2016 16th International Conference on Control, Automation and Systems (ICCAS 2016)
Oct. 16–19, 2016 in HICO, Gyeongju, Korea

Resource-Constrained Decentralized Active Sensing for Multi-Robot Systems using Distributed Gaussian Processes

Kshitij Tiwari[1*], Valentin Honoré[2], Sungmoon Jeong[1], Nak Young Chong[1], Marc Peter Deisenroth[3]

[1]School of Information Science, JAIST, Japan {kshitij.tiwari,jeongsm,nakyoung}@jaist.ac.jp
[2]Department of Computer Science, École Normale Supérieure de Lyon, France (valentin.honore@ens-lyon.fr)
[3]Department of Computing, Imperial College London, United Kingdom (m.deisenroth@imperial.ac.uk)

Abstract: We consider the problem of area coverage for robot teams operating under resource constraints, while modeling spatio-temporal environmental phenomena. The aim of the mobile robot team is to avoid exhaustive search and only visit the most important locations that can improve the prediction accuracy of a spatio-temporal model. We use a Gaussian Process (GP) to model spatially varying and temporally evolving dynamics of the target phenomenon. Each robot of the team is allocated a dedicated search area wherein the robot autonomously optimizes its prediction accuracy. We present this as a *Decentralized Computation and Centralized Data Fusion* approach wherein the trajectory sampled by the robot is generated using our proposed *Resource-Constrained Decentralized Active Sensing (RC-DAS)*. Since each robot possesses its own independent prediction model, at the end of robot's mission time, we fuse all the prediction models from all robots to have a global model of the spatio-temporal phenomenon. Previously, all robots and GPs needed to be synchronized, such that the GPs can be jointly trained. However, doing so defeats the purpose of a fully decentralized mobile robot team. Thus, we allow the robots to independently gather new measurements and update their model parameters irrespective of other members of the team. To evaluate the performance of our model, we compare the trajectory traced by the robot using active and passive (e.g., nearest neighbor selection) sensing. We compare the performance and cost incurred by a resource constrained optimization with the unconstrained entropy maximization version.

Keywords: Multiple Robots, Decentralized Active Sensing, Field Mapping, Gaussian Process.

1. INTRODUCTION

Environmental monitoring has been a core area of research in geostatistics [1], [2], [3]. Recently, the robotics community started aiding this research by providing autonomous mobile sensor nodes that can assist in making scrupulous models of complex spatio-temporal environmental phenomena [4], [5], [6], [7]. In the past, this problem was mostly posed as a static sensor placement problem [4]. In this case, the target phenomenon covers a wide area and due to intrinsic variations, the sensor nodes may have to be moved from time-to-time in order to obtain the most informative measurements [7]. This problem was later solved by using mobile sensor nodes like UAVs, AUVs, UGVs etc. With the involvement of robots, we get the benefit of covering wide areas and even wider domains of environmental sensing, e.g., oil spill monitoring, phytoplankton density monitoring [8], some urban problems like road traffic [9] and ambulance demand monitoring [10] alongside the classical weather forecasting problems and pollution management.

Formulating a parametric model and identifying its parameters are complicated for spatio-temporal dynamic phenomena. Therefore, we choose a non-parametric Bayesian framework for modeling spatio-temporal phenomena. In particular, we use Gaussian processes (GPs) [11] for this purpose. GPs exhibit the desirable property of increasing flexibility (as more data becomes available) while not being prone to over-fitting. Furthermore, low-level parameters (e.g., the degree of the polynomial for regression) are no longer necessary. Instead, high-level parameters (e.g., smoothness) are determined more easily. The GP's predictions are naturally equipped with variances, which allow us to determine whether a prediction is made with high confidence.

