Gaussian Process-Based Decentralized Data Fusion and Active Sensing for Mobility-on-Demand System

Jie Chen, Kian Hsiang Low, and Colin Keng-Yan Tan
Department of Computer Science, National University of Singapore, Republic of Singapore
{chenjie, lowkh, ctank}@comp.nus.edu.sg

***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. In this paper, we enhance the capability of a MoD system by deploying robotic shared vehicles that can autonomously cruise the streets to be hailed by users. A key challenge to managing the MoD system effectively is that of real-time, fine-grained mobility demand sensing and prediction. This paper presents a novel decentralized data fusion and active sensing algorithm for real-time, fine-grained mobility demand sensing and prediction with a fleet of autonomous robotic vehicles in a MoD system. Our Gaussian process (GP)-based decentralized data fusion algorithm can achieve a fine balance between predictive power and time efficiency. We theoretically guarantee its predictive performance to be equivalent to that of a sophisticated centralized sparse approximation for the GP model: The computation of such a sparse approximate GP model can thus be distributed among the MoD vehicles, hence achieving efficient and scalable demand prediction. Though our decentralized active sensing strategy is devised to gather the most informative demand data for demand prediction, it can achieve a dual effect of fleet rebalancing to service the mobility demands. Empirical evaluation on real-world mobility demand data shows that our proposed algorithm can achieve a better balance between predictive accuracy and time efficiency than state-of-the-art algorithms.**

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 [21] (e.g., Vélib system of over 20000 shared bicycles in Paris, experimental car-sharing systems described in [22]) 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 enhance the capability of a MoD system by deploying robotic shared vehicles (e.g., General Motors Chevrolet EN-V 2.0 prototype [2]) 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 [20]: (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 for MoD systems [22], dynamic traffic assignment problems [23], dynamic one-to-one pickup and delivery problems [4], and location recommendation and dispatch for cruising taxis [6, 9, 13, 30] have tackled variants of the second challenge by assuming the necessary input of mobility demand 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 are often subject to short-term random fluctuations and perturbations, in particular, due to frequent special events (e.g., storewide sales, exhibitions), unpredictable weather conditions, or emergencies (e.g., breakdowns in public transport services). So, in order for the active fleet management strategies to perform well, they require accurate, fine-grained information of the spatiotemporally varying mobility demand patterns in real time, which is the desired outcome of addressing the first challenge. To the best of our knowledge, there is little progress in the algorithmic development of the first challenge, which will be the focus of our work in this paper.

Given that the autonomous robotic vehicles are used directly as mobile probes to sample real-time demand data (e.g., pickup counts of different regions), how can the vacant ones actively cruise a road network to gather and assimilate the *most informative* demand data for predicting the mobility demand pattern? To solve this problem, a centralized approach to data fusion and active sensing [5, 12, 15, 16, 18, 19] 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 a novel decentralized data fusion and active sensing algorithm for real-time, fine-grained mobility demand sensing and prediction with a fleet of autonomous robotic vehicles in a MoD system. The specific contributions of our work include:

- Modeling and predicting the spatiotemporally varying mobility demand pattern using a rich class of Bayesian nonparametric models called the *log-Gaussian process* (ℓGP) (Section II);
- Developing a novel *Gaussian process-based decentralized data fusion* (GP-DDF$^+$) algorithm (Section III) that achieves a fine balance between predictive power and time efficiency, and theoretically guaranteeing its predictive performance to be equivalent to that of a sophisticated centralized sparse approximation for the Gaussian process (GP) model: The computation of such a sparse approximate GP model can thus be distributed among the MoD vehicles, hence achieving efficient and scalable demand prediction;
- Exploiting a *decentralized active sensing* (DAS) strategy (Section IV) to gather the most informative demand data for predicting the mobility demand pattern: Interestingly, we can analytically show that DAS exhibits 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;
- 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 algorithm on a real-world mobility demand pattern over the central business district of Singapore during evening hours (Section VI).

