Adaptive Sampling for Spatial Prediction in Wireless Sensor Networks

By

Van Linh NGUYEN

Thesis submitted as a requirement for the degree of

Doctor of Philosophy

School of Electrical, Mechanical and Mechatronic Systems

Faculty of Engineering and Information Technology

University of Technology, Sydney (UTS)

December 2014

Certificate of Original Authorship

I, Van Linh Nguyen, certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Student:

_____

Date:

_____

Acknowledgements

I would like to thank everybody who have helped me to accomplish this thesis. First and foremost, I would like to express my deepest gratitude to my principal supervisor, Associate Professor Sarath Kodagoda, who has always been a great mentor, giving me tremendously valuable suggestions, thoughtful guidance, and continuous support throughout past years. The accomplishment of this thesis would not have been possible without his invaluable advice.

I am sincerely grateful to my co-supervisor, Professor Gamini Dissanayake, for his research expertise and deep insight from face-to-face meetings at the beginning of my study having inspired me in selecting my research topic. I am thankful to Dr Ravindra Ranasinghe, who is also my co-supervisor, for informative comments and technical advice he has provided in group meetings and individual discussions.

I would like to express my deeply gratitude to my former supervisor at Vietnam National University of Agriculture, Dr. Duong Van Nguyen, now in Ministry of Education and Training, Vietnam for his support and advice on both research and a real life.

I would also like to thank the Vietnam International Education Development (VIED) for having offered me a PhD scholarship and having introduced me to the VIED-UTS joint scholarship. Without this scholarship, I would never have been able to create this thesis at this point.

I am indebted to the Vice-Chancellor's Postgraduate Research Student Conference Funding and the Faculty of Engineering and Information Technology (FEIT) Travel Funding committees for their generous financial support for my conference

trips. I would also send many thanks to the officers at the Research and Development Office, FEIT for their dedicated assistance.

I am very keen to appreciate the rest of the team at the Center for Autonomous Systems (CAS) for their help and support. Through many interesting discussions and seminars, I have learned so much about doing research and being part of scientific community.

I would like to extend further my thanks to the experts from the Sydney Water, including Heriberto Bustamante, Dammika Vitanage, Tung Nguyen, and Nur Nobi for giving me good advice from face-to-face and email meetings and providing me with the data sets for my simulated experiments.

Last but not least, I cannot thank enough my parents and my parents-in-law for always being there for me and for letting me choose my way. A very special thank goes to my beloved spouse, Huong, for her endless love, patience, and encouragement. In particular, my heartfelt gratitude is towards my daughter, Giang, for being part of my life.

Contents

List of Figures

Acronyms and Abbreviations

WSN	Wireless Sensor Network
WSNs	Wireless Sensor Networks
GP	Gaussian Process
GPs	Gaussian Processes
SWSN	Stationary Wireless Sensor Network
SWSNs	Stationary Wireless Sensor Networks
MRWSN	Mobile Robotic Wireless Sensor Network
MRWSNs	Mobile Robotic Wireless Sensor Networks
MRWS	Mobile Robotic Wireless Sensor
MRWSs	Mobile Robotic Wireless Sensors
GMRF	Gaussian Markov Random Field
RMSE	Root Mean Square Error
RMSEs	Root Mean Square Errors
ARMSE	Average Root Mean Square Error
ARMSEs	Average Root Mean Square Errors
JOR	Jacobi Over-Relaxation
DAC	Discrete-time Average Consensus
H_2S	Hydrogen Sulphide
SPDE	Stochastic Partial Differential Equation
MI	Mutual Information
SA	Simulated Annealing
OF	Objective Function
MAC	Medium Access Control
VIED	Vietnam International Education Development

CAS	Center for Autonomous Systems
FEIT	Faculty of Engineering and Information Technology
UTS	University of Technology, Sydney

UNIVERSITY OF TECHNOLOGY, SYDNEY

Abstract

Faculty of Engineering and Information Technology

Centre for Autonomous Systems

Doctor of Philosophy

Adaptive Sampling for Spatial Prediction in Wireless Sensor Networks

by Van Linh NGUYEN

Networks of wireless sensors are increasingly exploited in crucial applications of monitoring spatially correlated environmental phenomena such as temperature, rainfall, soil ingredients, and air pollution. Such networks enable efficient monitoring and measurements can be included in developing models of the environmental fields even at unobserved locations. This requires determining the number of sensors and their sampling locations which minimize the uncertainty of predictions. Therefore, the aim of this thesis is to present novel, efficient and practically feasible approaches to sample the environments, so that the uncertainties at unobserved locations are minimized. Gaussian process (GP) is utilized to statistically model the spatial field. This thesis includes both stationary wireless sensor networks (SWSNs) and mobile robotic wireless sensor networks (MRWSNs), and thus the issues are correspondingly formulated into *sensor selection* and *sensor placement* problems, respectively. In the first part of the thesis, a novel performance metric for the sensor selection in the SWSNs, named average root mean square error, which reflects the average uncertainty of each predicted location, is proposed. In order to minimize this NP-hard

and combinatorial optimization problem, a simulated annealing based algorithm is proposed; and the sensor selection problem is effectively addressed. Particularly, when considering the sensor selection in constrained environments, *e.g.* gas phase hydrogen sulphide in a sewage system, a modified GP with an improved covariance function is developed. An efficient mutual information maximization criterion suitable for this particular scenario is also presented to select the most informative gaseous sensor locations along the sewer system. The second part of this thesis introduces centralized and distributed methods for spatial prediction over time in the MRWSNs. For the purpose of finding the optimal sampling paths of the mobile wireless sensors to take the most informative observations at each time iteration, a sampling strategy is proposed based on minimizing the uncertainty at all unobserved locations. A novel and very efficient optimality criterion for the adaptive sampling problem is then presented so that the minimization can be addressed by a greedy algorithm in polynomial time. The solution is proven to be bounded; and computational time of the proposed algorithm is illustrated to be practically feasible for the resource-constrained MRWSNs. In order to enhance the issue of computational complexity, Gaussian Markov random field (GMRF) is utilized to model the spatial field exploiting sparsity of the precision matrix. A new GMRF optimality criterion for the adaptive navigation problem is also proposed such that computational complexity of a greedy algorithm to solve the resulting optimization is deterministic even with increasing number of measurements. Based on the realistic simulations conducted using the pre-published data sets, it has shown that the proposed algorithms are superior with appealing results.

To the memory of my dearest grandfathers and grandmother!

Chapter 1

Introduction

Environmental issues such as pollution of land, water and air, drastic climate change, natural disasters, and resource depletion have topped the agenda in recent years. These are crucial not only to governments but also to environmental scientists. Therefore, it would be prudent to monitor spatially correlated environmental phenomena so as to ameliorate the knowledge and understanding of their economic, environmental, and health impacts and implications. For instance, determining distributional patterns of ecological phenomena can be utilized to design resource-specific exploitation plans [1]. Better understanding of physics underlying occurrence process of earthquakes [2] can generate early vital hazard warnings to society. Results from observing sulphur dioxide SO_2 in the air [3], which may affect the respiratory system, should be given to alert community council to reduce burning of sulphur-containing fuels in factories. In the agricultural field, monitoring nitrogen density in the soil can be used to regulate farm inputs that leads to mitigation of environmental pollution due to over-application of nitrogen fertilizer [4]. As a part of this thesis, monitoring hydrogen sulphide H_2S in a gravity sewer system aims to mitigate

corrosion problems in a sewer network and to alleviate foul smells that adversely affect any exposed communities. In addition to concerns mentioned above, there is required ubiquitous observations of temperature, humidity, rainfall, soil ingredients, and mono-nitrogen oxides NO_x in natural and built-up habitats [5–9]. Nonetheless, since environmental measurements are a single scalar quantity that is only locally valid, it is required to make predictions about the process at unmeasured locations by using observations. Therefore, there are currently strong motivations to monitor, model and predict the environmental field of interest that is often represented as complex phenomena. More particularly, if sensing equipment is effectively used in monitoring, modelling and predicting the spatial phenomena, which is referred to as adaptive sampling, results from visually representing the physical field are useful in making decisions regarding environmental issues.

This chapter describes a basic background on monitoring and modelling environmental physical phenomena by the use of wireless sensor networks (WSNs). This background also furnishes two fundamental issues in the WSNs for efficiently estimating and predicting the spatial field, which ultimately form the research objectives for this thesis. Moreover, the chapter presents the key contributions and outlines the organization of the thesis in order to provide the reader with an overview of this thesis. Finally, some published works related to the thesis are summarised in this chapter.

1.1 Background

In the past, data was usually collected by humans who harvested measurements by walking through the field. Nevertheless, this manual procedure was tiresome and

time-consuming. 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 WSNs that are increasingly useful in crucial applications in environmental monitoring [10]. Not only do these systems provide a virtual connection with the physical field in general, the WSNs are particularly dominant in remote and hazardous environments where many essential phenomena have hardly been investigated due to their inaccessibility. In addition to collecting the data, combining the measurements with a model, a wireless sensor network (WSN) is also competent to estimate and predict the spatial phenomenon at unobserved locations. Any such estimation and prediction should be associated with some measure of uncertainty that demonstrates the quality of the prediction. In equivalent words, the most informative observations certainly lead to minimum uncertainty in resulting predictive inference. As a result, an important fundamental challenge in utilizing the WSNs for real-world applications of environmental monitoring is the identification of locations where wireless sensors should be deployed to take measurements in order to maximize the information gained from them.

In order to position the wireless sensors in the optimal locations of the environment to capture the environmental phenomena most effectively, which directly influences the accuracy of the predicted field, it is essential to model the spatial phenomena. Usually, these physical processes are described by deterministic and data-driven models [11]. The prime disadvantage of the deterministic model is that it requires model parameters and initial conditions to be known in advance. Furthermore, model complexity and various interactions in the deterministic models that are difficult to model tilt the balance in favor of data-driven approaches. In this work, it is particularly proposed to consider the Gaussian process data-driven

model [5, 12, 13] to statistically model spatial fields. The use of a Gaussian process (GP) allows prediction of the environmental phenomena of interest efficiently at any unobserved points.

The typical task of the WSN consists of gathering measurements of a spatial field over a region of interest. However, for instance in a stationary wireless sensor network (SWSN), where none of sensors are allowed to change their locations after their deployment, multiple wireless sensor nodes co-located within the vicinity of a phenomenon in a dense SWSN may generate similar data samples, which is called an *over-sampling* problem. This over-sampling problem has the potential to create a sizeable redundancy in sensed data that adversely affects goodput, that is effective usage to provide intelligence about the phenomenon, of the WSN as the redundant samples do not contribute to gaining 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 energy consumption, maintenance burden, and 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 stationary wireless sensor nodes out of all potential ones, which should participate in the sensing task. The selection procedure is known as a *sensor selection* problem in which the resulting prediction is required to conform requirement of the highly 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 sampling interval. In this research, the sensor selection problem in both unconstrained and constrained environments is to be considered.

Nevertheless, if the networks of stationary wireless sensors are deployed in a

changing environment, the stationary wireless sensor locations selected by a sensor selection algorithm are no longer the most informative over time. On the other hand, if a dense sampling of the environment or a large area to be monitored are needed, the use of a large SWSN is required, which leads to various challenges including managing and maintaining the network. As a consequence, a wireless sensor network incorporating mobile robotic platforms is desirable. With a set of networked mobile sensor nodes, the mobile robotic wireless sensor networks (MRWSNs) are capable of providing services required not only for monitoring but also for exploring the environment. Flexibility of the MRWSNs enables enhanced efficiency of data collection, adaptation to the changes in the environment, and production of a robust response to individual sensor failures. Moreover, the mobility of the robotic wireless sensors can be utilized to generate the optimal strategies for collecting measurements in the spatial field of interest. Effectively utilizing the MRWSNs to observe and predict the environmental field is defined in this thesis as a *sensor placement* problem. In equivalent words, resolving the sensor placement problem is to find the most informative sampling paths for the mobile robotic wireless sensors (MRWSs) to take future observations in order to minimize uncertainty in resulting predictions of the physical field at all unmeasured locations of interest.

1.2 Research Objectives

The aim of this thesis is to present a new class of approaches for finding the best sampling locations in the environment for wireless sensors to extract the most useful information out of the spatial field. The design objectives are based upon information-theoretic criteria such as root mean square error, conditional entropy

and mutual information so that the uncertainty at all unobserved locations of interest is minimized. Both sensor selection and sensor placement problems are stated to be data-driven optimization issues that are successively considered in both stationary and mobile wireless sensor networks in order to respond to various configurations in real systems. On the other hand, due to their small size, the WSNs are limited to resource constraints including energy, memory, and communication restrictions. Hence, the thesis is also a step forward in making sophisticated statistical inference based combinatorial optimization problems to be feasible in practice in the resource-constrained WSNs.

Correspondingly, the specific research objectives are:

1. Develop efficient methods to address the sensor selection problem in the SWSNs. In this objective, the SWSNs are utilized to observe both unconstrained and constrained environments.

2. Develop novel algorithms to find efficient and near-optimal sampling paths for the MRWSNs, where the spatial field is modelled by the use of the GP. Centralized and distributed Gaussian predictive inferences are also objectively designed to effectively estimate and predict the physical processes.

3. Develop computationally-efficient centralized and distributed approaches for the MRWSNs by employing the Gaussian Markov random field (GMRF). This aim is needed to make the resource-constrained MRWSNs feasible in practice.

1.3 Main Contributions

According to the objectives of the research defined above, the following presents the pivotal contributions of this thesis in the order of chapters.

In Chapter 4, the problem of selecting the most informative subset of potential stationary wireless sensors in observing an environmental phenomenon in a free space is comprehensively considered. The selective criterion is proposed based upon a novel performance metric that reflects the root mean square error (RMSE) for the each predicted location, which is named average RMSE (ARMSE). Although the proposed criterion for the sensor selection, which minimizes the ARMSE in all unobserved locations of interest, is a combinatorial optimization issue, an efficient algorithm which simulates the process of forming metal crystal from metal liquid, called annealing, is introduced to cope with this challenging minimization. Accompanied by a technique of replacing a member of the selected subset by its neighbours at each iteration, the simulated annealing based algorithm not only effectively lowers computational time but also fully covers the space. The proposed approach demonstrates considerable outperformance as compared with other conventional methods such as the greedy heuristic algorithm of which information-theoretic criteria including conditional entropy and mutual information are utilized.

Another key contribution in this chapter is to address the sensor selection problem in a constrained environment. The scenario in which the study was conducted was the gas phase hydrogen sulphide (H_2S) in a sewage system, where the H_2S data was generated by the wastewater model developed by Sydney Water. In order to model the gaseous H_2S content, a frequently-utilized GP is chosen for implementation. Note that the standard GP relies upon correlations among all nearby observations in a free space. Nonetheless, the primarily crucial specification of the

constrained environment is that spatial phenomena are often limited by conditions of the environment. For instance, in the gravity sewer system, the H_2S content is constrained by the sewer boundaries, where dependences among sensing measurements are more or less prevented in direct distances. As a consequence, in this work, a modified GP with an improved covariance function is proposed. The covariance value between two locations is computed employing the lengths along the sewage pipes. The proposed model is confirmed by comparing with a complex model developed by Sydney Water. Due to constraints of sewage pipes, every sensor locations has two upper-stream and down-stream neighbours, which prevents the simulated annealing based algorithm from finding the most informative gaseous sensor locations along the sewer system. Thus, an efficient mutual information maximization criterion is presented. The primary specification of this optimization is that the mutual information is computed between the H_2S phase at selected sensor locations and that at a set of unselected sensor and unmeasured locations. The proposed approach for the H_2S sensor selection outperforms [14].

In Chapter 5, resource-constrained MRWSNs optimally monitoring and predicting the spatial field are taken into account. The chapter first presents a centralized estimation and prediction approach to address the sensor placement problem. By the use of the GP, the Gaussian predictive inference method is derived over time, where unknown mean parameters and hyperparameters are estimated online by the maximum likelihood approach. For the purpose of finding out the optimal sampling paths on which mobile wireless sensors could take the most informative observations at each time iteration, the sampling strategies are proposed based on information-theoretic criteria. The premise behind the sampling criteria is to minimize uncertainty at all unobserved locations of interest. For a continuous space, a locational optimization based navigation scheme is proposed to drive mobile sensing agents

going through positions of high uncertainty density in the corresponding surface. The key contribution of this proposition is that the continuous density function is developed based on posterior variances. The optimal robotic sensor locations are delineated as the centroids of their Voronoi cells by the centroidal Voronoi tessellation approach. In a discrete space, an adaptive sampling optimization problem by the use of conditional entropy is formulated. The criterion aims to select the most informative locations for the MRWSs from a discretized set of potential sensing points. The chapter illustrates that the existing approximate greedy heuristic algorithms to deal with this combinatorial and NP-hard optimization are intractable if the networks search for a large number of unobserved locations in the field. A novel but very efficient optimality criterion for the adaptive sampling problem is proposed so that it can be addressed by the greedy algorithm in polynomial time. Although a near-optimal solution is obtained, computational time of the algorithm is proven to be feasible in practice for the resource-constrained MRWSNs. It is also proven that the proposed approach is guaranteed by a lower bound and an upper bound. The lower bound is within $1 - \frac{1}{e}$ of the optimal solution, which is then utilized as a stopping criterion for the sampling algorithm that stops the MRWSs from sampling the spatial field to conserve the limited energy.

In an effort to resolve the centrally-computed scheme of sophisticated statistical inference of the GP regression at each robotic sensor in the resource-constrained networks of mobile, wireless and noisy agents monitoring physical phenomena of interest, a distributed spatial prediction approach is presented by the use of distributed computation tools consisting of Jacobi over-relaxation (JOR) and discrete-time average consensus (DAC) algorithms. In this proposed method, every mobile sensor locally exchanges information with its neighbours within a certain range of communication. Based on its own measurements and its neighbours' observations, each

mobile sensing agent first estimates its own mean parameters and hyperparameters, then recursively obtains global model parameters via the distributed techniques. It is proposed that one mobile agent predicts the environmental process using all the measurements it has collected and received from the beginning to the current time. The globally predicted field of interest can also be obtained at each robotic sensor by the utilization of the JOR and the DAC computations. Finally, a distributed sampling scheme is designed to let the MRWSs navigate the most uncertain points in the environment.

In Chapter 6, it is proposed to utilize the GMRF, a discretely indexed approximation of the GP, to model the physically spatial phenomena. By employing a stochastic partial differential equation (SPDE) approach, the GMRF represents the spatial field on an irregular lattice of a triangulation. Due to the sparsity of its precision matrix, the GMRF spatial estimation and prediction approach are derived with a remarkable benefit in computation. By discussing a technique of the one-step-ahead forecast to predict the future measurements from the GMRF standpoint, an adaptive sampling strategy of which conditional entropy of latent variables at unmeasured locations is to be minimized is formulated. More importantly, it is thoughtfully proven that the existing approximate greedy algorithms to address this combinatorial and NP-hard problem of finding the sampling paths for the MRWSNs in a large number of potential locations in the field are often intractable. Consequently, an efficient novel optimality criterion for the adaptive navigation problem is proposed such that computational complexity of a greedy algorithm to solve the resulting optimization is deterministic even though either the number of measurements goes up or the unmeasured locations are dense on a very large mesh of a triangulation. For the case where each mobile sensor has a short wireless communication range, a distributed spatial prediction algorithm is developed to address the

centralized GMRF-based regression. A sampling scheme for each robotic sensing agent is also designed in a distributed fashion. By utilizing these distributed approaches, although each mobile sensor communicates with its neighbouring sensors only, it approximately reaches to global estimation and prediction values.

Therefore, more specifically, the contributions of the thesis are:

- **Sensor selection:** The first key contribution is a novel approach to select the most informative sensors in a SWSN for spatial prediction purposes. A new ARMSE optimization criterion for the sensor selection problem and an efficient simulated annealing based algorithm for solving this selection optimization are proposed in the unconstrained environment. Moreover, a new GP based model to represent the gaseous H_2S spatial distribution in the sewer networks and a mutual information based algorithm are also proposed to address the sensor selection issue in the constrained environment.

- **Sensor placement using the GP:** A new posterior variances based density function is presented for the locational optimization based method to find the optimal sampling locations in a continuous environment. In addition, a novel optimality criterion for the entropy based sampling strategy is proposed in a discrete space which can be near-optimally resolved by a greedy algorithm in polynomial time. This proposed scheme is proven to be guaranteed by a lower and an upper bounds. A distributed approach is also designed by using the JOR and DAC techniques for a network where a mobile sensor has a short communication range.

- **Sensor placement using the GMRF:** Novel GMRF based centralized and distributed sampling schemes are introduced to find the most informative sampling locations on a very dense and large lattice of a triangulation. Due to

sparsity of the precision matrix, it is proven that a simple greedy algorithm can address those combinatorial and NP-hard optimization problems of finding the sampling paths in deterministic time, which is not dependent on the number of either measurements or unobserved locations. The proposed approach is demonstrated to be practically feasible.

1.4 Organization of the Thesis

The structure of this thesis is arranged as follows. In Chapter 2, the basic mathematical notations that will be utilized throughout this thesis is first presented. Then, the random field models and the GP regression are sequentially introduced. The concepts for the MRWSNs are also discussed in this chapter before the notations of Voronoi partition and submodularity function are reviewed. Chapter 3 reviews up-to-date methods used for the spatial prediction in the WSNs in the literature. The existing criteria for sensor selection and sensor placement problems are introduced before discussion of the different algorithms that are utilized to address these issues. Three chapters, four to six, present the efficient approaches proposed for this study. Chapter 4 is commenced by modelling the spatial field. Next, the sensor selection problem is successively considered in unconstrained and constrained environments. The proposed approaches are evaluated in many unconstrained scenarios including indoor and outdoor environments. In particular, the gaseous H_2S phase, which is generated by a complex model developed by Sydney Water, is in the first stage utilized to verify the sensor selection algorithm in the sewerage system constrained by the sewer pipes. The sensor networks assisted by mobile robotic platforms for monitoring and predicting the GP modelled environmental processes are investigated in Chapter 5. Both centralized and distributed spatial estimation

and prediction approaches are proposed to optimize the sensing activities. The locational optimization, conditional entropy and posterior variance based methods are introduced to efficiently and effectively employ the MRWSs to sample the physical phenomena, which eventually enables not only significant improvement to the quality of prediction but also reduction of the model uncertainty. In order to overcome the computational complexity of factorizing the dense covariance matrix in the GP regression, Chapter 6 proposes the utilization of the GMRF to model the spatial field. The centralized and distributed spatial prediction approaches from a GMRF standpoint are fully derived. More interestingly, near-optimal but very effective algorithms are designed to find the efficient sampling paths for mobile sensing agents in both centralized and distributed manners. Conclusions and suggestions for future research are provided in Chapter 7.

1.5 Publications

This section lists journal and conference papers that have been published (or under review) and related to the topic of this thesis.

1. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Mobile robotic wireless sensor networks for efficient spatial estimation via Gaussian Markov random field" *IEEE Transactions on Robotics*, 2014 (*under review*).

2. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Information-driven adaptive sampling strategy for mobile robotic wireless sensor network" *IEEE Transactions on Control Systems Technology*, 2014 (*revision submitted*).

3. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Spatially-distributed prediction with mobile robotic wireless sensor networks" in *Proceedings of IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV2014)*, Marina Bay Sands, Singapore, December, 2014.

4. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, Gamini Dissanayake, Heriberto Bustamante, Dammika Vitanage, and Tung Nguyen "Spatial prediction of hydrogen sulphide in sewers with a modified Gaussian process combined mutual information" in *Proceedings of IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV2014)*, Marina Bay Sands, Singapore, December, 2014.

5. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Mobile robotic wireless sensor networks for efficient spatial prediction" in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2014)*, Chicago, USA, September, 2014.

6. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Locational optimization based sensor placement for monitoring Gaussian processes modelled spatial phenomena" in *Proceedings of IEEE International Conference on Industrial Electronics and Applications (ICIEA2013)*, Melbourne, Australia, June, 2013.

7. **Linh Van Nguyen**, Sarath Kodagoda, Ravindra Ranasinghe, and Gamini Dissanayake, "Simulated annealing based approach for near-optimal sensor selection in Gaussian processes" in *Proceedings of the IEEE International Conference on Control, Automation and Information Science (ICCAIS 2012)*, Ho Chi Minh city, Vietnam, November 2012.

