Spatial Sensor Selection via Gaussian Markov Random Fields

Linh V. Nguyen, *Member, IEEE,* Sarath Kodagoda, *Member, IEEE* and Ravindra Ranasinghe, *Member, IEEE*

Abstract—The paper addresses the problem of selecting the most informative sensor locations out of all possible sensing positions in prediction of spatial phenomena by using a wireless sensor network. The spatial field is modelled by Gaussian Markov random fields (GMRF), where sparsity of the precision matrix enables the network to benefit from computation. A new spatial sensor selection criterion is proposed based on mutual information between random variables at a selected locations and those at unselected locations and interested but unlikely sensor placed positions, which enhances resulting prediction. The GMRF based optimality criterion is then proved to be very computationally efficiently resolved, especially in a large-scale sensor network, by a polynomial time approximation algorithm. More importantly, with demonstrations of monotonicity and submodularity properties of the mutual information set function in the proposed selection criterion, our near-optimal solution is also guaranteed by at least within $(1 - 1/e)$ of the optimal performance. The effectiveness of the proposed approach is compared and illustrated using two real-life large data sets with promising results.

Index Terms—Wireless sensor networks, sensor selection, spatial prediction, Gaussian Markov random fields.

Fig. 1: The predicted field of rainfall in Switzerland on 8 May 1986 using 467 measurements. X Coord and Y Coord are in km.

I. INTRODUCTION

Recently, technological developments in micro-electro-mechanical systems and wireless communications, which involve the substantial evolution in reducing the size and the cost of components, have led to the emergence of a wireless sensor network (WSN) [1], [2] that are increasingly useful in crucial applications in environmental monitoring [3]. Not only do these systems provide a virtual connection with the physical field in general, the WSN is particularly dominant in remote and hazardous environments where many essential phenomena have hardly been investigated due to their inaccessibility. In addition to collecting data, combining the measurements with a model, the WSN is also competent to estimate and predict the spatial phenomenon at unobserved locations. For instance, the posterior rainfall field in Switzerland is predicted by using 467 measurements collected from 467 recording stations on 8 May 1986, as illustrated in Fig. 1.

Typical task of a WSN consists of gathering measurements of a spatial field over a region of interest. However, multiple wireless sensor nodes co-located within the vicinity of a phenomenon in a dense WSN may generate similar data samples, which is called an *over-sampling* problem. This over-sampling problem has the potential to create a sizable
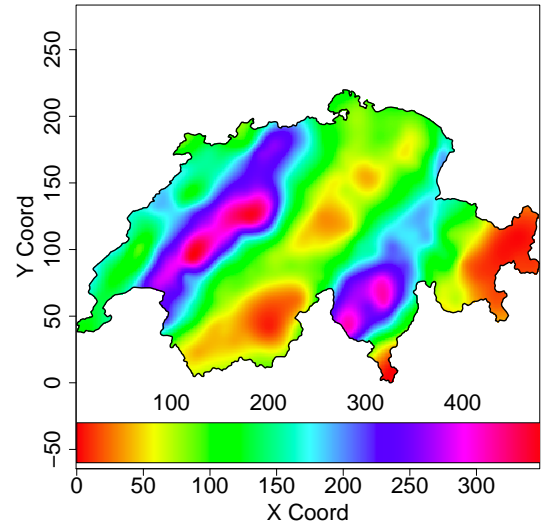
Linh V. Nguyen, Sarath Kodagoda and Ravindra Ranasinghe are with Centre for Autonomous Systems, University of Technology, Sydney, New South Wales 2007, Australia (e-mail: {vanlinh.nguyen, sarath.kodagoda, ravindra.ranasinghe}@uts.edu.au).

redundancy in sensed data that adversely affects good-put, i.e. effective usage to provide intelligence about the phenomenon, of the WSN as the redundant samples do not contribute to gain any additional information about the phenomenon. Furthermore, the over-sampling issue also causes data collection and analysis of long-term monitoring to be very expensive due to practical constraints like higher energy consumption, elevated maintenance burden, and increased computational cost. On the other hand, owing to resource constraints, the number of active sensors at one particular time is limited. Consequently, it is crucial to select the most informative subset of wireless sensor nodes out of all potential ones, which should participate in the sensing task. The selection procedure is called a *spatial sensor selection* problem in which the resulting prediction is required to conform requirement of the higher sensing quality in realistic applications. Additionally, the sensor selection problem enables the network to save energy since only the selected subset of sensors will be active in a particular sampling interval.

In statistics, selecting observations has been considered as an experimental design problem [4]–[7]. The design objective is to derive the deployment of sensing devices by the use of model uncertainty, which could be formulated by complicated statistical techniques. The optimality criteria are constructed based upon the properties of the inverse moment matrix.

For instance, D-optimality considers the determinant [8], A-optimality examines the trace and E-optimality calculates the maximum eigenvalue [5]. However, in the context of spatial prediction, the design objective frequently concerned is the quality of sensing, which is described as the accuracy of prediction or the uncertainty at unobserved locations of interest, after the observations are made. This requirement has been utilized to develop information-theoretic criteria [9]–[12]. The information-theoretic approaches such as entropy [13] or mutual information [14] were proposed to consider the prediction uncertainty of the random variables at unobserved locations in space, as demonstrated in the works [9], [15]–[18]. The primary principle of the entropy approach is to minimize the uncertainty of conditional entropy of unobserved locations, given measurements. As proposed by Ko *et al.* [16], a greedy yet near-optimal heuristic algorithm is used to maximize the joint entropy of random variables at selected sensor locations. Nonetheless, Guestrin *et al.* in in [17] have shown that the entropy based sensor selection approach causes sensed information waste since it tends to pick measuring locations along the border of interested space. Hence, to address the drawbacks of the entropy approach, the authors in [17] and then in [9], [18] have proposed a mutual information based sensor selection criterion. The premise behind the mutual information based approach is to maximize the mutual information between random variables at selected sensor locations and at an unselected sensor set. Up until now, all sensor selection optimality criteria have been proven to be a combinatorial NP-hard problem [16], [19]. Under assumption the modeling spatial field by Gaussian processes (GP) [20]–[22], a greedy but efficient algorithm can be utilized to obtain a near-optimal solution for the sensor selection problem.

In aforementioned existing works, the authors principally utilized the GP to statistically model the spatial fields. The specification of the GP is the covariance function that intuitively interprets the field properties. Nevertheless, the computational issues of the GP have always been a bottleneck, since the computational complexity of factorizing dense matrices is cubic in the dimension of observations. Therefore, even suppose that the spatial sensor selection problem is solved by a greedy heuristic algorithm, the GP based approaches are only reasonable in a small-scale sensor network but time consuming in a large-scale network.

Therefore, in this paper we first deal with the computational complexity associated with the GP regression under large-scale sensor networks in order to efficiently and effectively address the sensor selection problem in the spatial prediction. In the context of spatial statistics, the challenge in the GP computation has been addressed in [23]–[26] by replacing the GP by a computationally efficient Gaussian Markov random field (GMRF) [27], [28]. It is worthwhile to note that this discretely indexed Gaussian field (aka GMRF) can approximate the GP surprisingly well [23]. The computational benefits come from the sparsity of the precision matrix, an inverse of a dense covariance matrix, whose zero elements relate directly to conditional independence assumptions. The advantage of the precision matrix motivates us to propose a new spatial sensor selection optimality criterion based on the GMRF. The crite-rion is to maximize the mutual information between random variables collected at best selected sensor locations and those at unselected sensing positions and interested but unlikely placed sensor locations. Incorporating interested locations, where there is no availability of sensor deployment, allows the sensor networks to improve the accuracy of prediction results. The GMRF based mutual information maximization problem is resolved by an approximation algorithm, which is proved to be very computationally efficient. More importantly, by using properties of monotonicity and submodularity of the proposed mutual information set function in the sensor selection problem, the near-optimal solution obtained by our proposed algorithm is theoretically proven to be guaranteed by a lower bound, which is $(1 - 1/e)$ of the optimum. Our proposed approach is finally demonstrated to appealingly outperform the GP based method on two real-life large data sets.

Although the GMRF was used to design an adaptive sampling strategy for mobile sensor networks in [29], Xu *et al.* restricted their solutions to a regular lattice in which the hyperparameters were chosen *a priori* from a discrete support set. Hence, in this work we consider the GMRF model represented on an irregular lattice, where the model parameters are learned from all available observations. On the other hand, the sensor selection criterion based on the mutual information was studied in [9], where spatial field is modeled by the GP. Due to the bottleneck problem, the GP based methods are limited to small-scale networks. More importantly, the GP based mutual information is only monotonic under certain conditions [9], which leads to a resulting guarantee of $(1 - 1/e - \epsilon)$ as compared with the optimal performance for any $\epsilon > 0$. Nevertheless, by using the GMRF to model the spatial phenomena, we demonstrate that our approach not only deal with the sensor selection problem efficiently in large-scale networks but also enhances a better lower bound to $(1 - 1/e)$ of the optimal solution when the monotonicity of the GRMF based mutual information is comprehensively apparent.

The remainder of the paper is arranged as follows. Section II first introduces the concept of the GMRF, and then represents how to model the spatial field by using the GMRF. In Section III, the GMRF modeled spatial field is represented on a triangulation mesh by the use of stochastic partial differential equations (SPDE) approach, which is then utilized to formulate the mutual information based spatial sensor selection problem. This problem is solved by a near-optimal algorithm and its solution is proved to be bounded in Section IV. Section V illustrates the efficiency of our proposed approach on two real-world large data sets before Section VI concludes the paper.