II. MODELING A MOBILITY DEMAND PATTERN

The service area in an urban city 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'. Each region $s \in V$ is associated with a p-dimensional feature vector x_s representing its context information (e.g., location, time, precipitation), and a measurement y_s quantifying its mobility demand[1]. 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 [11] 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 [6, 13] and our real-world data (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. We like to consider using a rich class of Bayesian nonparametric models called *Gaussian process* (GP) [25] to model the demand pattern. But, the GP covariance structure is sensitive to strong positive skewness and easily destabilized by a few extreme measurements [28]. 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 [10], 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.

A. Gaussian Process (GP)

Each region $s \in V$ is associated with a realized (random) log-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. The GP is fully specified by its *prior* mean $\mu_s \triangleq \mathbb{E}[Z_s]$ and covariance $\sigma_{ss'} \triangleq \text{cov}(Z_s, Z_{s'})$ for all $s, s' \in V$, the latter of which is defined by the widely-used squared exponential covariance function:

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

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 a set $D \subset V$ of observed regions and a column vector z_D of corresponding log-measurements, the GP can be used to predict the log-measurements of any set $S \subset V$ of unobserved regions with the following Gaussian *posterior* mean vector and covariance matrix:

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

$$\Sigma_{SS|D} \triangleq \Sigma_{SS} - \Sigma_{SD}\Sigma_{DD}^{-1}\Sigma_{DS} \quad (2)$$

[1]The service area is represented as a grid of regions instead of a network of road segments like in [7] because we observe less smoothly-varying, noisier demand measurements (hence, lower spatial correlation) for the latter in our real-world data (Section VI) since many road segments do not permit stopping of vehicles.

where $\mu_S(\mu_D)$ is a column vector with mean components μ_s for all $s \in S(s \in D)$, $\Sigma_{SD}(\Sigma_{DD})$ 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}.

B. Log-Gaussian Process (ℓGP)

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 of 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)

$$\widehat{\mu}_{s|D} \triangleq \exp(\mu_{s|D} + \Sigma_{ss|D}/2) \tag{3}$$

where $\mu_{s|D}$ and $\Sigma_{ss|D}$ are simply the Gaussian posterior mean (1) and variance (2) of GP, respectively. The uncertainty of predicting the measurements of 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 DAS strategy (Section IV):

$$\mathbb{H}[Y_S|Y_D] \triangleq \frac{1}{2}\log(2\pi e)^{|S|}\left|\Sigma_{SS|D}\right| + \mu_{S|D}\cdot\mathbf{1} \tag{4}$$

where $\mu_{S|D}$ and $\Sigma_{SS|D}$ are the Gaussian posterior mean vector (1) and covariance matrix (2) of GP, respectively.

III. Decentralized Demand Data Fusion

The demand data are gathered in a distributed manner by the vacant MoD vehicles cruising the service area and have to be assimilated in order to predict the mobility demand pattern. A straightforward approach to data fusion is to fully communicate all the data to every vehicle, each of which then performs the same exact ℓGP prediction (3) separately. This approach, which we call full Gaussian process (FGP) [15, 16], unfortunately cannot scale well and be performed in real time due to its cubic time complexity in the size of the data.

Alternatively, the work of [7] has recently proposed a Gaussian process-based decentralized data fusion (GP-DDF) approach to efficient and scalable approximate GP and ℓGP prediction. Their key idea is as follows: Each vehicle summarizes all its local data, based on a common prior support set $U \subset V$, into a local summary (Definition 1) to be exchanged with every other vehicle. Then, it assimilates the local summaries received from the other vehicles into a global summary (Definition 2), which is then exploited for predicting the demands of unobserved regions. 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.

We propose a novel decentralized data fusion algorithm called GP-DDF$^+$ that combines the best of both worlds, that is, the predictive power of FGP and efficiency of GP-DDF. GP-DDF$^+$ is based on the intuition that a vehicle can exploit its local data to improve the demand predictions for unobserved regions "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, as reproduced below:

Definition 1 (Local Summary): Given a common support set $U \subset V$ known to all K vehicles, a set $D_k \subset V$ of observed regions and a column vector z_{D_k} of corresponding demand measurements local to vehicle k, its local summary is defined as a tuple $(\dot{z}^k_U, \dot{\Sigma}^k_{UU})$ where

$$\dot{z}^k_B \triangleq \Sigma_{BD_k}\Sigma^{-1}_{D_kD_k|U}(z_{D_k} - \mu_{D_k}) \tag{5}$$

$$\dot{\Sigma}^k_{BB'} \triangleq \Sigma_{BD_k}\Sigma^{-1}_{D_kD_k|U}\Sigma_{D_kB'} \tag{6}$$

such that $\Sigma_{D_kD_k|U}$ is defined in a similar manner as (2).

