unobserved set $S \subset V$, vehicle k first computes $|S|$ number of γ_{sU}^k for all unobserved regions $s \in S$ in $\mathcal{O}\big(|S||U|^2 + |S|(|D|/K)^2\big)$ time by (20). Then, after inverting $\ddot{\Sigma}_{UU}$ in $\mathcal{O}\big(|U|^3\big)$, the predictive means and variances for all $s \in S$ are computed in $\mathcal{O}\big(|S||U|^2 + |S|(|D|/K)^2\big)$ time by (18) and (22), respectively. Let $\Delta \triangleq \delta^L$ denote the number of possible walks of length L where δ is the maximum out-degree of graph G. In full DAS, to obtain the locally optimal walk, the log-Gaussian posterior entropies (32) of all possible walks are derived from (18) and (22), respectively, in $\mathcal{O}\big(\Delta L|U|^2\big)$ and $\mathcal{O}\big(\Delta(L|U|)^2\big)$ time. We assume $|S| < \delta\Delta$ where S denotes the set $\bigcup_{w_k} S_{w_k}$ of regions covered by any vehicle k's all possible walks of length L. Then, the time complexity for our GP-DDF$^+$ coupled with full DAS algorithm is $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K + \Delta(L^3 + (L|U|)^2 + (|D|/K)^2)\big)$. In contrast, the time incurred by full GP and GP-DDF coupled with full DAS algorithms are, respectively, $\mathcal{O}\big(|D|^3 + \Delta(L^3 + (L|D|)^2)\big)$ and $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K + \Delta(L^3 + (L|U|)^2)\big)$. It can be observed that our GP-DDF$^+$ coupled with full DAS algorithm can scale better with large size $|D|$ of data and fleet size K than full GP coupled with full DAS algorithm, and its increased computational load, as compared to GP-DDF coupled with full DAS algorithm, is well distributed among K vehicles.

2) Communication Complexity: In each iteration, each vehicle of the system running our GP-DDF$^+$ coupled with full DAS algorithm has to broadcast a $\mathcal{O}\big(|U|^2\big)$-sized local summary for constructing the global summary, exchange $\mathcal{O}(\Delta)$ scalar values for constructing the assignment function, and request $\mathcal{O}(\Delta)$ number of $\mathcal{O}(|U|)$-sized γ_{sU}^k components for evaluating the entropies of all possible local walks. In contrast, full GP coupled with full DAS algorithm needs to broadcast $\mathcal{O}(|D|/K)$-sized message comprising all its local data to handle communication failure, and GP-DDF coupled with full DAS algorithm only needs to broadcast a $\mathcal{O}\big(|U|^2\big)$-sized local summary.

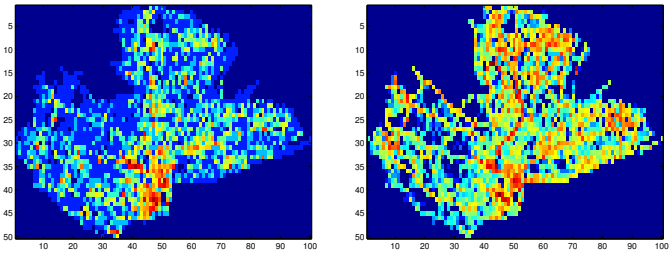
B. GP-DDF coupled with Partial DAS

In this section, the time and communication overheads (TABLE I) of our proposed GP-DDF coupled with partial DAS algorithm (Algorithm 2) are analyzed and compared to that of both full GP (Section II) and subset of data (Section II-A) coupled with *centralized active sensing* (CAS) (28) algorithms. We call Algorithm 2 D^{2}FAS to follow our original naming convention in [1]. As shown in Section VI-B later, our D^{2}FAS algorithm (Algorithm 2) will be run on each vehicle k to perform active sensing of a real-world traffic phenomenon featuring speeds of road segments over an urban road network being modeled with a relational GP. Note that partial DAS cannot be used for sampling a mobility demand pattern being modeled with a ℓGP, as explained earlier in Section IV-A.

1) Time Complexity: The data fusion component of D^{2}FAS (i.e., GP-DDF) involves computing the local and global summaries and the Gaussian predictive distribution. To construct the local summary using (9) and (10), each vehicle has to evaluate $\Sigma_{D_k D_k|U}$ in $\mathcal{O}\big(|U|^3 + |U|(|D|/K)^2\big)$ time and invert it in $\mathcal{O}\big((|D|/K)^3\big)$ time, after which the local summary is obtained in $\mathcal{O}\big(|U|^2|D|/K + |U|(|D|/K)^2\big)$ time. The global summary is computed in $\mathcal{O}\big(|U|^2K\big)$ by (11) and

Algorithm 2: D^{2}FAS$(U, K, L, k, D_k, z_{D_k})$

while *true* **do**
- /* Data fusion (Section III) */
- Construct local summary by (9) & (10)
- Exchange local summary with every vehicle $i \neq k$
- Construct global summary by (11) & (12)
- Predict measurements for unobserved road segments by (13) & (14)
- /* Active Sensing (Section IV-A) */
- Construct Φ_k by (34)
- Exchange Φ_k with every vehicle $i \neq k$
- Compute adjacency vector a_k by (35) & (36)
- Exchange adjacency vector with every vehicle $i \neq k$
- Construct adjacency matrix of coordination graph
- Find vertex set $\mathcal{V}_n$ of its residing connected component
- Compute maximum-entropy joint walk $\widehat{w}_{\mathcal{V}_n}$ by (38)
- Execute walk $\widehat{w}_k$ and observe road segments $S_{\widehat{w}_k}$
- Update local information D_k and z_{D_k}



(a) Demand (b) Supply

Fig. 1. Historic (a) demand and (b) supply distribution: 'Hotter' regions indicate larger numbers of (a) pickups and (b) cruising taxis within a 30-minute time slot in the central business district of Singapore.

(12). Finally, the Gaussian predictive distribution is derived in $\mathcal{O}\big(|U|^3 + |U||S|^2\big)$ time using (13) and (14). Supposing $|S| \leq |U|$ for simplicity, the time complexity of GP-DDF is then $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2 K\big)$.

Let the maximum out-degree of G be denoted by δ. Then, each vehicle has to consider $\Delta \triangleq \delta^L$ possible walks of length L. The active sensing component of D^{2}FAS (i.e., partial DAS) involves computing Φ_k in $\mathcal{O}\big(\Delta L|U|^2\big)$ time, a_k in $\mathcal{O}\big(\Delta^2 L^2|U|K\big)$ time, its residing connected component in $\mathcal{O}\big(\kappa^2\big)$ time, and the maximum-entropy joint walk by (14) and (38) with the following incurred time: The largest connected component of κ vehicles in $\mathcal{G}$ has to consider Δ^κ possible joint walks. Note that $\overline{\Sigma}_{S_{w_{\mathcal{V}_n}} S_{w_{\mathcal{V}_n}}} = \mathrm{diag}\Big((\Sigma_{S_{w_k} S_{w_k}|U})_{k \in \mathcal{V}_n}\Big) + \Sigma_{S_{w_{\mathcal{V}_n}} U} \ddot{\Sigma}_{UU}^{-1} \Sigma_{U S_{w_{\mathcal{V}_n}}}$ where $\mathrm{diag}(B)$ constructs a diagonal matrix by placing vector B on its diagonal. By exploiting Φ_k, the diagonal and latter matrix terms for all possible joint walks can be computed in $\mathcal{O}\big(\kappa\Delta(L|U|^2 + L^2|U|)\big)$ and $\mathcal{O}\big(\kappa^2\Delta^2 L^2|U|\big)$ time, respectively. For each joint walk $w_{\mathcal{V}_n}$, evaluating the determinant of $\overline{\Sigma}_{S_{w_{\mathcal{V}_n}} S_{w_{\mathcal{V}_n}}}$ incurs $\mathcal{O}\big((\kappa L)^3\big)$ time. Therefore, the time complexity of partial DAS is $\mathcal{O}\big(\kappa\Delta L|U|^2 + \Delta^2 L^2|U|(K + \kappa^2) + \Delta^\kappa(\kappa L)^3\big)$.

Hence, the time complexity of our D^{2}FAS algorithm is $\mathcal{O}\big((|D|/K)^3 + |U|^2(|U| + K + \kappa\Delta L) + \Delta^2 L^2|U|(K + \kappa^2) + \Delta^\kappa(\kappa L)^3\big)$. In contrast, the time incurred by full GP and subset of data coupled with CAS are, respectively, $\mathcal{O}\big(|D|^3 + \Delta^K KL(|D|^2 + (KL)^2)\big)$ and $\mathcal{O}\big(|U|^3|D| + \Delta^K KL(|U|^2 + (KL)^2)\big)$. It can be observed that D^{2}FAS can scale better with large size $|D|$ of data and fleet size K. The scalability of D^{2}FAS vs. full GP and subset of data will be further evaluated empirically in Section VI-B.

2) Communication Complexity: Let the communication overhead be defined as the size of each broadcast message. Recall the data fusion component of D^{2}FAS (i.e., GP-DDF) in Algorithm 2 that, in each iteration, each vehicle broadcasts a $\mathcal{O}\big(|U|^2\big)$-sized summary encapsulating its local data, which is robust against communication failure. In contrast, full GP and subset of data require each vehicle to broadcast, in each iteration, a $\mathcal{O}(|D|/K)$-sized message comprising exactly its local data to handle communication failure. If the size of local data grows to be larger in size than a local summary of predefined size, then GP-DDF is more scalable than full GP and subset of data in terms of communication overhead. For the active sensing component of D^{2}FAS (i.e., partial DAS), each vehicle broadcasts $\mathcal{O}(\Delta L|U|)$-sized Φ_k and $\mathcal{O}(K)$-sized a_k messages.