8. **Linh Van Nguyen**, Ravindra Ranasinghe, Sarath Kodagoda, and Gamini Dissanayake, "Sensor selection based routing for monitoring Gaussian processes modelled spatial phenomena" in *Proceedings of the Australian Conference on Robotics and Automation (ACRA 2012)*, Wellington, New Zealand, December, 2012.

9. Zulkarnain Zainudin, Sarath Kodagoda, and **Linh Van Nguyen**, "Mutual information based data selection in Gaussian processes for people tracking" in *Proceedings of the Australian Conference on Robotics and Automation (ACRA 2012)*, Wellington, New Zealand, December, 2012.

1.6 Summary

This chapter has provided an introduction to utilizing the WSNs to observe the spatial processes. Two principal problems of the sensor selection and the sensor placement have been presented to meet the requirement of efficiently sampling the environment to improve the quality of resulting prediction. In fact, these issues aim to find the most informative locations in the environment to deploy the wireless sensors. Therefore, finding novel and efficient approaches to address these problems in practically feasible time for the resource-constrained WSNs are the objectives of this doctoral study research.

Chapter 2

Notations and Preliminaries

2.1 Introduction

This chapter introduces basic concepts and mathematical notations that are employed throughout the thesis. Models for the spatial field such as the GP and the GMRF are first considered and the GP regression is then discussed. Additionally, the properties of the MRWSNs are also highlighted for design of sampling strategies. Definitions of Voronoi partition and submodularity function are also delineated, which will be utilized to develop navigation and its bounds for the MRWSs.

2.2 Mathematical Notations

Here, mathematical notations that are used throughout this thesis are defined. Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{>0}$ and $\mathbb{Z}_{>0}$ denote the sets of real numbers, non-negative real numbers, positive real numbers and non-negative integer numbers, respectively.

The Euclidean distance function is defined by $\|\cdot\|$. The $|\cdot|$ denotes cardinality. Let $\mathcal{Q}$ denote a convex polytope in $\mathbb{R}^d$. A convex polytope is the convex hull of a finite point set.

In terms of a matrix, let $\det(A)$ and $\mathrm{tr}(A)$ denote the determinant and trace of a square matrix A, respectively. Let $A^T \in \mathbb{R}^{n\times m}$ be the transpose of a matrix $A \in \mathbb{R}^{m\times n}$. The positive definiteness of a square matrix A is by $A \succ 0$.

Let $\mathbb{E}$ denote the operator of an expectation and $\mathcal{O}(\cdot)$ denote the operator of running time of an algorithm. $\mathrm{prec}(\cdot)$ and $\mathrm{corr}(\cdot)$ are also defined as operators of the precision and the correlation. The inner product operator is defined by $\langle ., . \rangle$. $\bigtriangledown$ and $\log(\cdot)$ denote the gradient and the logarithm operators, respectively.

A random variable z that is distributed in a normal distribution with a mean μ and a covariance matrix Σ is defined by $z \sim \mathcal{N}(\mu, \Sigma)$. If two random variables x and y are independent, it is written as $x \perp y$.

$a_t[i]$ is defined as the i^{th} element of the vector a_t, and $b_t[i, j]$ is referred as the element at the i^{th} row and j^{th} column of the matrix b_t.

An undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a vertex set $\mathcal{V}$ and a edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The relative complement of a set $\mathcal{A}$ in a set $\mathcal{B}$ is denoted by $\mathcal{B} \setminus \mathcal{A}$.

Other notations will be explained as and when they occur.

2.3 Random Field Models

In this section, dominant concepts and properties are introduced on the Gaussian process (GP) [12, 13] and the Gaussian Markov Random fields (GMRF) [15] that

are used to model the random field in this work.

2.3.1 Gaussian process

A GP is a very popular non-parametric Bayesian technique for modeling spatially correlated data. Initially known as *kriging*, the technique has its roots in geostatistics where it is mainly used for estimation of mineral resources [16]. The Gaussian processes (GPs) extend multivariate Gaussian distributions over a finite vector space to function space of infinite dimensionality. It is formally defined as follows.

Definition 2.1. *A Gaussian process is a collection of random variables, any finite number of which have a consistent joint Gaussian distribution.*

Consider a spatial location $v \in \mathbb{R}^d$, a random variable $z(v)$ at v is modeled as a GP and written as

$$z(v) \sim \mathcal{GP}(\mu(v), cov(v, v')), \tag{2.1}$$

where $v, v' \in \mathbb{R}^d$ are the inputs. $\mu(v)$ is a mean function and $cov(v, v')$ is a covariance function, often called a kernel function. These functions are defined as

$$\mu(v) = \mathbb{E}[z(v)],$$
$$cov(v, v') = \mathbb{E}[(z(v) - \mu(v))(z(v') - \mu(v'))]$$

A spatial GP is stationary if $cov(v, v') = cov(v - v')$. That is, the covariance depends only on the vector difference between v and v'. Furthermore, if $cov(v, v') = cov(\|v - v'\|)$, the stationary process is isotropic. Hence, the covariance between a pair of variables of $z(v)$ at any two locations is only dependent on the distance between them.

The covariance function is a vital ingredient in a GP. In fact, there is a practical family of parametric covariance functions proposed in [17]. For example, one of the frequently used kernel functions is squared exponential, that is,

$$cov(v, v') = \sigma^2 \exp\left(-\frac{\|v - v'\|^2}{2\phi^2}\right), \tag{2.2}$$

where σ^2 is the marginal variance (also known as the maximum allowable covariance), ϕ is the range parameter (also called the length scale) that is referred to as the reduction rate of the correlation between $z(v)$ and $z(v')$ when $\|v - v'\|$ increases. These both σ^2 and ϕ parameters can be varied; therefore, these parameters are referred to as hyperparameters since they correspond to the hyperparameters in neural networks.

2.3.2 Gaussian Markov random field

A GMRF is a spatial process whose realizations contain random variables observed in a spatial field. Moreover, the GMRF not only follows GP [13] but also satisfies additional conditional independence concepts, which is also known as the Markovian property. For a more detailed discussion about GMRF, readers are referred to [15].

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

$$p(z) = (2\pi)^{-\frac{n}{2}} (\det(Q))^{\frac{1}{2}} \exp\{-\frac{1}{2}(z - \mu)^T Q (z - \mu)\}. \tag{2.3}$$

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

$$z_i \perp z_{-\{i,N_i\}}|z_{N_i}, \tag{2.4}$$

for $i = 1, ..., n$, where $z_{-\{i,N_i\}}$ is all the elements in z but not z_i and z_{N_i}. Rue *et al.* [15] derived that the mean μ is not related to the pairwise conditional independence properties of z, so the result is 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 influences the computational efficiency of a GMRF substantially. Thus, the GMRF can be greatly exploited by the resource-constrained mobile robotic wireless sensor networks. In particular, the prediction can be directly obtained by the precision matrix Q with conditional interpretation.

Theorem 2.2. *([15], Theorem 2.3) Let z be a GMRF with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean μ and precision matrix $Q \succ 0$, then*

$$\begin{aligned} \mathbb{E}(z_i|z_{-i}) &= \mu_i - \frac{1}{Q_{ii}} \sum Q_{ij}(z_j - \mu_j), \\ prec(z_i|z_{-i}) &= Q_{ii}, \\ corr(z_i, z_j|z_{-i}) &= -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \quad i \neq j. \end{aligned}$$

2.4 Gaussian Process Regression

After reviewing the random field models in the previous section, the regression technique is now described, which is utilized to predict continuous quantities of the physical process.

Consider a data set of n observations $\mathcal{D} = \{(v_i, y_i)|i = 1, ..., n\}$ collected by the wireless sensor network, where v_i is a location vector of dimension D and y_i is a scalar value of noise corrupted output. For simplicity of notations, let $v := (v_1^T, v_2^T, ..., v_n^T)^T \in \mathbb{R}^{nD}$ denote as an $n \times D$ location matrix, and the outputs are aggregated in a vector $y := (y_1, y_2, ..., y_n)^T \in \mathbb{R}^n$. The corresponding vector of noise-free observations is referred to as $z := (z(v_1), z(v_2), ..., z(v_n))^T \in \mathbb{R}^n$. As discussed in Section 2.3.1, the prior z can be described as

$$z \sim \mathcal{N}(\mu, \Sigma_{zz}), \tag{2.5}$$

where $\mu \in \mathbb{R}^n$ is the mean vector obtained by $\mu_i = \mu(v_i)$, and Σ_{zz} is an $n \times n$ covariance matrix whose elements can be computed by $\Sigma_{zz}[i, j] = cov(v_i, v_j)$. The advantage of the GP formulation is that the combination of the prior and noise can

be implemented exactly by matrix operations [18]. Therefore, the noisy observations can be normally distributed as

$$y \sim \mathcal{N}(z, \tau^2 I), \tag{2.6}$$

where τ^2 is a noise variance and I is an $n \times n$ identity matrix. Note that the GP models and all formulas are always conditional on the corresponding locations. In the following, the explicit conditioning on the matrix v will always be neglected.

Given the observations, the objective of probabilistic regression is to compute the predictive prediction of the real values $z_* = z(v_*)$ at some interested points v_*. In [13], Rasmussen *et al.* demonstrated that the GP has a marginalization property, which implies that the joint distribution on random variables at v and v_* is Gaussian, specified by

$$\begin{bmatrix} y \\ z_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} \Sigma_{zz} + \tau^2 I & \Sigma_{zz_*} \\ \Sigma_{z_*z} & \Sigma_{z_*z_*} \end{bmatrix} \right), \tag{2.7}$$

where μ_* and $\Sigma_{z_*z_*}$ are the mean vector and the covariance matrix of z_*. $\Sigma_{zz_*} (= \Sigma_{z_*z}^T)$ is the cross-covariance matrix between y and z_*.

In probabilistic terms, the conditional distribution at predicted positions of v_* given y is derived as follows.

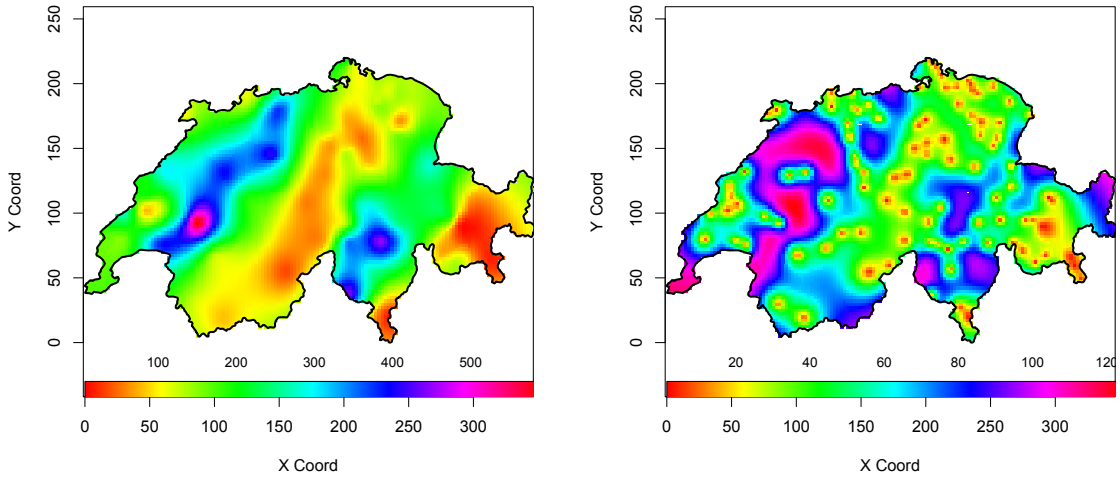
$$\mu_{z_*|y} = \mu_* + \Sigma_{z_*z}(\Sigma_{zz} + \tau^2 I)^{-1}(y - \mu), \tag{2.8}$$

and

$$\Sigma_{z_*|y} = \Sigma_{z_*z_*} - \Sigma_{z_*z}(\Sigma_{zz} + \tau^2 I)^{-1}\Sigma_{zz_*}, \tag{2.9}$$

where $\mu_{z_*|y}$ and $\Sigma_{z_*|y}$ are posterior mean vector and covariance matrix of z_*, given y.

As a consequence, using observations at locations in set v, quantities at unobserved locations, v_*, can be predicted. For instance, there have been used 100 observations of the rainfall data set that was collected by 100 gauges deployed in Switzerland to predict the amount of the rainfall throughout the whole country, which is illustrated in Fig. 2.1.



(a) *(b)*

Figure 2.1: Predicted rainfall (1/10th of a mm) in Switzerland: (a) Posterior mean, (b) Posterior variance. Colour bars represent range of rainfall. X Coord and Y Coord are in km.

2.5 Voronoi Partition

In this section, the concept of Voronoi partitions [19] that will be employed to design the sampling paths for mobile robotic sensors is reviewed.

Consider a set of n distinct points $\mathcal{P} = \{p_1, ..., p_n\}$ in $\mathcal{Q} \subset \mathbb{R}^d$. Let $\mathbb{P}(\mathcal{Q})$ define the collection of finite subsets of $\mathcal{Q}$. At fixed locations, the points could be divided by the optimal partitions of $\mathcal{Q}$, which are a set of Voronoi partitions

$\mathcal{V}(\mathcal{P}) = \{V_1(\mathcal{P}), ..., V_n(\mathcal{P})\} \subset \mathbb{P}(\mathcal{Q})$ defined by, for each $i \in \{1, ..., n\}$

$$V_i(\mathcal{P}) = \{q \in \mathcal{Q} | \parallel q - p_i \parallel \leq \parallel q - p_j \parallel, \quad \forall j \neq i\}.$$

In equivalent words, $\mathcal{V}_i(\mathcal{P})$ is the set of the positions of $\mathcal{Q}$ that are closer to p_i than to any of the other points in $\mathcal{P}$. $\mathcal{V}(\mathcal{P})$ is also called the Voronoi diagram. Two points p_i and p_j are Voronoi neighbours if their Voronoi cells share an edge. In this concept, p_i is also referred as the centroid of its own Voronoi cell, $V_i(\mathcal{P})$.

In terms of robotic networks, it is assumed that in a given period of time, a mobile sensor can travel a distance $r \in \mathbb{R}_{>0}$, which forms a concept of range-limited Voronoi partition.

Definition 2.3. (*[19], Definition 2.5*) *Given a set of n distinct points* $\mathcal{P} = \{p_1, ..., p_n\}$ *in* $\mathcal{Q}$*, and a positive real number* $r \in \mathbb{R}_{>0}$*, the r-limited Voronoi partitions inside* $\mathcal{Q}$ *is the collection of sets* $\mathcal{V}_r(\mathcal{P}) = \{\mathcal{V}_{1,r}(\mathcal{P}), ..., \mathcal{V}_{n,r}(\mathcal{P}) \subset \mathbb{P}(\mathcal{Q})\}$ *specified by*

$$\mathcal{V}_{i,r}(\mathcal{P}) = \mathcal{V}_i(\mathcal{P}) \cap \overline{B}(p_i, r), \quad i \in \{1, ..., n\},$$

where $\overline{B}(p_i, r)$ *is the closed sphere of radius* r *centered at* p_i.

2.6 Mobile Robotic Wireless Sensor Network

In this section, a model for the MRWSNs and an allowable movement region of individual MRWS to avoid physical collisions are described. These properties are utilized to optimize an adaptive sampling scheme that allows the MRWSNs to estimate and predict efficiently the physical process of interest.

2.6.1 Network model

Consider N spatially distributed mobile sensing agents indexed by $i \in \{1, ..., N\}$. Agents are equipped with identical sensors, and take measurements at times $t \in \mathbb{Z}_{>0}$. At time t, agent i takes a noisy measurement at its current location $s_{t,i} \in \mathcal{Q} \subset \mathbb{R}^d$, specified by

$$y_{t,i} = z(s_{t,i}) + \varepsilon(s_{t,i}), \tag{2.10}$$

where $z(s_{t,i})$ is a random variable, and $\varepsilon(s_{t,i})$ is a noise with a zero mean and a variance τ^2.

For notational simplicity, locations of all N mobile robotic sensors at time t are to be denoted by

$$s_t := (s_{t,1}^T, ..., s_{t,N}^T)^T \in \mathcal{Q}^N,$$

and the measurements made by all N robotic sensors at time t by

$$y_t := (y_{t,1}, ..., y_{t,N})^T \in \mathbb{R}^N.$$

Moreover, the collection of all mobile sensors' locations and the collective measurements from time 1 to t are defined as

$$s_{1:t} := (s_1^T, ..., s_t^T)^T \in \mathcal{Q}^{Nt},$$

and

$$y_{1:t} := (y_1, ..., y_t)^T \in \mathbb{R}^{Nt}.$$

Between measurement instants, each mobile sensor moves according to the discrete dynamics

$$s_{t+1,i} = s_{t,i} + l_{t,i}, \tag{2.11}$$

where $\| l_{t,i} \| \leq l_{max} \in \mathbb{R}_{>0}$, in which l_{max} is the maximum distance an agent can move between time instances.

Following are the major assumptions associated with resource-constrained MR-WSNs considered in this thesis:

A.1: A robotic sensor has a communication range R and is a neighbour of other sensors iff it can communicate with them.

A.2: The interference between mobile sensor nodes is insignificant and there exists a Medium Access Control (MAC) protocol to resolve the collisions on the network.

A.3: In the centralized scheme, at a particular instant, all robotic sensors can transmit their observations to the sink (base station) through multihop paths specified by a specific routing structure.

Remark 2.4. *A.1* and *A.2* stem from the technical issues in the wireless sensor networks. *A.3* is required for performing the estimation in a centralized strategy. ■

2.6.2 Physical collision avoidance

This scheme originally presented in [20] is adopted to preserve the safe navigation of the MRWSs. Let ∂V denote the boundary of a set V, and *dist* denotes the distance

function. The $\omega-$contraction of a set V, with $\omega > 0$, is the set

$$V_\omega = \{q \in V | dist(q, \partial V) \geq \omega\}.$$

This subsection takes into account the regions in which the mobile sensors are allowed to move. In other words, one imposes a minimum distance requirement between MRWSs to avoid physical collisions.

Suppose that $\omega \in \mathbb{R}_{>0}$ is a desired safety buffer width and is small enough compared to the size of $\mathcal{Q}$. To make sure that each mobile agent never collides with each other, it is proposed to utilize the contraction concept in geometry. Take the robotic sensors' locations into account at time step of t, s_t, let $(V_i(s_t))_{\frac{\omega}{2}}$ denote the $\frac{\omega}{2}$-contraction of Voronoi cell of $s_{t,i}$, $V_i(s_t)$. Therefore, the allowable movement region of each mobile sensor is defined by

$$\Omega_i^{(t)} = (V_i(s_t))_{\frac{\omega}{2}} \cap \overline{B}(s_{t,i}, l_{max}), \tag{2.12}$$

where $\overline{B}(s_{t,i}, l_{max})$ is the sphere of radius l_{max} centered at $s_{t,i}$ and $\Omega_i^{(t)} \subset \mathcal{Q}$.

Alternatively, one has $dist(\Omega_i^{(t)}, \Omega_j^{(t)}) \geq \omega$. An illustration of the concept of allowable movement for a 5 robotic sensors network is shown in $\mathbb{R}^2$ in Fig. 2.2.

2.7 Submodularity Function

In this section, the concept of submodular function [21] and its properties are introduced, which will be exploited to guarantee solutions of the adaptive sampling

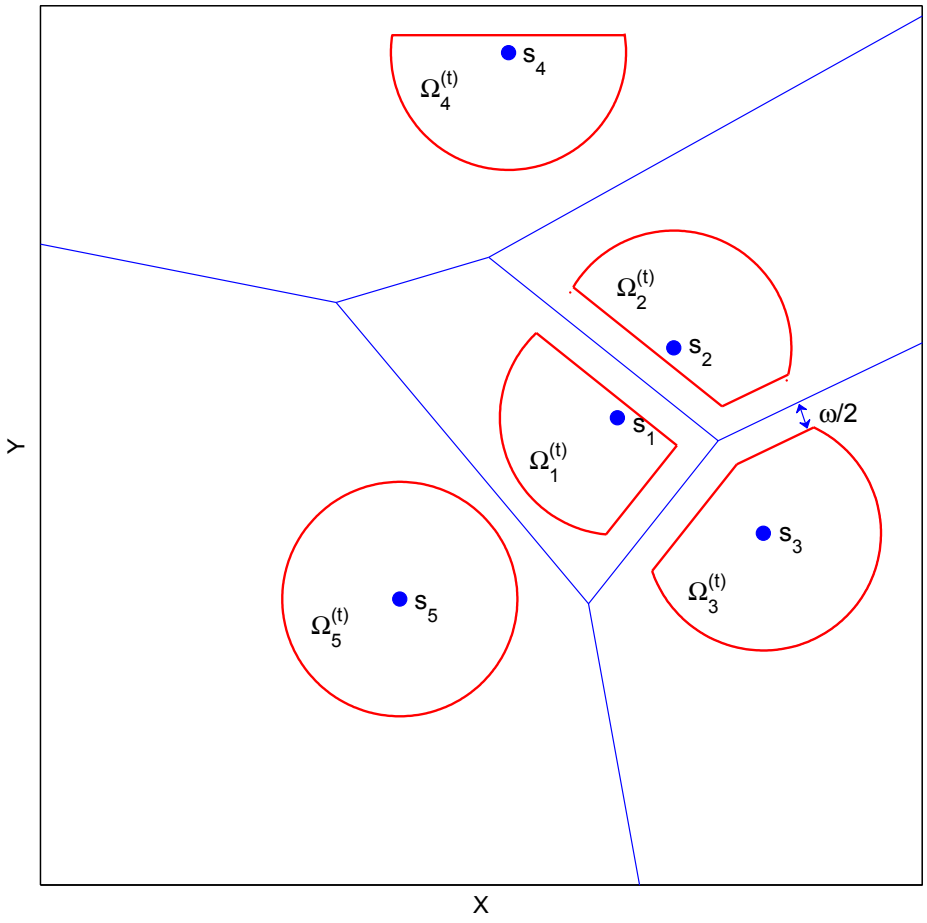


Figure 2.2: Allowable movement regions of sensors at time step t (red) versus Voronoi cells (blue).

optimization problem. Submodularity is a characteristic of set functions. For instance, consider the function $f : 2^{\mathcal{Q}} \rightarrow \mathbb{R}$, for each subset $S \subseteq \mathcal{Q} \subset \mathbb{R}^d$, f returns a value $f(S)$. Hereby, $\mathcal{Q}$ is a finite set, generally called the ground set. The submodularity can be also regarded as a discrete analog to convexity in the continuous optimization [22]. Like convexity functions, submodular functions arise in many applications as well. Let us suppose that $f(\oslash) = 0$, that is, the empty set bears no value.

Definition 2.5. (*Submodularity*) *A function* $f : 2^{\mathcal{Q}} \to \mathbb{R}$ *is submodular if for every subset* $S_1, S_2 \subseteq \mathcal{Q}$,

$$f(S_1 \cap S_2) + f(S_1 \cup S_2) \leq f(S_1) + f(S_2). \tag{2.13}$$

In other words, as described by Nemhauser *et al.* [23], a set function $f : 2^{\mathcal{Q}} \to \mathbb{R}$ is *submodular* if for all subsets $S_1 \subseteq S_2 \subseteq \mathcal{Q}$, and $s \in \mathcal{Q} \setminus S_2$ it holds that $f(S_1 \cup s) - f(S_1) \geq f(S_2 \cup s) - f(S_2)$. There are primarily two types of the submodular function: *monotone* and *non-monotone*. In this work, the monotone submodular function will be taken into account.

Definition 2.6. (*Monotonicity*) *A function* $f : 2^{\mathcal{Q}} \to \mathbb{R}$ *is monotone if for every subset* $S_1 \subseteq S_2 \subseteq \mathcal{Q}$, $f(S_1) \leq f(S_2)$.

Note that a function f is monotone iff its first derivative (which is not necessarily continuous) does not change the sign [24].

2.8 Summary

In this chapter, the fundamental theories of the GP, the GMRF, Voronoi cell, and submodularity function have been covered. Furthermore, the GP regression, a key basis for predicting the environmental field, has been reported. A network model of mobile sensing agents and cooperative movements of these sensors have been successively outlined. These preliminaries are necessary for reading and understanding of the following chapters.

Chapter 3

Literature Review

3.1 Introduction

This chapter presents an up-to-date review of the spatial prediction problem by utilizing stationary and mobile robotic wireless sensor networks. First, the problem of selecting the best subset of stationary wireless sensors monitoring environmental phenomena in terms of sensing quality is surveyed. Then, predictive inference approaches and sampling algorithms for mobile sensing agents to optimally observe spatially physical processes in existing works are analysed.