Definition 2 (Global Summary): Given a common support set $U \subset V$ known to all K vehicles and the local summary $(\dot{z}^k_U, \dot{\Sigma}^k_{UU})$ of every vehicle $k = 1, \ldots, K$, the global summary is defined as a tuple $(\ddot{z}_U, \ddot{\Sigma}_{UU})$ where

$$\ddot{z}_U \triangleq \sum_{k=1}^{K}\dot{z}^k_U \tag{7}$$

$$\ddot{\Sigma}_{UU} \triangleq \Sigma_{UU} + \sum_{k=1}^{K}\dot{\Sigma}^k_{UU}\ . \tag{8}$$

Using GP-DDF [7], each vehicle exploits the global summary to compute a globally consistent predictive Gaussian distribution of the log-measurements of any set of unobserved regions. The resulting predictive Gaussian mean and variance can then be plugged into (3) to obtain the log-Gaussian posterior mean for predicting the demand of any unobserved region in the original scale. To improve the predictive power of GP-DDF, we develop the following novel GP-DDF$^+$ algorithm that is further augmented by local information.

Definition 3 (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 $(\dot{z}^k_U, \dot{\Sigma}^k_{UU})$, a set $D_k \subset V$ of observed regions and a column vector z_{D_k} of corresponding measurements local to vehicle k, its GP-DDF^+_k algorithm computes a predictive Gaussian distribution $\mathcal{N}(\overline{\mu}^k_S, \overline{\Sigma}^k_{SS})$ of the demand measurements of any set $S \subset V$ of unobserved regions where $\overline{\mu}^k_S \triangleq \left(\overline{\mu}^k_s\right)_{s\in S}$ and $\overline{\Sigma}^k_{SS} \triangleq \left(\overline{\sigma}^k_{ss'}\right)_{s,s'\in S}$ such that

$$\overline{\mu}^k_s \triangleq \mu_s + \left(\gamma^k_{sU}\ddot{\Sigma}^{-1}_{UU}\ddot{z}_U - \Sigma_{sU}\Sigma^{-1}_{UU}\dot{z}^k_U\right) + \dot{z}^k_s \tag{9}$$

$$\begin{aligned}\overline{\sigma}^k_{ss'} \triangleq \sigma_{ss'} - \Big(&\gamma^k_{sU}\Sigma^{-1}_{UU}\Sigma_{Us'} - \Sigma_{sU}\Sigma^{-1}_{UU}\dot{\Sigma}^k_{Us'} \\ &- \gamma^k_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^k_{Us'}\Big) - \dot{\Sigma}^k_{ss'}\end{aligned} \tag{10}$$

and

$$\gamma^k_{sU} \triangleq \Sigma_{sU} + \Sigma_{sU}\Sigma^{-1}_{UU}\dot{\Sigma}^k_{UU} - \dot{\Sigma}^k_{sU}\ . \tag{11}$$

Remark 1. Both the predictive Gaussian mean $\overline{\mu}^k_s$ (9) and covariance $\overline{\sigma}^k_{ss'}$(10) of GP-DDF$^+_k$ exploit summary information (i.e., bracketed term) derived from global and local summaries and local information (i.e., last term) derived from local data.

Remark 2. The predictive Gaussian mean $\overline{\mu}^k_s$ (9) and variance $\overline{\sigma}^k_{ss}$ (10) can be plugged into (3) to obtain the log-Gaussian posterior mean for predicting the demand 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 mobility demand pattern.