VI. Experiments and Discussion

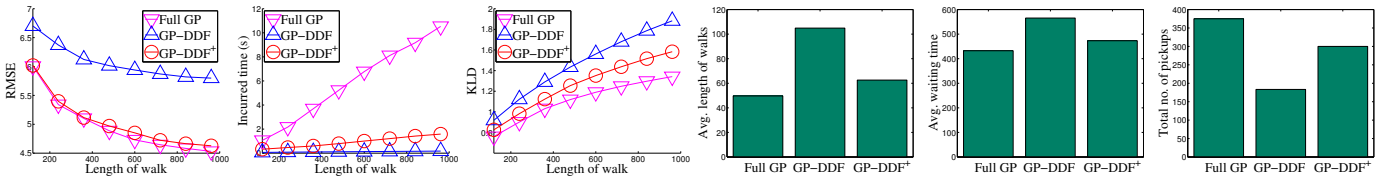
This section evaluates the performance of our proposed Gaussian process decentralized data fusion and active sensing algorithms empirically on two datasets featuring real-world traffic phenomena, which include a mobility demand pattern over the central business district of Singapore (Section VI-A) and speeds of road segments over an urban road network in Singapore (Section VI-B).

A. Mobility Demand Pattern Modeled with ℓGP

We use a real-world taxi trajectory dataset taken from the central business district of Singapore between 9:30 p.m. and 10 p.m. on August 2, 2010. The 29.4 km $\times$ 11.9 km service area is gridded into 100×50 regions such that 2506 regions are included into the dataset as the remaining regions contain no road segment for cruising vehicles to access. The maximum out-degree δ of graph G over these regions is 8. The feature vector of each region is specified by its corresponding location. In any region, the demand (supply) measurement is obtained by counting the number of pickups (taxis cruising by) from all historic taxi trajectories generated by a major taxi company in a 30-minute time slot. After processing the taxi trajectories, the historic demand and supply distributions are obtained, as shown in Fig. 1. Then, a number C of users are randomly distributed over the service area with their locations drawn from the demand distribution (Fig. 1a). Similarly, a fleet of K vacant MoD vehicles are initialized at locations drawn from the supply distribution (Fig. 1b).

In our simulation, when a vehicle enters a region with users, it picks up one of them randomly. Then, the MoD system removes this vehicle from the fleet of vacant cruising vehicles and introduces a new vacant vehicle drawn from the supply distribution. Similarly, a new user appears at a random location drawn from the demand distribution. The MoD system operates for T time steps and each vehicle plans a walk of length $L = 4$ at each time step, with all vehicles running a data fusion algorithm coupled with our full DAS algorithm.

We will compare the performance of our GP-DDF$^+$ algorithm with that of full GP and GP-DDF algorithms when



(a) Accuracy (b) Efficiency (c) Balance (d) Vehicles (e) Users (f) Pickups

Fig. 2. Performance of MoD systems in sensing, predicting, and servicing mobility demands.

coupled with our full DAS algorithm. Note that our partial DAS algorithm cannot be used for sampling a mobility demand pattern being modeled with a ℓGP, as explained earlier in Section IV-A. The experiments are conducted on a Linux system with Intel® Xeon® CPU E5520 at 2.27 GHz.

The tested algorithms are evaluated with two sets of performance metrics. The performance of sensing and predicting mobility demands is evaluated using (a) *root mean square error* (RMSE) $\sqrt{|V|^{-1}\sum_{s\in V}\left(y_s-\widehat{\mu}_{s|D}\right)^2}$ where y_s is the demand measurement and D is the set of regions observed by the MoD vehicles, and (b) incurred time of the algorithms. The performance of servicing mobility demands is evaluated by comparing the *Kullback-Leibler divergence* (KLD) $\sum_{s\in V} P_c(s)\log\left(P_c(s)/P_d(s)\right)$ between the fleet distribution P_c of vacant MoD vehicles controlled by the tested algorithms and historic demand distribution P_d (i.e., lower KLD implies better balance between fleet and demand), average cruising length of MoD vehicles, average waiting time of users, and total number of pickups resulting from the tested algorithms.

For notational simplicity, we will use GP-DDF$^+$, full GP, and GP-DDF to represent the algorithms of their corresponding data fusion components coupled with full DAS algorithm.

1) Performance: The MoD system comprises $K = 20$ vehicles running three tested algorithms for $T = 960$ time steps in a service area with $C = 200$ users. All results are taken from the average of 40 random instances.

The performance of MoD systems in sensing and predicting mobility demands is illustrated in Figs. 2a-2b. Fig. 2a shows that the demand data collected by MoD vehicles using GP-DDF$^+$ can achieve predictive accuracy comparable to that of using full GP and significantly better than that of using GP-DDF. This indicates that exploiting the local data of vehicles for predicting demands of nearby unobserved regions can improve the prediction of the mobility demand pattern. Fig. 2b shows the average incurred time of each vehicle using three algorithms. GP-DDF$^+$ is significantly more time-efficient (i.e., one order of magnitude) than full GP, and only slightly less time-efficient than GP-DDF. This can be explained by the time analysis in Section V-A1. The above results indicate that GP-DDF$^+$ is more practical for real-world deployment due to a better balance between predictive accuracy and time efficiency.

The performance of MoD systems in servicing the mobility demands is illustrated in Figs. 2c-2f. Fig. 2c shows that a MoD system using GP-DDF$^+$ can achieve better fleet rebalancing of vehicles to service mobility demands than GP-DDF, but worse rebalancing than full GP. This implies that a better prediction of the underlying mobility demand pattern (Fig. 2a) can lead to better fleet rebalancing. Note that KLD (i.e., imbalance between mobility demand and fleet) increases over time because we assume that when a vehicle picks up a user, its local data is removed from the fleet of cruising vehicles, and a new vehicle is introduced at a random location that may be distant from a demand hotspot, hence worsening the imbalance between demand and fleet. It can also be observed that an algorithm generating a better balance between fleet and demand will also perform better in servicing the mobility demands, that is, shorter average cruising trajectories of vehicles (Fig. 2d), shorter average waiting time of users (Fig. 2e), and larger total number of pickups (Fig. 2f). These observations imply that exploiting an active sensing algorithm to collect the most informative demand data for predicting the mobility demand pattern achieves a dual effect of improving performance in servicing the mobility demands since these vehicles have higher chance of picking up users in demand hotspots or sparsely sampled regions (Section IV).

2) Scalability: We vary the number $K = 10, 20, 30$ of vehicles in the MoD system, and keep the total length of walks of all the vehicles to be the same, that is, these vehicles will walk for $T = 960, 480, 320$ steps, respectively. All three algorithms are tested in a service area with $C = 600$ users. All results are obtained by averaging over 40 random instances.

From Figs. 3a-3c, it can be observed that all three algorithms can improve their prediction accuracy with an increasing number of vehicles in the MoD system because more vehicles indicate less walks when the total length of walks are the same, thus suffering less from the myopic planning ($L = 4$) and gathering more informative demand data. Figs. 3d-3f show that, with more MoD vehicles, GP-DDF$^+$ and GP-DDF incur less time, while full GP incurs more time. This is because the computational load in decentralized data fusion algorithms are distributed among all vehicles, thus reducing the incurred time with more vehicles.

Figs. 4a-4c show that all three algorithms can achieve better balance between mobility demand and fleet with larger number of vehicles. It can also be observed that all three algorithms can improve the performance of servicing the mobility demand with more vehicles, that is, shorter average cruising trajectories of vehicles (Fig. 4d), shorter average waiting time of users (Fig. 4e), and larger total number of pickups (Fig. 4f). This is because MoD vehicles can collect more informative demand data with larger number of vehicles sampling demand hotspots or sparsely sampled regions, which are the regions with higher chance of picking up users than the rest of the service area.

The above results indicate that more vehicles in MoD system result in better accuracy in predicting the mobility demand pattern, and achieve a dual effect of better performance in

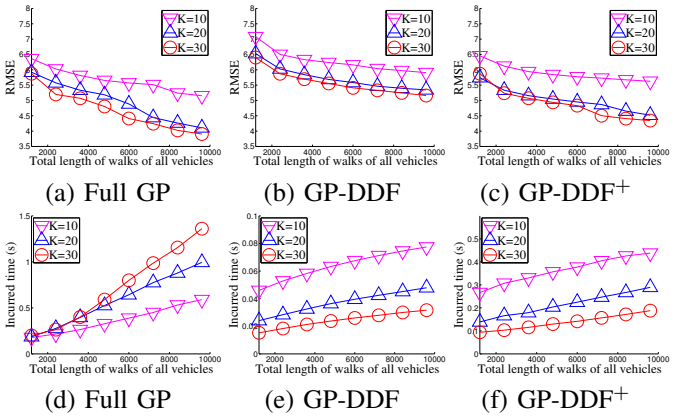


Fig. 3. Scalability of MoD systems in sensing and predicting mobility demands.

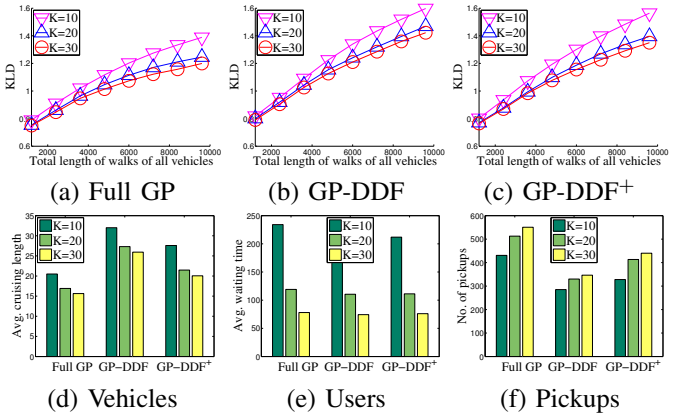


Fig. 4. Scalability of MoD systems in servicing mobility demands.