II. PRELIMINARIES

In this section, we introduce Gaussian Markov random fields and a spatial field model, which are utilized as a basis for the spatial sensor selection proposed in this paper.

A. Notations

Let $\mathbb{R}$ denote the set of real numbers. The Euclidean distance function is defined by $\| \cdot \|$. We let $\mathcal{Q}$ be a convex

polytope in $\mathbb{R}^d$. A convex polytope is the convex hull of a finite point set. If two random variables x and y are independent, it is written as $x \perp y$. Let $cov(\cdot)$ denote the operator of the covariance and $\det(\cdot)$ denote the operator of the determinant. We let $\mathcal{O}(\cdot)$ denote the operator of running time of an algorithm. The inner product operator is defined by $\langle .,. \rangle$. $\bigtriangledown$ and $\log(\cdot)$ denote the gradient and the logarithm operators, respectively. In addition, $\mathbb{E}(\cdot)$ defines expectation operator, and $|\cdot|$ is used for a cardinality. Other notations will be explained as and when they occur.

B. Gaussian Markov Random Fields

A Gaussian Markov random field (GMRF) is a spatial process whose realizations contain random variables observed in a spatial field. Moreover, the GMRF not only follows Gaussian processes (GP) [21] but also satisfies additional conditional independence concepts, which is also known as Markovian property. For a more detailed discussion about GMRFs, readers are referred to [27].

Let $z = (z_1, ..., z_n)^T$ with $z \sim \mathcal{N}(\mathbf{m}, Q^{-1})$ denote a GMRF that is specified by a mean $\mathbf{m}$ and a symmetric and positive definite precision matrix Q, an inverse of the covariance matrix, Σ, of the GP. Therefore, the density of z has the form

$$p(z) = (2\pi)^{-\frac{n}{2}} (\det(Q))^{\frac{1}{2}} \exp\{-\frac{1}{2}(z - \mathbf{m})^T Q(z - \mathbf{m})\}.$$

The advantage of the Markovian property enables the GMRF to represent the conditional distributions for each component given all the others in terms of the neighbourhood structure of the process. In other words, the full conditional distribution of z_i $(i = 1, ..., n)$ is only dependent on the elements set of its neighbours and is designed by

$$p(z_i|z_{-i}) = p(z_i|z_{N_i}),$$

where z_{-i} denotes all the elements in z excluding z_i, and z_{N_i} is the neighbour elements of z_i. Thus, it is stated that given the neighbour elements, z_i is independent on any others in z except for z_{N_i}, which constitutes the conditional independence definition as

$$z_i \perp z_{-\{i,N_i\}}|z_{N_i},$$

for $i = 1, ..., n$. Rue *et al.* [27] derived that the mean $\mathbf{m}$ is not related to the pairwise conditional independence properties of z, so it results in that this characteristic must be specified only in the precision matrix Q.

In general, if z_i and z_j are conditionally independent,

$$z_i \perp z_j|z_{-\{i,j\}}$$

is equivalent to $Q_{ij} = 0$. This specification leads to $Q_{ij} \neq 0$ if $j \in \{i, N_i\}$ and derives the sparsity of the precision matrix Q, which benefits substantially in the computation.

C. Spatial Field Model via GMRF

In this subsection, we introduce the basic concepts and results on the spatial field model that are used in this paper. The discussion here follows the models represented in [22], [20].
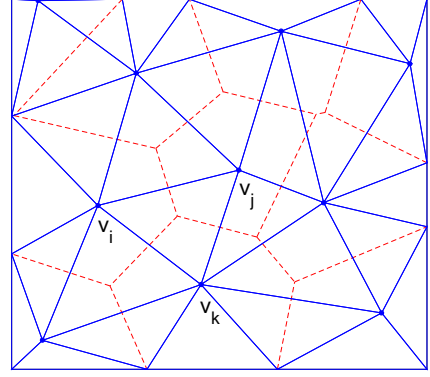

Fig. 2: The spatial field $\mathcal{Q} \subset \mathbb{R}^2$. Voronoi cells are partitioned by dash red lines and a mesh of a triangulation is created by solid blue lines.

We consider the spatial field of interest, $\mathcal{Q} \subset \mathbb{R}^d$, that is discretized into n Voronoi partitions [30] in which their centroids are n vertices of a mesh. In this special consideration, the mesh is represented as an irregular lattice of a triangulation as illustrated in Fig. 2. In a real application, a much denser mesh will be considered. We denote spatial locations at the vertices of the mesh as $\mathbf{v} = (v_1^T, v_2^T, ..., v_n^T)^T$ and a corresponding vector $y(\mathbf{v})$ denoted by $y(\mathbf{v}) = (y(v_1), y(v_2), ..., y(v_n))^T$ describes a vector of noisy measurements of the field. In this study, the random variables are modelled as a summation of a large scale component, a random field and an independent and identically distributed (i.i.d.) noise. The model is defined by

$$y(\mathbf{v}) = X(\mathbf{v})\beta + z(\mathbf{v}) + \varepsilon(\mathbf{v}), \tag{1}$$

where $X(\mathbf{v})\beta = \mathbb{E}(y(\mathbf{v}))$ is the expectation of $y(\mathbf{v})$. $X(\mathbf{v})$ denotes a $n \times p$ matrix of spatially referenced non-random variables (known as covariates) at location $\mathbf{v}$ and β is a p unknown column vector of mean parameters. z is a GMRF with a n zero mean vector and a $n \times n$ precision matrix Q. Note that the precision matrix consists of a small number of non-zero elements. Construction of Q will be considered in Section III-A. $\varepsilon(\mathbf{v})$ is a noise with a zero mean vector and a known covariance matrix $\sigma_\varepsilon^2 I$. Here I is a $n \times n$ identity matrix.

As discussed in [26], the GMRF can be constructed explicitly and accurately represents a Gaussian random field when the continuous domain SPDE has a solution of a stationary GP with a Matérn covariance function as given

$$cov(h) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}}(\kappa h)^\nu K_\nu(\kappa h),$$

where h denotes the Euclidean spatial distance $h = \parallel v_i - v_j \parallel$, σ^2 is the marginal variance, κ denotes the spatial scale parameter, ν is the Matérn smoothness, K_ν denotes the modified Bessel function of the second kind and order $\nu > 0$.

In this case, the term $X(v_i)\beta$ is referred to as the mean function of the GP.

III. Problem Statement

The spatial sensor selection focuses on finding the most informative subset of sensor locations out of all potential sensor positions, which can be utilized in monitoring, modelling and predicting spatial phenomena. It has been shown that addressing the sensor selection problem in the spatial field modelled by the GP is time consuming when the number of observations increases, especially in a large-scale sensor network. In this section, we propose a new computationally efficient sensor selection criterion for environmental monitoring applications. In this proposition, we first model the physical field by using Gaussian Markov random fields. The sparsity of the precision matrix then allows us to productively develop a mutual information based criterion.

A. Sensor Measurement Modelled by GMRF using SPDE Approach

When dealing with the spatial field, the stochastic partial differential equations (SPDE) approach proposed by Lindgren *et al.* [26] particularly illustrates a computational effectiveness. This method represents the random field models as solutions to the continuous domain SPDEs. The premise behind the SPDE approach is to represent a continuously indexed spatial process, e.g. GP, as a discretely indexed GMRF. In other words, this technique is to construct the GMRF with local neighbourhood and the sparse precision matrix Q to address the "big n problem" [31].

As shown in [26], the SPDE approach utilizes the Finite Element method [32] to project the SPDE onto a basis representation that consists of piece-wise linear basis functions described by a triangulation of the domain of interest. The triangulation is constructed by a set of non-intersecting triangles. Assume that a realization of the spatial process $z(s)$ is observed at N spatial locations $s = (s_1^T, s_2^T, ..., s_N^T)^T$; then the triangle initial vertices are imposed at these spatial locations. Moreover, for spatial prediction, additional vertices are added to complete a fully large triangulation.

Given the triangulation of the domain $\mathcal{Q}$, the GMRF model is built on the basis function representation

$$z(s) = \sum_{i=1}^{n} f_i(s) w_i, \tag{2}$$

where $\{f_i(s)\}$ represents the basis function, and $\{w_i\}$ represents Gaussian distributed weight. The functions $\{f_i(s)\}$ are piece-wise linear on each triangle, which denotes that $f_i(s)$ is 1 at the ith vertex of the mesh and 0 at all other vertices. w_i is the value of the spatial field at each triangle vertex i. Therefore, in the perspective of the continuous domain, the GP can be developed by the joint distribution of the weights $\{w_i\}$. It can be concluded that the finite element representation (2) enables not only the description of the continuous interpretation of the spatial field but also the implementation of the practical computations effectively. The SPDE method provides the explicit link between the GP and the GMRF that maps from the parameters of the covariance function to the elements of the precision matrix Q.

The precision matrix Q with the size of $n \times n$ is calculated by

$$Q = \tau^2(\kappa^4 C + 2\kappa^2 G + GC^{-1}G), \tag{3}$$

where C and G are the $n \times n$ matrices with $C_{ij} = \langle f_i, f_j \rangle$, $G_{ij} = \langle \bigtriangledown f_i, \bigtriangledown f_j \rangle$, and τ is utilized to control the variance. The dimension of the precision matrix is defined by the number of the vertices of the triangulation in the domain of interest. It can be clearly seen that Q expresses as a function of τ and κ. Here, we define $\theta = (\log(\tau), \log(\kappa))$ as a hyperparameter vector. We can say that the GMRF representation constructed by the linear basis functions holds the sparse property of the precision matrix.