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

Definition 4 (Assignment Function): An assignment function $\tau : V \mapsto \{1 \ldots K\}$ is defined as

$$\tau(s) \triangleq \underset{k\in\{1\ldots K\}}{\arg\min}\ \overline{\sigma}^k_{ss} \tag{12}$$

for all $s \in S$ where the predictive variance $\overline{\sigma}^k_{ss}$ is defined in (10). 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 predictive Gaussian distribution, as detailed in Theorem 1A below:

Theorem 1 (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 predictive Gaussian distribution $\mathcal{N}(\overline{\mu}_S, \overline{\Sigma}_{SS})$ of the demand measurements of any set $S \subset V$ of unobserved regions where $\overline{\mu}_S \triangleq (\overline{\mu}^{\tau_s}_s)_{s\in S}$ (9) and $\overline{\Sigma}_{SS} \triangleq (\overline{\sigma}_{ss'})_{s,s'\in S}$ such that

$$\overline{\sigma}_{ss'} \triangleq \begin{cases} \overline{\sigma}^{\tau_s}_{ss'} & \text{if } \tau_s = \tau_{s'}, \\ \Sigma_{ss'|U} + \gamma^{\tau_s}_{sU}\ddot{\Sigma}^{-1}_{UU}\gamma^{\tau_{s'}}_{Us'} & \text{otherwise}\ , \end{cases} \tag{13}$$

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

B. Let $\mathcal{N}(\mu^{\text{PIC}}_{S|D}, \Sigma^{\text{PIC}}_{SS|D})$ be the predictive Gaussian distribution computed by the centralized sparse partially independent conditional (PIC) approximation of GP model [26] where $\mu^{\text{PIC}}_{S|D} \triangleq \left(\mu^{\text{PIC}}_{s|D}\right)_{s\in S}$ and $\Sigma^{\text{PIC}}_{SS|D} \triangleq \left(\sigma^{\text{PIC}}_{ss'|D}\right)_{s,s'\in S}$ such that

$$\mu^{\text{PIC}}_{s|D} \triangleq \mu_s + \widetilde{\Gamma}_{sD}\left(\Gamma_{DD} + \Lambda\right)^{-1}(z_D - \mu_D) \tag{14}$$

$$\sigma^{\text{PIC}}_{ss'|D} \triangleq \sigma_{ss'} - \widetilde{\Gamma}_{sD}\left(\Gamma_{DD} + \Lambda\right)^{-1}\widetilde{\Gamma}_{Ds'} \tag{15}$$

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

$$\Gamma_{BB'} \triangleq \Sigma_{BU}\Sigma^{-1}_{UU}\Sigma_{UB'} \tag{16}$$

$$\widetilde{\Gamma}_{sD} \triangleq (\widetilde{\Gamma}_{s\bar{s}})_{\bar{s}\in D} \tag{17}$$

$$\widetilde{\Gamma}_{s\bar{s}} \triangleq \begin{cases} \sigma_{s\bar{s}} & \text{if } \tau_s = \tau_{\bar{s}}, \\ \Gamma_{s\bar{s}} & \text{otherwise}\ , \end{cases} \tag{18}$$

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 $D = \bigcup^K_{k=1} D_k$, and let $\tau_{\bar{s}} \triangleq k$ for all $\bar{s} \in D_k$. Then, $\overline{\mu}_s = \mu^{\text{PIC}}_{s|D}$ and $\overline{\sigma}_{ss'} = \sigma^{\text{PIC}}_{ss'|D}$ for all $s, s' \in S$.

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

Remark 1. In Theorem 1A, if $\tau_s = \tau_{s'} = k$, then vehicle k can compute $\overline{\mu}^{\tau_s}_s$ (9) and $\overline{\sigma}_{ss'}$ (10) 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 $\gamma^{\tau_s}_{sU}$ and $\gamma^{\tau_{s'}}_{s'U}$ from the respective vehicles τ_s and $\tau_{s'}$ to compute $\overline{\sigma}_{ss'}$ (13).

Remark 2. The equivalence result of Theorem 1B implies that the computational load of the centralized PIC approximation of GP can be distributed among K vehicles, hence improving the time efficiency of demand 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 $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K\big)$ time of running GP-DDF$^+$ 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 real-world mobility demand data (Section VI) also show that GP-DDF$^+$ achieves predictive accuracy comparable to FGP and significantly better than GP-DDF, thus justifying the practicality of such an assumption for predicting a mobility demand pattern.