3.2 Sensor Selection in SWSNs

There is a rich literature on the selection of sensor observations for the purpose of coverage [25–27], target tracking and localization [28–30], signal estimation [31], event detection [32], and spatial prediction.

3.2.1 Criteria for sensor selection

In statistics, selecting observations has been considered as an experimental design problem [33–36]. 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 were constructed based upon the properties of the inverse moment matrix. For instance, D-optimality considers the determinant [37], A-optimality examines the trace and E-optimality calculates the maximum eigenvalue [34].

Recently, research attention to the sensor selection has concentrated on selecting observations in order to maximize the quality of parameter estimation [38–42] with a special focus on linear models that are often coupled with a stochastic measurement error term. The sensor selection metric can be formulated in a Bayesian framework [43] in which it is supposed to have the knowledge prior to carrying out the experiments. In equivalent words, combining the prior probability distribution of the parameter space with the observations, the design criteria can be derived. The sensor selection criteria are also defined based upon scalar functions of the Fisher information matrix or the Bayesian Fisher information matrix [40]. 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 [14, 44–50].

3.2.2 Algorithms for sensor selection

In terms of sensor selection algorithms, one can simply process all direct enumerations of $\binom{n}{k}$ possible choices and pick the best subset of k sensors out of n potential ones having the minimal prediction error. It can be seen that this straightforward approach has practical implications depending on the values of n and k and thus motivates more structured methods. In [51, 52], the global optimization techniques such as branch and bound were employed to exactly solve this problem. Nevertheless, since the sensor selection problem, which can be viewed as a combinatorial optimization problem, is NP-hard [53, 54], these accurate approaches are often computationally intensive [55], even with modest values of n and k, and not attractive in real world solutions.

In an effort to improve the model parameter estimation, there have been some interesting methods proposed. For instance, in [38], Joshi *et al.* proposed the heuristic method based on convex optimization [33] for the sensor selection problem. The heuristic approach in [38] utilizes a relaxation technique to convert a discrete optimization problem of sensor selection into a continuous optimization problem. Gupta *et al.* [56] in their work represented a stochastic sensor selection algorithm that selects sensor locations randomly by the use of a probability distribution. Maximum information in the estimation of the state variables can be obtained by a mixed-integer semi-definite program approach [48]. In [41], a binary particle swarm optimization technique was employed to choose a subset of sensors so that the error in parameter estimation is minimized.

Based on Bayesian experimental design, the information-theoretic approaches such as entropy [57] or mutual information [58] were proposed to consider the prediction uncertainty of the random variables at unobserved locations in space. The

greedy heuristic algorithms based on these information-theoretic models together with Gaussian processes proposed by Cressie [12] can obtain near-optimal solutions for the sensor selection problem. These algorithms were demonstrated in the works [14, 44–46, 53, 59, 60]. The premise behind the entropy approach is to minimize the uncertainty of conditional entropy of unobserved locations, given observations. Under Gaussian assumption, Ko *et al.* [53] proposed a greedy suboptimal algorithm by reorganizing the maximization of joint entropy of a chosen set as maximizing the determinant of the covariance matrix of random variables at chosen locations. However, as shown in [60], the entropy method tends to pick locations along the border of interested space causing sensed information waste. To address the drawbacks of the entropy approach, Guestrin *et al.* [60] and Krause *et al.* [14, 44–46] proposed a new method based on the mutual information. In this method a subset k from potential n sensor locations is selected such that the mutual information between the selected subset and the rest of the sensor locations is maximal. In the context of the sensor selection problem, this maximum mutual information is obtained indirectly.

The greedy approximation [61] is a simple, but often used, algorithm for combinatorial optimization problems which can be directly applied in the sensor selection problem. By building up a solution piece by piece, the greedy approximation algorithm may complete the computational tasks in a few seconds. In every iteration, one sensor is chosen and moved to the selected set. Once a sensor is moved to the selected set it is impossible to remove it from the selected set during later iterations. Nevertheless, this prime disadvantage of the algorithm voids the solution to reach optimal values. Another prominent heuristic algorithm to solve combinatorial optimization problem is the genetic algorithm. The genetic algorithm imitates the evolutionary process of nature in which a solution deputizes for the organisms' genetic string. Yao *et al.* [62] in their work illustrated the use of the genetic algorithm

to solve the sensor placement problem. Such an approach, however, can be very expensive for some computational costs when going over a population of individuals [63].

3.3 Sensor Placement in MRWSNs

In the wireless sensor network and robotic research community, mobile robotic sensor nodes have attracted much recent attention due to their vital impact on applications such as surveillance, environmental monitoring, wildlife detection and urban search and rescue operations [64]. The mobile ability of robotics can be utilized to improve performance in WSNs such as node localization, data collection, data aggregation and detection and reaction of failed nodes. For instance, in [65], with characteristics of miniature size, lower cost, low power and simple fabrication, Robomote, a tiny mobile robot, is considered as a necessity for establishing a large-scale distributed network of mobile sensors. An inexpensive and mobile modular robot named CotsBots developed by Bergbreiter *et al.* in their work [66] can also be used for the purpose of integrating the robots into a sensor network environment. Another robotic platform, MICAbot, was designed for large-scale mobile sensor networks, where considerations involved focus on cost, size, functionality, modularity and structural stability [67].

With robots embedded in the sensor network, mobile robotic wireless sensor nodes provide a considerable benefit to connectivity, cost, reliability, and energy efficiency throughout the network as compared with a stationary wireless sensor network [68, 69]. Furthermore, the mobility of wireless sensors allows enhancement of the connectivity in a sparse WSN. For instance, in [70], mules are proposed to pick

up data from wireless sensors when they are in close range, to buffer it and to drop off data to the wired access points. Authors in [71] propose to move mobile sensors in the area of deployment as intermediate locations. However, the most widely utilized advantage of mobile robots in the WSNs is to efficiently improve data collection [72–75]. The combination of the new paradigm brings in new opportunities to reduce and better distribute the energy usage within the network. Particularly, mobile agents have the potential to decline the number of hops in data transmission, which justifies the reduction of energy used to transmit the data and avoids the funnelling effect in centralized WSNs. In addition, there are other benefits of employing mobile robotics to address many of the critical issues in the WSNs, such as service [64, 76, 77], data aggregation [78], coverage [79, 80], localization [81], target detection [82], fault-tolerance [83], navigation [81] and path planning [84].

3.3.1 Sensor placement in WSNs

The term sensor placement has been used in various contexts of the WSNs. For instance, in the work by Fletcher *et al.* [76], an algorithm named Randomized Robot-assisted Relocation of Static Sensors (R3S2) is proposed to utilize mobile robots for the purpose of servicing the WSNs. Specifically, in R3S2, robots travel around the network to discover sensing holes that are not being covered due to unpredictable node failure, then move redundant sensors to the uncovered area. Similarly to the work [76] in terms of servicing systems, authors in [77] examine robot task allocation and robot task fulfilment in wireless sensor and robot networks. For example, the network will organize a group of robots to achieve a desired goal, while other moving robots will recharge batteries on the nodes in the regions of the network. Moreover, there are algorithms proposed in [85, 86] for placement of

relocatable nodes in order to improve network connectivity. Considering a large-scale static WSN, [83, 87] propose approaches that also employ mobile robots to detect and report failed sensors and then replace these broken nodes.

3.3.2 Sensor placement for spatial prediction

In the context of spatial fields, the sensor placement problem for predicting the spatial phenomena has been investigated, which has considerably contributed to a number of interesting approaches and algorithms. In what follows, related works of sampling strategies for the MRWSs in the spatial prediction are summarized. In [88], locational optimization had been proposed in optimizing the mobile sensor network locations with respect to a known event probability density in the spatial environment. However, physical processes are not known a priori, and a density function can be only established when measurements are to be taken. Leonard *et al.* [89] employed a linear model to predict an ocean field and proposed a performance metric that minimizes uncertainty in a model estimate of the sampled field to derive a parameterized family of paths for the mobile sensor networks. By combining with coverage control [88], Martínez [90] derived a distributed prediction scheme based on a nearest neighbour interpolation approach for field estimation in mobile sensor networks. The primary disadvantage of the linear models in both [89] and [90] is that the model parameters must be known a priori. In terms of the compressive sensing framework, Huang *et al.* [91] maximized the entropy of next measurements to find the next most informative positions for networked mobile sensors to reconstruct an unknown sensing field. By defining a graph whose vertices and edges are considered as a single robot's visiting locations and moving paths, respectively, a path planning algorithm for a mobile robot was proposed in [92] so as to maximize information

gained from measurements of a spatio-temporal phenomenon. Ouyang *et al.* [93] utilized a Dirichlet process mixture of Gaussian processes to model the continuous-valued spatial field and then proposed a decentralized multi-robot Bayes-optimal active learning policy for multiple robotic sensors so that the most informative partitions of a non-stationary phenomenon are to be sampled. The policy is then resolved by a greedy algorithm. In [94], La *et al.* used consensus filters to propose a distributed sensor fusion algorithm for multiple mobile sensor nodes to automatically adjust their movements to obtain quasi uniform confidence of estimating and mapping a scalar field. By utilizing a Kalman filter for a downsampled system, [8] has developed the optimal sampling strategies in order to balance the estimation quality and the sensor network lifetime.

In terms of statistics, the spatially environmental phenomena are efficiently and effectively modelled by the Gaussian processes (GPs). Therefore, Suh *et al.* by their work [95] represented an environmental monitoring navigation strategy for a sensing robot, in which the information gain along the robot's trajectory is maximized. Considering a team of sensing agents, Graham *et al.* described the random field models by tools from geostatistics, that is Kriging, and proposed to utilize either a known [20] or an unknown [96] covariance function. The sensor network includes static computing nodes and mobile sensing agents taking measurements of a random process. The static nodes compute the gradient of variance and send control commands to robotic sensors. Nevertheless, in [97], maximizing joint entropy of measurements in a distributed fashion was investigated to consider the adaptive sampling paths, where motion coordination was designed based upon Voronoi partitions.

Popa *et al.* proposed extended Kalman filter [98] and non-linear extended Kalman filter [99] based adaptive sampling approaches to optimally estimate the

parameters of distributed variable field models. These schemes also aim to decline the uncertainty in the knowledge of a linear-in-parameters field distribution (linear model). Choi *et al.* [100] introduced a Kalman filter based technique to learn the parameters of a physical spatio-temporal process model and then presented criteria to navigate mobile sensors throughout an environment in order to maximize a specified performance. In [101], the authors delineated an objective function of a trajectory optimization problem for robotic wireless sensors as a deterministic optimal control problem. In other words, the objective function is imposed on minimizing the variance of the estimate of the environment, which is eventually resolved by a dynamic programming algorithm. Euler *et al.* in their work [102] introduced a sampling navigation scheme for a group of unmanned aerial vehicles to simultaneously observe multiple concentration levels of an atmospheric plume. The authors incorporated the estimations of the concentration into the uncertainty at these levels to find out the optimal sampling locations. Wu *et al.* [103] proposed a switching scheme for a team of mobile sensors to switch between individual exploration and cooperative exploration as they were exploring an unknown environment. It is proposed that the density field is defined by mapping uncertainty, it changes with every measurement taken; a symmetry-preserving coordinated motion strategy for sensing agents was delineated in [104] to provide optimal measurements.

Cortés in [105] developed a distributed Kriged Kalman filter for robotic wireless sensors to predict the field of interest. A consensus algorithm is implemented on new measurements to calculate state predictions of the field. A gradient based controller was designed to drive the mobile wireless sensors to take optimal samples so that the variance of the estimate error is decreased. In [106], Oh *et al.* proposed a distributed learning algorithm for robotic sensing systems, called cross validation.

In this proposition, each mobile sensor learns model parameters using its own measurements and sends the learned parameters to other sensors to validate until all mobile sensors share the best fitness results.

Xu *et al.* primarily used the GP regression for estimating and predicting the generally scalar field and designed optimality criteria based on the Fisher information matrix [107] and the average of the prediction error variances [108], [109] for the optimal sampling paths of the MRWSNs. A maximum likelihood recursive filter is proposed to learn unknown model parameters for the covariance function as well as the basis functions. More specially, in [109], a theoretical foundation of the GP regression with a subset of measurements is derived for the MRWSNs. Based on this proposition, a gradient descent based algorithm is delineated to manage the sensing robot coordination. The authors in [110] introduced the Bayesian optimization based technique for the purpose of choosing the much more relevant informative locations for the MRWSs in the GP modelled field. Moreover, a new utility function based on travelled distances of the MRWSs was also proposed to be used indirectly in trade-off between the exploration and the exploitation of the mobile sensors. However, so far, to the best of available knowledge, there have been no existing works which analyse the bounds of solutions and the computational complexity of the algorithms in resource-constrained MRWSNs.

3.4 GMRF for Spatial Prediction

With respect to the GP model, the computational issues have always been a bottleneck, since the computational complexity of factorizing dense covariance matrices is cubic in dimension, which is known as "the big n problem" [111]. In the context

of statistics, this challenge has been dealt with by a reduced-rank approximation of the Gram matrix [112], a sparse greedy approach [113], and a sparse GP [114]. Recently, [109] proposed an approach to diminish the computational complexity of a large dense covariance matrix by shortening mobile sensor observations. The disadvantage of this method is that the model parameters need to be known in advance. Other efforts are to represent a continuously spatial process by a discretely indexed Gaussian field. In other words, [115–118] have attempted to enhance the computational complexity in modelling the spatial field by replacing the GP by a computationally efficient Gaussian Markov random field (GMRF) [15]. The GMRF is considerably specified by a sparse precision matrix that makes it substantially advantageous to effective computation. The sparsity property of the precision matrix is constituted by a conditional independence concept in which conditional distribution of every random variable only depends on its neighbours. As a consequence, the sparsity of the precision matrix allows the GMRF to have received more and more attention for resource-constrained MRWSNs as compared to the standard GP [101, 119–122]. In their work [101], Ny *et al.* introduced an estimator by the use of the Kalman filter, where the sensor trajectory optimization problem based upon an information criterion is in turn a deterministic optimal control problem. In addition, [119] represented an approach in which the physical phenomenon of interest is regularly discretized and modelled by a GMRF. In this proposition, the hyperparameters of the GMRF model are supposed to be known and chosen with a support. The authors in [122] introduced a new class of a GP built on a GMRF for modelling the spatial process. Nonetheless, this technique proposes a model with a known precision matrix. More interestingly, Jadaliha *et al.* [120] investigated the GMRF to tackle the simultaneous localization and spatial prediction problem in a fully Bayesian fashion. These authors also considered localization uncertainty of the

mobile sensing agent in the spatial prediction utilizing the GMRF in another work [121]. However, the approaches proposed by the works [119–122] are limited to a representation of a regular lattice, which requires the model parameters to be known a priori.

3.5 Summary

In this chapter, a detailed literature review on the spatial prediction by the utilization of the WSNs has been given. Various criteria for the sensor selection problem in the SWSNs have been proposed, which have been then addressed by different approaches, in the literature. Many other methods to resolve centralized and distributed spatial prediction issues and to find the sampling paths for the MRWSs in the sensor placement problems have been introduced. Some disadvantages of the existing techniques have been also analysed. However, there are some main problems that have not been comprehensively considered in the literature such as the accuracy of the prediction results, the computational time of finding the efficient sampling paths for the MRWSNs, and the computational complexity of factorizing the covariance matrix in the GP regression. Meeting these challenges is the main focus behind the contribution of proposals in this thesis. In the next chapters, these contributions are contemplated in more detail.

Chapter 4

Sensor Selection for Spatial Prediction

4.1 Introduction

In most existing works of the spatial prediction in the literature, the sensor selection problem has been indirectly resolved based upon various criteria such as minimizing the errors of the parameter estimation [38, 43, 56], maximizing the mutual information gained from measurements [14, 44–46, 60] and maximizing the largest entropy reduction in the estimates of the state variables [48]. However, the direct criterion in estimating the quality of sensing which demonstrates the prediction errors at unobserved locations of interest has not been fully considered. In this chapter, a novel performance metric for the sensor selection problem is proposed which reflects the root mean square error (RMSE) for each predicted location and is defined as an *average RMSE* (ARMSE). Moreover, a simulated annealing based algorithm for selecting sensor locations that minimizes the ARMSE in all unobserved locations of

interest is also presented. This novel and efficient algorithm starts by selecting a potential subset of cardinality k from the entire population of sensor locations. After that, it iteratively attempts to substitute the members of the selected subset by its neighbours according to the optimization criterion. Importantly, in each iteration, the algorithm selects the neighbours of a given member of the selected subset based on Euclidean distances which guarantees the full coverage of the space.

Another important contribution in this chapter is the first stage of considering the sensor selection problem in constrained environments. Particularly, a case study was conducted on the data generated by the wastewater model developed by Sydney Water. In this work, it is proposed that the gas phase H_2S monitored is modelled using a Gaussian process (GP) [13], [12], a nonparametric Bayesian technique for modelling spatially correlated data. In fact, GP relies on correlations among all nearby observations in a free space. Nevertheless, in the gravity sewer system, the H_2S in the air is constrained by the sewer pipe itself that prevents it from interacting with some of its spatial neighbours in direct distances. Therefore, in this study, an improved covariance function is computed based on a distance between two locations computed along the lengths of a sewer system. More importantly, since there has been a limited number of gas sensors, it is essential to position the H_2S sensors in the optimal locations in the sewers to capture the H_2S most effectively, which eventually influences the accuracy of predicting the gas phase H_2S. Since each gaseous sensor location has only two upper-stream and down-stream neighbouring sensor locations affecting directly on it, the simulated annealing based algorithm, which is characterized by the neighbouring replacement, is not suitable to resolve the problem of selecting the H_2S sensor locations. Therefore, an efficient approach based on mutual information, a concept from information theory, is proposed to find

the best (most informative) sensor locations out of possible points along the sewer pipes.

The structure of this chapter is arranged as follows: In Section 4.2, a model is introduced for spatially physical fields that are developed based on a GP basis. A technique which illustrates how model parameters and hyperparameters to be learned is also described in this section. Section 4.3 firstly discusses a new criterion of how to choose a subset of sensors among important but expensive observations gathered in an environment that is represented in a free space. A novel algorithm based upon simulated annealing is then proposed to address the sensor selection problem in a near-optimal yet very efficient and effective fashion. In addition, an approach which is proposed to select the best sensor locations in a boundary constrained field of the gas H_2S in the sewer system is considered in Section 4.4. A modified GP based model for this particular spatial problem is also presented and analysed in this section. The proposed sensor selection approaches for unconstrained and constrained environments are extensively validated in Section 4.5. Section 4.6 summarises the chapter.

4.2 Spatial Field Model

In the spatial field of interest $\mathcal{Q} \subset \mathbb{R}^d$, let spatial locations within $\mathcal{Q}$ denote as $v := (v_1^T, v_2^T, ..., v_n^T)^T \in \mathbb{R}^{dn}$. The data consists of one measurement taken at each observed location in v. Let a random vector $y(v)$ denoted by $y(v) := (y(v_1), y(v_2), ..., y(v_n))^T \in \mathbb{R}^n$ describe a vector of measurements. In this study, it is supposed that $v_i, i = 1, ..., n$ varies continuously through $\mathcal{Q}$. The spatial field model is a summation of a large scale component, a random field and a noise. The noise is supposed to be

independent and identically distributed (i.i.d.). Hence, the model is defined by

$$y(v_i) = X(v_i)\beta + \xi(v_i) + \varepsilon(v_i), \tag{4.1}$$

where $X(v_i)\beta$ is the expectation of $y(v_i)$, which is also referred to as a spatial trend function. Particularly, this function is frequently derived through a polynomial regression model, for example a constant, first, or second order polynomial function. $X(v_i) = (1, X_1(v_i), ..., X_{p-1}(v_i) \in \mathbb{R}^p$ is a spatially referenced non-random variable (known as covariate) at location v_i. $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})^T$ is an unknown vector of mean parameters. For instance, it is assumed that $v_i \in \mathbb{R}^2$, that is $v_i = (v_{i1}, v_{i2})$, the second order polynomial expectation is dependent on the coordinates of a sensor location, specified by

$$X(v_i)\beta = \beta_0 + \beta_1 v_{i1} + \beta_2 v_{i2} + \beta_3 v_{i1}^2 + \beta_4 v_{i2}^2 + \beta_5 v_{i1} v_{i2}. \tag{4.2}$$

Therefore, $X(v_i) = (1, v_{i1}, v_{i2}, v_{i1}^2, v_{i2}^2, v_{i1}v_{i2})$. $\xi(v_i) \sim \mathcal{N}(0, cov(v_i, v_j))$ is a stationary GP as presented in the Section 2.3.1, and $\varepsilon(v_i)$ is a noise with a zero mean and an unknown variance τ^2. $cov(v_i, v_j)$ is the covariance function. There is a family of the covariance functions [17]; nonetheless, there are most frequently utilized covariance functions including the squared exponential function that has been introduced in Section 2.3.1 and the Matérn function specified by

$$cov(v_i, v_j) = \frac{\sigma^2}{2^{\nu-1}} \left(\frac{\| v_i - v_j \|}{\phi} \right)^{\nu} K_\nu \left(\frac{\| v_i - v_j \|}{\phi} \right), \tag{4.3}$$

where ν is the Matérn smoothness, K_ν denotes the modified Bessel function of second kind and order $\nu > 0$.

Let $\theta := (\sigma^2, \phi, \tau^2) \in \mathbb{R}^3_{\geq 0}$ denote a hyperparameter vector. The mean parameters β and hyperparameters θ that are hereafter called model parameters of the spatial field model can be estimated by utilizing the maximum likelihood approach [5]. Rewriting the marginal distribution of $y(v)$ given model parameters yields

$$y(v)|\sigma^2, \phi, \tau^2, \beta \sim \mathcal{N}(X(v)\beta, \Sigma_{vv} + \tau^2 I), \tag{4.4}$$

where $X(v)$ is an $n \times p$ matrix of covariates. Σ_{vv} is an $n \times n$ covariance matrix whose elements can be obtained by $cov(v_i, v_j)$. For the sake of simplicity, it is denoted $\Sigma := \Sigma_{vv} + \tau^2 I$. As a result, the log-likelihood function can be obtained by

$$\mathcal{L}(\theta, \beta) = -\frac{1}{2}\{(y(v) - X(v)\beta)^T \Sigma^{-1}(y(v) - X(v)\beta) + \log\det(\Sigma) + n\log(2\pi)\}. \tag{4.5}$$

Suppose that θ is given, β will be the only variable. Thus, it is recognized that the log-likelihood function obtains the maximum value at

$$\hat{\beta} = (X(v)^T \Sigma^{-1} X(v))^{-1} X(v)^T \Sigma^{-1} y(v). \tag{4.6}$$

However, in fact θ is not known; by substituting $\hat{\beta}$ into the log-likelihood function and numerically optimizing this function with respect to σ^2, ϕ and τ^2, the estimated $\hat{\theta}$ can be obtained. Eventually, $\hat{\beta}$ can be computed by the back substitution of $\hat{\theta}$.

4.3 Sensor Selection in Unconstrained Environment

In this section, selecting the best subset of sensor measurements of a spatial field in a free space will be discussed. By exploiting the rich properties of the GP, an efficient metric, the ARMSE, is first derived to estimate the prediction error. Then, the proposed simulated annealing based efficient search and replace algorithm seamlessly changes the members of the selected sensors to orderly represent the entire space. These two together establish an efficient and highly accurate solution to the sensor selection problem in an unconstrained environment.

4.3.1 Metric for sensor selection

Consider a two dimensional sensor network that has a $\mathcal{V}$, $|\mathcal{V}| = n$, set of possible locations, which provide point measurements of some physical quantities. Sensor selection addresses the problem of choosing a subset, $S \subset \mathcal{V}$, at which the subset of the measurements can closely represent the distribution of the physical quantity in the whole space.