Suppose that the latent random field at the n vertices of the triangulation is represented by a GMRF as

$$z|\theta \sim \mathcal{N}\left(\mathbf{0}, Q^{-1}\right). \tag{4}$$

In order to map between the basis function representation responded at n vertex locations of the triangulation mesh and the random field at sensor locations with a dimension of N, we let A denote a projector matrix that projects the latent random field modelled at the triangulation vertices to the data locations. The size of the projector matrix, A, is $N \times n$. Under the spatial field model as described in subsection II-C, the measurements at N locations are given

$$y|z, \theta, \beta, \sigma_\varepsilon^2, A \sim \mathcal{N}\left(X(s)\beta + Az, \sigma_\varepsilon^2 I_N\right), \tag{5}$$

where $X(s)$ denotes a $N \times p$ matrix of covariates, β, θ are parameters estimated by the maximum likelihood approach [22] from all the available observations, and I_N denotes the $N \times N$ identity matrix. If all model parameters are learned, the joint distribution of y and z, which is computed by employing the technique in [33] is specified by

$$z, y|\theta, \beta, \sigma_\varepsilon^2, A \sim \tag{6}$$

$$\sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ X(s)\beta \end{bmatrix}, \begin{bmatrix} Q^{-1} & Q^{-1}A^T \\ A^T Q^{-1} & \sigma_\varepsilon^2 I + AQ^{-1}A^T \end{bmatrix}\right).$$

In probabilistic terms, the full conditional distribution of z given y is also Gaussian, derived from (6) by using the block-wise inversion approach [34] and the Schur complement, as follows

$$z|y, \theta, \beta, \sigma_\varepsilon^2, A \sim \mathcal{N}\left(\mu_{z|y}, Q_{z|y}^{-1}\right), \tag{7}$$

where the vector of posterior means and the posterior precision matrix are given as

$$\mu_{z|y} = X(\mathbf{v})\beta + Q_{z|y}^{-1}A^T(\sigma_\varepsilon^2 I_N)^{-1}(y - X(s)\beta), \tag{8}$$

$$Q_{z|y} = Q + A^T(\sigma_\varepsilon^2 I_N)^{-1}A. \tag{9}$$

The primary computation of (8) is to factorize the sparse matrix $Q_{z|y}$, which can be obtained in a short time. On the other hand, $Q_{z|y}$ is not dependent on collection of measurements. This contrasts to the standard GP model [21], where computational complexity is cubic in dimension of dataset. And the GP-based computational issues become intractable with the increase in the number of observations.

B. Sensor Selection Problem

Now we describe the sensor selection problem in terms of optimizing the spatial field monitoring. In other words, we must deploy a limited number of sensors so that the spatial field at unobserved locations can be effectively and efficiently predicted. To the best of our knowledge, the most efficient criterion for the sensor selection problem up to now is based on mutual information [15], [9]. In information theory, mutual information (MI) is a quantity that measures the dependency between two random variables [35]–[37]. That is, it measures how much information is acknowledged in one random variable about another. In a learning task of a selection, the MI between two physical quantities can be utilized to reflect the amount of information redundancy between them. Therefore, when the mutual dependence between two random variables is high, removing one of them from a selected set would not have much effect on the solution.

As discussed in the works of Caselton *et al.* [15] and Krause *et al.* [9], the sensor selection criterion is formulated so that the MI between random variables gathered by selected subset of sensors and those at unselected sensor locations is maximized. Let us consider a two dimensional sensor network that has a set of possible sensor locations S, which provide point measurements of some physical quantities. Sensor selection addresses the problem of choosing a subset, $C \subset S$, at which the subset of the measurements can closely represent the distribution of the physical quantity in the whole space. It is then given the MI between the selected subset C and the rest of the space $S \setminus C$ as follows

$$MI(C, S \setminus C) = H(S \setminus C) - H(S \setminus C | C), \quad (10)$$

where $H(S \setminus C)$ and $H(S \setminus C | C)$ are the entropy and the conditional entropy of random variables $y_{S \setminus C}$ and $y_{S \setminus C | C}$ at unmeasured locations $S \setminus C$ before and after deploying sensors at the locations C, respectively. If the spatial field is modelled by Gaussian processes, $MI(C, S \setminus C)$ can be computed by

$$MI(C, S \setminus C) = \frac{1}{2} \{ \log \det(\Sigma_{S \setminus C}) - \log \det(\Sigma_{S \setminus C | C}) \}, \quad (11)$$

where $\Sigma_{S \setminus C}$ and $\Sigma_{S \setminus C | C}$ are covariance matrices of $y_{S \setminus C}$ and $y_{S \setminus C | C}$, respectively.

It can be clearly seen that computing the mutual information $MI(C, S \setminus C)$ is dependent on a number of the possible sensor locations. Since cost of logarithm of determinant of a covariance matrix is cubic in dimension of the data set, if the objective is to select the k best locations out of n possible sensor positions, a greedy heuristic algorithm can near-optimally resolve the objective function (11) in $\mathcal{O}(kn^4)$. As a result, addressing the sensor selection problem in a large-scale sensor network, though by a polynomial technique, is really time consuming. For instance, in the typical example of the Swiss rainfall data set [38], there are 467 possible gauge positions which facilitate placing rainfall sensors. Selecting the 100 best gauge locations out of these 467 positions by greedily adding sensors in sequence to maximize the MI introduced above in a modern computer (refer to Section V for more details) takes more than 6 hours.

On the other hand, if it is required to effectively improve resulting prediction quality at locations of interest, where for some reasons we cannot deploy any sensors in (for example, the locations are out of wireless communication range of a network, or the network is short of energy), it seemingly shows that the previous works have not embedded these interested positions in the sensor selection criterion. Therefore, in this paper, we propose a new spatial sensor selection criterion that computes the mutual information between the variables at selected sensor locations and those at a set of unselected sensor locations and unmeasured yet interested locations which are unlikely within sensor placements.

We now consider a spatial space of interest as a discrete set of locations $V = S \cup U$, where S is a set of possible sensor locations in which sensors can be deployed to observe the physical field, and U is a set of interested locations in which no sensor placements are available. Selecting the best subset C with k locations out of S via Gaussian Markov random fields is described as follows. We first create a triangulation mesh whose vertices consist of all locations $V \setminus C$. If z is defined a vector of latent random field at the n mesh vertices, we then compute the mutual information between z and y_C, $MI(z, y_C)$. Hence, the goal of the new spatial sensor selection criterion is to find the best measurements correspondingly gathered at the optimal subset C so that $MI(z, y_C)$ is maximized. This leads to most efficiently predict the field in whole space. Mathematically, we formulate the spatial sensor selection problem as follows.

$$C^{opt} = \underset{\substack{C \subseteq S \\ |C| = k}}{\textbf{argmax}} \quad MI(z, y_C), \quad (12)$$

where $MI(z, y_C) = H(z) - H(z | y_C)$ in which $H(z)$ and $H(z | y_C)$ are the entropy and the conditional entropy of z and $z | y_C$, respectively.

As discussed in the previous subsection, the entropy of Gaussian Markov random variables z and $z | y_C$ can be obtained as

$$H(z) = \frac{1}{2} \log \left((2\pi e)^n \det(Q^{-1}) \right) = -\frac{1}{2} \log \left((2\pi e)^n \det(Q) \right),$$

$$H(z | y_C) = \frac{1}{2} \log \left((2\pi e)^n \det(Q_{z|y_C}^{-1}) \right)$$
$$= -\frac{1}{2} \log \left((2\pi e)^n \det(Q + A_k^T (\sigma_\varepsilon^2 I_k)^{-1} A_k) \right),$$

where A_k is a $k \times n$ projector matrix that maps between the triangulation mesh vertices and the selected locations C, I_k is a $k \times k$ identity matrix. Therefore, (12) is equivalent to

$$C^{opt} = \underset{\substack{C \subseteq S \\ |C| = k}}{\textbf{argmax}} \quad \log \det(Q + A_k^T (\sigma_\varepsilon^2 I_k)^{-1} A_k) - \log \det(Q).$$
$$(13)$$

To select the best subset C out of the possible set S in order to maximize $MI(z, y_C)$ is a combinatorial NP-hard problem [16], [19]. In terms of the global optimization techniques, branch and bound method [39] was employed to resolve the kind of this problem [40]. Nevertheless, this accurate approach

is often computationally intensive [41], even with modest values of n and k, and not attractive in real world solutions. Thus, one of very often chosen methods to effectively solve the NP-hard optimization problem is the greedy algorithm, which will be discussed in the next section.

It can be obviously seen that computing the mutual information based on the GP (11) is subject to the number of the possible sensor locations while since the precision matrix Q is constructed on the triangulation mesh vertices, computing the GMRF based mutual information in (12) is free from the number of observations. The new spatial sensor selection criterion (13) allows the sensor network not only to incorporate all possible sensor locations and interested locations into one context but also to gain significant benefits in computation due to the use of the sparse property of the precision matrix.

IV. NEAR-OPTIMAL SOLUTION

Addressing the combinatorial NP-hard optimization problem (13) in polynomial time can only be achieved by a greedy algorithm. In this section, we first describe the algorithm in the context of the spatial sensor selection. A bound is then presented to guarantee the proposed near-optimal solution.