IV. Decentralized Active Demand Sensing

Suppose that there are K vacant MoD vehicles in the fleet actively cruising the service area and each vehicle $k \in \{1 \ldots K\}$ has observed a set $D_k \subset V$ of regions. In the active demand sensing problem, all vehicles have to jointly select the most informative walks $w^*_1, \ldots w^*_K$ of length H each along which demand data will be sampled:

$$(w^*_1, \ldots, w^*_K) \triangleq \underset{(w_1,\ldots,w_K)}{\arg\max}\ \mathbb{H}\left[Y_{\bigcup^K_{k=1} S_{w_k}} \middle| Y_{\bigcup^K_{k=1} D_k}\right] \tag{19}$$

where S_{w_k} denotes the set of unobserved regions to be visited by the walk w_k. To ease notation, let $w \triangleq (w_1 \ldots w_K)$ and $S_w = \bigcup^K_{k=1} S_{w_k}$ (similarly, for w^* and S_{w^*}). Then, each vehicle k executes its walk w^*_k while observing the demands of regions $S_{w^*_k}$, and updates its location and stored data.

To derive the most informative joint walk w^*, the posterior entropy (19) of every possible joint walk w has to be evaluated. Such a centralized strategy cannot be performed in real time due to the following two issues: (a) It relies on all the demand data that are gathered in a distributed manner 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 log-Gaussian posterior joint entropy using the decentralized GP-DDF$^+$ algorithm (Theorem 1A), thus distributing its computational load among all vehicles. Then, the active demand sensing problem (19) is approximated by

$$w^* = \arg\max_{w} \overline{\mathbb{H}}\left[Y_{S_w}\right] \quad (20)$$

$$\overline{\mathbb{H}}\left[Y_{S_w}\right] \triangleq \frac{1}{2}\log(2\pi e)^{|S_w|}\left|\overline{\Sigma}_{S_wS_w}\right| + \overline{\mu}_{S_w}\cdot\mathbf{1}\,. \quad (21)$$

To obtain $\overline{\mathbb{H}}\left[Y_{S_w}\right]$ (21), $\Sigma_{S_wS_w|D}$ and $\mu_{S_w|D}$ in $\mathbb{H}[Y_{S_w}|Y_D]$ ((4) & (19)) are replaced by $\overline{\Sigma}_{S_wS_w}$ and $\overline{\mu}_{S_w}$ defined in Theorem 1A, respectively.

To address the second issue, a partially decentralized active sensing strategy proposed by [7] partitions the vehicles into several small groups such that each group of vehicles selects its joint walk independently. This partitioning heuristic performs poorly when the largest group formed still contains many vehicles. In our work, this is indeed the case because many vehicles tend to cluster within hotspots, as explained later. To scale well in the fleet size, we therefore adopt a fully decentralized active sensing (DAS) strategy by assuming that the joint walk $w_1^* \ldots w_K^*$ is derived by selecting the locally optimal walk of each vehicle k:

$$w_k^* = \arg\max_{w_k} \overline{\mathbb{H}}\left[Y_{S_{w_k}}\right] \quad (22)$$

where $\overline{\mathbb{H}}\left[Y_{S_{w_k}}\right]$ is defined in the same way as (21). 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 $Y_{S_{w_1}}, \ldots, Y_{S_{w_K}}$.

More importantly, it can be observed from (21) and (22) that the cruising behavior of 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, DAS is able to achieve a dual effect of fleet rebalancing to service mobility demands.

V. Time and Communication Analysis

In this section, we analyze the time and communication overheads of our proposed GP-DDF$^+$ coupled with DAS algorithm (Algo. 1) and compare them with that of both FGP (Section II) and GP-DDF [7] coupled with DAS algorithms (Section IV).