Let $\mathcal{V} := \{v_1, v_2, ..., v_n\}$ denote a set of locations. The cardinality of $\mathcal{S}$ is also defined as k. For any subset $\mathcal{S}$, let $y_{\mathcal{S}}$ denote the collections of observations at locations in $\mathcal{S}$. In addition, take into account $\mathcal{U} = \mathcal{V} \backslash \mathcal{S}$ as the set of all elements in $\mathcal{V}$ but not in $\mathcal{S}$, and $z_{\mathcal{U}}$ is a vector of random variables over these unobserved locations. As described in Section 2.4, the joint distribution of $y_{\mathcal{S}}$ and $z_{\mathcal{U}}$ are Gaussian. As a result, the posterior distribution of $z_{\mathcal{U}}$ conditioned on $y_{\mathcal{S}}$, $z_{\mathcal{U}|\mathcal{S}}$, turns out to be

$$\mu_{\mathcal{U}|\mathcal{S}} = X(\mathcal{U})\beta + \Sigma_{\mathcal{U}\mathcal{S}}(\Sigma_{\mathcal{S}\mathcal{S}} + \tau^2 I)^{-1}(y_{\mathcal{S}} - X(\mathcal{S})\beta), \tag{4.7}$$

and

$$\Sigma_{\mathcal{U}|\mathcal{S}} = \Sigma_{\mathcal{U}\mathcal{U}} - \Sigma_{\mathcal{U}\mathcal{S}}(\Sigma_{\mathcal{S}\mathcal{S}} + \tau^2 I)^{-1}\Sigma_{\mathcal{S}\mathcal{U}}, \tag{4.8}$$

where $\mu_{\mathcal{U}|\mathcal{S}}$ and $\Sigma_{\mathcal{U}|\mathcal{S}}$ are mean vector and covariance matrix of $z_{\mathcal{U}|\mathcal{S}}$. $X(\mathcal{S})\beta$ and $X(\mathcal{U})\beta$ ($\Sigma_{\mathcal{S}\mathcal{S}}$ and $\Sigma_{\mathcal{U}\mathcal{U}}$) are expectation vectors (covariance matrices) of $y_{\mathcal{S}}$ and $z_{\mathcal{U}}$, respectively. $\Sigma_{\mathcal{U}\mathcal{S}}(= \Sigma_{\mathcal{S}\mathcal{U}}^T)$, are cross-covariance matrices between $z_{\mathcal{U}}$ and $y_{\mathcal{S}}$; and I is an $k \times k$ identity matrix.

The quality of prediction is generally measured by calculating the errors at unobserved locations. Therefore, the goal is to select a subset $\mathcal{S} \subset \mathcal{V}$ so that it will minimize a certain measure of prediction error at unobserved locations, given observations $y_{\mathcal{S}}$. A typically used function of estimation error is the mean square error. Specifically, it is given by

$$\text{MSE}(\mathcal{U}|\mathcal{S}) = \mathbb{E}[(z_{\mathcal{U}|\mathcal{S}} - z_{\mathcal{U}})^2]. \tag{4.9}$$

Assuming that to be an unbiased estimator (which is the case in GP regression), the mean square error can be estimated to be the variance of $z_{\mathcal{U}|\mathcal{S}}$ [123]. Therefore, $\text{MSE}(\mathcal{U}|\mathcal{S})$ can be calculated by

$$\text{MSE}(\mathcal{U}|\mathcal{S}) = \text{tr}(\Sigma_{\mathcal{U}\mathcal{U}} - \Sigma_{\mathcal{U}\mathcal{S}}(\Sigma_{\mathcal{S}\mathcal{S}} + \tau^2 I)^{-1}\Sigma_{\mathcal{S}\mathcal{U}}). \tag{4.10}$$

This equation can be reformulated to reflect the root mean square error (RMSE) for each predicted location. Further, an average RMSE (ARMSE) is defined, by

normalizing it with $|\mathcal{U}|$ that is,

$$\text{ARMSE}(\mathcal{U}|\mathcal{S}) = \sqrt{\frac{1}{|\mathcal{U}|}\text{tr}(\Sigma_{\mathcal{U}\mathcal{U}} - \Sigma_{\mathcal{U}\mathcal{S}}(\Sigma_{\mathcal{S}\mathcal{S}} + \tau^2 I)^{-1}\Sigma_{\mathcal{S}\mathcal{U}})}. \tag{4.11}$$

Now the aim is to minimize the ARMSE in unobserved locations. Therefore, the objective function for the sensor selection problem is specified as

$$\begin{aligned} &\textbf{minimize } \sqrt{\frac{1}{|\mathcal{U}|}\text{tr}(\Sigma_{\mathcal{U}\mathcal{U}} - \Sigma_{\mathcal{U}\mathcal{S}}(\Sigma_{\mathcal{S}\mathcal{S}} + \tau^2 I)^{-1}\Sigma_{\mathcal{S}\mathcal{U}})} \\ &\text{subject to } \mathcal{S} \subset \mathcal{V} \\ &\qquad\qquad |\mathcal{S}| = k. \end{aligned} \tag{4.12}$$

This cost function can be solved by employing optimization techniques for set function minimization, where $\mathcal{S}$ and $\mathcal{U}$ are set variables. According to [53], it can be seen that the minimization (4.12) is NP-hard. Therefore, it is proposed to address this optimization by a new and efficient simulated annealing based algorithm. Note that the objective function (4.12) depends only on the covariance matrices. However, (4.8) shows that covariance matrices are completely independent from the observations. As a result, the optimization problem in sensor selection can be performed even before deploying the sensors.

4.3.2 Simulated annealing based sensor selection

In the following discussion, the concept of mapping the metal annealing in the physical process to the simulated annealing (SA) approach is reviewed. Then, how the SA method can be utilized to address the sensor selection problem is analyzed and discussed .

4.3.2.1 Simulated annealing technique

Simulated annealing technique [63] is derived from the process of forming metal crystal from metal liquid, which is known as annealing. The annealing allows the metal to rearrange free atoms in a molten state into a crystal lattice when the temperature goes down. The quality of the crystal substantially relies on the cooling process. In particular, if the molten metal is cooled at a very fast rate, the crystallization gets out of equilibrium and causes defects inside the material. Nonetheless, if the temperature is reduced gradually in a carefully controlled manner, the solidification corresponds to the state of minimum internal energy for the metal.

Based upon an analogy of the thermodynamic process of the annealing, the SA algorithm is proposed to find a minimum in the optimization problem. Kirkpatrick *et al.* [55] employed the SA method to address combinatorial optimization problems, where a set of slowly controlled cooling iterations was exploited to make a minimized function obtain an optimal solution. Unlike a normal search approach such as the gradient descent that may lead the optimization problem to a local minimum since the solution is only disturbed in the direction of reducing cost, the SA approach can sometimes interchange the current solution with a new yet worse solution. This uphill movement probably helps the solution jump out of the local minimum and continue to find the global optimal value.

Despite the fact that the SA algorithm can be considered as a random heuristic search method that is not guaranteed to find the optimum, it still provides acceptable solutions within the limit of time and computer memory requirement. An acceptable solution relies considerably on how to simulate the thermal annealing of critically heated solids to reach to the lowest internal energy state for the metal, which is accompanied by a slowly sufficiently controlled cooling rate. The SA approach has

been used in solving nonlinear [124], multiobjective [125], and combinatorial [55] optimization problems with great success.

4.3.2.2 Simulated annealing based algorithm

The aim here is to choose the best subset $\mathcal{S}$ of k sensors from a given set, $\mathcal{V}$. This could be achieved by using the SA algorithm with the objective function (4.12) as given in the algorithm 1. In this case, let $\mathcal{S}$ be defined as $\mathcal{S} := \{s_1, s_2, ..., s_k\}$.

At the beginning, in physical annealing terms, the system is in a high internal energy state, which is equivalent to a large value of $ARMSE(\mathcal{U}|\mathcal{S})$. In each iteration, one neighbour subset of the current subset $\mathcal{S}$ will be chosen randomly, and the objective function will be repeatedly calculated to obtain a new solution for the minimum value. This procedure will be repeated for all the $k \times q$ iterations at each temperature, where q is a predetermined number of locations randomly chosen from all neighbours of $s_i \in \mathcal{S}$ in $\mathcal{U}$. After completing q iterations corresponding to point s_i, the algorithm executes another q iterations for location $s_{i+1} \in \mathcal{S}$. In every iteration, after calculating the new value of the objective function from the new subset $\mathcal{S}$, $\triangle$ of the objective function (OF) will be evaluated based on the difference between its new and old values as

$$\triangle = OF_{new} - OF_{old}. \tag{4.13}$$

If $\triangle$ is less than or equal to zero, then the new subset $\mathcal{S}$ is accepted as a starting solution subset in the next iteration. Otherwise, whether $\mathcal{S}$ is accepted or not depends on a probability function

$$p(\mathcal{S}_{new}) = \exp\left(-\frac{\triangle}{T}\right), \tag{4.14}$$

Algorithm 1 Proposed simulated annealing based algorithm for sensor selection

Input:

1. Set of potential locations $\mathcal{V}$
2. Number of selected sensors k
3. Temperature, T, cooling rate α
4. Choose randomly an initial solution subset $\mathcal{S}$ from $\mathcal{V}$, calculate the objective function (4.12) and keep the objective function value as an optimum

Output:

1. Selected subset $\mathcal{S}$

while T is not sufficiently small **do**
  for $i = 1$ to k **do**
    Find neighbour points of $s_i \in \mathcal{S}$ in $\mathcal{V}$
    for $j = 1$ to q **do**
      Choose randomly a neighbour point of s_i
      Replace s_i by this point in $\mathcal{S}$
      Repeatedly calculate the objective function (4.12)
      Evaluate the change in the objective function
      $\triangle \leftarrow OF_{new} - OF_{old}$
      if $\triangle \leq 0$ **then**
        Accept new subset $\mathcal{S}$
        if $OF_{new} <$ optimum **then**
          Optimum $\leftarrow OF_{new}$
          Optimal subset $\leftarrow$ new subset
        end if
      else
        Generate a random number p in $(0, 1)$
        if $p \leq p(\mathcal{S}_{new}) = \exp\left(-\frac{\triangle}{T}\right)$ **then**
          Accept new subset $\mathcal{S}$
        else
          Reject the randomly chosen point
          Keep the old subset $\mathcal{S}$
        end if
      end if
    end for
  end for
  $T \leftarrow \alpha \times T$
end while
ARMSE $\leftarrow$ optimum
$\mathcal{S} \leftarrow$ optimal subset

where T is a temperature parameter. This implies that if $p(\mathcal{S}_{new})$ is greater than the value of a randomly generated number in the interval $(0, 1)$, then the new subset $\mathcal{S}$ is accepted; otherwise, it is rejected, and the old subset $\mathcal{S}$ is kept for the next iteration.

In the SA approach, the convergence of the algorithm relies considerably on choosing and controlling the parameters such as temperature T and cooling rate α. For example, at the initial iterations, owing to high temperature, the probability for a solution escaping from a local minimum is high. When the temperature gets lower, the ability of an uphill move is less. Hence, the slow reduction of the temperature might help the solution jump out of many local minima. In the implementation, the initial temperature T is chosen by averaging five values of the objective functions calculated from five randomly selected subsets; the cooling rate is empirically chosen to be 0.8.

The effectiveness of the proposed approaches will be demonstrated in Section 4.5.

4.4 Sensor Selection in Constrained Environment

This section proposes a data driven sensor selection approach for spatial prediction of hydrogen sulphide (H_2S) in a gravity sewer system. The gaseous H_2S in the overhead of the gravity sewer is modelled using a GP with an improved covariance function due to the constraints of sewer boundaries. The covariance function is proposed based on the distance between two locations computed along the lengths of the sewer network. An efficient mutual information based sensor selection algorithm is also presented to estimate the number of the most informative sensor observations

and their locations to guarantee a required prediction error. The proof-of-concept study discussed here is the first stage of assessing the model.

4.4.1 Spatial modelling of Hydrogen sulphide

The production and emission of hydrogen sulphide (H_2S) in the air leads to corrosion problems in a sewer network [126] and also gives rise to foul smells adversely affecting any exposed communities [127]. In general, H_2S is generated under anaerobic conditions in which the sulphide is formed by bacterial reduction of sulfate in the wastewater [128]. The gas phase H_2S is emitted from dissolved sulphide in wastewater in gravity sewers and depends on a variety of factors such as velocity of wastewater flow, pH, discharge of industrial wastewater, sulfate, temperature and electron acceptors [129]. With geographical features and year round warm climate, many countries are facing sewer corrosion and odour problems [130]. In Australia, several millions of dollars per year are spent on asset renewal and mitigation [131]. Therefore, in order to mitigate the sulphide related issues in a sewage system, monitoring and predicting H_2S in the sewers are very important. In [132], Hvitved-Jacobsen *et al.* developed the Wastewater Aerobic/Anaerobic Transformation in Sewers (WATS) model that accounts for the impact of a carbon source on sulfate reduction. However, this model has not included coordinates of locations into the prediction function. Sharma *et al.* [128, 129] considered the WATS model under conditions of the biological, physical and chemical changes. The standard method in industry to monitor H_2S dissolved in the wastewater is to collect and determine soluble H_2S in the wastewater samples and on-line monitoring of the gas phase H_2S [130]. In general, the methods in literature require an in-depth understanding of the mathematical models. Additionally, these methods involve highly complex computations.

In this subsection, the data driven machine learning approach is described for predicting H_2S concentration in gravity sewer systems. The uniqueness in this method as compared with the traditional techniques of the spatial prediction such as in geostatistics is the constraints imposed by sewer boundaries (H_2S flow is constrained by a sewer pipe network as demonstrated in Fig. 4.13).

4.4.1.1 Hydrogen sulphide model as a multivariate normal distribution

Consider a network of n gas sensors that are deployed in a sewage system (Fig. 4.13) for monitoring gas phase H_2S. Let $S := \{s_1, s_2, ..., s_n\}$ denote a finite set of spatial sensor locations within a region $\mathcal{Q} \subset \mathbb{R}^2$. Let $y(S) := (y(s_1), y(s_2), ..., y(s_n))^T$ define a vector of the H_2S measurements. The sensor selection problem allows it to predict the H_2S concentration at one of the sensor locations given the other sensor observations. In order to do this, the joint probability distribution needs to be known. In this consideration, it is assumed that the H_2S contents have a multivariate Gaussian distribution [123], specified by

$$p(y(S)) = \frac{1}{(2\pi)^{\frac{n}{2}}(\det(\Sigma_{SS}))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y(S) - \mu_S)^T \Sigma_{SS}^{-1}(y(S) - \mu_S)\right), \qquad (4.15)$$

where μ_S is a mean vector and Σ_{SS} is a symmetric positive definite covariance matrix.

Moreover, since the multivariate normal distribution has a marginalization property, the joint probability distribution on the H_2S contents at any subset, $C \subseteq S$, is also normally distributed.

4.4.1.2 Spatial distribution of hydrogen sulphide model using Gaussian process

The sensor is required to predict H_2S concentration not only at sensor locations but also at unmeasured locations. To this end, it is proposed to utilize the GP regression approach (in Section 2.4). In this chapter, the model of the H_2S measurement consists of an unobserved random field and a noise as follows.

$$y(s_i) = z(s_i) + \varepsilon(s_i), \tag{4.16}$$

where $z(s_i) \sim \mathcal{GP}(m(s_i), cov(s_i, s_j))$ is a GP comprehensively defined by a mean function $m(s_i)$ and a covariance function $cov(s_i, s_j)$. $\varepsilon(s_i)$ is a noise with a zero mean and an unknown noise variance τ^2.

In this case, the mean function of second order polynomial on the coordinates is employed as defined in (4.2),

$$m(s_i) = \beta_0 + \beta_1 s_{i1} + \beta_2 s_{i2} + \beta_3 s_{i1}^2 + \beta_4 s_{i2}^2 + \beta_5 s_{i1} s_{i2}, \tag{4.17}$$

where $s_i := (s_{i1}, s_{i2})$.

The covariance function has been chosen from the family of the covariance functions [17], for example the Matérn function (4.3). It can be clearly seen that elements in the covariance matrix are principally dependent on correlations among all nearby measurements. Nevertheless, in this research, the amount of the gas phase H_2S in a large sewer network does not necessarily correlate in the same way. This can be explained by the constraints of the sewer boundaries in which H_2S is restricted to flow freely from one section to another if they are not directly linked as demonstrated

in Fig. 4.1 For instance, two points A and B are very close in terms of Euclidean

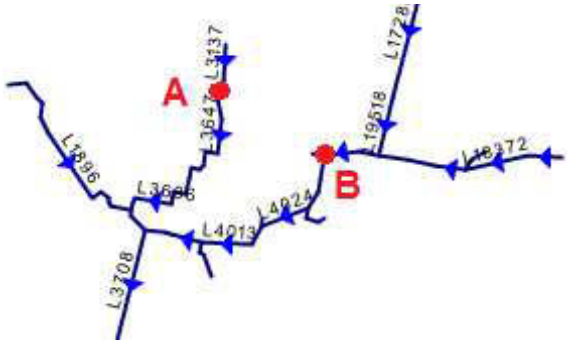


Figure 4.1: Boundary constraints in a sewer system: Although 'A' and 'B' are proximally located, the H_2S contents are "isolated" from each other.

distance in space; nonetheless, the dependence between two random variables at these locations is mainly trivial. As a result, it is improved a covariance function with which the GP can be utilized to train a model of the H_2S concentration and to spatially predict this gas phase H_2S at any unobserved locations in a sewage system. Given H_2S measurements, a GP model of the gas phase H_2S in a sewer network can be trained by employing the covariance function

$$cov(s_i, s_j) = \sigma^2 \exp\left(-\frac{l^2}{2\phi_l^2}\right), \tag{4.18}$$

where l is a distance between two locations s_i and s_j computed along the lengths of sewer pipes, and ϕ_l is the length-scale learned corresponding to l. Notice that in order to use the covariance function (4.18) in modeling the H_2S, it is assumed that the H_2S contents at points which are close in the pipes are likely similar and that the prediction at a point is dependent on information of measurements collected at points around it.

Remark 4.1. The distance l is always greater than or equal to the distance between two locations s_i and s_j which is computed in a free space. In this work, a covariance function based on the squared exponential function is used. However, covariance

functions based on other functions such as Cauchy, exponential, cubic, spherical can easily be derived. ■

Note that all the model parameters σ^2, ϕ_l, τ^2 and β are unknown a priori but can be obtained by maximizing the log-likelihood function as discussed in Section 4.2.

4.4.2 Mutual information based sensor selection

In Section 4.3, a novel approach based on the SA has been proposed to address the NP-hard sensor selection problem in the unconstrained environment. A typical characteristic of this SA based algorithm is to iteratively replace the members of the current selected subset by their neighbours based on Euclidean distance. The more neighbours each member of the selected subset has, the closer the SA based algorithm converges to the optimal solution. Nevertheless, considering locations of H_2S sensors inside the sewer pipes, it shows that each sensor location has only two upper-stream and down-stream neighbouring sensor locations impacting directly on it. Hence, the SA based approach is not appropriate to address the sensor selection problem in the constrained environment. In this section, another efficient maximization criterion for the mutual information technique [14] will be introduced. The proposed criterion 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 that are unlikely within sensor placements. Note that in [14], Krause *et al.* only considered the mutual information between the variables at selected sensor locations and those at unselected sensor locations.

4.4.2.1 Mutual information function

In information theory, mutual information (MI) is a quantity that measures the dependency between two random variables [133], [134]. In other words, it measures how much information is acknowledged in one random variable about another. The MI between two random variables A and B, denoted as $I(A;B)$, is described as the reduction of uncertainty about A when B is known and is specified by

$$I(A;B) = \int_{b\in B}\int_{a\in A} p(a,b)\log\left(\frac{p(a,b)}{p(a)p(b)}\right)\mathrm{d}a\mathrm{d}b$$

where $p(a,b)$ is now the joint probability density functions of A and B, and $p(a)$ and $p(b)$ are the marginal probability density functions of A and B, respectively. It is straightforward that the MI between two variables is 0 iff the two variables are statistically independent.

Moreover, the MI can be represented via Shannon entropy, which is a quantity of measuring uncertainty of random variables, given as follows

$$I(A;B) = H(A) - H(A|B) = H(B) - H(B|A),$$

where $H(A)$ and $H(B)$ are the entropies of A and B, and $H(A|B)$ and $H(B|A)$ are the conditional entropies of A given B and of B given A, respectively.

In a learning task of a selection, if A and B are two physical quantities, $I(A;B)$ can be utilized to reflect the amount of information redundancy between them. Therefore, when the mutual dependence between A and B is high, removing one of them from a selected set would not have much effect on the solution. This will be clarified in the following sensor selection problem.

4.4.2.2 Mutual information based algorithm

Consider the sewage system as a discrete set of locations $D = S \cup U$, where S is a set of possible locations in which sensors can be deployed to observe the H_2S, and U is a set of interested locations in which no H_2S sensor placements are available. The objective of the sensor selection problem is to find the best subset $C \subseteq S$ that can still represent the distribution of the physical quantity in the whole sewage system and effectively predict the gas phase H_2S at unmeasured locations $D \setminus C$. To this end, it is proposed a new evaluation approach based on the MI to select k most informative H_2S data locations along the sewer pipes that most significantly reduce the uncertainty about the predictions in the rest of the space of the sewage network.

Proposition 4.2. Given H_2S observations and the possible gas sensor locations, the best k sensor locations of the most informative subset $C^{opt} \subseteq S$ can be determined by addressing the following optimization problem

$$C^{opt} = \underset{C \subseteq S : |C| = k}{\textbf{argmax}} \det(\Sigma_C)\det(\Sigma_{D\setminus C}), \tag{4.19}$$

where Σ_C and $\Sigma_{D\setminus C}$ are the covariance matrices of $y(C)$ and $y(D \setminus C)$, respectively.

Proof: As the MI criterion discussed in [14], one has

$$\begin{aligned} C^{opt} &= \underset{C \subseteq S : |C| = k}{\textbf{argmax}} I(C; D \setminus C) \\ &= \underset{C \subseteq S : |C| = k}{\textbf{argmax}} H(D \setminus C) - H(D \setminus C|C). \end{aligned}$$

By the chain rule for entropies $H(D \setminus C|C) = H(D) - H(C)$, therefore

$$\begin{aligned} C^{opt} = & \quad \underset{C \subseteq S : |C| = k}{\textbf{argmax}} \quad - [H(D) - (H(C) + H(D \setminus C))] \\ = & \quad \underset{C \subseteq S : |C| = k}{\textbf{argmin}} \quad H(D) - (H(C) + H(D \setminus C)). \end{aligned}$$

Since $H(D)$ is a constant,

$$C^{opt} = \underset{C \subseteq S : |C| = k}{\textbf{argmax}} \quad H(C) + H(D \setminus C).$$

As discussed in Section 4.4.1.1, $y(C)$ is a multivariate Gaussian distribution, then the entropy of $y(C)$ can be computed by

$$H(C) = \frac{1}{2} \log(2\pi e)^k \det(\Sigma_C).$$

Therefore,

$$H(C) + H(D \setminus C) = \frac{1}{2} \log\{\det(\Sigma_C)\det(\Sigma_{D\setminus C})\} + \text{const}, \tag{4.20}$$

where Σ_C and $\Sigma_{D\setminus C}$ are symmetric, positive definite covariance matrices of the random variables at the locations of C and $D \setminus C$. Since the number of the locations of C is given, the logarithm function in (4.20) is monotonic, which completes the proof. ■

It can be clearly seen that the above combinatorial optimization problem as proved by Ko *et al.* [53] is NP-hard. Usually, this issue can be approximately addressed by employing a greedy heuristic algorithm as shown in Algorithm 2.

Algorithm 2 Approximately selecting H_2S sensor locations in sewers

Input:

1. Set of potential locations D
2. Number of selected sensor locations k
3. Hyperparameters, σ^2, ϕ_l, τ^2

Output:

1. Selected subset C

$C \leftarrow \oslash$

while $|C| \leq k$ **do**
  for i = 1 to length of $S \setminus C$ **do**
    $s_i \in S \setminus C$
    C temporarily receives s_i
    Calculate Σ_C and $\Sigma_{D\setminus C}$
    $MI \leftarrow \det(\Sigma_C)\det(\Sigma_{D\setminus C})$
  end for
  $s^* \leftarrow \underset{s\in S\setminus C}{\operatorname{argmax}}\ MI$
  $C \leftarrow C \cup s^*$
end while

4.5 Experiments and Results

In order to illustrate the efficiency of the proposed algorithms, the following experiments were conducted using: *(a)* three real world data sets in unconstrained environments including dataset 1 - an indoor temperature distribution data set from the Intel Berkeley research lab [135], dataset 2 - an outdoor rainfall data set from Switzerland [5], and dataset 3 - an outdoor solid calcium ingredients data set in Brazil [5]; *(b)* the H_2S data sets generated by the integrated sewer network model of Sydney Water.

Note that the experimental data does not include temporal variations.