Using a standard GP framework incurs computation and memory costs, which scale cubically and quadratically in the size of the data set, respectively. Therefore, many researchers are proposing new covariance structures [12] or trying to propose scalable GPs [13], [14], which can handle a large number of training instances. This is particularly useful for the case of mobile sensor nodes that gather large amount of measurements, which must be analyzed. Also, instead of using an offline model, researchers like to use an Online version of GP with Sparse Approximations, such that the model is updated as the new data comes in without incurring any additional memory cost [15], [16], [9]. In this paper, we use the Distributed Gaussian Process (DGP) framework [14], an offline GP method. In the DGP, the training points are distributed among multiple GPs, which reduces the burden of computation. Since in the DGP framework the computation is Decentralized we harness this aspect of the model to endow our team of mobile robots with a fully Decentralized active sensing scheme. Here, we extend the work of [14] by using active sensing, such that we can choose inputs to improve model performance rather than relying on a fixed training set.

In order to optimize the resource utilization, e.g., the

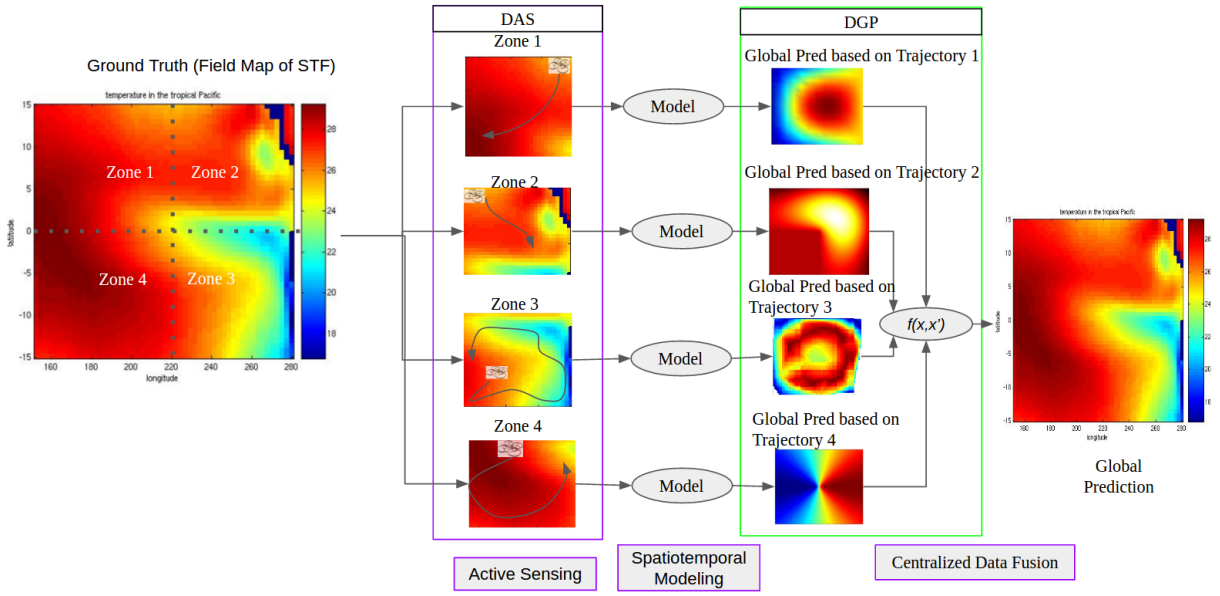

Fig. 1 **Sensing Scenario.** Illustration of centralized data fusion of the prediction model generated by all GPs in order to fuse them into a global prediction [14]. Images from http://tiny.cc/jkpfby, http://tiny.cc/qkpfby, http://tiny.cc/ykpfby

battery life of the robot, we propose a Resource Constrained Decentralized Active Sensing called RC-DAS. Our sensing scenario is illustrated in Fig. 1 for four robots: We partition the sensing area into four zones, i.e., one dedicated sensing zone per robot. The robots are only allowed to gather data within their own sensing zones but they can predict for any location in any zone. All locations available for sensing and predicting are known to all robots. The robots then use either full DAS or RC-DAS active sensing schemes, in order to move around and collect the most informative measurements. Upon termination of exploration the robots (GPs) make a prediction for the inputs that remain unobserved. Since each robot visited different locations in its respective sensing zone, it may predict differently for the unobserved set. Thus, we need to fuse all such predictions into a single global prediction model, which is done as a terminal step of exploration by the base station.