Algorithm 1: GP-DDF$^+$+DAS$(U, K, H, k, D_k, z_{D_k})$

while *true* **do**
- `/* Data fusion (Section III) */`
- Construct local summary by (5) & (6)
- Exchange local summary with every vehicle $i \neq k$
- Construct global summary by (7) & (8)
- Construct assignment function by (12)
- Predict demand measurements of unobserved regions by (9) & (13)
- `/* Active Sensing (Section IV) */`
- Compute local maximum-entropy walk w_k^* by (22)
- Execute walk w_k^* and observe its demand measurements $Y_{w_k^*}$
- Update local information D_k and y_{D_k}

A. Time Complexity

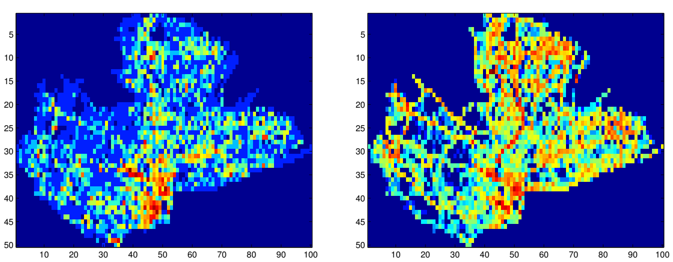
Firstly, each vehicle k has to evaluate $\Sigma_{D_kD_k|U}$ in $\mathcal{O}\big(|U|^3 + |U|(|D|/K)^2\big)$ time and invert it in $\mathcal{O}\big((|D|/K)^3\big)$ time. After that, the data fusion component constructs the local summary in $\mathcal{O}\big(|U|^2|D|/K + |U|(|D|/K)^2\big)$ time by (5) and (6), and subsequently the global summary in $\mathcal{O}\big(|U|^2K\big)$ time by (7) and (8). 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 (11). Then, after inverting $\ddot{\Sigma}_{UU}$ in $\mathcal{O}(|U|^3)$, 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 (9) and (13), respectively. Let $\Delta \triangleq \delta^H$ denote the number of possible walks of length H where δ is the maximum out-degree of graph G. In the active sensing component, to obtain the locally optimal walk, the log-Gaussian posterior entropies (22) of all possible walks are derived from (9) and (13), respectively, in $\mathcal{O}\big(\Delta H|U|^2\big)$ and $\mathcal{O}\big(\Delta(H|U|)^2\big)$ time. We assume $|S| \leq \delta\Delta$ where S denotes the set $\bigcup_{w_k} S_{w_k}$ of regions covered by any vehicle k's all possible walks of length H. Then, the time complexity for our GP-DDF$^+$ coupled with DAS algorithm is $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K + \Delta(H^3 + (H|U|)^2 + (|D|/K)^2)\big)$. In contrast, the time incurred by FGP and GP-DDF coupled with DAS algorithms are, respectively, $\mathcal{O}\big(|D|^3 + \Delta(H^3 + (H|D|)^2)\big)$ and $\mathcal{O}\big((|D|/K)^3 + |U|^3 + |U|^2K + \Delta(H^3 + (H|U|)^2)\big)$. It can be observed that our GP-DDF$^+$ coupled with DAS algorithm can scale better with large size $|D|$ of data and fleet size K than FGP coupled with DAS algorithm, and its increased computational load, as compared to GP-DDF coupled with DAS algorithm, is well distributed among K vehicles.

B. Communication complexity

In each iteration, each vehicle of the system running our GP-DDF$^+$ coupled with DAS algorithm has to broadcast a $\mathcal{O}(|U|^2)$-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^k_{sU} components for evaluating the entropies of all possible local walks. In contrast, FGP coupled with 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 DAS algorithm only needs to broadcast a $\mathcal{O}(|U|^2)$-sized local summary.

VI. Experiments and Discussion

This section evaluates the performance of our proposed algorithm using 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 service area is gridded into 50×100 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).



(a) Demand (b) Supply

Fig. 1. Historic demand and supply distributions.

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 introduce 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 L time steps and each vehicle plans a walk of length 4 at each time step, with all vehicles running a data fusion algorithm coupled with our DAS strategy. We will compare the performance of our GP-DDF$^+$ algorithm with that of FGP and GP-DDF algorithms when coupled with our DAS strategy. The experiments are conducted on a Linux system with Intel® Xeon® CPU E5520 at 2.27 GHz.

A. Performance Metrics

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.