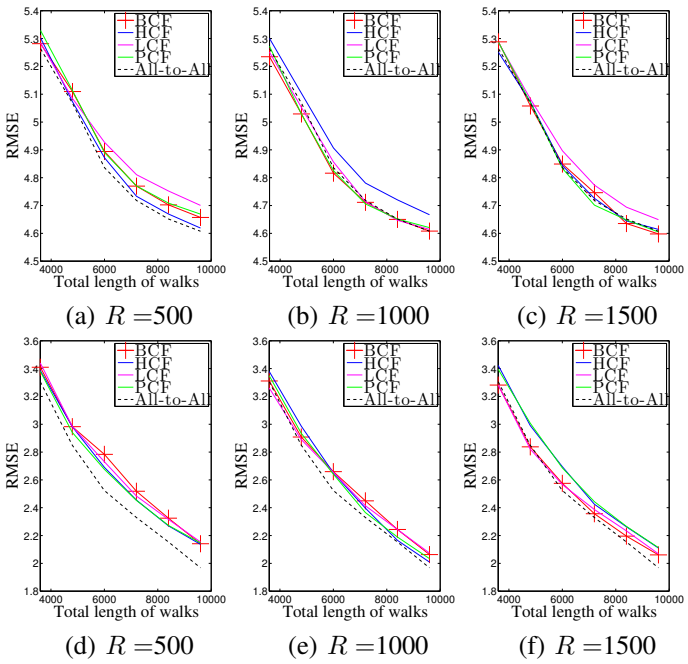


Fig. 5. Predictive performance of different consensus filtering-based GP-DDF$^+$ (8 iterations per time step) and all-to-all communication-based GP-DDF$^+$ coupled with full DAS algorithms with varying communication range R (meters). The experimental setting for the results in (d-f) differs from that in (a-c) in that all vehicles are fully dedicated to mobility demand sensing and are not removed by the MoD system, as elaborated in the main text.

servicing mobility demands.

3) Predictive Performance of Consensus Filtering-Based GP-DDF$^+$: Fig. 5 compares results of the performance of MoD systems in sensing and predicting mobility demands between the all-to-all communication-based GP-DDF$^+$ and different consensus filtering-based GP-DDF$^+$ (Section III-A) coupled with full DAS (Section IV) algorithms. In particular, four consensus filtering variants of GP-DDF$^+$ are implemented: Band-pass consensus filter (BCF) [34], low-pass consensus filter (LCF) [37], high-pass consensus filter (HCF) [36], and proportional consensus filter (PCF) [38]. PCF is an extension of HCF to resolve the incorrect initialization of internal estimators caused by splitting and merging of communication network. For notational simplicity, we will use BCF, LCF, HCF, and PCF to represent their respective implementations of consensus filtering-based GP-DDF$^+$ coupled with full DAS algorithms.

The MoD system comprises $K = 30$ vehicles with varying communication range[6] $R = 500, 1000, 1500$ meters running the tested algorithms for $T = 320$ time steps in a service area with $C = 200$ users. All results are taken from the average of 40 random instances.

When the total length of walks of all vehicles is less than 3600 (i.e., $T \leq 120$), the predictive performance of the consensus filtering-based GP-DDF$^+$ coupled with full DAS algorithms are comparable to that of all-to-all communication-based GP-DDF$^+$ coupled with full DAS algorithm and their results are hence omitted from Fig. 5.

From Figs. 5a and 5c, it can be observed that HCF and PCF outperform LCF. This is due to a rapid change in local data: When a vacant MoD vehicle picks up a user, the MoD system removes it from the fleet and a new randomly deployed vehicle is introduced into the fleet. Then, the new vehicle can gather demand data that is distinct from the previous demand data. This rapid change in local data will inevitably result in a rapidly-varying local summary, which favors the HCF and PCF. Fig. 5b shows that HCF performs worse than LCF while PCF still achieves similar predictive performance as LCF. This is because HCF suffers from splitting and merging of communication network at $R = 1000$. Furthermore, we design an additional experimental setup that uses the same setting as above, except that all vehicles are fully dedicated to mobility demand sensing and are not removed by the MoD system. This makes the local data change slowly, hence yielding slowly-varying local summaries. Figs. 5d-5f show the corresponding results: It can be observed that LCF achieves slightly better performance than HCF and PCF, as expected. In both of the above setups, the performance of BCF is stable and comparable to that of the best-performing consensus filter, thus implying that BCF can handle local summaries with wider range of variations.

[6]It is reported in [39] that 802.11b compliant equipments can achieve connectivity range of up to 1000 meters under suitable driving conditions.

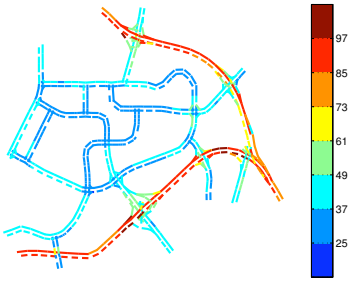


Fig. 6. Speeds (km/h) of road segments over an urban road network.

B. Traffic Flow Modeled with Relational GP

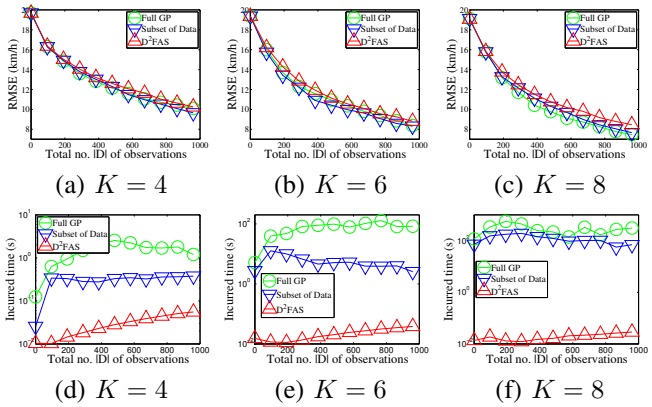
Fig. 6 shows a real-world traffic phenomenon (i.e., speeds (km/h) of road segments) over an urban road network in the 4.9 km $\times$ 4.3 km Tampines area of Singapore during evening peak hours on April 20, 2011. It comprises 775 road segments including highways, arterials, slip roads, etc. The mean speed is 48.8 km/h and the standard deviation is 20.5 km/h.

In our simulation, a network of K MoD vehicles is tasked to actively cruise the road network to gather a total data size of up to 960. To reduce computational time, each vehicle repeatedly computes and executes maximum-entropy walks of length $L = 2$ (instead of computing a very long walk), unless otherwise stated.

The performance of D^{2}FAS (i.e., GP-DDF coupled with partial DAS as shown in Algorithm 2) is compared to that of both state-of-the-art full GP (Section II) and subset of data (Section II-A) coupled with *centralized active sensing* (CAS) (28) algorithms. Note that GP-DDF$^+$ cannot be coupled with partial DAS, as explained earlier in Section IV-A. For D^{2}FAS and subset of data, $|U|$ is set to 64. For the active sensing component of D^{2}FAS (i.e., partial DAS), ε is set to 0.1, unless otherwise stated. The experiments are run on a Linux PC with Intel® Core™2 Quad CPU Q9550 at 2.83 GHz.

The first performance metric evaluates the predictive performance of a tested algorithm: It measures the *root mean squared error* (RMSE) $\sqrt{|V|^{-1}\sum_{s\in V}\left(z_s-\widehat{\mu}_s\right)^2}$ over the entire domain V of the road network that is incurred by the predictive mean $\widehat{\mu}_s$ of the tested algorithm, specifically, using (2) of full GP, (5) of subset of data, or (13) of D^{2}FAS. The second metric evaluates the time efficiency and scalability of a tested algorithm by measuring its incurred time; for D^{2}FAS, the maximum of the time incurred by all subsets $\mathcal{V}_1,\ldots,\mathcal{V}_\mathcal{K}$ of vehicles is recorded.

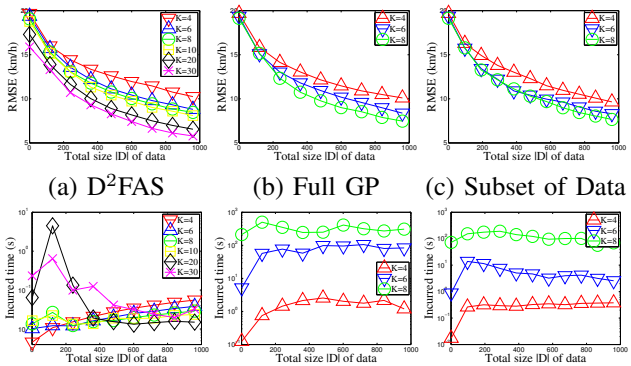
1) Predictive Performance and Time Efficiency: Fig. 7 shows results of the performance of the tested algorithms averaged over 40 randomly generated starting vehicle locations with varying number $K = 4, 6, 8$ of vehicles. It can be observed that D^{2}FAS is significantly more time-efficient and scales better with increasing size $|D|$ of data (Figs. 7d to 7f) while achieving predictive performance close to that of full GP and subset of data coupled with CAS (Figs. 7a to 7c). Specifically, D^{2}FAS is about $1, 2, 4$ orders of magnitude faster than full GP and subset of data coupled with CAS for $K = 4, 6, 8$ vehicles, respectively.



(a) $K = 4$ (b) $K = 6$ (c) $K = 8$

(d) $K = 4$ (e) $K = 6$ (f) $K = 8$

Fig. 7. Graphs of (a-c) predictive performance and (d-f) time efficiency vs. total size $|D|$ of data gathered by varying number K of vehicles.