The contribution of our paper is a novel Active Sensing scheme, the *Resource Constrained Decentralized Active Sensing (RC-DAS)*. In this approach, the robots actively select the most informative sensing locations to enhance the prediction performance of the model being generated whilst simultaneously conserving the utilization of resources like battery, travel distance, etc.

1.1 Problem Statement

The problem addressed in this paper is as follows: *Given a team of robots and an unknown spatio-temporal phenomenon, we seek a decentralized strategy to improve their respective predictive models under resource constraints.* We propose to actively explore the spatio-temporal phenomenon. The local models of this decentralized robotic system are fused to a global model so that we can predict measurements at any unobserved location.

We wish to model a spatio-temporal phenomenon $z = f(x) + \epsilon$, where $x \in D \subset \mathbb{R}^d$ are tuples <Latitude,Longitude,Time> and $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$ is i.i.d. Gaussian measurement noise. We place a Gaussian process (GP) prior on the spatio-temporal phenomenon f and write $f \sim GP$. For each input $x \in D$, the associated measurement is denoted by z_x if the input was visited and the measurement was observed. Otherwise we define Z_x as a random variable, which we wish to predict at unobserved locations. Then, $\{Z_x\}_{x \in D}$ is a GP, and any of its finite subsets is a multivariate Gaussian distribution [11].

We propose to (a) model the spatio-temporal phenomenon using a probabilistic GP and (b) exploit the GP's predictive variance for active sensing, such that the robot only visits the most informative inputs rather than exhaustively searching through the whole zone. Active sensing usually encourages the robot to visit the most uncertain inputs [9], [8], but for a real robot setup, it could mean that the robot might end up choosing locations, which are prohibitively far away. We pose this as a fully Decentralized Active Sensing (DAS, similar to [9]) problem and propose a Resource Constrained Decentralized Active Sensing (RC-DAS) wherein the robot trades-off the model performance to travel distance.

2. MODELING A SPATIO-TEMPORAL ENVIRONMENTAL PHENOMENON

In this Section, we explain how to model the spatio-temporal field as GP and how to distribute the computation over multiple robots and fuse the prediction models in order to create a global prediction model using DGP.

2.1 Gaussian Process

GPs are a rich class of probabilistic non-parametric Bayesian models, which allow us to model spatio-

temporal phenomenon f. GPs consistently quantify the uncertainty associated with predictions, which can be exploited by active sensing schemes for exploration and obtaining the *Next Best Location* for each robot.

A GP is a generalization of a Gaussian distribution and fully defined by a mean function $\mu(\cdot) = \mathbb{E}[f(x)] = \mathbb{E}[Z_x]$ and covariance function $k(\cdot,\cdot)$. The covariance function (kernel), defines the spatio-temporal correlation structure of the function to be modeled and is parametrized by a set of hyper-parameters θ. A common covariance function is the squared exponential

$$\sigma_{xx'} = \sigma_{sig}^2 \exp\big(-\tfrac{1}{2}(x-x')^T L^{-1}(x-x')\big) + \sigma_n^2 \delta_{xx'}$$

where $x, x' \in D$, $L = \text{diag}(l_1^2, \ldots, l_D^2)$ and the l_i are characteristic length scales, which determine the relevance of the corresponding input dimension for modeling the spatio-temporal phenomenon. σ_{sig} corresponds to the amplitude of the signal to be modeled whereas σ_n describes the magnitude of the noise. The hyper-parameters $\theta \triangleq \{\sigma_{sig}^2, \sigma_n^2, l_1, l_2, \ldots, l_d\}$ are trained using the standard procedure of evidence (type-II marginal likelihood) maximization [11]. Evidence maximization avoids overfitting by trading off data fit and model complexity.