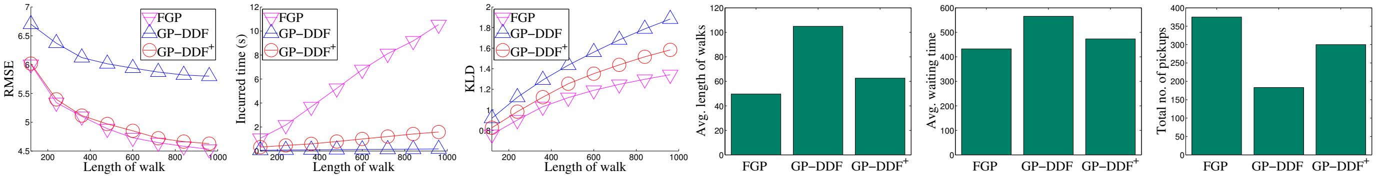
B. Results and Analysis

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

1) Performance: The MoD system comprises $K = 20$ vehicles running three tested algorithms for $L = 960$ time steps in a service area with $C = 200$ users. All results are obtained by averaging over 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 FGP 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 FGP, and only slightly less time-efficient than GP-DDF. This can be explained by the time analysis in Section V. 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 FGP. 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



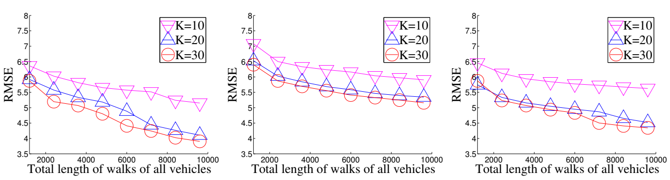
(a) Accuracy (b) Efficiency (c) Balance (d) Vehicles (e) Users (f) Pickups

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

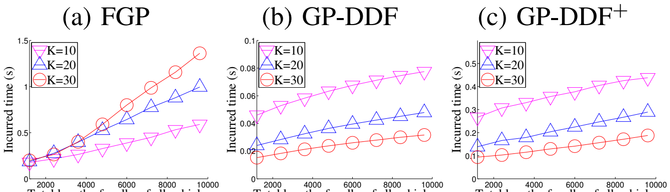
is, shorter average cruising trajectories of vehicles (Fig. 2d), shorter average waiting time of users (Fig. 2e), and larger total number of pickups (Fig. 2f). These observations imply that exploiting an active sensing strategy to collect the most informative demand data for predicting the mobility pattern 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 $L = 960, 480, 320$ 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 ($H = 4$) and gathering more informative demand data. Figs. 3d-3f show that, with more MoD vehicles, GP-DDF$^+$ and GP-DDF incur less time, while FGP 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.



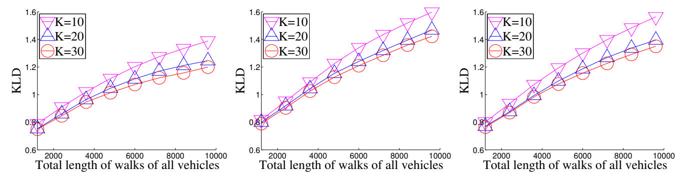
(a) FGP (b) GP-DDF (c) GP-DDF$^+$



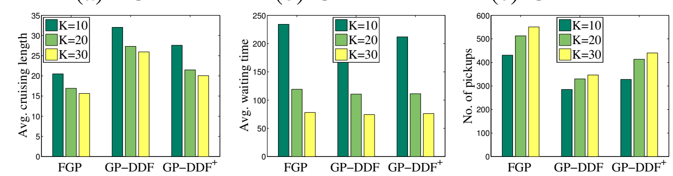
(d) FGP (e) GP-DDF (f) GP-DDF$^+$

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

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.



(a) FGP (b) GP-DDF (c) GP-DDF$^+$



(d) Vehicles (e) Users (f) Pickups

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

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 servicing mobility demands.

VII. Conclusion

This paper describes a novel GP-based decentralized data fusion (GP-DDF$^+$) and active sensing (DAS) algorithm for real-time, fine-grained mobility demand sensing and prediction with a fleet of autonomous robotic vehicles in a MoD system. We have analytically and empirically demonstrated that GP-DDF$^+$ can achieve a better balance between predictive accuracy and time efficiency than the state-of-the-art GP-DDF [7] and FGP [15, 16]. The practical applicability of GP-DDF$^+$ is not restricted to mobility demand prediction; it can be used in other urban and natural environmental sensing applications like monitoring of traffic, ocean and freshwater phenomena [5, 8, 14, 17, 18, 19, 24, 27, 29]. We have also analytically and empirically shown that even though DAS is devised to gather the most informative demand data for predicting the mobility demand pattern, it can achieve a dual effect of fleet rebalancing to service the mobility demands. For our future work, we will relax the assumption of all-to-all communication such that each sensor may only be able to communicate locally with its neighbors and develop a distributed data fusion approach to efficient and scalable approximate GP and ℓGP prediction.