4.5.1 Unconstrained Environment

For each of the experiments, the parameters $(\beta, \sigma^2, \phi, \tau^2)$ of the GP model were first learned using the likelihood maximization technique as discussed in the Section 4.2. In all the implementations, it was assumed that the spatial trend can be captured by a second order polynomial function that is dependent on the coordinates of the sensor location, as in (4.2). For the Dataset 1 and the Dataset 3, the isotropic squared exponential covariance function (2.2) was used; and for the Dataset 2 the isotropic Matérn covariance function (4.3) was proposed to be utilized.

The proposed SA based algorithm was compared with entropy [53], MI [14] and direct greedy heuristic methods to find out which method has selected k sensor locations out of n possible locations giving minimal ARMSE value.

4.5.1.1 Indoor environment

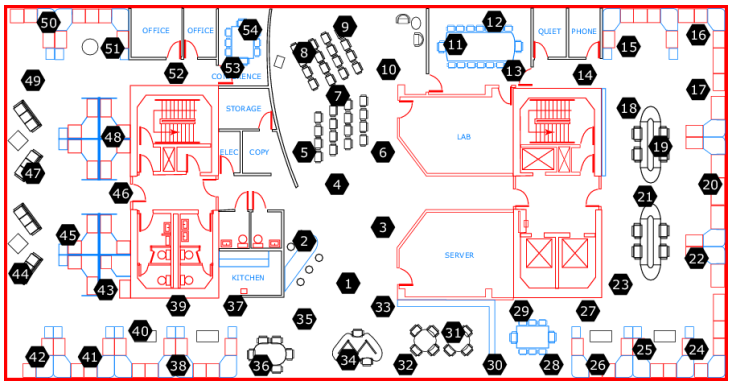


Figure 4.2: A deployment of a sensor network with 54 nodes at the Intel Berkeley research lab.

The Dataset 1 comprises of 54 sensors deployed in the Intel Berkeley research lab according to the diagram illustrated in Fig. 4.2. All 54 sensors were utilized to predict the best possible GP model. The MI and the entropy methods are first used to select the best subsets with the number of sensors ranging from 1 to 20. The

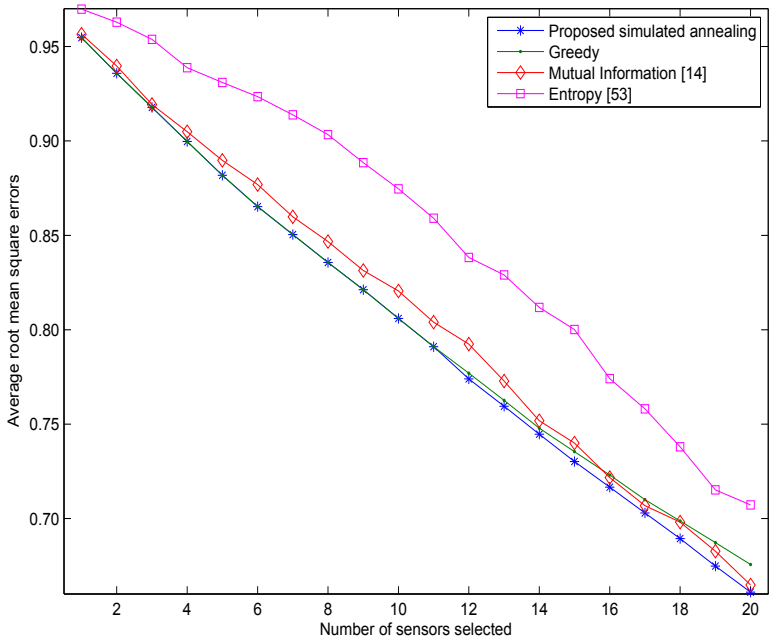


Figure 4.3: Average root mean square errors of temperature distribution of the Intel Berkeley research lab.

computed average root mean square errors (ARMSEs) for these subsets are shown in Fig. 4.3 with the direct greedy heuristic algorithm and the proposed SA based algorithm. Fig. 4.3 clearly shows that the SA based algorithm outperforms the entropy and MI approaches. That is because the SA based approach gradually sweeps the entire space by selecting the best k sensor locations to accurately represent the whole neighbourhood. The SA based method also outperforms the greedy algorithm especially with the increase of a number of sensors. As explained in the Section 4.3.2, it is possible for any sensor location in a neighbourhood to substitute a member in the selected set satisfying the minimal ARMSE criteria. So during this iterative process, the SA based algorithm most likely retains the sensor locations that represent the neighbourhoods better. Note that performance of the greedy heuristic approach degrades as the cardinality of the selected subset increases. Intuitively, with larger selected subsets, there is a higher chance that the algorithm chooses less important

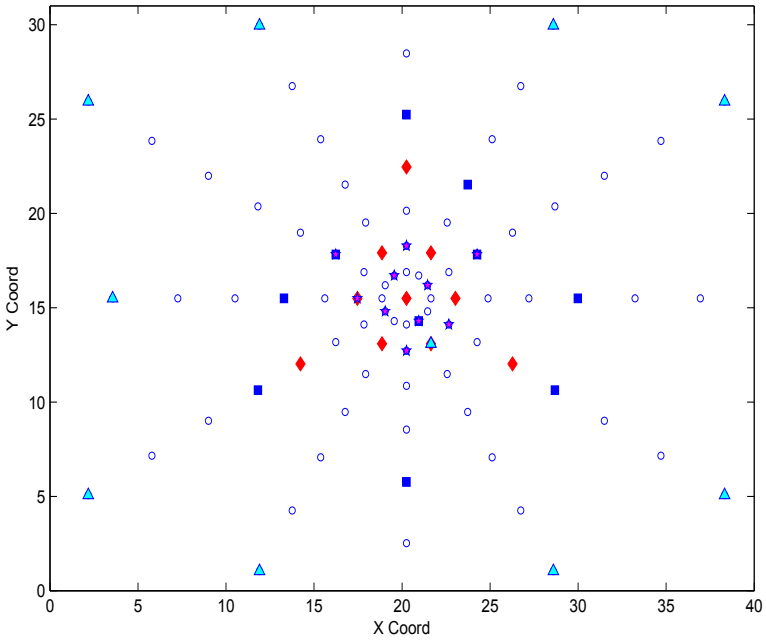


Figure 4.4: Deploying sensors densely in the center: Square - the locations selected by using SA based algorithm; diamond - the locations selected by using direct greedy method; pentagon - the locations selected by using mutual information approach; and triangle - the locations selected by using entropy technique. X Coord and Y Coord are in m.

sensor locations during the early iterations to the selected subset resulting in an apparent performance reduction.

Consequently, the proposed algorithm was tested for two special scenarios generated from the GP model surface corresponding to the Dataset 1 to analyse the effects of spatial distribution of sensors. The experiments aimed to evaluate how the algorithms select sensor locations in the extremely changed deployments. Then, the distribution of selected sensors was used to compute the accuracy in those studied cases. For instance, the first scenario was generated by deploying 87 sensors densely in the center of the space and sparsely in the boundary, as shown in Fig. 4.4. Therefore, as expected, Fig. 4.5 illustrates that the entropy approach suffers significantly due to the fact that it tends to select the subset along the border leading to sensed

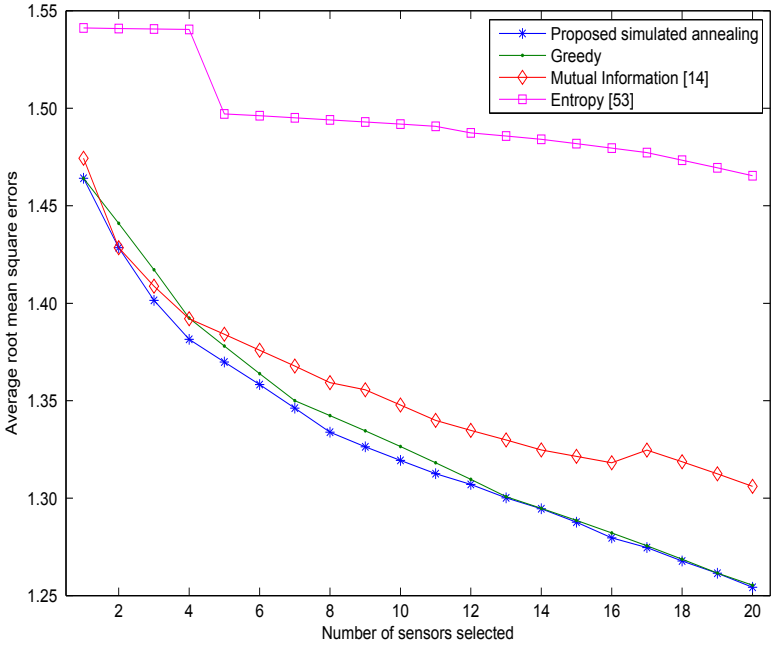


Figure 4.5: Average root mean square errors of densely central deployment.

information waste [60]. In contrast to entropy, MI tends to select sensor locations around the centre. However, SA selects sensor locations considerably sparsely resulting in better collection of sensor measurements to represent the region more accurately.

The second generated scenario shown in Fig. 4.6 has 54 sensors with most of the sensors distributed closer to the border. Once again, as illustrated in Fig. 4.6, the SA based approach picks sensor locations in a distributed fashion to cover the entire region. Fig. 4.7 shows that the SA based approach significantly reduces the ARMSE based prediction error compared to all the other tested sensor selection approaches. This shows that the proposed algorithm is more invariant to the type of spatial distribution of sensors.

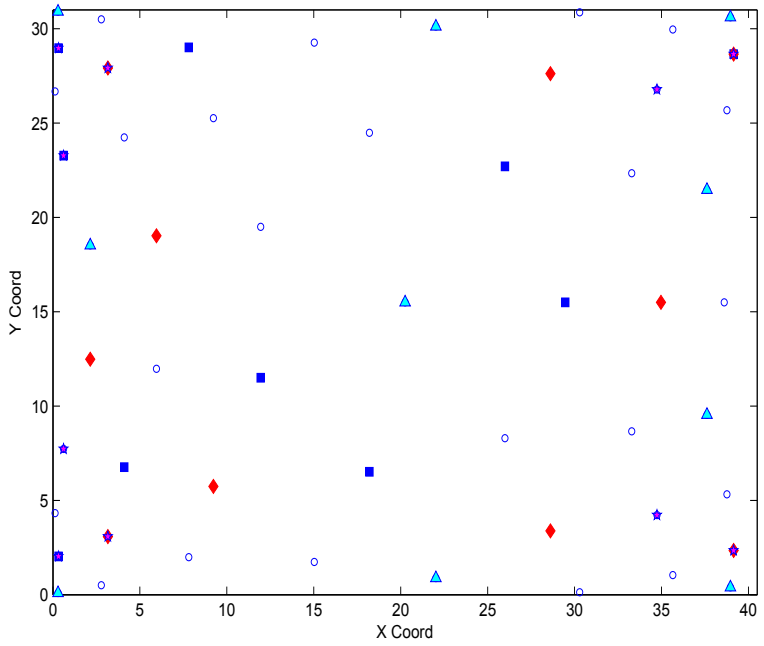


Figure 4.6: Deploying sensors densely in the boundary: Square - the locations selected by using SA based algorithm; diamond - the locations selected by using direct greedy method; pentagon - the locations selected by using mutual information approach; and triangle - the locations selected by using entropy technique. X Coord and Y Coord are in m.

4.5.1.2 Outdoor environment

In these experiments, it is interested to evaluate the performance of the algorithm using the outdoor data sets (Dataset 2 and Dataset 3) which exhibit considerable variations. The rainfall data in Dataset 2 varies in the interval of [10, 585] and the solid calcium ingredients data in Dataset 3 varies in the interval of [28, 78].

Fig. 4.8 shows the ARMSE performance metric under the four different sensor selection approaches in the rainfall phenomenon. While the accuracy obtained by the SA based approach is significantly better than that obtained by the entropy based method, it can be clearly seen that the SA based algorithm outperforms the direct

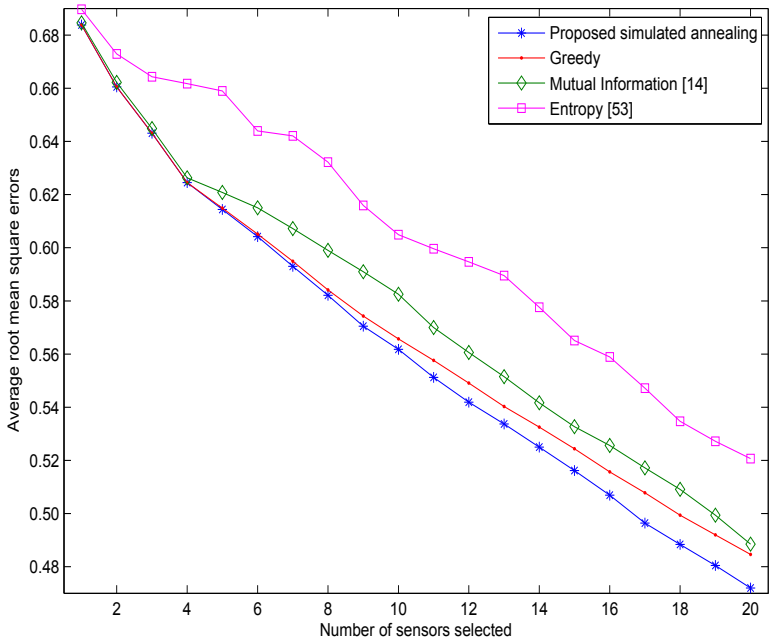


Figure 4.7: Average root mean square errors of densely border deployment.

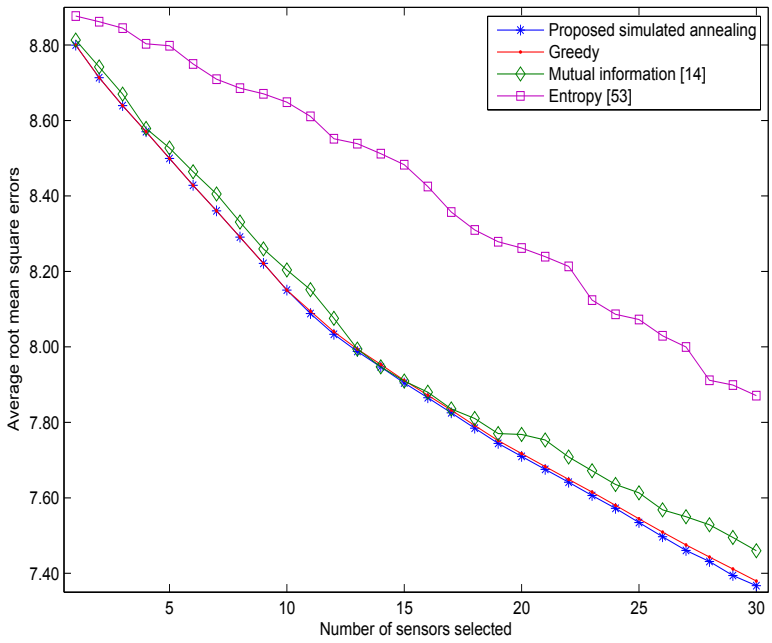


Figure 4.8: Average root mean square errors of Swiss rainfall data.

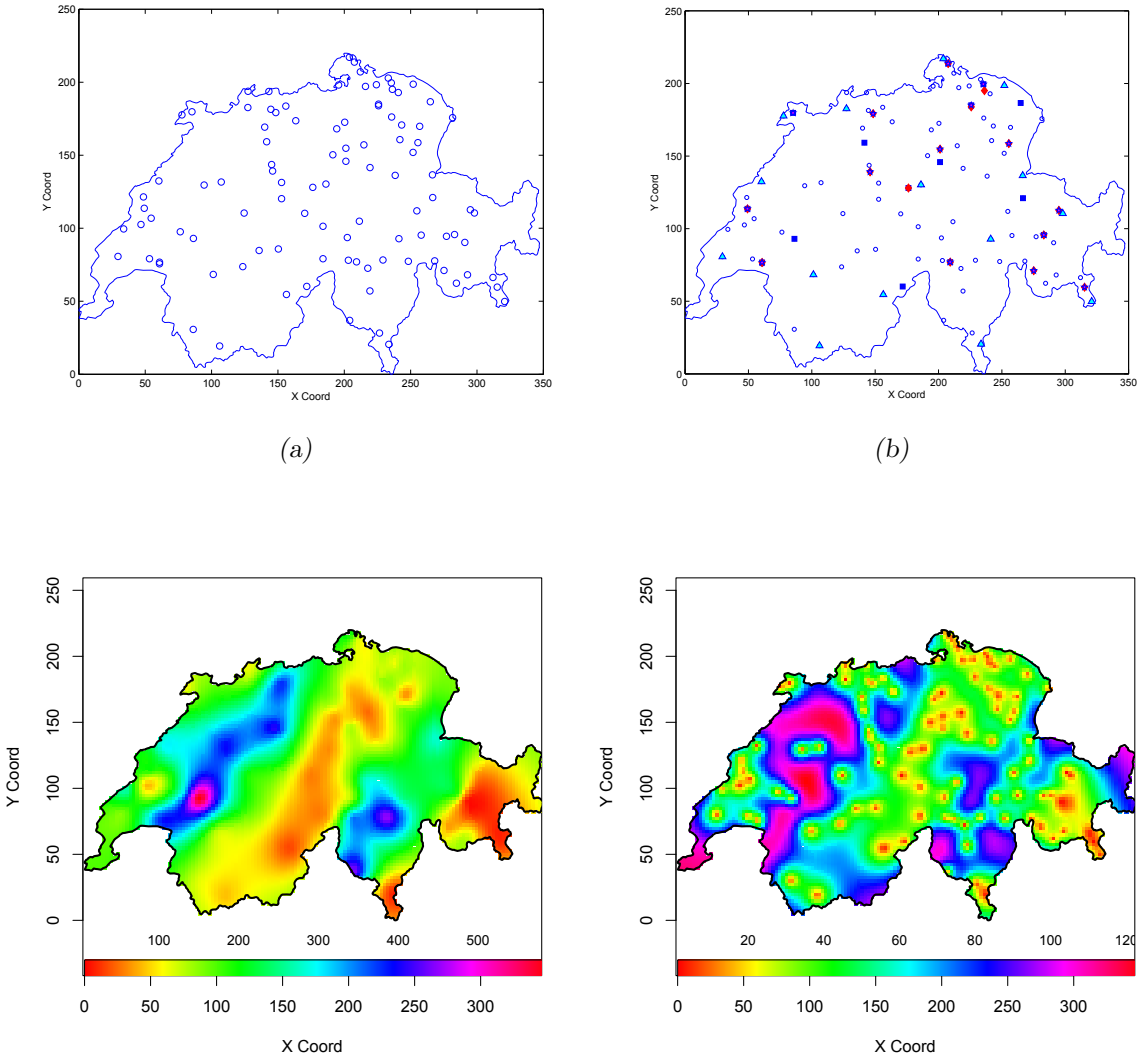


Figure 4.9: Deploying gauges and prediction of rainfall (1/10th of a mm) using all gauges. (a) Entire 100 gauges deployment throughout Switzerland; (b) Chosen gauges deployment of 30 sensor locations selected by using SA based algorithm (square), using direct greedy method (diamond), using mutual information approach (pentagon), and using entropy technique (triangle); (c) Predicted rainfall means using all gauges; (d) Predicted rainfall variance using all gauges. Colour bar in (c) represents range of rainfall, but colour bar in (d) describes range of errors. X Coord and Y Coord are in km.

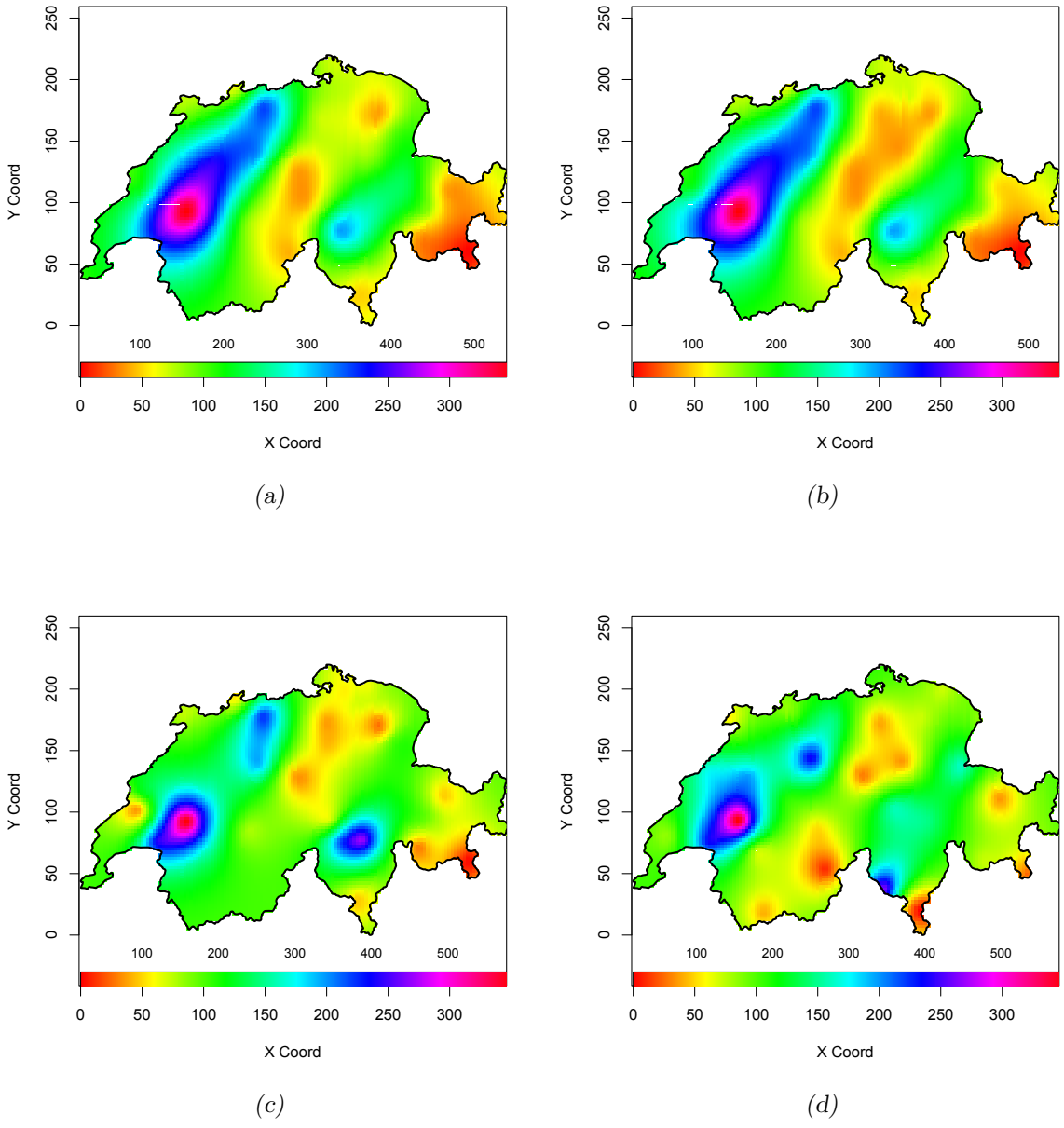
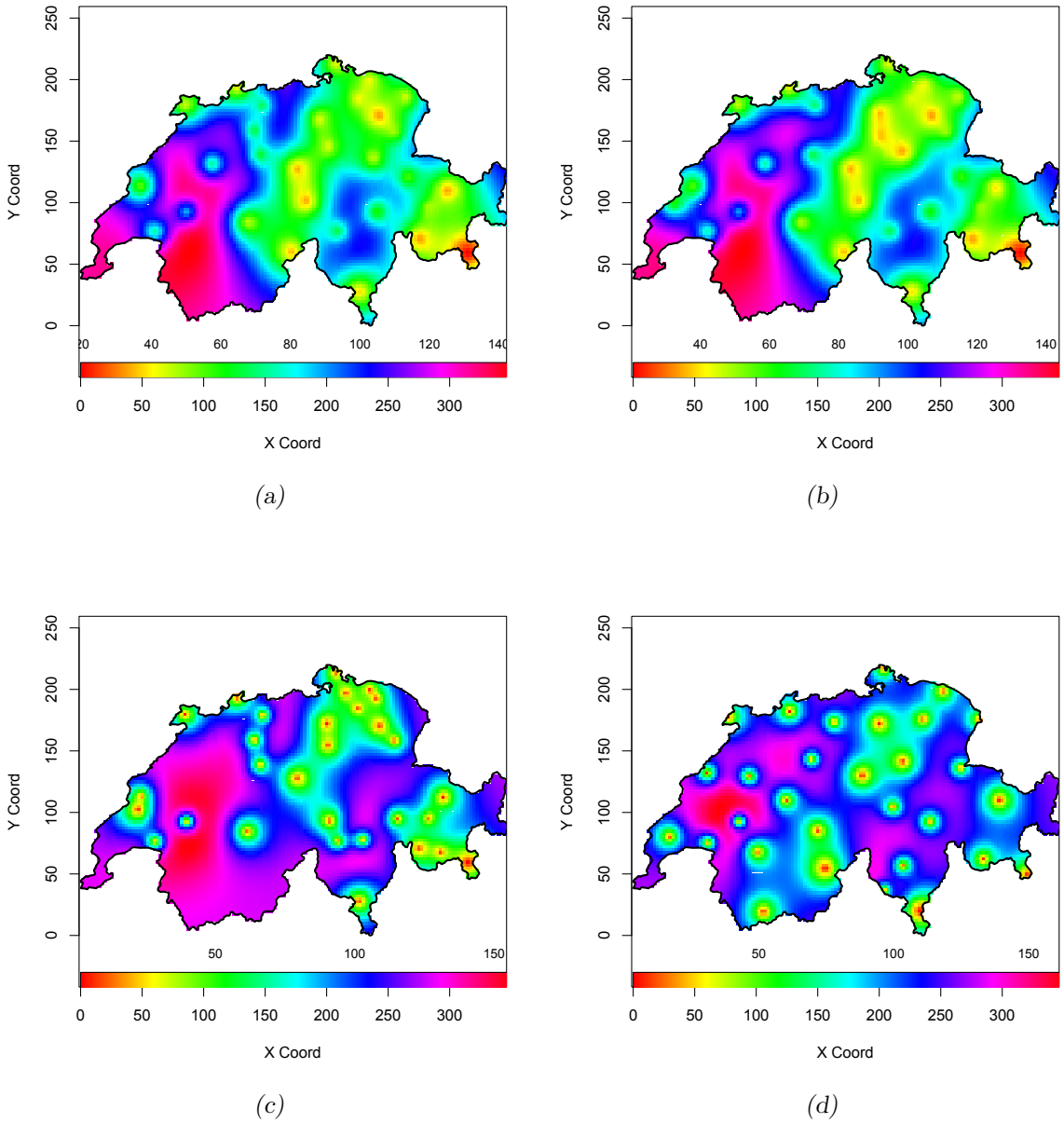


Figure 4.10: Comparisons of the Swiss rainfall (1/10th of a mm) predicted means using 30 best sensors obtained by (a) SA based algorithm, (b) direct greedy approximation method, (c) mutual information approach, and (d) entropy technique. Colour bars illustrate range of rainfall. X Coord and Y Coord are in km.