When a column vector z_O of realized measurements becomes available for a set $O \subset D$ of inputs, we can exploit these measurements to train the GP and predict the measurements at a set of unobserved inputs $U \subseteq D$ [9], [8][1]. The corresponding *Gaussian posterior predictive distribution* is

$$\mu_{U|O,\theta} \triangleq \mu_U + \Sigma_{UO|\theta}\Sigma_{OO|\theta}^{-1}(z_O - \mu_O) \quad (1)$$

$$\Sigma_{UU|O,\theta} \triangleq \Sigma_{UU|\theta} - \Sigma_{UO|\theta}\Sigma_{OO|\theta}^{-1}\Sigma_{OU|\theta} \quad (2)$$

where $\mu_{U|O,\theta}$ is a column vector of means of the predicted (posterior) measurements of spatio-temporal field and $\Sigma_{UU|O,\theta}$ is the corresponding predictive posterior covariance matrix. Here, we define the short-hand notation $\Sigma_{UO} = k(O,U)$, $\Sigma_{OO} = k(O,O)$, $\Sigma_{UU} = k(UU)$ for the corresponding covariance matrices.

2.2 Distributed Prediction Model

We learn the hyper-parameters of the GP by evidence maximization. For training the GP, we need to invert the covariance matrix Σ_{OO}, which requires $\mathcal{O}(|O|^3)$ time, where $|O|$ is the size of the training data set. At prediction time, we exploit a cached Σ_{OO}^{-1}, but the computational demand is still $\mathcal{O}(|O|^2)$ time for the predictive variance in Eq. (2). To lower this computational burden, we use an approximation proposed in the DGP framework [14]. The intuition of this approach is to split the training data set into M subsets and to train individual *GP experts* using these subsets. Then, if the k^{th} GP gets n_k data points for training, the computation time incurred is just $\mathcal{O}(|n_k|^3) \ll \mathcal{O}(|O|^3)$. The process can be parallelized straightforwardly. Besides computational and memory advantages, the DGP allows us to model local variations of the spatio-temporal function using a "local GP expert" and decentralize predictions.

Model. The DGP architecture is a framework for *product-of-GP-experts models*, which unifies the *Product of GP experts (PoE)* [17], the *generalized Product of GP Experts (gPoE)* [18], the *Bayesian Committee Machine* (BCM) [19] and the *robust BCM* [14], where the posterior prediction is a weighted product of the posterior predictions of all GP experts. The weights are determined based on each expert's confidence.

Training. In [14], the authors jointly trained all GP experts, which shared a single set of hyper-parameters θ. In the context of our decentralized robotics scenario, we assume that each robot independently optimizes its own set of hyper-parameters, such that we obtain a super-set of hyper-parameters $\Theta = [\theta_1, \theta_2, \ldots, \theta_K]$. This approach is more flexible than a shared set of hyper-paremeters and allows the robots to model local variations.

Prediction. In our model, the posterior predictive mean (predicted measurements over all the unobserved locations in set U) and the posterior covariance for the rBCM model for K robots (GP experts) are given by

$$\mu_{U|O,\Theta}^{rBCM} \triangleq \Sigma_{UU|O,\theta_k}^{rBCM} \sum_{k=1}^{K} \beta_k \Sigma_{UU|O_k,\theta_k}^{-1} \mu_{U|O_k}\,, \quad (3)$$

$$(\Sigma_{UU|O,\Theta}^{rBCM})^{-1} \triangleq \sum_{k=1}^{K} \beta_k \Sigma_{UU|O_k,\theta_k}^{-1} + \big((1 - \sum_{k=1}^{K} \beta_k)\Sigma_{UU|\theta_k}^{-1}\big)\,, \quad (4)$$

respectively, where the subscript k denotes a quantity related to the kth expert (e.g., the predictive mean or variance). In Eq.(3) and Eq.(4), the confidence of the ith model is encoded in its weight $\beta_i \triangleq 0.5(\mathbb{H}_{Z_{U_k}} - \mathbb{H}_{Z_{U_k}|z_{O_k},\theta_k})$ where $\mathbb{H}_{Z_{U_k}}$ refers to the prior entropy and $\mathbb{H}_{Z_{U_k}|Z_{O_k},\theta_k}$ refers to the posterior entropy of the prediction model.

2.2.1 Centralized Data Fusion

In our experiment, we allow each robot to gather unique training samples from its dedicated sensing zone and the unvisited/unobserved locations at the end of mission time are classified as testing set for the corresponding agent. Since, the number of feasible locations are fixed a priori, each robot's test set will have some locations which intersect with the other robot's test set. Thus, we need to fuse the predictions from all robots to make a global model of prediction such that we have a prediction over unobserved locations never visited by any robot.