A. Algorithm

In order to find the near-optimal informative subset of sensor locations C^* from all possible locations set S, the proposed algorithm starts from an empty set $C^* = \oslash$ which means that the corresponding vector of the random variables y_{C^*} is also empty. We first choose one location $s_i \in S$ and its corresponding measurement y_{s_i} and then compute

$$MI(z, y_{s_i}) = \log \det(Q + \frac{1}{\sigma_\varepsilon^2} A_1^T A_1) - \log \det(Q),$$

where A_1 is a n row vector that projects the triangulation mesh vertices to s_i. Select the maximum value from the sequential computations of $MI(z, y_{s_i})$, we equivalently obtain the first near-optimal sensor locations s_i^* and the corresponding measurement $y_{s_i^*}$. Both s_i^* and $y_{s_i^*}$ are then added to C^* and y_{C^*}, respectively; and s_i^* is removed from S. Repeatedly, we now run this iteration of the algorithm with given non-empty sets of C^* and y_{C^*}. In other words, each chosen y_{s_i} at this stage is temporarily added to y_{C^*} to calculate

$$MI(z, y_{C^*}) = \log \det(Q + A_m^T (\sigma_\varepsilon^2 I_m)^{-1} A_m) - \log \det(Q),$$

where m is the size of the temporary y_{C^*} in the current iteration. At each iteration, the solutions s_i^* and $y_{s_i^*}$, corresponding to the maximum $MI(z, y_{C^*})$, are sequentially added to C^* and y_{C^*}, respectively. Similarly, s_i^* is removed from S. The algorithm stops when $m = k$, or the number of selected sensor locations is reached. As a consequence, the cost of the polynomial approximation algorithm to near-optimally address (13) can be described by the following theorem.

Theorem 1: The approximate solution of the problem (13) can be obtained by a greedy algorithm in time $\mathcal{O}(kNn^{\frac{3}{2}})$, where N is a number of possible sensor locations S.

Proof: Note that the size of Q is $n \times n$. As discussed in [42],

$$\log \det(Q) = 2 \sum_{i=1}^{n} \log(L_{ii}),$$

where $Q = LL^T$. This Cholesky decomposition is calculated in time $\mathcal{O}(n^{\frac{3}{2}})$ in two dimensional space. Since size of S is N, in order to find one best locations out of all possible sensor positions, the greedy algorithm needs running such logarithm of determinant in $\mathcal{O}(N)$ times. Therefore, finding the k best sensor locations requires $\mathcal{O}(k)$ iterations. So, the optimization problem (13) can be finalized in running time $\mathcal{O}(kNn^{\frac{3}{2}})$. ∎

The proposed greedy algorithm for finding approximately maximum informative subset of sensor locations is summarized in Algorithm 1.

Algorithm 1 Near-optimal algorithm for maximizing the GMRF based mutual information

Input:
1) Set of potential sensor locations S
2) Set of interested locations U
3) Number of selected sensor locations k
4) Hyperparameters θ and σ_ε^2

Output:
1) Near-optimal set of sensor locations C^*

At beginning, do
$C^* \leftarrow \oslash$
$y_{C^*} \leftarrow \oslash$

1: **for** i = 1 to length of S **do**
2:      $s_i \in S$
3:      find the corresponding random variable at s_i, y_{s_i}
4:      $MI(z, y_{s_i}) = \log \det(Q + \frac{1}{\sigma_\varepsilon^2} A_1^T A_1) - \log \det(Q)$
5: **end for**
6: $s_i^* \leftarrow \underset{s_i \in S}{\operatorname{argmax}} MI(z, y_{s_i})$
7: $y_{s_i^*} \leftarrow \underset{s_i \in S}{\operatorname{argmax}} MI(z, y_{s_i})$
8: $C^* \leftarrow C^* \cup s_i^*$
9: $y_{C^*} \leftarrow y_{C^*} \cup y_{s_i^*}$
10: $S \leftarrow S \setminus s_i^*$
11: **while** $|C^*| \leq k$ **do**
12:      **for** i = 1 to length of S **do**
13:          $s_i \in S$
14:          C^* temporarily receives s_i
15:          y_{C^*} temporarily receives y_{s_i}
16:          $MI(z, y_{C^*}) = \log \det(Q + A_m^T (\sigma_\varepsilon^2 I_m)^{-1} A_m) - \log \det(Q)$
17:      **end for**
18:      $s_i^* \leftarrow \underset{s_i \in S}{\operatorname{argmax}} MI(z, y_{C^*})$
19:      $y_{s_i^*} \leftarrow \underset{s_i \in S}{\operatorname{argmax}} MI(z, y_{C^*})$
20:      $C^* \leftarrow C^* \cup s_i^*$
21:      $y_{C^*} \leftarrow y_{C^*} \cup y_{s_i^*}$
22:      $S \leftarrow S \setminus s_i^*$
23: **end while**

B. A Solution Bound

As previously discussed, the solution of the spatial sensor selection problem (13) obtained by the algorithm proposed in Subsection IV-A is near-optimal. Although it has been shown that the greedy algorithm is very effective to solving a combinatorial NP-hard issue in spatial prediction using mobile robotic wireless sensor networks [43], a theoretical bound on the approximated solution of the spatial sensor selection problem is paramount. Mathematically, Nemhauser *et al.* in their work [44] presented that the solution obtained by a greedy heuristic algorithm can be guaranteed by a specific level of the optimum if the objective optimization function is monotonic and submodular. In this subsection, we first prove that the set function

$$F(C) = MI(z, y_C) \tag{14}$$

holds both the monotonicity and submodularity properties. And, we then present a lower bound of our near-optimal solution of the objective optimization problem (13).

In the context of information theory, it is apparent that adding one measurement to a small set of observations provides more information to the sensor network than adding it to a large set. In other words, as described in [9], the mutual information holds a property of diminishing returns, or the function $F(C)$ is submodular.

On the other hand, it has been presented that the GP based mutual information is not totally monotone [9], [43]. Nonetheless, by representing the mutual information via the GMRF, we prove that the proposed function $F(C)$ satisfies monotonicity as following lemma.

Lemma 2: The set function $C \longmapsto MI(z, y_C)$ with

$$MI(z, y_C) = \log \det(Q + A_k^T (\sigma_\varepsilon^2 I_k)^{-1} A_k) - \log \det(Q)$$

is monotonic.

Proof: Suppose that we have two subsets C_1 and C_2 selected from the potential sensor locations S. Here cardinalities of C_1 and C_2 are k and m, respectively. Let A denote a $k \times n$ projector matrix that maps between the selected sensor locations C_1 and the triangulation mesh vertices. Similarly, a matrix B with a dimension of $m \times n$ projects locations C_2 to all mesh vertices. It is straightforward to have

$$MI(z, y_{C_1}) = \log \det(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A) - \log \det(Q).$$

If we define $C = C_1 \cup C_2$, then

$$MI(z, y_C) = \log \det \left(Q + \begin{bmatrix} A^T & B^T \end{bmatrix} (\sigma_\varepsilon^2 I_{k+m})^{-1} \begin{bmatrix} A \\ B \end{bmatrix} \right) - \log \det(Q),$$

where I_{k+m} is a $(k+m) \times (k+m)$ identity matrix.

In order to prove the monotonicity of the mutual information in (12), it is required to show that $F(C_1 \cup C_2) \geq F(C_1)$. That is, $MI(z, y_C) - MI(z, y_{C_1}) \geq 0$. Let us define

$$\Delta I = MI(z, y_C) - MI(z, y_{C_1}).$$

It can be clearly seen that

$$\Delta I = \log \det(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A + B^T (\sigma_\varepsilon^2 I_m)^{-1} B)$$
$$- \log \det(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A).$$

By the use of the matrix determinant lemma, we have

$$\log \det(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A + B^T (\sigma_\varepsilon^2 I_m)^{-1} B) =$$
$$= \log \det(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)$$
$$+ \log \det \left(I_m + \frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T \right),$$

where I_m is a $m \times m$ identify matrix. Therefore,

$$\Delta I = \log \det \left(I_m + \frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T \right).$$

Since Q and $(\sigma_\varepsilon^2 I_k)^{-1}$ are positive semidefinite, both $A^T (\sigma_\varepsilon^2 I_k)^{-1} A$ and $Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A$ are also positive semidefinite. Then, $\frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T$ is positive semidefinite. As represented by Horn *et al.* in their work [45], if $\{\lambda\}_1^m$ are eigenvalues of the matrix $\frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T$, then

$$\det \left(I_m + \frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T \right) = \prod_{i=1}^m (1 + \lambda_i)$$
$$\geq 1 + \prod_{i=1}^m \lambda_i.$$

As a result,

$$\Delta I \geq \log(1 + \prod_{i=1}^m \lambda_i).$$

Because the positive semidefinite property of the matrix $\frac{1}{\sigma_\varepsilon^2} B(Q + A^T (\sigma_\varepsilon^2 I_k)^{-1} A)^{-1} B^T$,

$$\prod_{i=1}^m \lambda_i \geq 0.$$

This leads to

$$\Delta I \geq 0,$$

which completes the proof. ∎

That the GMRF based mutual information function in (12) and (13) satisfies properties of submodularity and monotonicity allows the near-optimal solution of the spatial sensor selection problem (13), which is resolved by a greedy algorithm, to be bounded. Using the fundamental results in [44], we can say that our near-maximum mutual information solution, which is a basis to select the most informative spatial sensor locations, is guaranteed by a lower bound within $(1 - 1/e)$ of the optimum. In equivalent words,

$$F(C^*) \geq \left(1 - \frac{1}{e} \right) F(C^{opt}).$$

This demonstrates that approximate solutions in the spatial sensor selection with the mutual information criterion representing via the GMRF is better bounded when compared with those in the problem representing via the GP. For instance, Krause *et al.* in [9] show that the approximation algorithm can provide a resulting guarantee of $(1 - 1/e - \epsilon)$ as compared with the optimal solution for any $\epsilon > 0$.