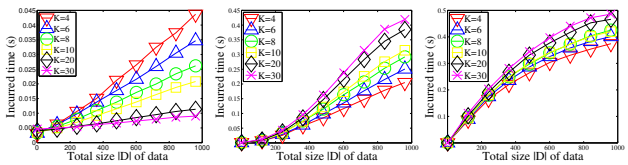
(a) D^{2}FAS (b) Full GP (c) Subset of Data

(d) D^{2}FAS (e) Full GP (f) Subset of Data

Fig. 8. Graphs of (a-c) predictive performance and (d-f) time efficiency vs. total size $|D|$ of data gathered by varying number K of vehicles.

2) Scalability of D^{2}FAS: Using the same results as that in Fig. 7, Fig. 8 plots them differently to reveal the scalability of the tested algorithms with increasing number K of vehicles. Additionally, we provide results of the performance of D^{2}FAS for $K = 10, 20, 30$ vehicles; such results are not available for full GP and subset of data coupled with CAS due to extremely long incurred time. It can be observed from Figs. 8a to 8c that the predictive performance of all tested algorithms improve with a larger number of vehicles because each vehicle needs to execute fewer number of walks and its performance is therefore less adversely affected by its myopic selection (i.e., $L = 2$) of maximum-entropy walks. As a result, more informative unobserved road segments are explored.

As shown in Fig. 8d, when the randomly placed vehicles gather their initial data (i.e., $|D| < 400$), the time incurred by D^{2}FAS is higher for greater K due to larger subsets of vehicles being formed to coordinate their walks (i.e., larger κ). As more data are gathered (i.e., $|D| \geq 400$), its partial DAS component directs the vehicles to explore further apart from each other in order to maximize the entropy of their walks. This consequently decreases κ, thus leading to a reduction in incurred time. Furthermore, as K increases from 4 to 20, the incurred time decreases due to its decentralized data fusion component (i.e., GP-DDF) that can distribute the computational load among a greater number of vehicles. When the road



(a) D^{2}FAS (b) Full GP (c) Subset of Data

Fig. 9. Graphs of time efficiency vs. total size $|D|$ of data gathered by varying number K of vehicles.

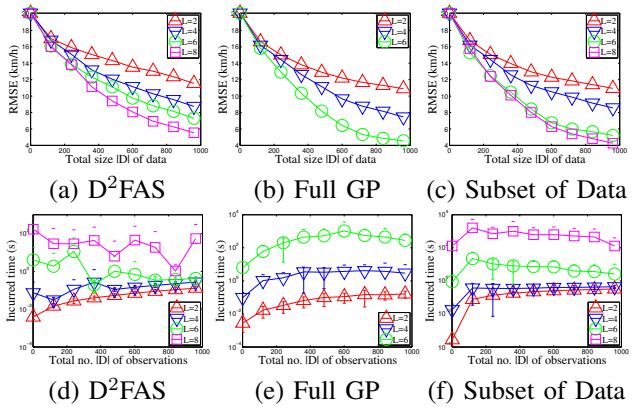
network becomes more crowded from $K = 20$ to $K = 30$ vehicles, the incurred time increases slightly due to slightly larger κ. In contrast, Figs. 8e and 8f show that the time taken by full GP and subset of data increase significantly primarily due to their centralized active sensing incurring exponential time in K. Hence, the scalability of our D^{2}FAS algorithm in the number of vehicles allows the deployment of a larger-scale vehicular sensor network (i.e., $K \geq 10$) to achieve more accurate traffic modeling and prediction (Figs. 8a to 8c).

3) Scalability of Data Fusion: Fig. 9 shows results of the scalability of the tested data fusion methods with increasing number K of vehicles. In order to produce meaningful results for fair comparison, the same active sensing component has to be coupled with the data fusion methods and its incurred time kept to a minimum. As such, we impose the use of full DAS component to be performed by each vehicle k: $w_k^* = \arg\max_{w_k} \mathbb{H}[Z_{S_{w_k}}|Z_D]$. For D^{2}FAS, this corresponds exactly to (32) by setting a large enough ε (in our experiments, $\varepsilon = 2$) to yield $\kappa = 1$; consequently, computational and communicational operations pertaining to the coordination graph can be omitted.

It can be seen from Fig. 9a that the time incurred by the decentralized data fusion component of D^{2}FAS (i.e., GP-DDF) decreases with increasing K, as explained previously. In contrast, the time incurred by full GP and subset of data increase (Fig. 9b and 9c): As discussed above, a larger number of vehicles result in a greater quantity of more informative unique data to be gathered (i.e., fewer repeated data), which increase the time needed for data fusion. When $K \geq 10$, D^{2}FAS is at least 1 order of magnitude faster than full GP and subset of data. It can also be observed that D^{2}FAS scales better with increasing size of data. So, the real-time performance and scalability of D^{2}FAS's decentralized data fusion (i.e., GP-DDF) enable it to be used data and fleet (including static sensors and passive mobile probes) are expected to be available.

4) Varying Length L of Walk: Fig. 10 shows results of the performance of the tested algorithms with varying length $L = 2, 4, 6, 8$ of maximum-entropy joint walks; we choose to experiment with just 2 vehicles since Figs. 8 and 9 reveal that a smaller number of vehicles produce poorer predictive performance and higher incurred time with large size of data for D^{2}FAS[7]. It can be observed that the predictive performance of all tested algorithms improve with increasing walk length L

[7]Furthermore, in practice, when the length L of walks is increased, the number K of vehicles has to be decreased in order to preserve time efficiency (see TABLE I), which is especially necessary for obtaining experimental results of those coupled with CAS algorithm in a reasonable amount of time.



(a) D^{2}FAS (b) Full GP (c) Subset of Data

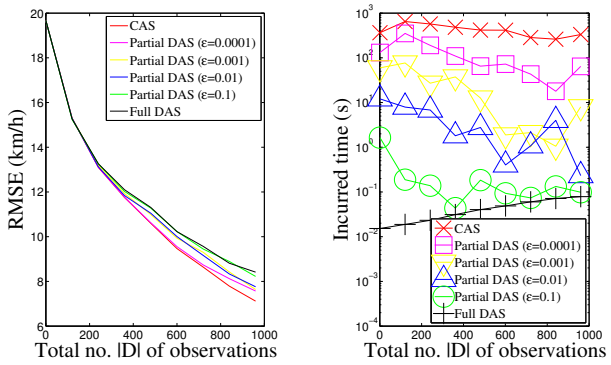
(d) D^{2}FAS (e) Full GP (f) Subset of Data

Fig. 10. Graphs of (a-c) predictive performance and (d-f) time efficiency vs. total size $|D|$ of data gathered by 2 vehicles with varying length L of maximum-entropy joint walks. (When the lower bound value is negative, the lower error bar cannot be shown in log-scale.)

because the selection of maximum-entropy joint walks is less myopic. The time incurred by D^{2}FAS increases due to larger κ but grows more slowly and is lower than that incurred by full GP and subset of data coupled with CAS. Specifically, when $L = 8$, D^{2}FAS is at least 1 order of magnitude faster (i.e., average of 60 s) than subset of data (i.e., average of > 732 s) and full GP (i.e., not available due to excessive incurred time) coupled with CAS. Also, notice from Figs. 8a and 8d that if a large number of vehicles (i.e., $K = 30$) is available, D^{2}FAS can select shorter walks of $L = 2$ to be significantly more time-efficient (i.e., average of > 3 orders of magnitude faster) while achieving predictive performance comparable to that of subset of data with $L = 8$ and full GP with $L = 6$.

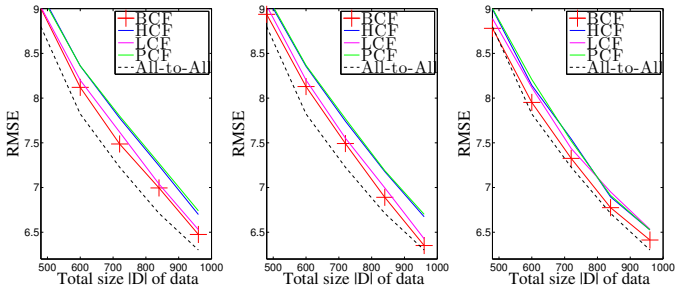
Fig. 10d shows fluctuations of time (e.g., when $L = 8$ and $|D| > 720$) incurred by D^{2}FAS algorithm due to variations in the number of possible walks that need to be evaluated in order to select the maximum-entropy walks at every time step. These variations are caused by a dynamically changing coordination graph $\mathcal{G}$ being formed with the size κ of its largest connected component varying over time as well as different out-degrees (i.e., ranging from 2 to 6) of the segments in the road network G. The different out-degrees of the road segments can also be used to explain the small fluctuations of time incurred by those coupled with CAS algorithm in Figs. 10e and 10f.

5) Varying Degrees of Coordination in Partial DAS: Fig. 11 shows the predictive performance and time efficiency of D^{2}FAS with varying degrees of coordination between $K = 4$ vehicles with walk length $L = 4$ to gather a size of up to $|D| = 960$ of data. The degree of coordination in the active sensing component of D^{2}FAS (i.e., partial DAS) is varied by setting the minimum required correlation ε (33) between walks of adjacent vehicles to different values: $0, 0.1, 0.01, 0.001, 0.0001$ and 2. As explained previously in Section IV-A, when $\varepsilon = 0$, partial DAS becomes *centralized active sensing* (CAS). On the other hand, when ε is set to be sufficiently large (in this case, $\varepsilon = 2$), partial DAS reduces to full DAS. From Fig. 11a, it can be observed that though the predictive performance of partial DAS decreases (i.e., RMSE increases) with an increasing ε, it can achieve a performance comparable to that of CAS by setting a small enough ε, thus corroborating the effect



(a) Predictive Performance (b) Time Efficiency

Fig. 11. Graphs of (a) predictive performance and (b) time efficiency of D^{2}FAS vs. total size $|D|$ of data gathered by 4 vehicles with varying minimum required correlation ε between walks of adjacent vehicles in partial DAS.