To combine these predictions into a single prediction, we exploit the idea of [14] and build a global prediction at the base station by a weighted combination of individual predictions. This is illustrated in Fig. (1): Every GP expert i transmits its predicted mean μ_i, variance, Σ_i and

[1]For the current time instance, the predictions are only made for the locations not visited, but the predictions could also include the training locations should we choose to make the prediction for the future

weight β_i to the base station, which combines individual predictions to a global prediction.

When combining the prediction models at the base station, we are interested in only these mutually common locations thus, we define $U_{global} \triangleq \{U_1 \cap U_2 \cap \ldots \cap U_K\}$ as the super set of all unobserved inputs that were never visited by any robot. Similarly, we define $O_{global} \triangleq \{O_1 \cup O_2 \cup \ldots \cup O_K\}$ as the super set of all observed inputs that were visited by all robots. Then the global prediction for any unobserved input in U_{global} is given by the fused predictions from all robots:

$$p(Z_{U_{global}}|z_{O_{global}}) = \frac{\prod_{k=1}^{K}\left(p_k^{\beta_k}(Z_{U_k}|z_{O_k},\theta_k)\right)}{p^{\sum_{k=1}^{K}\beta_k - 1}(Z_{U_{global}})}. \quad (5)$$

Note that in Eq. (5) predictions from *every* robot are taken into account to form a global prediction without boundary effects. Each robot's prediction p_k is weighted by β_k, which depends on the confidence of the prediction.

3. ACTIVE SENSING

The DGP framework, which we use for predictions, assumes batch data for training. In this paper, we generalize this framework to the sequential setting by using active sensing under resource constraints. We use the uncertainty of the global GP model to determine the next-best location to visit while taking limited resources (e.g., battery) into account. In particular, we propose an Active Sensing approach with two variants: *fully Decentralized Active Sensing (full DAS) [9]* and *Resource-Constrained DAS (RC-DAS)*, which are detailed below.

Active Sensing integrates *Path Planning* and *Mapping*. In active sensing, the robot (or GP) is endowed with a decision making capability, such that it can choose the most informative training data, instead of visiting all available locations sequentially. This ensures that a robot does not end up performing exhaustive search, which is of particular interest if we have resource constraints, e.g., sampling budget. Once the robot has evaluated the next best location it wants to move to, a path planner helps the robot to reach the desired target. Upon successfully reaching the target, the new measurement observed and the corresponding input location are added to the training set (z_O, O), and the GP model of the map is updated. Note that in order to maximize the model the *Next Best Location* may be prohibitively far away from the robot, which could critically reduce the number of measurements that the robot can gather in mission time. In order to effectively allocate resources, whilst not compromising the model performance, we propose optimizing the travel distance alongside the choice of *Next Best Location*. This gives rise to Resource Constrained Decentralized Active Sensing (RC-DAS), which is detailed in Section 3.3

3.1 Multi-Robot Centralized Active Sensing (CAS)

In a centralized active sensing approach, the base station must jointly coordinate all robot trajectories. Consider a network of K mobile sensor nodes (robots), which are actively moving and gathering data to accurately predict a target phenomenon, e.g., the variation in ozone concentration, which is modeled with a Distributed GP [14]. In a fully centralized active sensing approach, the base node selects the most informative walks $w_1^*, w_2^*, \ldots, w_K^*$ of length L of all K robots, such that each robot can gather the most informative measurements that affect the *overall* performance of the fused GP model, i.e.,

$$(w_1^*, w_2^*, \ldots, w_K^*) = \underset{w_1,\ldots,w_K}{\arg\max}\ \mathbb{H}\left[Z_{U_k}|z_{O_k}\right]. \quad (6)$$