Notice that an upper bound can be found as $(e/(e - 1)) F(C^*)$ for any other solutions obtained by other methods.

V. EXPERIMENTAL RESULTS

In this section, we provide experimental performances of our proposed approach on selecting the best spatial sensors out of all potential measuring locations in two real-world data sets. In both real-life data sets, they had utilized two large-scale sensor networks to take daily rainfall measurements in Switzerland [38] and average winter rainfall in Paraná State, Brazil [46], respectively. Illustrations of the results in this section demonstrate the effectiveness of our method to efficiently employ a sensor network in real environmental monitoring applications.

Fig. 3a illustrates 467 recording stations (in dot points) in Switzerland to gauge the rainfall on 8 May 1986. In this demonstration, the rainfall values at the sensor locations are also shown by the different levels of grey colours of the dots. Note that the unit of the rainfall measurements is $1/10$ *mm*. Likewise, 143 rainfall measuring stations (in triangles) were deployed in Paraná State, Brazil as shown in Fig. 3b. In addition, the average values of winter rainfall in this area of Brazil is represented by the grey colours of the triangles.

As can be seen in Fig. 3, the rainfall values range in Switzerland data set from 0.5 to 585 and in Paraná data set from 163 to 414. Therefore, the Gaussian processes can hardly model these drastically-changed data sets. In other words, the rainfall fields in the illustrative experiments are non-Gaussian. In these scenarios, we first transform the data sets into spatially Gaussian fields before implementing our proposed approach. One of very frequently used classes of transformations is the Box-Cox family [47]. If we assume that a random variable at measurement stage is y, then the transformed field is given

$$\overline{y} = \begin{cases} \frac{y^{\alpha}-1}{\alpha}, & \alpha \neq 0 \\ \log y, & \alpha = 0 \end{cases}$$

where α is a transformation parameter, which is chosen within the Box-Cox class. In this work, we chose $\alpha = 0.5$.

As introduced in Section III, in order to represent the rainfall spatial fields by the use of the GMRF, we need to construct a triangulation lattice mesh for the whole space. In this implementation, the lattice of triangles is created by the INLA package [48], installed in R. Since in our proposed method all possible sensor locations and locations of interest are imposed on the mesh vertices, the number of the lattice vertices is dependent on size of a sensor network and total of interested locations. For instance, in the example of Swiss rainfall data set, a mesh with 467 sensor positions and 200 unmeasured and randomly chosen locations in the space is sized to 803 vertices. Note that some vertices need to be added to construct a proper triangulation. One can also customize the mesh by setting the maximum allowed triangle edge length in the INLA package; however, this may cause increase in number of the mesh vertices. As analysed previously, cost of computing our proposed mutual information criterion is mainly subject to the size of the precision matrix Q that is created based on triangulation mesh vertices. Thus, increasing the size of the triangulation mesh definitely results in increment of computation time, as illustrated in Table I. In this analysis, we examine how choosing the mesh of triangles affect computational time and results. The implementations

TABLE I: COMPARISON OF RUNNING TIME AT DIFFERENT SIZES OF TRIANGULATION MESHES IN SWISS RAINFALL DATASET

Running time (in seconds) of GMRF based MI criteria on meshes			
Mesh 1	Mesh 2	Mesh 3	Mesh 4
2641	3271	3992	9490

were conducted on R V3.0 with a PC of 3.1GHz Intel Core i5-2400 Processor. It is assumed that the noise variances were estimated by the GP method, which are 6.0 and 1.5 for Swiss and Paraná rainfall data sets, respectively.

In these experiments, for each dataset case, we created four triangulation meshes with different sizes, representing the rainfall physical fields. In Swiss rainfall, four meshes, named mesh 1, mesh 2, mesh 3 and mesh 4, are sized to 803, 1017, 1309 and 2370 vertices, respectively. Similarly, in Paraná rainfall, four meshes with similar names to those of Swiss rainfall are sized to 444, 627, 766 and 1177 vertices, respectively. Notice that we have only 143 possible locations for measuring rainfall in Paraná.

Corresponding to each studied case, from four triangulation meshes, running our proposed algorithm 1, for each mesh, we selected the best subset of sensor locations out of 467 in Switzerland and 143 in Paraná possible recording positions. For the purpose of comparison, one best sensor subset in each real-life application was also chosen by using the GP based mutual information criterion [15], [9]. For instance, subsets of 100 best sensing locations from 467 possible positions in the Swiss data set and subsets of 30 best sensing locations from 143 possible stations in the Paraná data set were selected by our proposed approach and the method in [15], [9]. Figures 4 and 5 demonstrate the results of mutual information and root mean square errors obtained by the selected sensor subsets.

In Fig. 4a, the resulting mutual information values between the rainfall values at 100 selected sensor location subsets and those at the rest of the space, including unselected sensor locations and interested locations, were computed. There are 8 MI curves in this figure. It can be seen that on each created triangulation mesh two corresponding MI curves were computed by the use of 100 best sensor subsets, which were selected by our proposed method and the GP based technique. The Paraná results are illustrated in the same way, but note that the selected sensor subsets are sized 30. In overall, the MI results obtained by our approach outperforms those obtained by the GP based method. This accounts for why we incorporate the interested but unlikely sensor placed locations into the selection criterion. More importantly, Fig. 4 shows that the MI values with respect to each approach mostly does not depend on the size of the triangulation meshes. However, if we look at Table I, the running time of our algorithm is nearly proportional to number of the lattice vertices. As a consequence, in our spatial sensor selection criterion, the mesh of triangles only need to be designed as simple as it is enough to include all possible sensor locations and interested locations.

In terms of prediction quality, we consider the root mean square errors (RMSE), which are shown in Fig. 5. In this
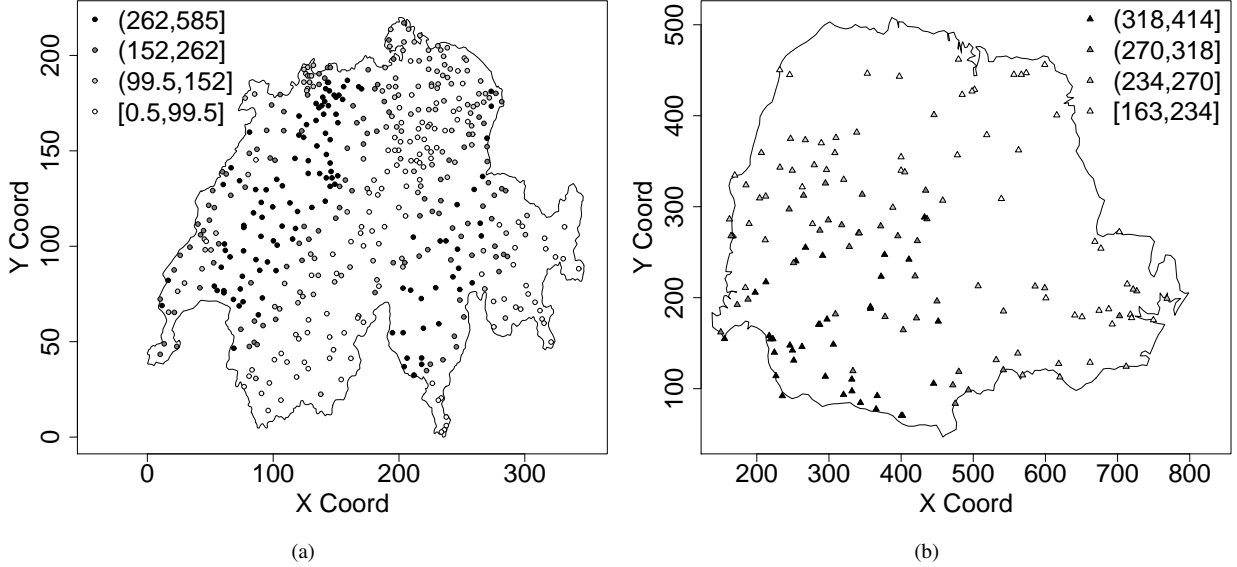
Fig. 3: Maps of Switzerland (a) and Paraná State, Brazil (b) illustrating the locations of the sensor stations and the corresponding logged rainfall values. X Coord and Y Coord are in km.

TABLE II: COMPARISON OF RUNNING TIME AT SELECTING DIFFERENT NUMBERS OF BEST SENSORS FROM 467 POSSIBLE LOCATIONS IN SWISS RAINFALL DATASET

Criterion	Running time (in seconds) of selecting ns best sensors from 467 possible locations				
	$ns = 100$	$ns = 75$	$ns = 50$	$ns = 25$	$ns = 1$
GP based MI	22642	18637	13665	7443	325
GMRF based MI	2641	2118	1394	740	31

calculation, the measurements at the best selected sensor locations were first utilized to model the spatial rainfall fields. The model was then used to predict the field at unselected sensing locations. In other words, the measurements at unselected sensor locations were employed for the purpose of validation. For instance, in the Swiss rainfall data set, if 100 sensors were chosen from 467 possible locations, then the data at 367 other sensing stations were used to compute the RMSE. As mentioned previously, we have four selected sensor subsets at four triangle meshes created in each consideration. Therefore, we computed four RMSE value curves in Fig. 5a for Swiss rainfall and in Fig. 5b for Paraná rainfall. It can be apparently seen that these curves in each figure, 5a and 5b respectively, are mostly close. In equivalent words, the resulting prediction quality is not subject to choosing the triangulation lattice as well. More interestingly, we also computed the RMSE where the best sensors were selected by the GP based MI criterion, which are illustrated by the blue curves in Fig. 5. This illustrative comparison shows that the prediction in terms of the RMSE values obtained by our method and the previous technique are fairly similar. Nonetheless, if we have a look at Table II where we compare the computing time of our proposed algorithm and the GP based algorithm in [9] in selecting the best sensors out of 467 possible sensing locations in Switzerland, the proposed approach based on the GMRF gains significant benefit in computation. There are five particular numbers of selected sensor locations are utilized in

this comparison. Table II shows that our technique improves 10 times faster than the GP based method.