(a) *(b)*

(c) *(d)*

Figure 4.11: Comparisons of the Swiss rainfall (1/10th of a mm) predicted variances using 30 best sensors obtained by (a) SA based algorithm, (b) direct greedy approximation method, (c) mutual information approach, and (d) entropy technique. Colour bars demonstrate range of errors. X Coord and Y Coord are in km.

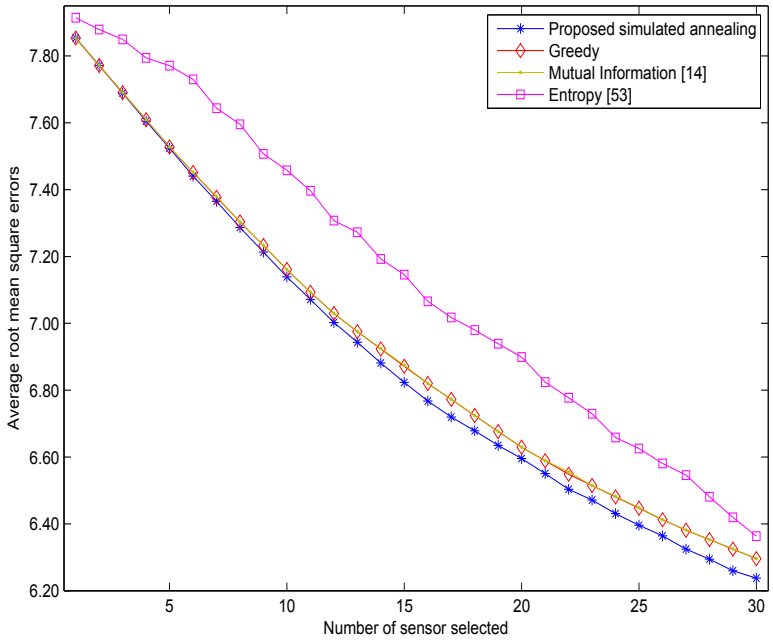


Figure 4.12: Average root mean square errors of solid calcium ingredient. Results of the greedy and MI based approaches mostly overlap.

greedy heuristic approach and the MI based method. This result clearly reinforces the benefits of the SA based algorithm.

In order to understand the rainfall distribution over the entire space, the GP model was used to produce the posterior mean and variance surfaces. For this, the geoR package [136] with the isotropic Matérn covariance functions was used. Fig. 4.9a shows the entire sensor population (100 sensors) throughout Switzerland, while Fig. 4.9b illustrates the selected best subsets of 30 sensors by SA, entropy, MI and direct greedy heuristic approaches. Fig. 4.9c and 4.9d show the predicted values of rainfall means and variances for the whole of Switzerland considering all the sensor readings. Fig. 4.10 (from *a* to *d*) illustrates the predicted mean rainfall under the four sensor selection approaches studied in this work. Fig. 4.11 (from *a* to *d*) shows the corresponding predicted variances for each approach. It can be seen that the

mean and variance results predicted by 30 selected sensors that are obtained by the SA based algorithm are very similar when compared to those predicted by all the sensors. Moreover, the proposed method outperforms the other three approaches

For Dataset 3, the best subsets varying from 1 to 30 sensors were selected among 178 potential sensors by the four methods, which were deployed to measure the calcium content in the soil. The average root mean square errors were then computed corresponding to each subset and are shown in Fig. 4.12. It is illustrated that the SA based algorithm generated better solutions to the sensor selection problem than the other three approaches.

4.5.2 Constrained Environment

This section demonstrates the effectiveness of the proposed approach on spatially modelling and predicting the gas phase H_2S in sewers. As a pilot study, only the forced ventilation zone in the North Head system of the Sydney Water sewer system, illustrated in Fig. 4.13, was chosen with H_2S at 75 percentile. This zone was organized into 121 sections with a total of 242 upper-stream and down-stream nodes that contain the corresponding H_2S observations spanning approximately along 5km. The levels of H_2S in the sewer network are illustrated in Fig. 4.14a.

The proposed GP based model was utilized to train the H_2S model using the collected gas phase H_2S data and mean and covariance functions proposed in Subsection 4.4.1.2. The entire estimated covariance matrix was employed to choose the best sensor locations in the chosen sewer system.

As described earlier, the MI was used for selecting the optimal number of H_2S sensors and their locations. Fig. 4.14b presents the predicted field of the H_2S in the

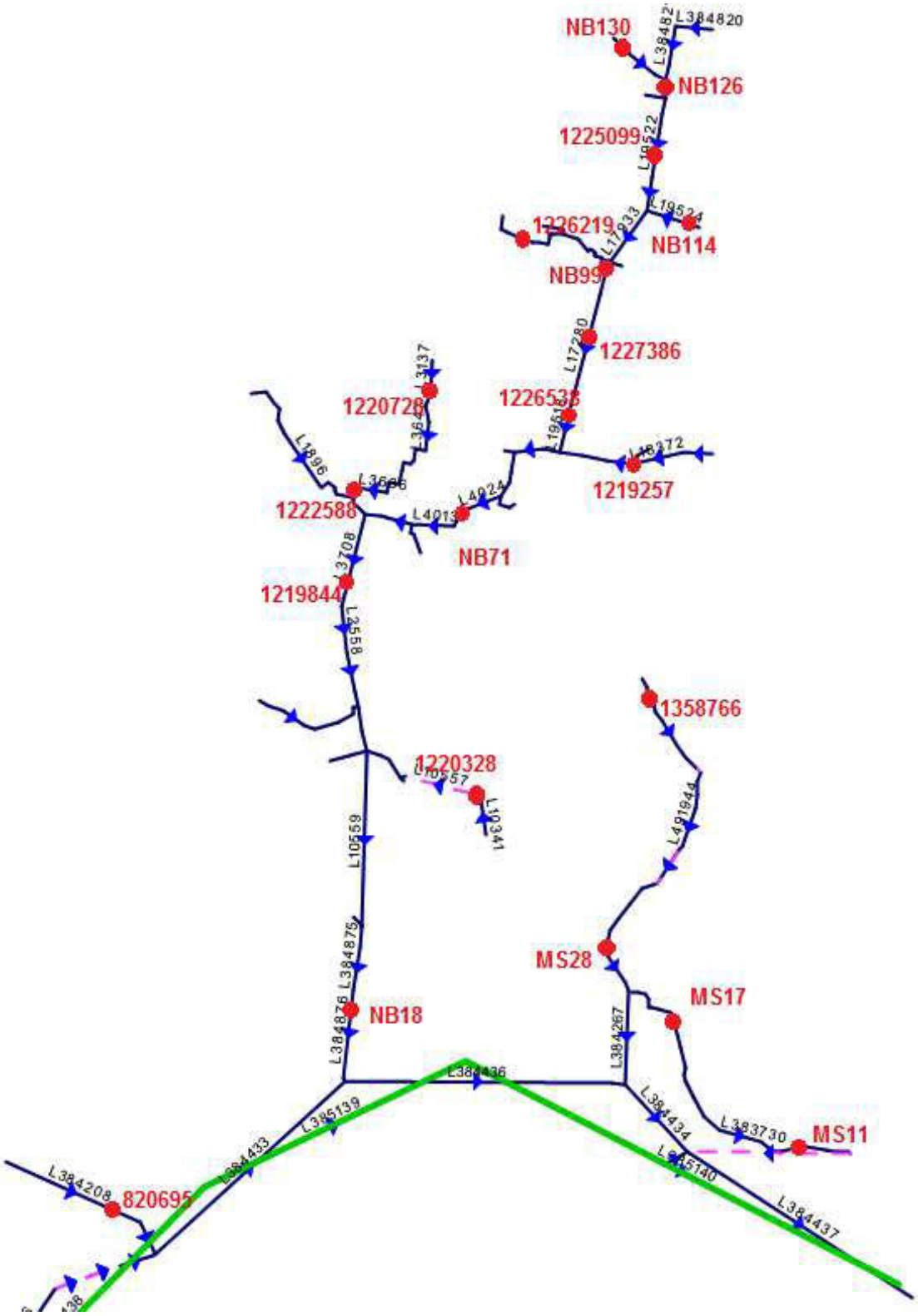


Figure 4.13: Forced ventilation zone in the North Head system, Sydney Water.

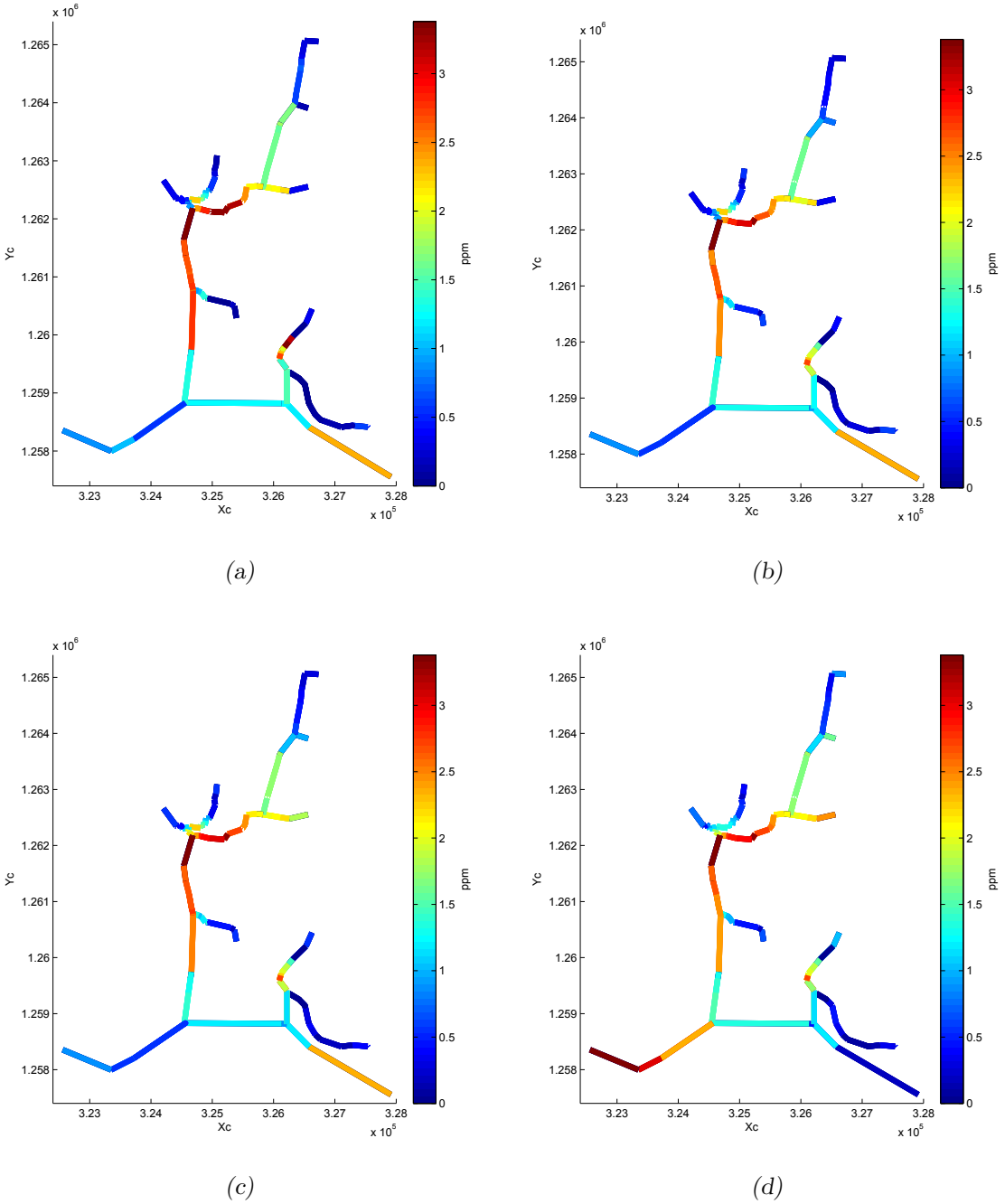
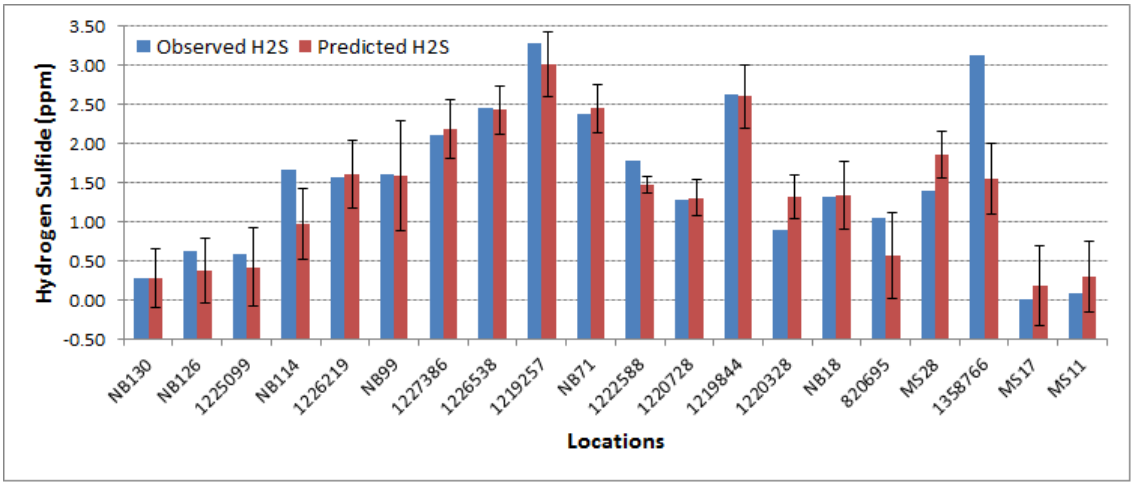
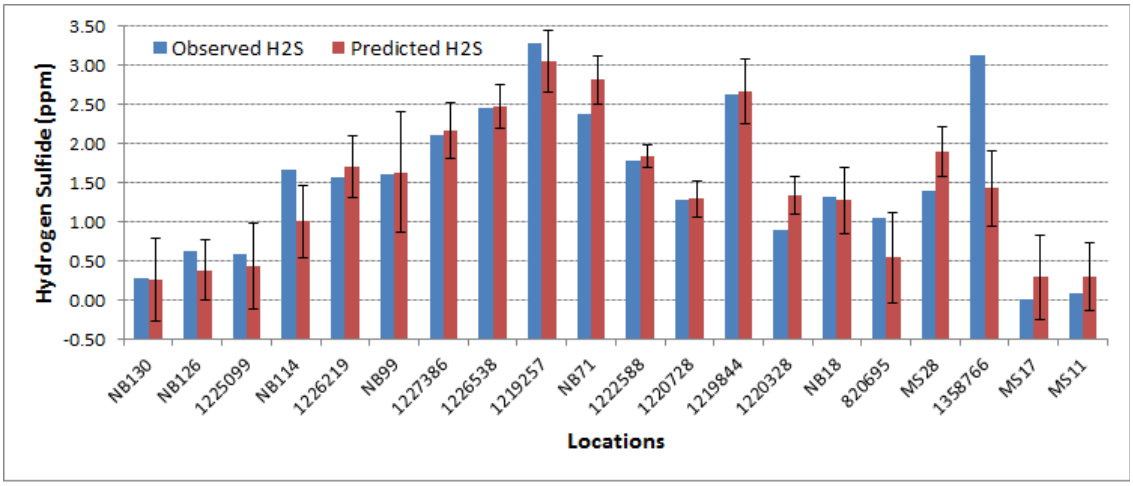


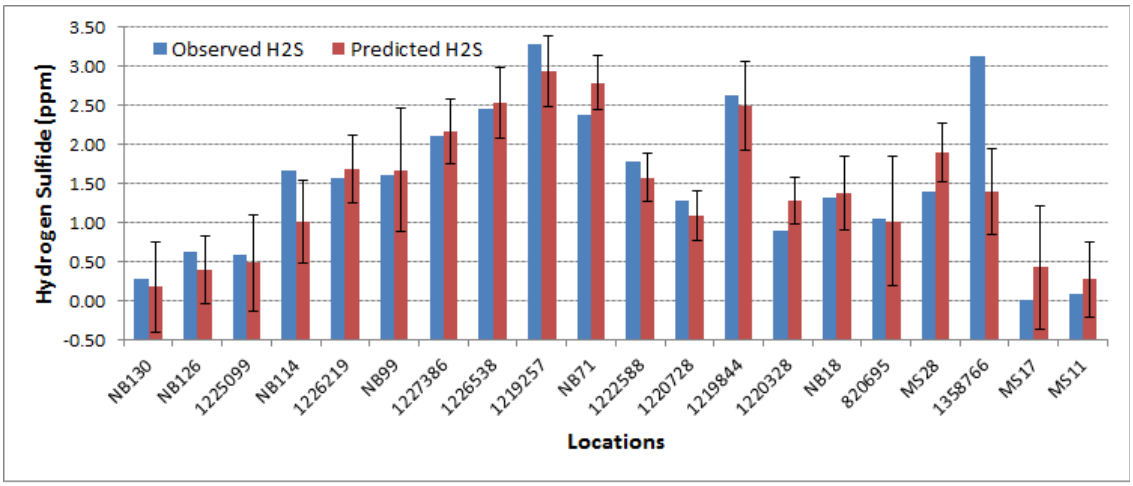
Figure 4.14: (a) The gas phase H_2S field. The predicted field using (b) 50, (c) 40, (d) 30 gas sensors, respectively. Colour bars represent range of hydrogen sulfide in ppm.



(a)



(b)



(c)

Figure 4.15: The H_2S predictions at particular locations using (a) 50, (b) 40, (c) 30 gas sensors, respectively.

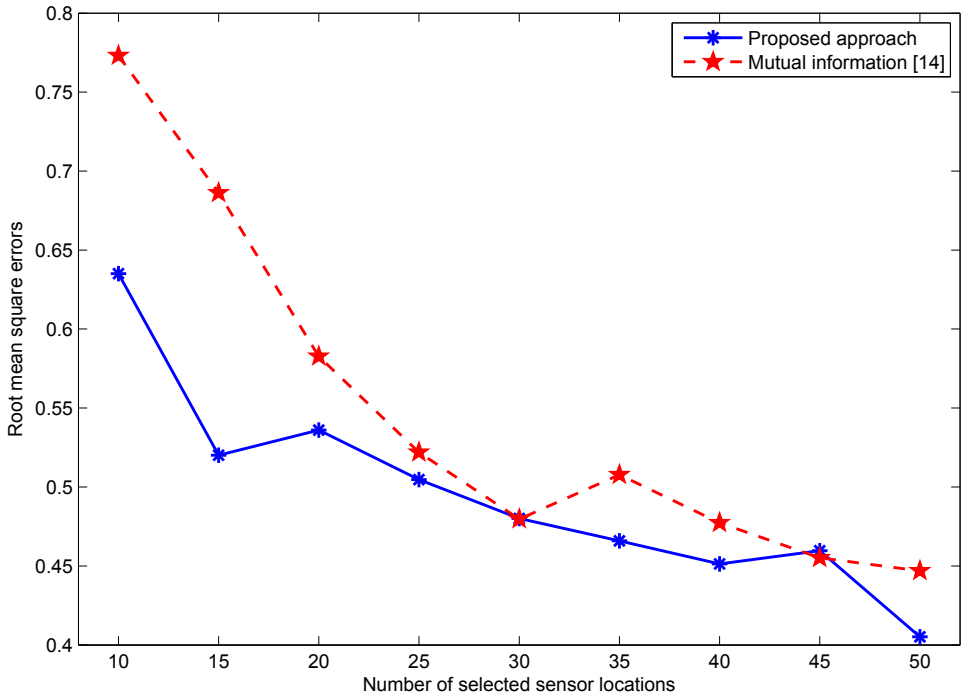


Figure 4.16: The results for H_2S prediction. Root mean square errors.

forced ventilation zone of the North Head system considering 50 most informative H_2S sensor locations with their observations. When comparing with Fig. 4.14a, it can be seen that the results predicted by the new GP model learned only from 50 best measurements closely represents the field in Fig. 4.14a. Further, even when only 40 and 30 most informative H_2S sensor readings were selected, the trained GP models show rewarding results in Fig. 4.14c and Fig. 4.14d, respectively.

For the purpose of individual validation, 20 particularly unmeasured locations, which are identified in red dots in Fig. 4.13, were chosen. The predicted H_2S concentrations at these particular points were compared with their available measurements and demonstrated in Fig. 4.15. Figures 4.15a, 4.15b and 4.15c present the H_2S predictions at these 20 locations estimated by the models with parameters and hyperparameters learned from 50, 40 and 30 most informatively selected

measurements, respectively. The corresponding standard deviation errors are also illustrated in these bar plots. It can be seen that the prediction results obtained by the proposed approach are very promising and mostly comparable to the realistic measurements at the studied locations. There is only one location named 1358766 at which the difference between the predicted H_2S value and the realistic observation is considerable. That is because, as referred to Fig. 4.13, the point 1358766 is located at the furthest point along the sewer network, and in fact there are not many sensors selected in this section. In other words, correlation between itself and other chosen locations is trivial.

Fig. 4.16 shows the root mean square errors (RMSEs) of the predicted field, which are computed via

$$\text{RMSE} = \sqrt{\frac{1}{|U|}\sum_{i=1}^{|U|}\left(m_{u_i|S} - y(u_i)\right)^2}, \tag{4.21}$$

where $y(u_i)$ is a measurement at $u_i \in U$, and $m_{u_i|S}$ is the predicted value at u_i given the observations at S. As expected, Fig. 4.16 shows a gradual decrement of the RMSE and hence increment of prediction accuracy with increased number of H_2S sensor placements. Therefore, this graph can be used to decide the number of sensors and their locations, based on a given RMSE, which is important for the water utility. The proposed algorithm outperforms the approach in [14]. For instance, in order to achieve an RMSE of 0.55 it only requires 15 sensors.