(a) R=300 (b) R=500 (c) R=1000

Fig. 12. Predictive performance of different consensus filtering-based GP-DDF (4 iterations per time step) and all-to-all communication-based GP-DDF coupled with full DAS algorithms with varying communication range R (meters).

of ε on the performance guarantee in Theorem 3. Fig. 11b shows that the time efficiency of partial DAS can be improved significantly by increasing ε due to a decreasing κ (i.e., largest subset of vehicles being formed to coordinate their walks). So, in contrast to CAS and full DAS, partial CAS can provide the flexibility in trading off between predictive performance vs. time efficiency by setting a suitable ε.

6) Predictive Performance of Consensus Filtering-Based GP-DDF: Fig. 5 compares results of the predictive performance between the all-to-all communication-based GP-DDF$^+$ and different consensus filtering-based GP-DDF$^+$ (Section III-A) coupled with full DAS (Section IV) algorithms. We vary the vehicles' communication range $R = 300, 500, 1000$ meters. All results are obtained by averaging over 40 random instances.

Fig. 12 shows that the predictive performance of the consensus filtering-based GP-DDF coupled with full DAS algorithms approach that of all-to-all communication-based GP-DDF coupled with full DAS algorithm with an increasing communication range R. This is expected since a better connectivity of the communication network allows consensus filters to better approximate the global summary. The results also show that LCF tends to outperform HCF and PCF. This is because LCF performs better with slowly-varying local summaries that result from gradually increasing local data. Similar to Fig. 5, the performance of BCF is stable and slightly better than that of LCF.

VII. Related Works

A. Models for Predicting Spatiotemporally Varying Traffic Flow

The spatiotemporal correlation structure of traffic flow can be exploited to predict the traffic flow condition of any unobserved road segment at any time using the data taken along the sensors' paths. To achieve this, existing Bayesian filtering frameworks [16], [17], [40] utilize various handcrafted parametric models predicting traffic flow along a highway stretch that only correlate adjacent segments of the highway. So, their predictive performance will be compromised when the current data are sparse and/or the actual spatial correlation spans multiple segments. Their strong Markov assumption further exacerbates this problem. It is also not shown how these models can be generalized to work for arbitrary road network topologies and more complex correlation structure. Existing multivariate parametric traffic prediction models [41], [42] do not quantify uncertainty estimates of the predictions and impose rigid spatial locality assumptions that do not adapt to the true underlying correlation structure.

In contrast, we assume the traffic flow over an urban road network (i.e., comprising the full range of road types like highways, arterials, slip roads) to be realized from a relational GP (Section II-B) that can formally characterize its spatiotemporal correlation structure and be refined with growing size of data. More importantly, GP can provide formal measures of predictive uncertainty (e.g., based on variance or entropy criterion) for directing the sensors to explore highly uncertain areas of the road network. The work of [15] used GP to represent the traffic flow over a network of only highways and defined the correlation of speeds between highway segments to depend only on the geodesic (i.e., shortest path) distance of these segments with respect to the network topology; their features are not considered. The work of [43] maintained a mixture of two independent GPs for flow prediction such that the correlation structure of one GP utilized road segment features while that of the other GP depended on manually specified relations (instead of geodesic distance) between segments with respect to an undirected network topology. Different from the above works, we propose a relational GP whose correlation structure exploits the geodesic distance between segments based on the topology of a directed road network with vertices denoting road segments and edges indicating adjacent segments weighted by dissimilarity of their features, hence tightly integrating the features and relational information.

B. Data Fusion

The data are gathered distributedly by each vehicle along its path in the road network and have to be assimilated in order to predict the traffic phenomenon. Since large data are expected to be collected, a centralized approach to GP prediction cannot be performed in real time due to its cubic time complexity.

To resolve this, we propose decentralized data fusion approaches to efficient and scalable approximate GP prediction (Section III). Existing decentralized and distributed Bayesian

filtering frameworks for addressing non-traffic related problems [34], [44]–[47] will face the same difficulties as their centralized counterparts described above if applied to predicting traffic phenomena, thus resulting in loss of predictive performance. Distributed regression algorithms [48], [49] for static sensor networks gain efficiency from spatial locality assumptions, which cannot be exploited by vehicles whose paths are not constrained by locality. The work of [50] proposed a distributed data fusion approach to approximate GP prediction based on an iterative Jacobi overrelaxation algorithm, which incurs some critical limitations: (a) The past data taken along the vehicles' paths are assumed to be uncorrelated, which greatly undermines its predictive performance when they are in fact correlated and/or the current data are sparse, (b) when the number of vehicles grows large, it converges very slowly, and (c) it assumes that the range of positive correlation has to be bounded by some factor of the communication range. Our proposed decentralized data fusion algorithms do not suffer from these limitations and can be computed exactly with efficient time bounds.

C. Active Sensing

The vehicles have to coordinate to actively gather the most informative data for minimizing the uncertainty of modeling and predicting the traffic phenomenon. Existing centralized [21], [22] and decentralized [51] active sensing algorithms scale poorly with large data and fleet. We propose decentralized active sensing algorithms that overcome these issues of scalability (Section IV).

VIII. Conclusion

This paper describes novel Gaussian process decentralized data fusion (GP-DDF and GP-DDF$^+$) and active sensing (full DAS and partial DAS) algorithms for real-time, fine-grained traffic sensing, modeling, and prediction with a fleet of autonomous robotic vehicles in a MoD system. Analytical and empirical results have demonstrated that (a) GP-DDF and GP-DDF$^+$ are significantly more time-efficient and scalable in the size of the data while achieving predictive performance close to that of state-of-the-art full GP [21], [22] and subset of data, (b) GP-DDF$^+$ can achieve a better balance between predictive performance and time efficiency than GP-DDF and full GP, (c) consensus filtering-based GP-DDF and GP-DDF$^+$ requiring only local communication between neighboring MoD vehicles can achieve predictive performance comparable to that of all-to-all communication-based GP-DDF and GP-DDF$^+$, respectively, (d) GP-DDF coupled with partial DAS is significantly more time-efficient and scalable in the size of the data and fleet while achieving predictive performance close to that of full GP and subset of data coupled with centralized active sensing, and (e) when full DAS is used to gather the most informative demand data for predicting a mobility demand pattern, it can achieve a dual effect of fleet rebalancing to service the mobility demands. Hence, our proposed algorithms are practical for deployment in a large-scale vehicular sensor network to achieve persistent and accurate traffic sensing, modeling, and prediction. Interestingly, GP-DDF and GP-DDF$^+$ can be adapted to parallel implementations to be run on a cluster of machines for achieving efficient and scalable probabilistic prediction with large data [52].

A limitation of our proposed algorithms is that the decentralized data fusion components assume independence between multiple traffic phenomena (e.g., mobility demand pattern and traffic flow) while the decentralized active sensing components only work for a single traffic phenomenon. So, in our future work, we plan to generalize our algorithms to perform active sensing of multiple traffic phenomena and remove the assumption of independence between traffic phenomena by exploiting techniques like multi-output GPs and co-kriging for modeling their correlation. We also like to generalize our partial DAS algorithm to be coupled with GP-DDF$^+$ and to perform active sensing of a traffic phenomenon being modeled by a ℓGP. Since the data gathered by each MoD vehicle streams in over time, we want to develop online learning variants of GP-DDF and GP-DDF$^+$ based on the online sparse GP model of [53]. Lastly, as mentioned in Section II, the hyperparameters of the relational GP and ℓGP are learned using the data by maximizing the log marginal likelihood. The sparse approximation method employed by PITC to improve the scalability of full GP model can be similarly applied to computing such a log marginal likelihood scalably, as explained in [32] (i.e., equation 30 in Section 9). Since GP-DDF is the decentralized version of PITC (Theorem 1B), the log marginal likelihood can be computed and maximized in a decentralized manner as well, albeit mathematically more tedious. This will be investigated in our future work.

Acknowledgment

The authors would like to thank Ali Oran and Max Timmons for our discussions on consensus filters. Many thanks to John Dolan,Gaurav Sukhatme and Colin Keng-Yan Tan for your valuable comments. This work was supported by Singapore-MIT Alliance Research and Technology Subaward Agreements No. 41 and 52.

References

[1] J. Chen, K. H. Low, C. K.-Y. Tan, A. Oran, P. Jaillet, J. M. Dolan, and G. S. Sukhatme, "Decentralized data fusion and active sensing with mobile sensors for modeling and predicting spatiotemporal traffic phenomena," in *Proc. 28th Conference on Uncertainty in Artificial Intelligence*, 2012, pp. 163–173.

[2] J. Chen, K. H. Low, and C. K.-Y. Tan, "Gaussian process-based decentralized data fusion and active sensing for mobility-on-demand system," in *Proc. Robotics: Science and Systems Conference*, 2013.

[3] "*Hong Kong in Figures*. Census and Statistics Department, Hong Kong Special Administrative Region (http://www.censtatd.gov.hk); *Singapore Land Transport: Statistics in Brief*. Land Transport Authority of Singapore (http://www.lta.gov.sg)," 2012.

[4] W. J. Mitchell, C. E. Borroni-Bird, and L. D. Burns, *Reinventing the Automobile: Personal Urban Mobility for the 21st Century*. Cambridge, MA: MIT Press, 2010.

[5] M. Pavone, S. L. Smith, E. Frazzoli, and D. Rus, "Robotic load balancing for mobility-on-demand systems," *Int. J. Robot. Res.*, vol. 31, no. 7, pp. 839–854, 2012.

[6] "*GM Shows Chevrolet EN-V 2.0 Mobility Concept Vehicle*. General Motors Co. (http://media.gm.com/media/us/en/gm/news.detail.html/content/Pages/news/us/en/2012/Apr/0423_EN-V_2_Rendering.html)," 2012.