One pivotal advantage in our proposed spatial sensor selection criterion as compared with the criterion based on the GP is that the computing time of finding the greedy algorithm based solution is not much affected by the size of a sensor network. For instance, in the experiment of the Swiss rainfall data set, for the purpose of comparison, we selected 30 best sensing locations from a small-scale and a large-scale networks of possible sensor locations, respectively. As results are shown in Table III, 30 best sensors were first chosen from 100 potential measuring locations, and then they were selected from 467 possible positions, by both criteria based on the GP and the GMRF. The running time of both approaches in the small-scale network of 100 sensing locations is practically reasonable, while those in the large-scale network of 467 measuring positions are much different. Our proposed method took around 14 minutes to find the best subset of sensor locations, meanwhile, the GP based approach required more than 2.3 hours to obtain the solution. Notice that, as discussed previously, the prediction results, for example the RMSE, obtained by using the selected subsets in both methods are highly comparable. Therefore, our approach is considerably advantageous to the large-scale sensor network in terms of computation time of the spatial sensor selection.

For the purpose of comparing the predicted fields in the whole space, including unselected sensor locations and inter-
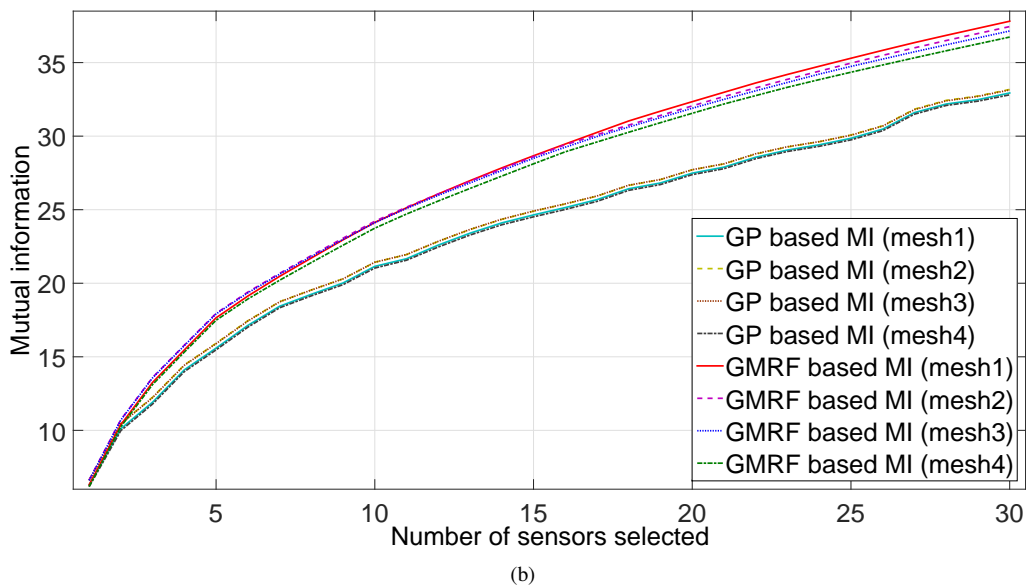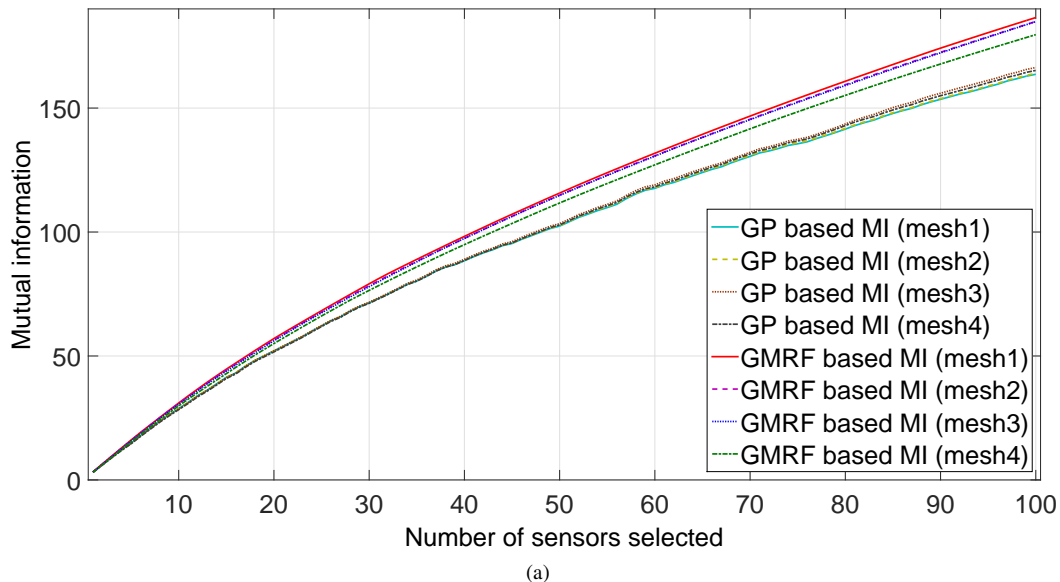
Fig. 4: Mutual information between random variables at selected sensor subsets and at the rest of the space, corresponding to rainfall Swiss (a) and Paraná (b) data sets, respectively.

TABLE III: COMPARISON OF RUNNING TIME AT DIFFERENT NUMBERS OF POSSIBLE SENSOR LOCATIONS IN SWISS RAINFALL DATASET

Criterion	Running time (in seconds) of selecting 30 best sensor locations from	
	100 possible locations	467 possible locations
GP based MI	36	8394
GMRF based MI	106	884

ested but unlikely deployed locations, we implemented the GP regression and the GMRF posterior prediction as described in Subsection III-A. Figures 6 and 7 demonstrate the posterior means and posterior variances of the rainfall on 8 May 1986 in the whole Switzerland, using 10, 50 and 100 sensor locations and measurements, respectively, selected out of 467 possible sensor positions and measurements. Fig. 7 also illustrates the locations of the sensors selected by two methods. The left columns in these figures were obtained by the GMRF posterior prediction technique given particular 10, 50 and 100 sensing locations and measurements selected by our proposed approach. Similarly, the right columns in these figures are results of the GP regression, utilizing the sensor locations and measurements chosen by the GP based MI criterion [15], [9]. It can be apparently seen that while the predicted rainfall fields obtained by the use of 50 and 100 chosen sensor locations
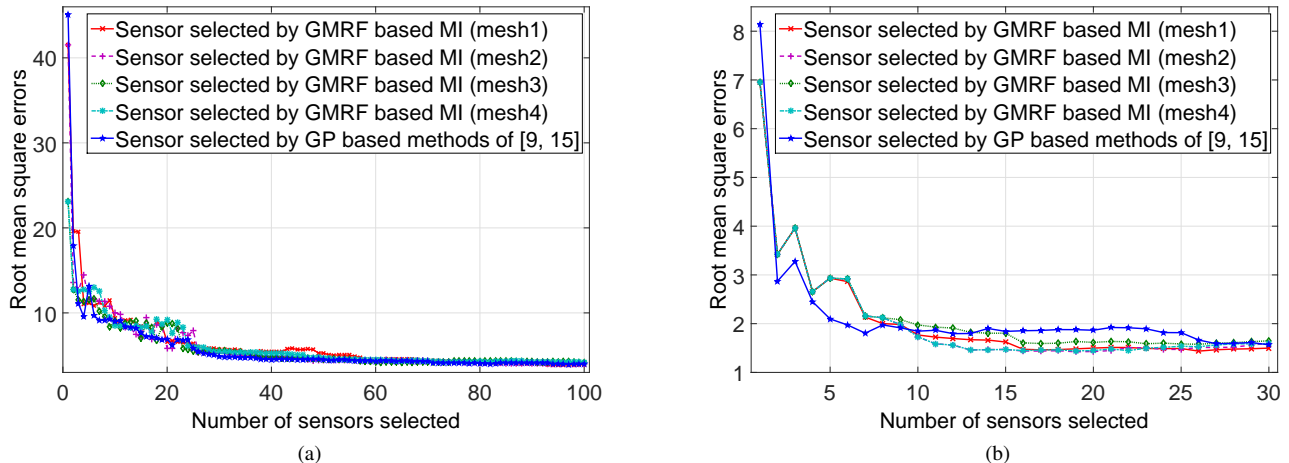
Fig. 5: Root mean square errors computed at validation locations, corresponding to rainfall Swiss (a) and Paraná (b) data sets, respectively.

and measurements as shown in Figures 6c, 6d, 6e and 6f are highly comparable to the predicted field obtained by using 467 locations and measurements, illustrated in Fig. 1, the predicted fields obtained from 10 selected locations and measurements in Figures 6a and 6b are visibly different. The values of the predicted rainfall obtained by our method range around from 50 to 550 (one tenth mm), whereas those obtained by the GP based approach are around 200 (one tenth mm). This difference is accounted for that our 10 selected measurements have full information representing entire space, meanwhile 10 measurements chosen by the GP based technique are about 200 (one tenth mm). Furthermore, the range of the posterior rainfall variances in the entire Switzerland obtained by our proposed GMRF based MI criterion in the left column of Fig. 7 is lower than that obtained by the GP based MI criterion as shown in the right column of the same figure. This is a result of that our proposed method presents the interested locations in the spatial sensor selection criterion.