In Eq. (6), $U_k \triangleq \bigcup_{k=1}^{K} U_{w_k}$ represents the set of (unobserved) locations visited by robot k during its walk w_k and similarly the observations are represented by $O_k \triangleq \bigcup_{k=1}^{K} O_{w_k}$. $\mathbb{H}\left[Z_{U_k}|z_{O_k}\right]$ refers to the Entropy (a measure of prediction uncertainty) as defined in Eq. (7) Other measures of uncertainty are available, e.g., $\mathrm{trace}(\Sigma_{\mathrm{UU|O},\theta})$, but these measures lose the correlations between measurements. Hence, we will be relying on entropy as a measure of uncertainty [9].

$$\mathbb{H}_{Z_{U_k}|z_{O_k},\theta_k} \triangleq \tfrac{|U_k|}{2}\ln(2\pi e) + \tfrac{1}{2}\ln(|\Sigma_{U_kU_k|O_k,\theta_k}|) \quad (7)$$

However, such a Centralized Active Sensing (CAS) approach requires the evaluation of an exponential number of combinations of all possible joint walks. Furthermore, it relies on all measurements collected by all robots in this distributed system. Thus, to ease this computational burden, a widely used approach is decentralized active sensing.

3.2 Multi-Robot Decentralized Active Sensing (DAS)

In order to remedy the exponential computation of a fully centralized approach, we employ a fully decentralized approach wherein each robot optimizes its own walk at its own digression. In such a setup, the active sensing problem can be reformulated as

$$w_K^* = \underset{w_K}{\arg\max}\ \mathbb{H}[Z_{U_{w_k}}|z_{O_{w_k}}]. \quad (8)$$

As opposed to Eq. (6), in Eq. (8), each robot k *locally* optimizes its own trajectory w_k based on its observations O_{w_k}. Since the robots do not communicate and coordinate with each other, their walks may become correlated, and robots may end up exploring overlapping areas, which leads to resource wastage. This induces correlations in the measurements being gathered, which violates the conditional independence between observations of each GPs, which is important for us to fuse the outputs obtained from multiple GPs. To harness the strengths of full decentralization while avoiding violations of conditional independence assumption such that no two robots

visit the same input, we partition the sensing area, such that each robot is bounded within its own sensing zone that does not interfere or overlap with that of its neighbors.This also ensures robots avoid collision. However, every robot can predict in any other region, which was is used by the full DAS.

3.3 Multi-Robot Resource-Constrained Decentralized Active Sensing (RC-DAS)

Entropy maximization is a widely used greedy criterion to choose the next-best location to move to. However, using such a cost function in a real robot implementation could force the robot to select new locations far away, which will cost the robot a substantial time and energy to reach. Thus, while selecting the next location to move to, the robot must trade-off the informativeness of the location and its distance from the current location.Unfortunately, this problem is NP-hard [20].

To find an approximate solution to this problem we propose a new cost function, which is a combination of Entropy Maximization and travel distance. For this, we first define two costs that are incurred every time a new measurement is taken: 1) A *sensing cost* $C_S(x)$, which is the cost incurred by the robot when taking a measurement at location x; 2) a *travel cost* $C_T(x, x')$, which is the cost incurred by the robot when moving from current location x to the *Next-Best-Location* x' [21]. Our sampling budget has an upper-limit B.

The total cost (called as *Path Cost* from here on) incurred during a walk w_k of length L by robot k is

$$C(w_k) = \sum_{i=1}^{L} C_S(x_i) + \sum_{i=2}^{L} C_T(x_{i-1}, x_i) \qquad (9)$$

With the sampling budget B and the path costs in Eq. (9), we redefine our DAS problem as a *Resource-Constrained DAS (RC-DAS)*, which is formulated as

$$w_K^* = \arg\max_{w_K} \left(\alpha \, \mathbb{H}[Z_{U_{w_k}} | z_{O_{w_k}}] - (1-\alpha) \ln \|\mathbf{x} - \mathbf{x}_*\| \right) \qquad (10)$$

where $\mathbf{x} \in \mathbf{O}$ and $\mathbf{x}_* \in \mathbf{U}$. If $\alpha = 1$ Eq. (10) becomes similar to our full DAS setup in Eq. (8). For $\alpha = 0$, the equation reduces to nearest neighbor selection routine without any active sensing. Thus, at $\alpha = 0.75$[2] our cost function trades-off the travel distance against the Entropy Maximization.