[7] W. J. Mitchell, "Mobility on demand: Future of transportation in cities," MIT Media Laboratory, Cambridge, MA, Technical Report, 2008.

[8] S. Peeta and A. K. Ziliaskopoulos, "Foundations of dynamic traffic assignment: The past, the present and the future," *Networks and Spatial Economics*, vol. 1, pp. 233–265, 2001.

[9] G. Berbeglia, J.-F. Cordeau, and G. Laporte, "Dynamic pickup and delivery problems," *Eur. J. Oper. Res.*, vol. 202, no. 1, pp. 8–15, 2010.

[10] L. Agussurja and H. C. Lau, "Toward large-scale agent guidance in an urban taxi service," in *Proc. UAI*, 2012, pp. 36–43.

[11] H. Chang, Y. Tai, and J. Y. Hsu, "Context-aware taxi demand hotspots prediction," *Int. J. Business Intelligence and Data Mining*, vol. 5, no. 1, pp. 3–18, 2010.

[12] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani, "An energy-efficient mobile recommender system," in *Proc. KDD*, 2010, pp. 899–908.

[13] X. Li, G. Pan, Z. Wu, G. Qi, S. Li, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Front. Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[14] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-Finder: A recommender system for finding passengers and vacant taxis," *IEEE Trans. Knowl. Data Eng.*, 2012.

[15] A. Krause, E. Horvitz, A. Kansal, and F. Zhao, "Toward community sensing," in *Proc. IPSN*, 2008, pp. 481–492.

[16] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transport. Res. B-Meth.*, vol. 39, no. 2, pp. 141–167, 2005.

[17] D. B. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. Bayen, "A traffic model for velocity data assimilation," *AMRX*, vol. 2010, no. 1, pp. 1–35, 2010.

[18] K. K. Srinivasan and P. P. Jovanis, "Determination of number of probe vehicle required for reliable travel time measurement in urban network," *Transport. Res. Rec.*, vol. 1537, pp. 15–22, 1996.

[19] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdener, "Travel time data collection handbook," Federal Highway Administration, Office of Highway Information Management, Washington, DC, Technical Report FHWA-PL-98-035, 1998.

[20] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, pp. 235–284, 2008.

[21] K. H. Low, J. M. Dolan, and P. Khosla, "Adaptive multi-robot wide-area exploration and mapping," in *Proc. AAMAS*, 2008, pp. 23–30.

[22] ——, "Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing," in *Proc. ICAPS*, 2009, pp. 233–240.

[23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[24] N. Lawrence, M. Seeger, and R. Herbrich, "Fast sparse Gaussian process methods: The informative vector machine," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003, pp. 609–616.

[25] M. Seeger and C. Williams, "Fast forward selection to speed up sparse Gaussian process regression," in *Proc. AISTATS*, C. M. Bishop and B. J. Frey, Eds., 2003.

[26] S. Oh, Y. Xu, and J. Choi, "Explorative navigation of mobile sensor networks using sparse Gaussian processes," in *Proc. CDC*, 2010, pp. 3851–3856.

[27] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. NY: Springer, 2005.

[28] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, 1st ed. Cambridge, MA: MIT Press, 2002.

[29] M. Jelasity, A. Montresor, and O. Babaoglu, "Gossip-based aggregation in large dynamic networks," *ACM Trans. Comput. Syst.*, vol. 23, no. 3, pp. 219–252, 2005.

[30] R. Webster and M. Oliver, *Geostatistics for Environmental Scientists*. NY: John Wiley & Sons, Inc., 2007.

[31] M. E. Hohn, *Geostatistics and Petroleum Geology*, 2nd ed. Springer, 1998.

[32] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *JMLR*, vol. 6, pp. 1939–1959, 2005.

[33] E. Snelson, "Local and global sparse Gaussian process approximations," in *Proc. AISTATS*, 2007.

[34] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *Proc. CDC*, 2005, pp. 8179–8184.

[35] ——, "Distributed tracking for mobile sensor networks with information-driven mobility," in *Proc. ACC*, 2007, pp. 4606–4612.

[36] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, "Dynamic consensus for mobile networks," in *Proc. 16th IFAC World Congress*, 2005.

[37] R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. CDC*, 2005, pp. 6698–6703.

[38] R. A. Freeman, P. Yang, and K. M. Lynch, "Stability and convergence properties of dynamic average consensus estimators," in *Proc. CDC*, 2006, pp. 338–343.

[39] J. P. Singh, N. Bambos, B. Srinivasan, and D. Clawin, "Wireless LAN performance under varied stress conditions in vehicular traffic scenarios," in *Proc. Vehicular Technology Conference*, 2002, pp. 743–747.

[40] H. Chen, H. A. Rakha, and S. Sadek, "Real-time freeway traffic state prediction: A particle filter approach," in *Proc. IEEE ITSC*, 2011, pp. 626–631.

[41] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transport. Res. Rec.*, vol. 1857, pp. 74–84, 2003.

[42] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transport. Res. C-Emer.*, vol. 19, no. 4, pp. 606–616, 2011.

[43] M. Neumann, K. Kersting, Z. Xu, and D. Schulz, "Stacked Gaussian process learning," in *Proc. ICDM*, 2009, pp. 387–396.

[44] T. H. Chung, V. Gupta, J. W. Burdick, and R. M. Murray, "On a decentralized active sensing strategy using mobile sensor platforms in a network," in *Proc. CDC*, 2004, pp. 1914–1919.

[45] M. Coates, "Distributed particle filters for sensor networks," in *Proc. IPSN*, 2004, pp. 99–107.

[46] M. Rosencrantz, G. Gordon, and S. Thrun, "Decentralized sensor fusion with distributed particle filters," in *Proc. UAI*, 2003, pp. 493–500.

[47] S. Sukkarieh, E. Nettleton, J. Kim, M. Ridley, A. Goktogan, and H. Durrant-Whyte, "The ANSER project: Data fusion across multiple uninhabited air vehicles," *IJRR*, vol. 22, no. 7-8, pp. 505–539, 2003.

[48] C. Guestrin, P. Bodik, R. Thibaus, M. Paskin, and S. Madden, "Distributed regression: an efficient framework for modeling sensor network data," in *Proc. IPSN*, 2004, pp. 1–10.

[49] M. A. Paskin and C. Guestrin, "Robust probabilistic inference in distributed systems," in *Proc. UAI*, 2004, pp. 436–445.

[50] J. Cortes, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Trans. Automat. Contr.*, vol. 54, no. 12, pp. 2816–2827, 2009.

[51] R. Stranders, A. Farinelli, A. Rogers, and N. R. Jennings, "Decentralised coordination of mobile sensors using the max-sum algorithm," in *Proc. IJCAI*, 2009, pp. 299–304.

[52] J. Chen, N. Cao, K. H. Low, R. Ouyang, C. K.-Y. Tan, and P. Jaillet, "Parallel Gaussian process regression with low-rank covariance matrix approximations," in *Proc. UAI*, 2013, pp. 152–161.

[53] N. Xu, K. H. Low, J. Chen, K. K. Lim, and E. B. Özgül, "GP-Localize: Persistent mobile robot localization using online sparse Gaussian process observation model," in *Proc. AAAI*, 2014, pp. 2585–2592.

[54] G. H. Golub and C.-F. Van Loan, *Matrix Computations*, 3rd ed. Johns Hopkins Univ. Press, 1996.

[55] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*. Academic Press, 1990.

[56] I. C. F. Ipsen and D. J. Lee, "Determinant approximations," Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC, Technical Report CRSC-TR03-30, 2003.

APPENDIX A
PROOF OF THEOREM 1B

We have to first simplify the $\Gamma_{SD}\left(\Gamma_{DD}+\Lambda\right)^{-1}$ term in the expressions of $\mu^{\text{PITC}}_{S|D}$ (15) and $\Sigma^{\text{PITC}}_{SS|D}$ (16).

$$\begin{aligned}
&\left(\Gamma_{DD}+\Lambda\right)^{-1}\\
&=\left(\Sigma_{DU}\Sigma^{-1}_{UU}\Sigma_{UD}+\Lambda\right)^{-1}\\
&=\Lambda^{-1}-\Lambda^{-1}\Sigma_{DU}\left(\Sigma_{UU}+\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}\right)^{-1}\Sigma_{UD}\Lambda^{-1}\\
&=\Lambda^{-1}-\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\ .
\end{aligned} \tag{40}$$

The second equality follows from matrix inversion lemma. The last equality is due to

$$\begin{aligned}
&\Sigma_{UU}+\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}\\
&=\Sigma_{UU}+\sum_{k=1}^{K}\Sigma_{UD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_kU}\\
&=\Sigma_{UU}+\sum_{k=1}^{K}\dot{\Sigma}^{k}_{UU}=\ddot{\Sigma}_{UU}\ .
\end{aligned} \tag{41}$$

Using (17) and (40),

$$\begin{aligned}
&\Gamma_{SD}\left(\Gamma_{DD}+\Lambda\right)^{-1}\\
&=\Sigma_{SU}\Sigma^{-1}_{UU}\Sigma_{UD}\left(\Lambda^{-1}-\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\right)\\
&=\Sigma_{SU}\Sigma^{-1}_{UU}\left(\ddot{\Sigma}_{UU}-\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}\right)\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\\
&=\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}
\end{aligned} \tag{42}$$

The third equality is due to (41).