VI. CONCLUSION

This paper proposed a new approach to select the most informative spatial sensor locations in monitoring the spatial phenomena that are modelled by the Gaussian Markov random fields. By maximizing the MI between random variables at selected locations and those at potential yet unselected locations and interested but impossibly deployed locations, the GMRF based MI criterion improves the quality of prediction. Despite its NP-hard completeness, the spatial sensor selection problem is computationally efficiently solved by a greedy algorithm. Particularly, in terms of computational time, the proposed approach outperforms the GP based method in a large-scale sensor network. The GMRF based sensor selection solution is also proved to be bounded by at least a $(1 - 1/e)$ level of the optimum. The performance of our method was evaluated and compared with that of the GP based technique in two real-world data sets.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.

[2] A. Rogers, E. David, and N. R. Jennings, "Self-organized routing for wireless microsensor networks," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35(3), pp. 349–359, 2005.

[3] P. Corke, T. Wark, R. Jurdak, W. Hu, P. Valencia, and D. Moore, "Environmental wireless sensor networks," *Proceedings of the IEEE*, vol. 98(11), pp. 1903–1917, 2010.

[4] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[5] F. Pukelsheim, *Optimal design of experiments.* Society for Industrial and Applied Mathematics, 2006.

[6] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Science*, vol. 10(3), pp. 273–304, 1995.

[7] P. Müller, B. Sanso, and M. D. Iorio, "Optimal Bayesian design by inhomogeneous Markov chain simulation," *Journal of the American Statistical Association*, vol. 99(467), pp. 788–798, 2004.

[8] N. Nguyen and A. J. Miller, "A review of some exchange algorithms for constructing discrete D-optimal designs," *Computational Statistics and Data Analysis*, vol. 14(4), pp. 489–498, 1992.

[9] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithm and empirical studies," *Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[10] H. Choi, J. P. How, and P. I. Barton, "An outer-approximation approach for information-maximizing sensor selection," *Optimization Letter*, vol. 7, pp. 745–764, 2013.

[11] M. S. Uddin, A. Kuh, A. Kavcic, and T. Tanaka, "Nested performance bounds and approximate solutions for the sensor placement problem," *APSIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1–13, 2014.

[12] J. Yun and J. Kim, "Deployment support for sensor networks in indoor climate monitoring," *International Journal of Distributed Sensor Networks*, pp. 1–10, 2013.

[13] P. Sebastiani and H. P. Wynn, "Maximum entropy sampling and optimal bayesian experimental design," *Journal of the Royal Statistical Society*, vol. 62, pp. 145–157, 2000.

[14] L. Paninski, "Asymptotic theory of information-theoretic experimental design," *Neural Computation*, vol. 17, pp. 1480–1507, 2005.
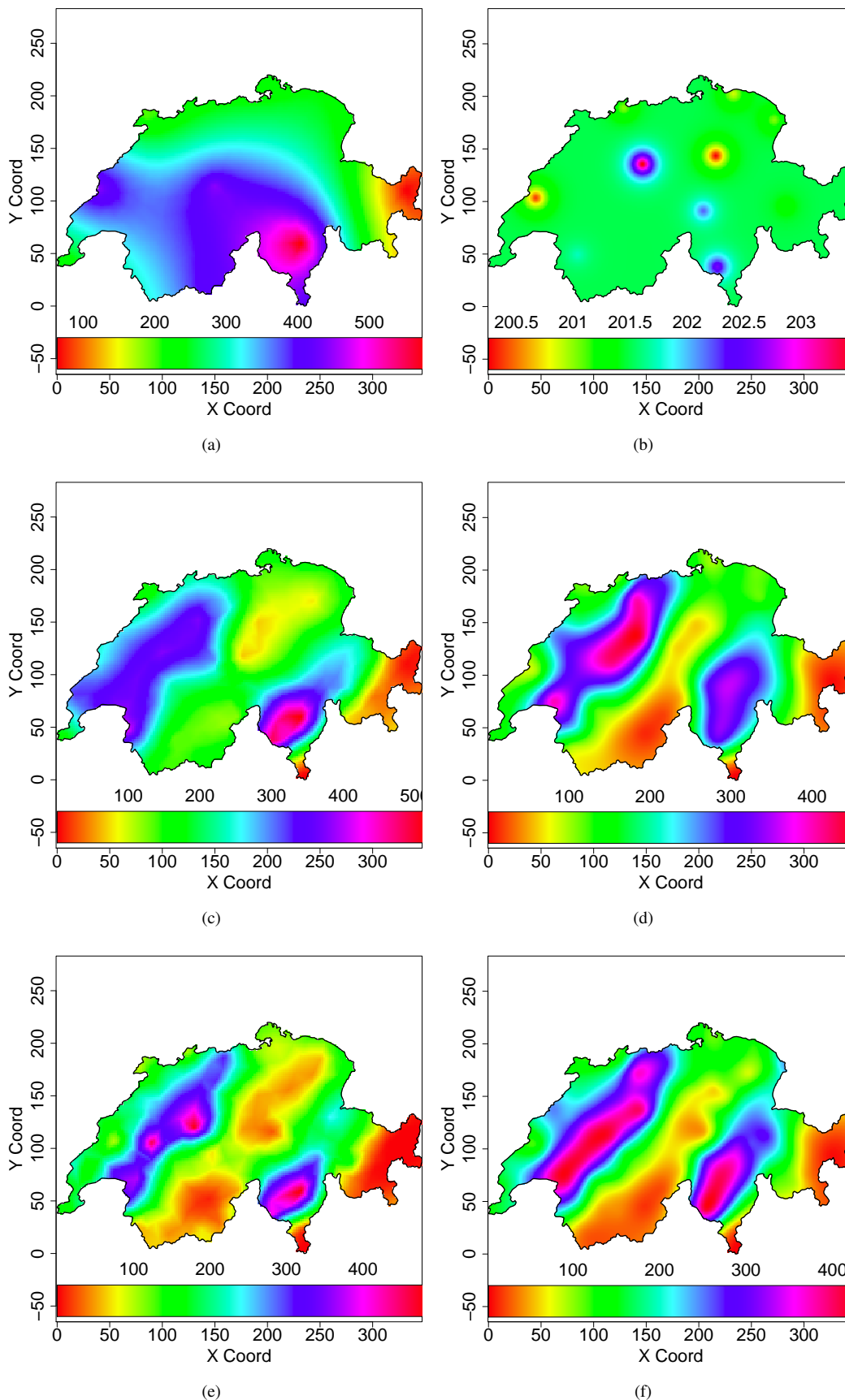
Fig. 6: Predicted rainfall fields in Switzerland using 10 (a) and (b), 50 (c) and (d), 100 (e) and (f) selected sensor locations and measurements. In left column, the selection procedure is conducted by our proposed criterion based on the GMRF. In right column, the selection procedure is based on the GP criterion [15], [9]. X Coord and Y Coord are in km.
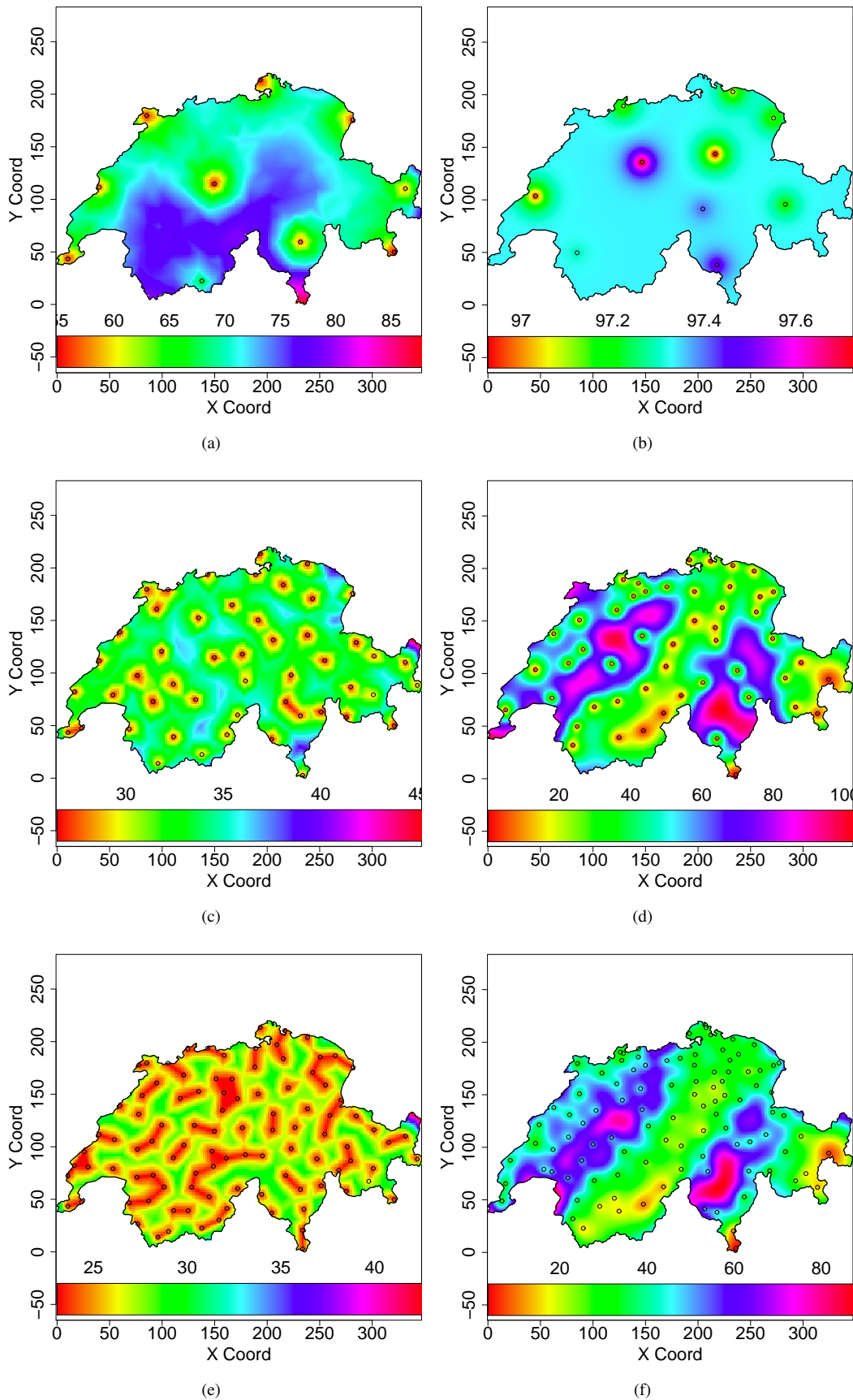
Fig. 7: Prediction error variances in Switzerland using 10 (a) and (b), 50 (c) and (d), 100 (e) and (f) selected sensor locations and measurements. In left column, the selection procedure is conducted by our proposed criterion based on the GMRF. In right column, the selection procedure is based on the GP criterion [15], [9]. X Coord and Y Coord are in km.