4.6 Summary

This chapter details the sensor selection problem and presents novel approaches to address this NP-hard combinatorial optimization problem in the constrained and unconstrained environments. In these considerations, it is proposed that the spatial phenomena to be monitored are modelled using the Gaussian processes that can be efficiently and effectively utilized to estimate and predict these physical processes at any unobserved locations. For the unconstrained environments, a simulated annealing based approximately heuristic algorithm has been proposed to resolve the new resulting minimization criteria in the sensor selection problem. Characteristics of the proposed algorithms have been extensively tested using both the indoor and outdoor data sets with different spatial distributions of observations. The simulated results have shown that the proposed simulated annealing based approach outperforms other conventional methods in many scenarios in the unconstrained environments. Furthermore, for the constrained environment of the gas phase H_2S concentration in a sewer network, a Gaussian process with an improved covariance function based on lengths along sewer pipes has been derived. Particularly, an efficient maximization criterion for the H_2S sensor selection problem has been proposed. This criterion aims to maximize the mutual information between the selected locations and the locations that are not selected or do not have any sensor deployments. Modelling and predicting the H_2S spatial concentrations in sewers are shown to be effective as compared with a more complex model developed by the utility participating in this study. More particularly, the proposed technique based on the mutual information provides a more accurate solution to the H_2S sensor selection problem than the existing method [14].

However, consider the case where it is required either to take environmental

measurements in a high density or to monitor the environment in a large region. If there is a very limited number of sensors, the stationary wireless sensor network will not be able to respond to these requirements. In another situation, if a large wireless sensor network is present, there will be many challenging situations related to managing and maintaining the network, such as selecting network topology or responding to individual sensor failures. It is also known that wireless transmissions are very power consuming, so data transmission needs to be as efficient as possible. Hence, a wireless network with a deterministic number of sensors that are based on mobile robotic platforms seems to be a positive solution. The mobile robotic wireless sensor networks will be considered in the following chapters to solve the spatial prediction problem. In other words, the sensor placement will be studied in the next chapters.

Chapter 5

Sensor Placement for Spatial Prediction using Gaussian Processes

5.1 Introduction

Recently, there has been flourishing interest in mobile robotic wireless sensor networks (MRWSNs) in which wireless sensor nodes are enriched with mobile capability by robotic platforms. In addition to features of wireless communication that provide the MRWSNs the ability of exchanging information, mobility in the MRWSNs enables the sensor nodes to reposition themselves to respond to a required goal. For instance, in the context of spatial prediction, the mobile wireless sensors will be required to redeploy themselves until uncertainty at predicted locations matches a criterion. Therefore, the new paradigm of a ubiquitous sensor network assisted by the mobile robots is evaluated to be more flexible for designing a data collection

scheme, more adaptive to the changes in the environment and more intelligent as compared with the stationary wireless sensor networks [68] as represented in Chapter 4.

It has been found that the MRWSNs have significant impact on environmental monitoring applications [8, 89, 90, 105, 108, 137–145]. After collecting measurements, the network is required to model the spatial phenomena and predict the field at unobserved locations of interest. In the works by [8, 89, 90, 98, 99, 140, 142, 143], the spatial processes are represented by a deterministic model that is often accompanied by a stochastic measurement error term. Nevertheless, the deterministic model requires that its model parameters and initial conditions are known a priori, which is intractable when the space of the parameters has high dimensions. As a result, Gaussian process (GP) [5, 12, 13] has been chosen to statistically model the physical spatial fields in this chapter. Furthermore, the GP regression is significantly powerful on statistical inference from environmental measurements [105, 108, 145, 146]. For example, near-optimal static sensor placements for spatial prediction were also addressed by using mutual information criterion in [60], [14]. More interestingly, in [105], Cortés brought a Kalman filter in the Kriging technique for predictive inference of the spatial field.

In designing the optimal sampling paths for the mobile robotic wireless sensors (MRWSs), objective function is usually derived from information-theoretic criteria such as conditional entropy maximization [91, 93, 95, 97], mutual information maximization [92], D- or A-optimality criterion using the Fisher information matrix [107], or minimization of prediction error variances [108, 109]. Then, optimal measurement locations for mobile sensing agents can be obtained by various algorithms that include active learning [93], Kalman filter [98, 99], Bayesian optimisation [110],

gradient descent [20, 96, 97], and gradient ascent [105]. More importantly, due to the NP-hard property of the sampling optimization problem, the mobile sensor coordination in the existing works has been fundamentally addressed by greedy heuristic algorithms [20, 92, 93, 98, 99, 101, 104]. However, so far, to the best of available knowledge, there have been no existing works that analyse the bounds of solutions and the computational complexity of the algorithms in resource-constrained MRWSNs. In this chapter, by predicting the future measurements at potentially new locations, an adaptive sampling optimization problem in terms of information-theoretic criteria is formulated by the use of conditional entropy. It is then demonstrated that the existing approximate greedy algorithms to address the combinatorial and NP-hard problem of optimizing the sampling paths for the MRWSNs are intractable if the networks search for a large number of unobserved locations in the field. As a result, an efficient novel optimality criterion for the adaptive sampling problem is proposed so that the issue can be addressed by the greedy algorithm in practically feasible time. Over and above, the bounds of solutions in optimizing the adaptive sampling problem in the MRWSNs have not been comprehensively taken into account. In this work, it is proved that under a certain condition the approximate entropy obtained by the proposed algorithm is guaranteed by a lower bound that is within $1 - \frac{1}{e}$ of the optimal solution. Then, it is proposed to use this lower bound as a stopping criterion for the sampling algorithm that stops mobile agents from further sampling the spatial field to conserve the limited energy. This lower bound is shown to be a practically feasible approach leading to acceptable prediction errors. An upper bound to the solution of the conditional entropy is also derived.

In the modern statistics, all the scalar field at spatial locations is correlated, which leads to modeling, estimation and prediction approaches to require full availability of all measurements and centralized calculation. This is a very costly option

for a MRWSN in which every mobile sensor has constrained resources such as a short wireless communication range. Therefore, the chapter also develops distributed estimation and prediction approaches, as well as designs distributed sampling strategies for resource-constrained MRWSNs, which requires each robotic sensor to interact only with its neighbours [147], [148] within a certain region of space. In fact, distributed methods make a MRWSN robust to individual sensor failures, scalable to a large number of agents and capable of easily dealing with changing conditions in the environment. This work is a step forward in addressing the sophisticated statistical inference in a distributed manner in the MRWSNs.

This chapter is organized as follows: In Section 5.2, a Gaussian predictive inference method is proposed to predict the field over time at locations of interest. These statistical tools are then utilized in Section 5.3 to resolve the sampling optimization problem. In this section, the algorithms based on locational optimization and conditional entropy are proposed to find out the most informative sampling paths for the MRWSs in continuous and discrete spaces, respectively. Section 5.4 presents a distributed spatial estimation and prediction approach to efficiently address the centrally-computed scheme of GP regression at each robotic sensor in resource-constrained networks of mobile, wireless and noisy agents. The efficiency of the proposed approaches is delineated via experiments in the real-world data sets in Section 5.5. Summaries are provided in Section 5.6.

5.2 Gaussian Predictive Inference

In this section, a Gaussian inference approach is proposed to consider the problem of using all the available data (observations) to predict the unobserved variables at a

set of locations of interest in the spatial field. More particularly, by the utilization of the spatial field model in Section 4.2, the spatial estimation and prediction formulae over time will be derived.

Let $Z(\mathbf{u}) := (z_1, z_2, ..., z_n)^T \in \mathbb{R}^n$ (hereafter Z) denote the variables to be predicted at locations $\mathbf{u} = (u_1^T, u_2^T, ..., u_n^T)^T \in \mathbb{R}^{dn}$, where d is a number of spatial dimensions. As represented in Section 2.6.1, the collection of all sensor measurements made by a N robotic sensor network up to time t can be rewritten as follows

$$y_{1:t} = X(s_{1:t})\beta_t + \xi(s_{1:t}) + \varepsilon(s_{1:t}), \tag{5.1}$$

where $X(s_{1:t})$ is a $tN \times (p+1)$ matrix of covariates. $\xi(s_{1:t}) \sim \mathcal{N}(\mathbf{0}, \Sigma_{y_{1:t}y_{1:t}})$ with a tN mean vector of $\mathbf{0}$. $\Sigma_{y_{1:t}y_{1:t}}$ is a $tN \times tN$ covariance matrix, whose elements can be computed by a function of $\sigma_t^2 corr(\phi_t, h, t, t')$. Here, $corr(\phi_t, h, t, t')$ is the spatio-temporal separable correlation function proposed by

$$corr(\phi_t, h, t, t') = \begin{cases} corr(\phi_t, h) & \text{if } t' \leq t+1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

$\varepsilon(s_{1:t}) \sim \mathcal{N}(\mathbf{0}, \tau_t^2 I)$, and I is a $tN \times tN$ identity matrix. $\beta_t, \sigma_t, \phi_t, \tau_t$ are model parameters learned by using all the available observations from time 1 to t. The hyperparameters vector is defined as $\theta_t := (\sigma_t^2, \phi_t, \tau_t^2) \in \mathbb{R}^3_{>0}$.

We extend the model specified in (5.1) to include $y_{1:t}$ and Z. If all model parameters are learned, by employing GP regression [13], the joint distribution is given by

$$y_{1:t}, Z|\beta_t, \theta_t \sim \mathcal{N}\left(\begin{bmatrix} X(s_{1:t}) \\ X(\mathbf{u}) \end{bmatrix} \beta_t, \begin{bmatrix} \Sigma_{1:t} & \Sigma_{y_{1:t}Z} \\ \Sigma_{Zy_{1:t}} & \Sigma_{ZZ} \end{bmatrix}\right), \tag{5.3}$$

where for simplicity, we set $\Sigma_{1:t} = \Sigma_{y_{1:t}y_{1:t}} + \tau_t^2 I$. Σ_{ZZ} is an $n \times n$ covariance matrix of Z, and $\Sigma_{y_{1:t}Z} (= \Sigma_{Zy_{1:t}}^T)$ is an $tN \times n$ cross-covariance matrix between $y_{1:t}$ and Z. The covariates at unmeasured locations $\mathbf{u}$ is defined as $X(\mathbf{u})$.

In probabilistic terms, the full conditional distribution of Z, given $y_{1:t}$, is also Gaussian, specified by

$$Z|y_{1:t}, \beta_t, \theta_t \sim \mathcal{N}\left(\mu_{Z|y_{1:t}}, \Sigma_{Z|y_{1:t}}\right), \tag{5.4}$$

where

$$\mu_{Z|y_{1:t}} = X(\mathbf{u})\beta_t + \Sigma_{Zy_{1:t}}\Sigma_{1:t}^{-1}\left(y_{1:t} - X(s_{1:t})\beta_t\right), \tag{5.5}$$

$$\Sigma_{Z|y_{1:t}} = \Sigma_{ZZ} - \Sigma_{Zy_{1:t}}\Sigma_{1:t}^{-1}\Sigma_{y_{1:t}Z}. \tag{5.6}$$

Note that the mean parameters β_t and the hyperparameters θ_t of this conditional distribution are unknown and can be estimated a priori by utilizing the generalized least squares technique [12] and the maximum likelihood approach [5], which will be delineated in the following sections.

5.3 Centralized Sensor Placement

In the first stage of considering the sensor placement problem in predicting spatially physical phenomena, the centralized schemes for mobile wireless sensor networks will be presented. In this consideration, at every time step t, all mobile sensing agents make new observations and transmit them to the sink via a specific routing tree. Then the base station conducts all centralized activities such as learning model parameters, estimating and predicting the spatial field at unmeasured locations of

interest, especially computing the sampling strategy that is then utilized to drive the robotic sensors navigating through the environment. Obviously, each control command is sent back to the corresponding mobile agent. In this discussion, finding the optimal sampling paths for the MRWSs in both continuous and discrete spaces is investigated. An algorithm based on a continuous density function, named locational optimization, is proposed, which enables the sensors to locate around positions of high uncertain density in the corresponding variance-built surface. From a discrete set of possible sensing positions in the environment, an efficient novel optimal criterion is also introduced to select the most informative locations for the sensing robots to take future observations in order to minimize uncertainty at all unobserved locations of interest. The solution is proven to be within bounds. The computational complexity of this proposition is shown to be practically feasible. More importantly, the lower bound is then utilized as a stopping criterion for the sampling algorithm. The criterion enables the prediction results to be within user-defined accuracies by controlling the number of mobile sensors.

5.3.1 Learning model parameters

Suppose that at each time instant t, the N mobile sensor network may collect N measurements. Hence, the collective measurement up to time t is tN observations available. In the following, a recursive algorithm for estimating simultaneously the mean parameters β_t and hyperparameters θ_t at time step t is described.

From (5.3), it is derived

$$y_{1:t} \sim \mathcal{N}(X(s_{1:t})\beta_t, \Sigma_{1:t}) \tag{5.7}$$

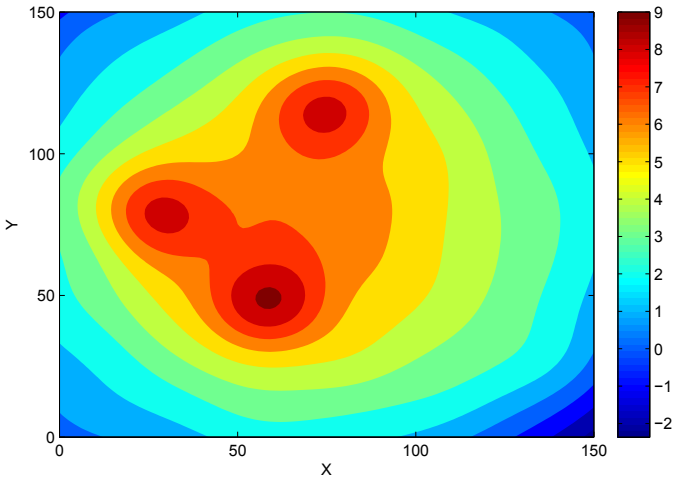


Figure 5.1: True field of spatial diffusion

Firstly, in the best linear unbiased estimator [12], β_t can be obtained by minimizing the function

$$f(\beta_t) = (y_{1:t} - X(s_{1:t})\beta_t)^T \Sigma_{1:t}^{-1} (y_{1:t} - X(s_{1:t})\beta_t) .$$

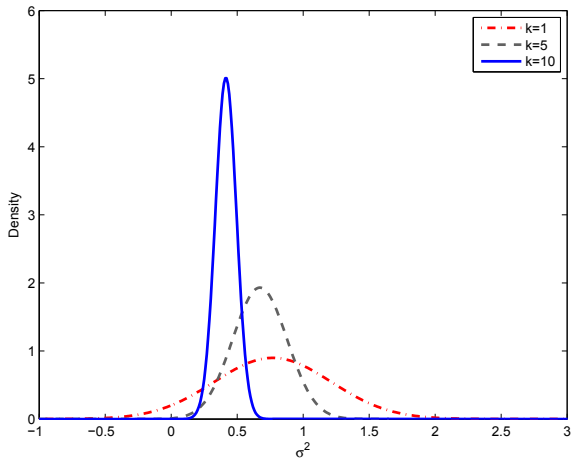
If given θ_t, that is $\Sigma_{1:t}$ known, the estimated β_t can be specified by

$$\widehat{\beta}_t = \left(X(s_{1:t})^T \Sigma_{1:t}^{-1} X(s_{1:t})\right)^{-1} X(s_{1:t})^T \Sigma_{1:t}^{-1} y_{1:t}. \tag{5.8}$$

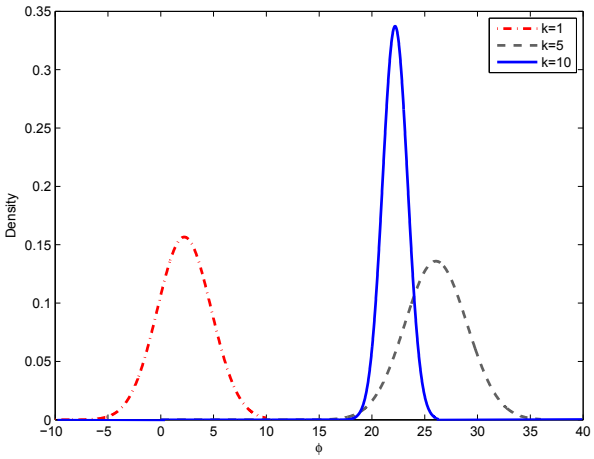
Secondly, from (5.7), the log-likelihood function can be written as

$$\mathcal{L}(\beta_t, \theta_t) = -\frac{1}{2}\{(y_{1:t} - X(s_{1:t})\beta_t)^T \Sigma_{1:t}^{-1} (y_{1:t} - X(s_{1:t})\beta_t) + \log\det(\Sigma_{1:t}) + tN\log(2\pi)\}. \tag{5.9}$$

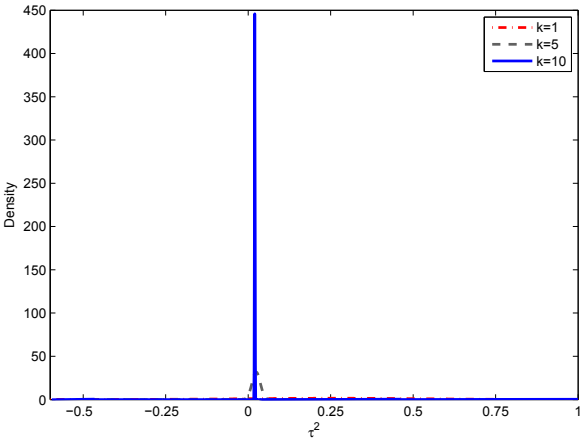
By substituting $\widehat{\beta}_t$ into (5.9) and then numerically maximizing this log-likelihood function with respect to θ_t, the hyperparameters $\widehat{\theta}_t$ can be obtained. Iteratively, these hyperparameters are employed to compute $\widehat{\beta}_t$ in the next iteration. The algorithm converges to the model parameters estimation as the error of two estimated values of each parameter at two consecutive iterations is acceptable.



(a)



(b)



(c)

Figure 5.2: Estimated model parameters: (a) σ^2, (b) ϕ and (c) τ^2.

To illustrate the hyperparameters learned over time, a scenario of a true field given in Fig. 5.1 that is generated by a set of hyperparameters $\theta_t = (0.52, 19.21, 0.01)$ is considered. A team of 10 robotic sensors is deployed in the field to gather data at every time step t. Note that each sensing agent moves to a new location after each measurement. Figures from Fig. 5.2a to Fig. 5.2c demonstrate the posterior probability density functions of the hyperparameters $(\sigma_t^2, \phi_t, \tau_t^2)$ at $t = 1, 5$, and 10. The illustrations show that the estimated hyperparameters converge at $t = 10$ to $\widehat{\sigma}_t^2 = 0.45$, $\widehat{\phi}_t = 21.00$, and $\widehat{\tau}_t^2 = 0.02$, which are within around 90 per cent of the true hyperparameters θ_t.

5.3.2 Locational optimization based sampling approach

An algorithm that allows sensors to concentrate around locations of high event density while satisfying the minimal area coverage constraints was investigated by [149]. For the purpose of reducing uncertainty at predicted locations of interest in the spatial prediction issue, a locational optimization based approach is proposed in this section, where the density function is developed on the posterior variances.

5.3.2.1 Centroidal Voronoi tessellation

Take a density function (also called bounded measurable function) $\Phi : \mathbb{R}^d \to \mathbb{R}_{>0}$ into consideration in a bounded environment of interest $\mathcal{Q} \subset \mathbb{R}^d$. Φ is considered as a function of measuring the information of probability of the event happening over $\mathcal{Q}$. Hence, the greater the value of $\Phi(p)$ is, the more significant the point p is. Let a non-increasing and piece-wise continuously differentiable function $f_s : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ denote a sensing performance function. This function characterizes the utility at

position p taken from i^{th} mobile robotic wireless sensor at the location s_i. One can easily see that the value of this function downgrades with the distance $\|p - s_i\|$ between p and s_i. Therefore, let this function be defined by a quadratic function $f_s := \|p - s_i\|^2$.

Given a density function Φ, and a sensing performance function f_s, the expected-value function of a N mobile sensor network deployed over the region $\mathcal{Q}$ is formulated as follows

$$\mathcal{H}(S) = \sum_{i=1}^{N} \int_{V_i} f_s \Phi(p) dp$$

or

$$\mathcal{H}(S) = \sum_{i=1}^{N} \int_{V_i} \|p - s_i\|^2 \Phi(p) dp, \tag{5.10}$$

where V_i is the Voronoi cell of s_i. From the notation of the Voronoi partition, introduced in Section 2.5, it can be easily seen that

$$\begin{aligned} &\textbf{minimize} \;\; \|p - s_i\| = \|p - s_j\| \;\; \text{for all } p \in V_j, \\ &\text{subject to} \;\; i \in \{1, ..., N\}. \end{aligned} \tag{5.11}$$

Therefore, the task here is to minimize the locational optimization function (5.10)

$$\textbf{minimize} \;\; \sum_{i=1}^{N} \int_{V_i} \|p - s_i\|^2 \Phi(p) dp. \tag{5.12}$$

In other words, this optimization problem is to identify an optimal robotic sensor network configuration that makes (5.10) minimum.

Carrying out the partial derivative of $\mathcal{H}(S)$ with respect to the i^{th} robotic sensor yields

$$\frac{\partial \mathcal{H}(S)}{\partial s_i} = \sum_{i=1}^{N} \int_{V_i} \frac{\partial}{\partial s_i} [\|p - s_i\|^2 \Phi(p)] dp. \tag{5.13}$$

In Euclidean space, the generalized area and the first mass moment of a set $\mathcal{Q} \subset \mathbb{R}^d$ with respect to Φ are specified by

$$A(\mathcal{Q}) = \int_{\mathcal{Q}} \Phi(p)dp \tag{5.14}$$

and

$$FM(\mathcal{Q}) = \int_{\mathcal{Q}} p\Phi(p)dp. \tag{5.15}$$

Thus, the centroid (center of mass) of a set $\mathcal{Q}$ is formulated by

$$C(\mathcal{Q}) = \frac{FM(\mathcal{Q})}{A(\mathcal{Q})}. \tag{5.16}$$

Moreover, the polar moment of inertia of $\mathcal{Q}$ about $s \in \mathcal{Q}$ is defined by

$$J(\mathcal{Q}, s) = \int_{\mathcal{Q}} \|p - s\|^2 \Phi(p)dp. \tag{5.17}$$

From [150], by using Parallel Axis Theorem, it can be easily stated that

$$J(\mathcal{Q}, s) = J(\mathcal{Q}, C(\mathcal{Q})) + A(\mathcal{Q})\|s - C(\mathcal{Q})\|^2. \tag{5.18}$$

From (5.10) and (5.17), using the Parallel Axis Theorem again, one simplifies (5.10) to

$$\mathcal{H}(S) = \sum_{i=1}^{N} J(V_i, C(V_i)) + \sum_{i=1}^{N} A(V_i)\|s_i - C(V_i)\|^2, \tag{5.19}$$

and (5.13) to

$$\frac{\partial \mathcal{H}(S)}{\partial s_i} = 2A(V_i)(s_i - C(V_i)). \tag{5.20}$$

Equation (5.20) demonstrates that the solution of the locational optimization problem (5.12) is the set of centroidal Voronoi configurations in $\mathcal{Q}$ [19]. In other words,

the optimal mobile robotic wireless sensor location is the centroid of its Voronoi cell.

5.3.2.2 Locational optimization based algorithm

As discussed in the Section 5.3.2.1, the premise behind the locational optimization approach is to develop a density function that describes the measurement of information or probability of the event over the environment. In the context of the spatial prediction, the quality of sensing is usually concerned with the uncertainty at predicted locations. Therefore, in this consideration, a density function based on the variances of the random variables is proposed.