4. EXPERIMENTS

To demonstrate our RC-DAS approach on a real-life scenario, we used the **US Ozone Dataset**. This dataset includes ozone concentration (in parts per billion) collected by US Environmental Protection Agency [3][22]. For data pre-processing, we followed the steps similar to those proposed by the authors in [12]. However, in the light of active sensing perspective of a real robot implementation, we do not split the data into testing and training set. Instead, we initialize an empty training set, which we call as the Observed set $O \subset D$, and put all admissible locations into another set called Unobserved set $U \subset D$ such that $U \cap O = \varnothing$. As the robot moves from one location to its *Next Best Sensing Location*, we move locations from U and add them to O. Since the measurement signal is noisy, we have to gather a few measurements and take their average as a measurement sample for a specific geo-location. In terms of ozone dataset, this meant taking a single measurement per month per location and computing the average over the year.

Table 1 **Performance Evaluation:** Performance of DGP when running on full DAS vs RC-DAS.

		Robot 1	Robot 2	Robot 3	Robot 4
RMSE	***NN***	17.4360	20.1437	13.0039	18.3455
	RC-DAS	16.3071	20.1437	13.0039	17.2355
	full DAS	14.5139	20.1412	18.9157	17.3427
Path Cost	***NN***	1.9185	6.3260	6.0406	3.6612
	RC-DAS	2.0083	6.3260	6.0406	3.5652
	full DAS	8.4230	7.9660	5.4497	9.1475

To allocate the maximum budget B we evaluate the maximum available spatial separation between any pair of sensing locations and assign the natural logarithm of this distance as B. Similarly, the minimum separation is assigned as the sensing cost C_S. We then evaluate the performance of Active Sensing versus that of Shortest Distance based *Next Best Sensing Location* selection criterion as shown in Table 1. The RMSE results show that if we just choose the different cost functions to deduce the *Next Best Sensing Location* then the GP prediction accuracy is the affected but since the results are presented on a normalized scale, the values remain in close proximity. We also contrast the trajectories sampled by the robots when running on full DAS v/s RC-DAS as shown in Fig. 2. The figure shows that our new RC-DAS cost function penalizes the travel cost binding the robot to visit the closest most uncertain locations first. This is clear from the trajectories of Robots 1 and 4. However, there are too few locations in the sensing zones of robots 2 and 3 so the trajectories remain unaffected. Globally, RC-DAS incurs lesser path cost than full DAS.

5. CONCLUSION

In this paper, we considered a wide-area coverage problem under resource constraints in the context of mobile robots. In particular, we considered the problem of exploring and modeling a spatio-temporal phenomenon. For this purpose, we utilized a decentralized team of mobile robots, each of which modeled the target phenomenon locally using an independent Gaussian process (GP) expert. Since GPs are data-driven models, the task at hand was to obtain the best possible model performance while conservatively utilizing the resource constraints (battery, travel distance, etc.) and efficiently trad-

[2] arbitrary choice for this paper
[3] https://www.epa.gov/castnet

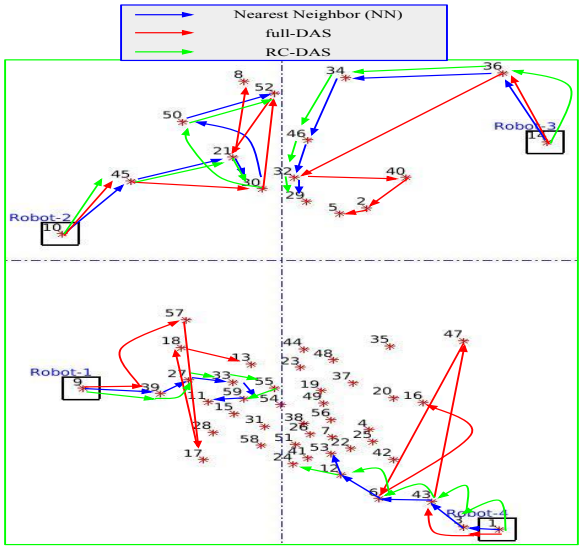

Fig. 2 **Full DAS v/s Nearest Neighbor v/s RC-DAS:** Trajectory of robots while using different cost functions to move to *Next Best Location*

ing off model quality with resource utilization.