[15] W. F. Caselton and J. V. Zidek, "Optimal monitoring network designs," *Statistics and Probability Letters*, vol. 2, pp. 223–227, 1984.

[16] C. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Operations Research*, vol. 43, pp. 684–691, 1995.

[17] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in Gaussian processes," in *Proc. 22th International Conference on Machine Learning*, Bonn, Germany, 2005, pp. 265–272.

[18] A. Krause, R. Rajagopal, A. Gupta, and C. Guestrin, "Simultaneous optimization of sensor placements and balanced schedules," *IEEE Transactions on Automatic Control*, vol. 56, pp. 2390–2405, 2011.

[19] F. Bian, D. Kempe, and R. Govindan, "Utility based sensor selection," in *Proc. 17th International Conference on Information Processing Sensor Networks*, Tennessee, USA, April 2006, pp. 11–18.

[20] N. A. Cressie, *Statistics for spatial data*. Wiley, 1991.

[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2006.

[22] P. J. Diggle and P. J. Ribeiro, *Model-based geostatistics*. Springer, New York, USA, 2007.

[23] H. Rue and H. Tjelmeland, "Fitting Gaussian Markov random fields to Gaussian fields," *Scandinavian Journal of Statistics*, vol. 29, pp. 31–49, 2002.

[24] N. Cressie and N. Verzelen, "Conditional-mean least-squares fitting of Gaussian Markov random fields to Guassian fields," *Computational Statistics and Data Analysis*, vol. 52, pp. 2794–2807, 2008.

[25] L. Hartman and O. Hössjer, "Fast kriging of large data sets with Gaussian Markov random fields," *Computational Statistics and Data Analysis*, vol. 52, pp. 2331–2349, 2008.

[26] F. Lindgren, H. Rue, and J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach," *Journal of the Royal Statistical Society (Series B)*, vol. 73, pp. 423–498, 2011.

[27] H. Rue and L. Held, *Gaussian Markov random field: Theory and applications*. Chapman and Hall - CRC Press, London, England, 2005.

[28] R. Chellappa, S. Chatterjee, and R. Bagdazian, "Texture synthesis and compression using Gaussian-Markov random field models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15(2), pp. 298–303, 1985.

[29] Y. Xu, J. Choi, S. Dass, and R. Maiti, "Efficient Bayesian spatial prediction with mobile sensor networks using Gaussian Markov random fields," *Automatica*, vol. 49, pp. 3520–3530, 2013.

[30] F. Bullo, J. Cortes, and S. Martínez, *Distributed control of robotic networks*. Princeton University Press, 2009.

[31] S. Banerjee, B. Carlin, and A. Gelfand, *Hierarchical modeling and analysis for spatial data*. Chapman and Hall, New York, 2004.

[32] A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations*. Springer, 2008.

[33] C. M. Bishop, *Pattern recognition and machine learning*. Springer, New York, USA, 2006.

[34] D. S. Bernstein, *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*. Princeton University Press, 2005.

[35] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley and Sons, Lnc, Hoboken, New Jersey, 2006.

[36] L. Paninski, "Estimation of entropy and mutual information," *Neural Computation*, vol. 15(6), pp. 1191–1253, 2003.

[37] B. Guo and M. S. Nixon, "Gait feature subset selection by mutual information," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 39(1), pp. 36–46, 2009.

[38] O. F. Christensen, P. J. Diggle, and P. J. R. Jr, *Analysing positive-valued spatial data: the transformed Gaussian model*. In P. Monestiez, D. Allard and Froidevaux (eds), GeoENV III - Geostatistics for environmental applications. Quantitative Geology and Geostatistics, Kluwer, 2001.

[39] E. L. Lawler and D. E.Wood, "Branch-and-bound methods: A survey," *Operations Research*, vol. 14, pp. 699–719, 1966.

[40] W. Welch, "Branch and bound search for experimental designs based on D optimality and other criteria," *Technometrics*, vol. 24, pp. 41–48, 1982.

[41] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[42] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society (Series B)*, vol. 71, pp. 319–392, 2009.

[43] L. V. Nguyen, S. Kodagoda, R. Ranasinghe, and G. Dissanayake, "Information-driven adaptive sampling strategy for mobile robotic wireless sensor network," *IEEE Transaction on Control Systems Technology*, p. accepted, 2015.

[44] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximation for maximizing submodulor set functions - I," *Mathematical Programming*, vol. 14(1), pp. 265–294, 1978.

[45] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985.

[46] P. J. Diggle and P. J. R. JR, "Bayesian inference in Gaussian model based geostatistics," *Geographical and Environmental Modelling*, vol. 6(2), pp. 129–146, 2002.

[47] G. E. P. Box and D. R. Cox, "An analysis of transformations (with discussion)," *Journal of the Royal Statistical Society, Series B*, vol. 26, pp. 211–252, 1964.

[48] [Online]. Available: http://www.r-inla.org

Linh V. Nguyen (M'15) received B.Eng. and M.Eng. degrees in electrical engineering and automation from Vietnam National University of Agriculture, Vietnam, in 2006 and 2010, respectively. He received Ph.D. degree in electrical engineering from University of Technology Sydney, Australia, in 2015. He was a Lecturer at Vietnam National University of Agriculture from 2006 to 2011 and a Visiting Lecturer at Chu Van An University and Hung Yen Institute of Technology, Vietnam from 2007 to 2011. Currently, he is a Postdoctoral Research Fellow at Centre for Autonomous Systems, University of Technology Sydney, Australia. He is a regular reviewer of Australian Control Conferences. His current research interests include spatial adaptive sampling methods, Gaussian processes, statistical learning algorithms for static and mobile sensor networks and robotics, sensors and signal processing, and infrastructure robotics.

Sarath Kodagoda (M'10) received B.Sc.Eng.Hons. degree in 1995, specializing in Electrical Engineering, from the University of Moratuwa, Sri Lanka. He received his M.Eng. (2000) and Ph.D. (2004) degrees specializing in robotics from the Nanyang Technological University, Singapore. Before joining the ARC Centre for Autonomous Systems (CAS) at University of Technology, Sydney (UTS), he worked as a Design Engineer in a reputed multinational company. He is currently an Associate Professor and Coordinator of the Mechanical and Mechatronics program at the University of Technology, Sydney. His main research contributions are in the areas of autonomous road vehicles and Human Robot Interaction. He is an Associated Editor, Programme Committee member and a regular reviewer of a number of top robotics journals and conference proceedings. He is a member of the IEEE. His research interests include infrastructure robotics, human robot Interaction, machine learning, perception, target tracking and mobile robotics.

Ravindra Ranasinghe (M'97) received B.Sc.Eng.Hons. degree in 1995, specializing in Computer Science and Engineering, from the University of Moratuwa, Sri Lanka. He received his Ph.D. (2002) degrees specializing in wireless communication protocols from the University of Melbourne, Australia. Before joining the Centre for Autonomous Systems at University of Technology, Sydney (UTS), he worked in several technology start-up companies in USA, Australia and Sri Lanka. He is currently a Senior Research Fellow at the Centre for Autonomous Systems, University of Technology, Sydney. He is a member of the IEEE since 1997. His research interests include wireless communication protocols, wireless sensor networks, mobile robotics wireless sensor networks, machine learning and assistive robotics.