From (15),

$$\begin{aligned}
\mu^{\text{PITC}}_{S|D}&=\mu_S+\Gamma_{SD}\left(\Gamma_{DD}+\Lambda\right)^{-1}\left(z_D-\mu_D\right)\\
&=\mu_S+\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\left(z_D-\mu_D\right)\\
&=\mu_S+\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U\\
&=\overline{\mu}_S\ .
\end{aligned}$$

The second equality is due to (42). The third equality follows from $\Sigma_{UD}\Lambda^{-1}\left(z_D-\mu_D\right)=\sum_{k=1}^{K}\Sigma_{UD_k}\Sigma^{-1}_{D_kD_k|U}\left(z_{D_k}-\mu_{D_k}\right)=\sum_{k=1}^{K}\dot{z}^{k}_{U}=\ddot{z}_U$.

From (16),

$$\begin{aligned}
&\Sigma^{\text{PITC}}_{SS|D}\\
&=\Sigma_{SS}-\Gamma_{SD}\left(\Gamma_{DD}+\Lambda\right)^{-1}\Gamma_{DS}\\
&=\Sigma_{SS}-\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}\Sigma^{-1}_{UU}\Sigma_{US}\\
&=\Sigma_{SS}-\Big(\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}\Sigma^{-1}_{UU}\Sigma_{US}\\
&\quad-\Sigma_{SU}\Sigma^{-1}_{UU}\Sigma_{US}\Big)-\Sigma_{SU}\Sigma^{-1}_{UU}\Sigma_{US}\\
&=\Sigma_{SS}-\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\left(\Sigma_{UD}\Lambda^{-1}\Sigma_{DU}-\ddot{\Sigma}_{UU}\right)\Sigma^{-1}_{UU}\Sigma_{US}\\
&\quad-\Sigma_{SU}\Sigma^{-1}_{UU}\Sigma_{US}\\
&=\Sigma_{SS}-\left(\Sigma_{SU}\Sigma^{-1}_{UU}\Sigma_{US}-\Sigma_{SU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{US}\right)\\
&=\Sigma_{SS}-\Sigma_{SU}\left(\Sigma^{-1}_{UU}-\ddot{\Sigma}^{-1}_{UU}\right)\Sigma_{US}\\
&=\overline{\Sigma}_{SS}\ .
\end{aligned}$$

The second equality follows from (17) and (42). The fifth equality is due to (41).

APPENDIX B
PROOF OF THEOREM 2B

We will first derive the expressions of four components useful for completing the proof later.

$$\begin{aligned}
&\widetilde{\Gamma}_{sD}\Lambda^{-1}(z_D-\mu_D)\\
&=\sum_{i\neq k}\Gamma_{sD_i}\Sigma^{-1}_{D_iD_i|U}(z_{D_i}-\mu_{D_i})+\Sigma_{sD_k}\Sigma^{-1}_{D_kD_k|U}(z_{D_k}-\mu_{D_k})\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\sum_{i\neq k}\left(\Sigma_{UD_i}\Sigma^{-1}_{D_iD_i|U}(z_{D_i}-\mu_{D_i})\right)+\dot{z}^{k}_{s}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\sum_{i\neq k}\dot{z}^{i}_{U}+\dot{z}^{k}_{s}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}(\ddot{z}_U-\dot{z}^{k}_{U})+\dot{z}^{k}_{s}\ .
\end{aligned} \tag{43}$$

The first two equalities expand the first component using the definition of Λ (Theorem 2B), (9), (25), (26), and (27). The last two equalities exploit (9) and (11).

$$\begin{aligned}
&\widetilde{\Gamma}_{sD}\Lambda^{-1}\Sigma_{DU}\\
&=\sum_{i\neq k}\Gamma_{sD_i}\Sigma^{-1}_{D_iD_i|U}\Sigma_{D_iU}+\Sigma_{sD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_kU}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\sum_{i\neq k}\left(\Sigma_{UD_i}\Sigma^{-1}_{D_iD_i|U}\Sigma_{D_iU}\right)+\Sigma_{sD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_kU}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\sum_{i\neq k}\dot{\Sigma}^{i}_{UU}+\dot{\Sigma}^{k}_{sU}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\left(\ddot{\Sigma}_{UU}-\dot{\Sigma}^{k}_{UU}-\Sigma_{UU}\right)+\dot{\Sigma}^{k}_{sU}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\ddot{\Sigma}_{UU}-\gamma^{k}_{sU}\ .
\end{aligned} \tag{44}$$

The first two equalities expand the second component by the same trick as that in (43). The third and fourth equalities exploit (10) and (12), respectively. The last equality is due to (20).

Let $\alpha_{sU}\triangleq\Sigma_{sU}\Sigma^{-1}_{UU}$ and its transpose is α_{Us}. By using similar tricks in (43) and (44), we can derive the expressions of the remaining two components.

If $\tau_s=\tau_{s'}=k$, then

$$\begin{aligned}
&\widetilde{\Gamma}_{sD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'}\\
&=\sum_{i\neq k}\Gamma_{sD_i}\Sigma^{-1}_{D_iD_i|U}\Gamma_{D_is'}+\Sigma_{sD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_ks'}\\
&=\Sigma_{sU}\Sigma^{-1}_{UU}\sum_{i\neq k}\left(\Sigma_{UD_i}\Sigma^{-1}_{D_iD_i|U}\Sigma_{D_iU}\right)\Sigma^{-1}_{UU}\Sigma_{Us'}+\dot{\Sigma}^{k}_{ss'}\\
&=\alpha_{sU}\sum_{i\neq k}\left(\dot{\Sigma}^{i}_{UU}\right)\alpha_{Us'}+\dot{\Sigma}^{k}_{ss'}\\
&=\alpha_{sU}\left(\ddot{\Sigma}_{UU}-\dot{\Sigma}^{k}_{UU}-\Sigma_{UU}\right)\alpha_{Us'}+\dot{\Sigma}^{k}_{ss'}\\
&=\alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'}-\alpha_{sU}\gamma^{k}_{Us'}-\alpha_{sU}\dot{\Sigma}^{k}_{Us'}+\dot{\Sigma}^{k}_{ss'}\ .
\end{aligned} \tag{45}$$

If $\tau_s = i$ and $\tau_{s'} = j$ such that $i \neq j$, then

$$\begin{aligned}
&\widetilde{\Gamma}_{sD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'}\\
&= \sum_{k\neq i,j} \Gamma_{sD_k}\Sigma^{-1}_{D_kD_k|U}\Gamma_{D_ks'}\\
&\quad + \Sigma_{sD_i}\Sigma^{-1}_{D_iD_i|U}\Gamma_{D_is'} + \Gamma_{sD_j}\Sigma^{-1}_{D_jD_j|U}\Sigma_{D_js'}\\
&= \Sigma_{sU}\Sigma^{-1}_{UU}\sum_{k\neq i,j}\left(\Sigma_{UD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_kU}\right)\Sigma^{-1}_{UU}\Sigma_{Us'}\\
&\quad + \Sigma_{sD_i}\Sigma^{-1}_{D_iD_i|U}\Sigma_{D_iU}\Sigma^{-1}_{UU}\Sigma_{Us'} + \Sigma_{sU}\Sigma^{-1}_{UU}\Sigma_{UD_j}\Sigma^{-1}_{D_jD_j|U}\Sigma_{D_js'}\\
&= \alpha_{sU}\left(\ddot{\Sigma}_{UU} - \dot{\Sigma}^i_{UU} - \dot{\Sigma}^j_{UU} - \Sigma_{UU}\right)\alpha_{Us'} + \dot{\Sigma}^i_{sU}\alpha_{Us'} + \alpha_{sU}\dot{\Sigma}^j_{Us'}\\
&= \alpha_{sU}\left(\ddot{\Sigma}_{UU} + \Sigma_{UU}\right)\alpha_{Us'} - \alpha_{sU}\left(\dot{\Sigma}^i_{UU} + \Sigma_{UU}\right)\alpha_{Us'}\\
&\quad - \alpha_{sU}\left(\dot{\Sigma}^j_{UU} + \Sigma_{UU}\right)\alpha_{Us'} + \dot{\Sigma}^i_{sU}\alpha_{Us'} + \alpha_{sU}\dot{\Sigma}^j_{Us'}\\
&= \alpha_{sU}\left(\ddot{\Sigma}_{UU} + \Sigma_{UU}\right)\alpha_{Us'} - \left(\alpha_{sU}\dot{\Sigma}^i_{UU} + \alpha_{sU}\Sigma_{UU} - \dot{\Sigma}^i_{sU}\right)\alpha_{Us'}\\
&\quad - \alpha_{sU}\left(\dot{\Sigma}^j_{UU}\alpha_{Us'} + \Sigma_{UU}\alpha_{Us'} - \dot{\Sigma}^j_{Us'}\right)\\
&= \alpha_{sU}\left(\ddot{\Sigma}_{UU} + \Sigma_{UU}\right)\alpha_{Us'} - \gamma^i_{sU}\alpha_{Us'} - \alpha_{sU}\gamma^j_{Us'}\\
&= \alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} + \Sigma_{sU}\Sigma^{-1}_{UU}\Sigma_{Us'} - \gamma^i_{sU}\alpha_{Us'} - \alpha_{sU}\gamma^j_{Us'}
\end{aligned} \tag{46}$$

If $\tau_s = k$, then

$$\begin{aligned}
&\mu^{\text{PIC}}_{s|D}\\
&= \mu_s + \widetilde{\Gamma}_{sD}\left(\Gamma_{DD} + \Lambda\right)^{-1}\left(z_D - \mu_D\right)\\
&= \mu_s + \widetilde{\Gamma}_{sD}\Lambda^{-1}(z_D - \mu_D)\\
&\quad - \widetilde{\Gamma}_{sD}\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}(z_D - \mu_D)\\
&= \mu_s + \widetilde{\Gamma}_{sD}\Lambda^{-1}(z_D - \mu_D) - \widetilde{\Gamma}_{sD}\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U\\
&= \mu_s + \Sigma_{sU}\Sigma^{-1}_{UU}(\ddot{z}_U - \dot{z}^k_U) + \dot{z}^k_s - \widetilde{\Gamma}_{sD}\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U\\
&= \mu_s + \Sigma_{sU}\Sigma^{-1}_{UU}(\ddot{z}_U - \dot{z}^k_U) + \dot{z}^k_s\\
&\quad - \left(\Sigma_{sU}\Sigma^{-1}_{UU}\ddot{\Sigma}_{UU} - \gamma^k_{sU}\right)\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U\\
&= \mu_s + \left(\gamma^k_{sU}\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U - \Sigma_{sU}\Sigma^{-1}_{UU}\dot{z}^k_U\right) + \dot{z}^k_s\\
&= \overline{\mu}^k_s \ .
\end{aligned}$$

The first equality is by definition (23). The second equality is due to (40). The third equality is due to the definition of global summary (11). The fourth and fifth equalities are due to (43) and (44), respectively.