We introduced Resource Constrained Decentralized Active Sensing (RC-DAS), which trades off the selection of most informative sensing location with that of the most energy-efficient location. We evaluated our approach empirically on a publicly available dataset for Ozone concentration and compared the model performance with that of an unconstrained entropy maximization approach.

ACKNOWLEDGEMENTS

This work was supported by the Industrial Convergence Core Technology Development Program(No. 10063172) funded by MOTIE, Korea. MPD has been supported by a Google Faculty Research Award.

REFERENCES

[1] Y. Q. Cui, M. Yoneda, Y. Shimada, and Y. Matsui, "Cost-effective strategy for the investigation and remediation of polluted soil using geostatistics and a genetic algorithm approach," *J. Env. Prot.*, 2016.

[2] R. Liu, F. Xu, W. Yu, J. Shi, P. Zhang, and Z. Shen, "Analysis of field-scale spatial correlations and variations of soil nutrients using geostatistics," *Env. Mon. Assess.*, 2016.

[3] X. Liu, P. C. Brookes, Q. Song, S. Jabeen, I. H. Shamsi, and Z. Zhu, "Spatio-temporal variability of heavy metal concentrations in soil-rice system and its socio-environmental analysis," *J. Agric. and Biol.*, 2016.

[4] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in Gaussian processes," in *ICML*, 2005.

[5] B. Lu, D. Gu, and H. Hu, "Tracking and modeling of spatio-temporal fields with a mobile sensor network," in *WCICA*, 2014.

[6] A. Singh, F. Ramos, H. D. Whyte, and W. J. Kaiser, "Modeling and decision making in spatio-temporal processes for environmental surveillance," in *ICRA*, 2010.

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

[8] R. Ouyang, K. H. Low, J. Chen, and P. Jaillet, "Multi-robot active sensing of non-stationary Gaussian process-based environmental phenomena," in *AAMAS*, 2014.

[9] J. Chen, K. H. Low, Y. Yao, and P. Jaillet, "Gaussian process decentralized data fusion and active sensing for spatiotemporal traffic modeling and prediction in mobility-on-demand systems," *IEEE ASE*, 2015.

[10] Z. Zhou, D. S. Matteson, D. B. Woodard, S. G. Henderson, and A. C. Micheas, "A spatio-temporal mixture model for point processes with application to ambulance demand," 2013.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. MIT Press, 2006.

[12] S. Garg, A. Singh, and F. Ramos, "Learning non-stationary space-time models for environmental monitoring," in *AAAI*, 2012.

[13] A. G. Wilson, C. Dann, and H. Nickisch, "Thoughts on massively scalable Gaussian processes," *arXiv preprint 1511.01870*, 2015.

[14] M. P. Deisenroth and J. W. Ng, "Distributed Gaussian processes," in *ICML*, 2015.

[15] J. Quiñonero Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, 2005.

[16] L. Csato, "Sparse on-line Gaussian processes," *Neur. Comp.*, 2002.

[17] J. W. Ng and M. P. Deisenroth, "Hierarchical mixture-of-experts model for large-scale Gaussian process regression," 2014. *arXiv* 1412.3078.

[18] Y. Cao and D. J. Fleet, "Generalized product of experts for automatic and principled fusion of Gaussian process predictions," 2014. *arXiv* 1410.7827.

[19] V. Tresp, "A Bayesian committee machine," *Neur. Comp.*, 2000.

[20] Z. W. Lim, D. Hsu, and W. S. Lee, "Adaptive informative path planning in metric spaces," in *WAFR*, 2015.

[21] A. Singh, A. Krause, C. Guestrin, W. Kaiser, and M. Batalin, "Efficient planning of informative paths for multiple robots," in *IJCAI*, 2007.

[22] L. Li, X. Zhang, and R. Piltner, "A spatiotemporal database for ozone in the conterminous us," in *(TIME'06)*, 2006.