ACKNOWLEDGMENTS

This work was supported by Singapore-MIT Alliance Research & Technology Subaward Agreements No. 28 R-252-000-502-592 & No. 33 R-252-000-509-592.

REFERENCES

[1] http://www.comp.nus.edu.sg/∼lowkh/pubs/rss2013e.pdf.

[2] *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.

[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] G. Berbeglia, J.-F. Cordeau, and G. Laporte. Dynamic pickup and delivery problems. *Eur. J. Oper. Res.*, 202 (1):8–15, 2010.

[5] N. Cao, K. H. Low, and J. M. Dolan. Multi-robot informative path planning for active sensing of environmental phenomena: A tale of two algorithms. In *Proc. AAMAS*, pages 7–14, 2013.

[6] H.-W. Chang, Y.-C. Tai, and J. Y.-J. Hsu. Context-aware taxi demand hotspots prediction. *Int. J. Business Intelligence and Data Mining*, 5(1):3–18, 2010.

[7] 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. UAI*, pages 163–173, 2012.

[8] J. M. Dolan, G. Podnar, S. Stancliff, K. H. Low, A. Elfes, J. Higinbotham, J. C. Hosler, T. A. Moisan, and J. Moisan. Cooperative aquatic sensing using the telesupervised adaptive ocean sensor fleet. In *Proc. SPIE Conference on Remote Sensing of the Ocean, Sea Ice, and Large Water Regions*, volume 7473, 2009.

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

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

[11] M. Jelasity, A. Montresor, and O. Babaoglu. Gossip-based aggregation in large dynamic networks. *ACM Trans. Comput. Syst.*, 23(3):219–252, 2005.

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

[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.*, 6(1):111–121, 2012.

[14] K. H. Low, G. J. Gordon, J. M. Dolan, and P. Khosla. Adaptive sampling for multi-robot wide-area exploration. In *Proc. IEEE ICRA*, pages 755–760, 2007.

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

[16] K. H. Low, J. M. Dolan, and P. Khosla. Information-theoretic approach to efficient adaptive path planning for mobile robotic environmental sensing. In *Proc. ICAPS*, pages 233–240, 2009.

[17] K. H. Low, G. Podnar, S. Stancliff, J. M. Dolan, and A. Elfes. Robot boats as a mobile aquatic sensor network. In *Proc. IPSN-09 Workshop on Sensor Networks for Earth and Space Science Applications*, 2009.

[18] K. H. Low, J. M. Dolan, and P. Khosla. Active Markov information-theoretic path planning for robotic environmental sensing. In *Proc. AAMAS*, pages 753–760, 2011.

[19] K. H. Low, J. Chen, J. M. Dolan, S. Chien, and D. R. Thompson. Decentralized active robotic exploration and mapping for probabilistic field classification in environmental sensing. In *Proc. AAMAS*, pages 105–112, 2012.

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

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

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

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

[24] G. Podnar, J. M. Dolan, K. H. Low, and A. Elfes. Telesupervised remote surface water quality sensing. In *Proc. IEEE Aerospace Conference*, 2010.

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

[26] E. Snelson. Local and global sparse Gaussian process approximations. In *Proc. AISTATS*, 2007.

[27] D. R. Thompson, N. Cabrol, M. Furlong, C. Hardgrove, K. H. Low, J. Moersch, and D. Wettergreen. Adaptive sampling of time series for remote exploration. In *Proc. IEEE ICRA*, 2013.

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

[29] J. Yu, K. H. Low, A. Oran, and P. Jaillet. Hierarchical Bayesian nonparametric approach to modeling and learning the wisdom of crowds of urban traffic route planning agents. In *Proc. IAT*, pages 478–485, 2012.

[30] 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.