If $\tau_s = \tau_{s'} = k$, then

$$\begin{aligned}
&\sigma^{\text{PIC}}_{ss'|D}\\
&= \sigma_{ss'} - \widetilde{\Gamma}_{sD}\left(\Gamma_{DD} + \Lambda\right)^{-1}\widetilde{\Gamma}_{Ds'}\\
&= \sigma_{ss'} - \widetilde{\Gamma}_{sD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'} + \widetilde{\Gamma}_{sD}\Lambda^{-1}\Sigma_{DU}\ddot{\Sigma}^{-1}_{UU}\Sigma_{UD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'}\\
&= \sigma_{ss'} - \widetilde{\Gamma}_{sD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'}\\
&\quad + \left(\alpha_{sU}\ddot{\Sigma}_{UU} - \gamma^k_{sU}\right)\ddot{\Sigma}^{-1}_{UU}\left(\ddot{\Sigma}_{UU}\alpha_{Us'} - \gamma^k_{Us'}\right)\\
&= \sigma_{ss'} - \alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} + \alpha_{sU}\gamma^k_{Us'} + \alpha_{sU}\dot{\Sigma}^k_{Us'} - \dot{\Sigma}^k_{ss'}\\
&\quad + \alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} - \alpha_{sU}\gamma^k_{Us'} - \gamma^k_{sU}\alpha_{Us'} + \gamma^k_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^k_{Us'}\\
&= \sigma_{ss'} - \left(\gamma^k_{sU}\alpha_{Us'} - \alpha_{sU}\dot{\Sigma}^k_{Us'} - \gamma^k_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^k_{Us'}\right) - \dot{\Sigma}^k_{ss'}\\
&= \overline{\sigma}^k_{ss'}\\
&= \overline{\sigma}_{ss'} \ .
\end{aligned}$$

The first equality is by definition (24). The second equality is due to (40). The third equality is due to (44). The fourth equality is due to (45). The last two equalities are by definition (22).

If $\tau_s = i$ and $\tau_{s'} = j$ such that $i \neq j$, then

$$\begin{aligned}
&\sigma^{\text{PIC}}_{ss'|D}\\
&= \sigma_{ss'} - \widetilde{\Gamma}_{sD}\Lambda^{-1}\widetilde{\Gamma}_{Ds'}\\
&\quad + \alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} - \alpha_{sU}\gamma^j_{Us'} - \gamma^i_{sU}\alpha_{Us'} + \gamma^i_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^j_{Us'}\\
&= \sigma_{ss'} - \left(\alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} + \Sigma_{sU}\Sigma^{-1}_{UU}\Sigma_{Us'} - \gamma^i_{sU}\alpha_{Us'} - \alpha_{sU}\gamma^j_{Us'}\right)\\
&\quad + \alpha_{sU}\ddot{\Sigma}_{UU}\alpha_{Us'} - \alpha_{sU}\gamma^j_{Us'} - \gamma^i_{sU}\alpha_{Us'} + \gamma^i_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^j_{Us'}\\
&= \sigma_{ss'} - \Sigma_{sU}\Sigma^{-1}_{UU}\Sigma_{Us'} + \gamma^i_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^j_{Us'}\\
&= \overline{\sigma}_{ss'} \ .
\end{aligned}$$

The first equality is obtained using a similar trick as when $\tau_s = \tau_{s'} = k$. The second equality is due to (46). The second last equality is by the definition of posterior covariance in GP model (3). The last equality is by definition (22).

APPENDIX C
PROOF OF THEOREM 3

Let $\widetilde{\Sigma}_{S_wS_w} \triangleq \overline{\Sigma}_{S_wS_w} - \widehat{\Sigma}_{S_wS_w}$ and ρ_w be the spectral radius of $\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}$. We have to first bound ρ_w from above.

For any joint walk w, $\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}$ comprises diagonal blocks of size $\left|S_{w_{\mathcal{V}_n}}\right| \times \left|S_{w_{\mathcal{V}_n}}\right|$ with components of value 0 for $n = 1, \ldots, \mathcal{K}$ and off-diagonal blocks of the form $\left(\overline{\Sigma}_{S_{w_{\mathcal{V}_n}}S_{w_{\mathcal{V}_n}}}\right)^{-1}\overline{\Sigma}_{S_{w_{\mathcal{V}_n}}S_{w_{\mathcal{V}_{n'}}}}$ for $n, n' = 1, \ldots, \mathcal{K}$ and $n \neq n'$. We know that any pair of vehicles $k \in \mathcal{V}_n$ and $k' \in \mathcal{V}_{n'}$ reside in different connected components of coordination graph $\mathcal{G}$ and are therefore not adjacent. So, by Definition 7,

$$\max_{i,i'}\left|\left[\overline{\Sigma}_{S_{w_{\mathcal{V}_n}}S_{w_{\mathcal{V}_{n'}}}}\right]_{ii'}\right| \leq \varepsilon \tag{47}$$

for $n, n' = 1, \ldots, \mathcal{K}$ and $n \neq n'$. Using (39) and (47), each component in any off-diagonal block of $\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}$ can be bounded as follows:

$$\max_{i,i'}\left|\left[\left(\overline{\Sigma}_{S_{w_{\mathcal{V}_n}}S_{w_{\mathcal{V}_n}}}\right)^{-1}\overline{\Sigma}_{S_{w_{\mathcal{V}_n}}S_{w_{\mathcal{V}_{n'}}}}\right]_{ii'}\right| \leq \left|S_{w_{\mathcal{V}_n}}\right|\xi\varepsilon \tag{48}$$

for $n, n' = 1, \ldots, \mathcal{K}$ and $n \neq n'$. It follows from (48) that

$$\max_{i,i'}\left|\left[\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}\right]_{ii'}\right| \leq \max_n\left|S_{w_{\mathcal{V}_n}}\right|\xi\varepsilon \leq L\kappa\xi\varepsilon \ . \tag{49}$$

The last inequality is due to $\max_n\left|S_{w_{\mathcal{V}_n}}\right| \leq L\max_n|\mathcal{V}_n| \leq L\kappa$. Then,

$$\begin{aligned}
\rho_w &\leq \left|\left|\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}\right|\right|_2\\
&\leq |S_w|\max_{i,i'}\left|\left[\left(\widehat{\Sigma}_{S_wS_w}\right)^{-1}\widetilde{\Sigma}_{S_wS_w}\right]_{ii'}\right|\\
&\leq KL^2\kappa\xi\varepsilon \ .
\end{aligned} \tag{50}$$

The first two inequalities follow from standard properties of matrix norm [54], [55]. The last inequality is due to (49).

The rest of this proof utilizes the following result of [56] that is revised to reflect our notations:

Theorem 4: If $|S_w|\rho_w^2 < 1$, then $\log\left|\overline{\Sigma}_{S_wS_w}\right| \leq \log\left|\widehat{\Sigma}_{S_wS_w}\right| \leq \log\left|\overline{\Sigma}_{S_wS_w}\right| - \log\left(1 - |S_w|\rho_w^2\right)$ for any joint walk w.

Using Theorem 4 followed by (50),

$$\begin{aligned}\log\left|\widehat{\Sigma}_{S_w S_w}\right| - \log\left|\overline{\Sigma}_{S_w S_w}\right| &\leq \log\frac{1}{1-|S_w|\rho_w^2} \\ &\leq \log\frac{1}{1-\left(K^{1.5}L^{2.5}\kappa\xi\varepsilon\right)^2}\end{aligned} \tag{51}$$

for any joint walk w.

$$\begin{aligned}&\overline{\mathbb{H}}[Z_{S_{w^*}}] - \overline{\mathbb{H}}[Z_{S_{\widehat{w}}}] \\ &= \frac{1}{2}\left((|S_{w^*}| - |S_{\widehat{w}}|)\log(2\pi e) + \log\left|\overline{\Sigma}_{S_{w^*}S_{w^*}}\right| - \log\left|\overline{\Sigma}_{S_{\widehat{w}}S_{\widehat{w}}}\right|\right) \\ &\leq \frac{1}{2}\left((|S_{w^*}| - |S_{\widehat{w}}|)\log(2\pi e) + \log\left|\widehat{\Sigma}_{S_{w^*}S_{w^*}}\right| - \log\left|\overline{\Sigma}_{S_{\widehat{w}}S_{\widehat{w}}}\right|\right) \\ &\leq \frac{1}{2}\left((|S_{\widehat{w}}| - |S_{\widehat{w}}|)\log(2\pi e) + \log\left|\widehat{\Sigma}_{S_{\widehat{w}}S_{\widehat{w}}}\right| - \log\left|\overline{\Sigma}_{S_{\widehat{w}}S_{\widehat{w}}}\right|\right) \\ &\leq \frac{1}{2}\log\frac{1}{1-\left(K^{1.5}L^{2.5}\kappa\xi\varepsilon\right)^2}\ .\end{aligned}$$

The first equality is due to (30). The first, second, and last inequalities follow from Theorem 4, (37), and (51), respectively.