Let $S, U \subset \mathbb{R}^2$ define the locations visited by a set of mobile sensing agents and unmeasured locations, respectively. At each sampling iteration t, according to the Gaussian predictive inference in the Section 5.2, the posterior variance at a location p, given the observations at S, is specified by

$$\Sigma_{p|S} = \Sigma_{pp} - \Sigma_{pS}(\Sigma_{SS} + \tau_t^2 I)^{-1}\Sigma_{Sp}, \tag{5.21}$$

where $\Sigma_{pp} = \sigma_t^2$, and Σ_{pS} $(= \Sigma_{Sp}^T)$ is the cross-covariance column vector between the random variables at S and the random variable at p. If $\mathcal{Q}$ is embedded into a two dimensional space, the density function for the sensor placement problem in the spatial prediction is formulated as

$$\Phi(p) = \sigma_t^2 - \Sigma_{pS}(\Sigma_{SS} + \tau_t^2 I)^{-1}\Sigma_{Sp}. \tag{5.22}$$

It can be clearly seen that the density function $\Phi(p)$ is substantially influenced by the environmental model hyperparameters. Consequently, after a new iteration

of sampling environment, the model hyperparameters will be updated and a new density function will be generated.

Nonetheless, there are two key issues associated with the implementation of the proposed approach in real mobile robotic wireless sensor networks: *(a)* limited computing resources to continuously compute the density function; *(b)* moving mobile sensors out of certain locations due to undesirable characteristics of the density function.

As it is required to compute the density function in real-time after each iteration, which depends on the number of observations, calculation of this density function becomes intractable with increasing number of sensors' readings. As a result, the first issue is addressed by employing the sensor selection techniques discussed in chapter 4. For instance, a strategy based upon the mutual information to select the most informative subset of measurements from among all of the potential sensor observations can be implemented. The best group of the measurements allows the effective elaboration of the density function and considerably diminishes the computational time.

As discussed in the Section 5.3.2.1, the centroid of a sensor's current location will be the sensor's new location in the next movement. More importantly, computation of the centroids is fundamentally dependent on the shape of the density function. For instance, if the shape of the current density function is quite similar to that of the previous one or the shape of the density function is nearly flat, the distance between current and new locations is negligibly small. This may lead the mobile wireless sensors to be trapped in stationary positions, which does not provide any further improvement to the sensing quality. In order to address the second issue, a circumcenter notion [151] is introduced to this scenario. In the event that all mobile

robotic wireless sensors are immobile during any iteration without reaching the criterion of sensing quality, mobile sensing agents are driven from current positions to circumcenters rather than to the centroids of their Voronoi cells. The circumcenter of a bounded set $D \subset \mathbb{R}^2$ is the center of the circle of minimum radius that contains D. Hence, the circumcenter is unique. The calculation of the circumcenter of a polygon $D \subset \mathbb{R}^2$ is to minimize the radius r of the circle centered at p subject to the constraints in which the distance between p and each of the polygon vertices $\mathrm{Ve}(D)$ is smaller than or equal to r. Mathematically, the problem can be formulated as

$$\begin{aligned} &\textbf{minimize } r, \\ &\text{subject to } \| p - d \|^2 \leq r^2, \text{ for all } d \in \mathrm{Ve}(D). \end{aligned} \tag{5.23}$$

Algorithm 3 Locational optimization based pseudocode for sensor placement

Input: Starting sensor locations s_t; collective measurements
Output: Next sensor locations s_{t+1}

loop
  Learn hyperparameters
  Develop density function
  Generate Voronoi partitions $V(s_t) = \{V_1, ..., V_N\}$
  Calculate all centroids $C(V_i)$, $i = 1, \cdots, N$
  $s_{t+1} \leftarrow \{C(V_i)\}$, $i = 1, \cdots, N$
  Calculate $d(i) = \| s_{t+1,i} - s_{t,i} \|$
  if Max(d) < predefined value **then**
    Remove s_{t+1}
    Calculate all circumcenters $CC(V_i)$, $i = 1, \cdots, N$
    $s_{t+1} \leftarrow \{CC(V_i)\}$, $i = 1, \cdots, N$
  end if
  Drive sensors to new locations s_{t+1}
  Get measurements at s_{t+1}
  Select best measurements out of all collective observations by using sensor selection approaches
end loop

Therefore, an algorithm which finds the next optimal sampling locations for mobile sensing agents at the next time instant is delineated as follows. Firstly,

the model hyperparameters are estimated from all the measurements collected from beginning to the current time step. The learned hyperparameters are then utilized to develop a density function that represents the surface of uncertainty in the whole environment. Secondly, the Voronoi partitions of all the current sensor locations need to be generated. By implementing the centroidal Voronoi technique introduced in the Section 5.3.2.1, all the centroids $C(V_i)$ of all the Voronoi cells corresponding to the agent locations s_i, where $i = 1, \cdots, N$, are presented. The solved centroids are considered as the new positions for mobile sensors if the gaps between the current positions and themselves satisfy an Euclidean distance criterion. Otherwise, the next sampling locations will be computed based upon the circumcenter technique. Finally, one of the sensor selection approaches is utilized to choose the most informative observations to enhance computation in the next iteration. The proposed algorithm for the sensor placement by the utilization of the locational optimization method is summarized in Algorithm 3.

Note that minimization of the expected-value function with a density function developed on the posterior variances means to minimize the uncertainty at predicted locations of interest.

5.3.3 Entropy based sampling approach

In the previous section, the locational optimization based approach was proposed to find out the optimal sampling locations for the mobile sensors in a continuous space. Nonetheless, due to the shape of the density function, mobile sensing agents can be potentially trapped in saddle points in some scenarios. On the other hand, this continuous method imposes the MRWSNs upon high difficulty of computing mathematical integration as the number of sensor measurements is increasing. In this

section, it is proposed to discretize the spatial field with small neighbours, and the most informative sampling paths are chosen from a discrete set of potential sensing locations. The proposed algorithm is proven to have a constant computational time, which is practically feasible for resource-constrained MRWSNs.

5.3.3.1 Conditional entropy

In information theory, the conditional entropy is a measure of the uncertainty in a random variable, given the knowledge of another random variable [133]. The conditional entropy of Z conditioned on Y, written as $H(Z|Y)$, quantifies the amount of information needed to describe the outcome of Z when Y is known and is defined by

$$H(Z|Y) = -\int p(Z,Y)\log\left(p(Z|Y)\right)\mathrm{d}Z\mathrm{d}Y,$$

where $p(Z,Y)$ is the joint probability distribution function of Z and Y, and $p(Z|Y)$ is the conditional probability distribution function of Z given Y.

Due to $p(Z|Y) = \frac{p(Z,Y)}{p(Y)}$, the conditional entropy can be rewritten as

$$H(Z|Y) = H(Z,Y) - H(Y),$$

where $H(Y)$ is the entropy of Y, and $H(Z,Y)$ is the joint entropy of Z and Y.

Suppose that the conditional distribution $Z|Y$ of cardinality of c has a Gaussian distribution with a covariance matrix $\Sigma_{Z|Y}$, then

$$H(Z|Y) = 0.5\log\{(2\pi e)^c \det\left(\Sigma_{Z|Y}\right)\} = 0.5\{\log\det\left(\Sigma_{Z|Y}\right) + c\log(2\pi) + c\}. \quad (5.24)$$

5.3.3.2 Adaptive sampling strategy

In this subsection, the properties of the entropy [133] are exploited to derive an efficient sampling algorithm for the robotic sensor networks to estimate the most informative sampling locations at time $t+1$ in order to improve the quality of the prediction and the model parameters estimation. Approximately, it is proposed to discretize the field $\mathcal{Q}$ into Q spatial sites of interest with small neighbours. Let M and Z denote the set of locations and the vector of latent variables at centres of Q spatial sites. In equivalent words, this approximation directs the mobile sensors to sample at the centres of these spatial sites. From this point, the entropy-based approach to drive the agents to the most informative spatial sites in the field is presented.

Given the collective observations $y_{1:t}$ up to time t, collected by N mobile sensors, as discussed in Section 5.2, the predicted measurements $\widehat{y}_{t+1}$ at next sensor locations s_{t+1} can be obtained by

$$\widehat{y}_{t+1}|y_{1:t} \sim \mathcal{N}\left(\mu_{\widehat{y}_{t+1}|y_{1:t}}, \Sigma_{\widehat{y}_{t+1}|y_{1:t}}\right), \tag{5.25}$$

where

$$\mu_{\widehat{y}_{t+1}|y_{1:t}} = X(s_{t+1})\beta_t + \Sigma_{\widehat{y}_{t+1}y_{1:t}}\Sigma_{1:t}^{-1}\left(y_{1:t} - X(s_{1:t})\beta_t\right), \tag{5.26}$$

$$\Sigma_{\widehat{y}_{t+1}|y_{1:t}} = \Sigma_{\widehat{y}_{t+1}\widehat{y}_{t+1}} - \Sigma_{\widehat{y}_{t+1}y_{1:t}}\Sigma_{1:t}^{-1}\Sigma_{y_{1:t}\widehat{y}_{t+1}}. \tag{5.27}$$

$\Sigma_{\widehat{y}_{t+1}\widehat{y}_{t+1}}$ is an $N \times N$ covariance matrix of $\widehat{y}_{t+1}$, and $\Sigma_{y_{1:t}\widehat{y}_{t+1}} (= \Sigma_{\widehat{y}_{t+1}y_{1:t}}^T)$ is an $tN \times N$ cross-covariance matrix between $y_{1:t}$ and $\widehat{y}_{t+1}$. Let Z_{-Nt} denote the subset of Z formed by removing Nt variables corresponding to the Nt locations the mobile sensors have visited. Notice that the observations $y_{1:t}$ are treated as constants, and $\widehat{y}_{t+1} = \{\widehat{y}_{t+1,1}, ..., \widehat{y}_{t+1,N}\}$ is a subset of Z_{-Nt}. Let $Z_{-N(t+1)}$ define as $Z_{-Nt} \setminus \widehat{y}_{t+1}$

and $W = Q - N(t+1)$ with $W \gg N$. Therefore, the conditional entropy for a multivariate Gaussian distribution of random variables at unobserved locations at time $t+1$ is given by a closed-form expression [133],

$$H\left(Z_{-N(t+1)}|y_{1:t}, \widehat{y}_{t+1}\right) = 0.5 \log\det(\Sigma_{-N(t+1)}) + \text{const}, \tag{5.28}$$

where $\Sigma_{-N(t+1)}$ is the covariance matrix of $Z_{-N(t+1)}$ conditioned on $y_{1:t}$ and $\widehat{y}_{t+1}$. This matrix can be computed by the use of (5.6).

Problem 5.1. The problem is to find the next locations for the MRWSs in the next time step, which leads to the lowest uncertainties at unmeasured locations. The minimization is given by

$$s_{t+1}^{opt} = \underset{s_{t+1,i} \in M}{\textbf{argmin}} \quad H\left(Z_{-N(t+1)}|y_{1:t}, \widehat{y}_{t+1}\right), \tag{5.29}$$

where $s_{t+1,i}$ is the next location for sensor i.

As proved by Ko *et al.* [53], the optimization problem (5.29) is combinatorial NP-hard and can be resolved by a greedily near-optimal algorithm. Nevertheless, in order to estimate the quality of sensing, a large number of unobserved locations (W) is usually considered. This results in computational complexity in calculating the log determinant of $\Sigma_{-N(t+1)}$, which makes even the greedy heuristic algorithm intractable in resource-constrained MRWSNs as in the following theorem.

Theorem 5.2. *The greedy algorithm can resolve the optimization problem (5.29) in time* $\mathcal{O}(NW^4)$.

Proof: Since the size of $\Sigma_{-N(t+1)}$ is $W \times W$, the cost of $\log\det(\Sigma_{-N(t+1)})$ is $\mathcal{O}(W^3)$ operations [152]. On the other hand, there have $O(W)$ such the logarithms

of the determinants to be calculated. Moreover, in the adaptive sampling strategy, it needs to find N next MRWS locations at every time step. Thus, the greedy algorithm has to run N iterations in time $\mathcal{O}(N)$. So, the optimization problem (5.29) can be finalized in running time $\mathcal{O}(NW^4)$. ■

As W is large, addressing (5.29) by the resource-constrained MRWSNs is intractable. Therefore, an efficiently novel optimality criterion for designing the mobile sensor navigation is proposed in the following.

Proposition 5.3. The optimal sampling locations at time step $t+1$ for the robotic sensors can be determined by solving the following optimization problem

$$s_{t+1}^{opt} = \underset{s_{t+1,i} \in M_i^{(t)}}{\textbf{argmax}} \; H\left(\widehat{y}_{t+1}|y_{1:t}\right), \tag{5.30}$$

where $M_i^{(t)} \subset M$ is a discrete set of locations bounded by $\Omega_i^{(t)}$ to where sensor i can move at time $t+1$. Here, $\Omega_i^{(t)}$ is the allowable movement region of a robotic sensor i, defined in Section 2.6.2.

Proof: By the chain rule for conditional entropy [133], it yields

$$H\left(Z_{-N(t+1)}|y_{1:t}, \widehat{y}_{t+1}\right) = H\left(Z_{-N(t+1)}, \widehat{y}_{t+1}|y_{1:t}\right) - H\left(\widehat{y}_{t+1}|y_{1:t}\right).$$

Due to $\{Z_{-N(t+1)} \cup \widehat{y}_{t+1}\} = Z_{-Nt}$ and $y_{1:t}$ is known, $H\left(Z_{-N(t+1)}, \widehat{y}_{t+1}|y_{1:t}\right)$ is a constant. Therefore, it can be clearly seen that

$$\begin{aligned} s_{t+1}^{opt} &= \textbf{argmin}\, H\left(Z_{-N(t+1)}|y_{1:t}, \widehat{y}_{t+1}\right) \\ &= \textbf{argmax}\, H\left(\widehat{y}_{t+1}|y_{1:t}\right) \end{aligned}$$

■

This proposition is to find the optimal set of sensor locations at time $t+1$, s^{opt}_{t+1}, from all potential sets of s_{t+1}, which corresponds to a maximum of $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$. The location $s^{opt}_{t+1,i}$ for the i^{th} mobile sensor at time $t+1$ must be selected from one of all potential $s_{t+1,i}$ points bounded by $\Omega^{(t)}_i$. The solution of (5.30) is obtained if all $s^{opt}_{t+1,i}$ locations are found in the corresponding regions $\Omega^{(t)}_i$. Note that $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ is specified by

$$H\left(\widehat{y}_{t+1}|y_{1:t}\right) = 0.5 \log\det(\Sigma_{\widehat{y}_{t+1}|y_{1:t}}) + \text{const.} \tag{5.31}$$

Since the size of $\Sigma_{\widehat{y}_{t+1}|y_{1:t}}$ is $N \times N$, computation of the (5.30) is practically feasible for resource-constrained MRWSNs.

Let $Z_{\Omega^{(t)}_i}$ denote the vector of the variables at unobserved spatial sites inside $\Omega^{(t)}_i$ and $Z_{\Omega^{(t)}} := \bigcup_{i=1}^{N} Z_{\Omega^{(t)}_i}$ denote the variables at unmeasured locations in all the allowable movement regions of an N mobile sensors network. These latent variables can be computed by using (5.4). It is also defined $M^{(t)} := \bigcup_{i=1}^{N} M^{(t)}_i$. Following [53], the problem (5.30) is still NP-hard. Therefore, an approximate algorithm which finds the near-optimal sampling points s^*_{t+1} for the mobile sensors by adding the next sensors' locations in sequence is described as follows.

At each time step t, it starts from an empty set of locations, $s^*_{t+1} = \oslash$, that corresponds to the empty set of predicted measurements at time step $t+1$, $\widehat{y^*}_{t+1} = \oslash$. It first requires to maximize the conditional entropy

$$H\left(\widehat{y}_{t+1,j}|y_{1:t}\right) = 0.5 \log\det(\sigma^2_{\widehat{y}_{t+1,j}|y_{1:t}}) + \text{const}$$

with respect to $\widehat{y}_{t+1,j} \in Z_{\Omega^{(t)}}$, where $\sigma^2_{\widehat{y}_{t+1,j}|y_{1:t}}$ is the variance of $\widehat{y}_{t+1,j}$ given $y_{1:t}$. From the solution of this maximization, $\widehat{y^*}_{t+1,j}$, it could be easy to obtain $s^*_{t+1,j}$. If

$s^*_{t+1,j}$ is inside $\Omega^{(t)}_i$, it is certainly the next location for sensor i. Both $\widehat{y^*}_{t+1,j}$ and $s^*_{t+1,j}$ are then added to $\widehat{y^*}_{t+1}$ and s^*_{t+1}, respectively; and the latent variable vector $Z_{\Omega^{(t)}_i}$ and the set $M^{(t)}_i$ in $\Omega^{(t)}_i$ are removed from $Z_{\Omega^{(t)}}$ and $M^{(t)}$. In the next iterations of the algorithm, it iteratively does

$$s^*_{t+1,j} = \underset{\substack{s_{t+1,j} \in M^{(t)} \\ \widehat{y}_{t+1,j} \in Z_{\Omega^{(t)}}}}{\textbf{argmax}} \; H\left(\widehat{y^*}_{t+1} \cup \widehat{y}_{t+1,j} | y_{1:t}\right). \tag{5.32}$$

Each solution $\widehat{y^*}_{t+1,j}$ and the corresponding location $s^*_{t+1,j}$ found are greedily added to $\widehat{y^*}_{t+1}$ and s^*_{t+1}, respectively. If the current resulting location $s^*_{t+1,j}$ is identified as an element of $M^{(t)}_i$, then $M^{(t)}_i$ and the corresponding vector $Z_{\Omega^{(t)}_i}$ are removed from $M^{(t)}$ and $Z_{\Omega^{(t)}}$. The algorithm stops when s^*_{t+1} reaches to cardinality, N. The cost of this proposed algorithm can be described as follows.

Theorem 5.4. *The near-optimal solution of the problem (5.30) can be obtained by the greedy heuristic algorithm in time $\mathcal{O}(N^4 m)$, where m is a number of latent variables $Z_{\Omega^{(t)}}$ and $m < W$.*

Proof: In this case, $\log\det(\Sigma_{\widehat{y}_{t+1}|y_{1:t}})$ can be computed in time $\mathcal{O}(N^3)$. The rest of this proof is similar and the reader is referred to the proof of the **Theorem 5.2**. Therefore, the computational complexity of (5.30) is $\mathcal{O}(N^4 m)$. ■

Since $N << W$, the **Proposition 5.3** benefits remarkably in computation for resource-constrained MRWSNs.

The proposed greedy algorithm for finding near-optimal sampling paths in the MRWSNs is summarized in Algorithm 4. Note that $\mathbb{Z}_{>0}$ is the set of nonnegative integer numbers. $\Sigma_{\widehat{y}_{t+1}\widehat{y}_{t+1}}$ is the covariance matrix of $\widehat{y}_{t+1}$. $\Sigma_{\widehat{y}_{t+1,j}y_{1:t}}$ $(= \Sigma^T_{y_{1:t}\widehat{y}_{t+1,j}})$

and $\Sigma_{\widehat{y}_{t+1}y_{1:t}}$ $(= \Sigma^T_{y_{1:t}\widehat{y}_{t+1}})$ are the cross-covariance vector between $\widehat{y}_{t+1,j}$ and $y_{1:t}$ and the cross-covariance matrix between $\widehat{y}_{t+1}$ and $y_{1:t}$, respectively.

Algorithm 4 Sequentially approximate algorithm for maximizing the conditional entropy

Input: $y_{1:t}$; $M^{(t)} = \bigcup_{i=1}^{N} M_i^{(t)}$; $Z_{\Omega^{(t)}} = \bigcup_{i=1}^{N} Z_{\Omega_i^{(t)}}$
Output: Next sensor locations s^*_{t+1}
At time $t \in \mathbb{Z}_{>0}$, do
$s^*_{t+1} \leftarrow \oslash$; $\widehat{y^*}_{t+1} \leftarrow \oslash$

for j = 1 to length of $Z_{\Omega^{(t)}}$ **do**
  $\widehat{y}_{t+1,j} \in Z_{\Omega^{(t)}}$
  $\sigma^2_{\widehat{y}_{t+1,j}|y_{1:t}} = \sigma^2_t - \Sigma_{\widehat{y}_{t+1,j}y_{1:t}}\Sigma^{-1}_{1:t}\Sigma_{y_{1:t}\widehat{y}_{t+1,j}}$
  $\widehat{y}^*_{t+1,j} \leftarrow \underset{\widehat{y}_{t+1,j}\in Z_{\Omega^{(t)}}}{\text{argmax}} \log(\sigma^2_{\widehat{y}_{t+1,j}|y_{1:t}})$
  $\widehat{y^*}_{t+1} \leftarrow \widehat{y^*}_{t+1} \cup \widehat{y}^*_{t+1,j}$
  $s^*_{t+1,j} \leftarrow \underset{s_{t+1,j}\in M^{(t)}}{\text{argmax}} \log(\sigma^2_{\widehat{y}_{t+1,j}|y_{1:t}})$
  if $s^*_{t+1,j} \in \Omega_i^{(t)}$ **then**
    $s^*_{t+1} \leftarrow s^*_{t+1} \cup s^*_{t+1,j}$
  end if
end for
$M^{(t)} \leftarrow M^{(t)} \setminus M_i^{(t)}$
$Z_{\Omega^{(t)}} \leftarrow Z_{\Omega^{(t)}} \setminus Z_{\Omega_i^{(t)}}$
while cardinality $|s^*_{t+1}| < N$ **do**
  for $\widehat{y}_{t+1,j} \in Z_{\Omega^{(t)}}$ **do**
    $\widehat{y}_{t+1} \leftarrow \widehat{y^*}_{t+1} \cup \widehat{y}_{t+1,j}$
    $\Sigma_{\widehat{y}_{t+1}|y_{1:t}} = \Sigma_{\widehat{y}_{t+1}\widehat{y}_{t+1}} - \Sigma_{\widehat{y}_{t+1}y_{1:t}}\Sigma^{-1}_{1:t}\Sigma_{y_{1:t}\widehat{y}_{t+1}}$
  end for
  $\widehat{y}^*_{t+1,j} \leftarrow \underset{\widehat{y}_{t+1,j}\in Z_{\Omega^{(t)}}}{\text{argmax}} \log\det\left(\Sigma_{\widehat{y}_{t+1}|y_{1:t}}\right)$
  $\widehat{y^*}_{t+1} \leftarrow \widehat{y^*}_{t+1} \cup \widehat{y}^*_{t+1,j}$
  $s^*_{t+1,j} \leftarrow \underset{s_{t+1,j}\in M^{(t)}}{\text{argmax}} \log\det\left(\Sigma_{\widehat{y}_{t+1}|y_{1:t}}\right)$
  if $s^*_{t+1,j} \in \Omega_i^{(t)}$ **then**
    $s^*_{t+1} \leftarrow s^*_{t+1} \cup s^*_{t+1,j}$
    $M^{(t)} \leftarrow M^{(t)} \setminus M_i^{(t)}$
    $Z_{\Omega^{(t)}} \leftarrow Z_{\Omega^{(t)}} \setminus Z_{\Omega_i^{(t)}}$
  end if
end while

In the following theorem, an upper bound to the solution of the (5.30) is demonstrated.

Theorem 5.5. *Let H be a conditional entropy in $\widehat{y}_{t+1}$ given $y_{1:t}$, and let H_i be the conditional entropy in $\widehat{y}_{t+1,i}$ given $y_{1:t}$. Then*

$$H\left(\widehat{y}_{t+1}|y_{1:t}\right) \leq \sum_{i=1}^{N} H_i\left(\widehat{y}_{t+1,i}|y_{1:t}\right). \tag{5.33}$$

Proof: The proof is a straightforward result of applying the theorem 17.1.5 in [133]. ■

The **Theorem 5.5** provides an upper bound on the value of the conditional entropy computed at the optimal sampling locations, which can be also utilized to bound the quality of the most informative sensors' locations by any other methods such as the local optimization approach. It can be clearly seen that if the measurements are independent, the upper bound is actually optimal for the problem (5.30).

5.3.3.3 Stopping Criterion Based on Lower Bound

The predictions based on the Algorithm 4 can lead to very high accuracies as the time goes to infinity, meaning the MRWSNs are allowed to operate long term. However, in this study, the resource-constrained wireless mobile sensor networks are considered. These networks have resource constraints such as limited energy and communication restrictions, which need to be considered in the algorithmic design stage. Therefore, in order to save the limited energy of the mobile sensory nodes, a stopping criterion is proposed. This is achieved by finding the lower bound for the entropy given in the (5.30).

It is proved by Nemhauser *et al.* [23] that the lower bound of the entropy given in (5.30) can be guaranteed, if the set function of the conditional entropy $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ holds both the monotonicity and submodularity properties. Since $\widehat{y}_{t+1}$ and $y_{1:t}$ are real-valued, it can be seen that $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ is a submodular function, which is defined in Section 2.7, as proved by Fujishige [153]. Nevertheless, in general, the function $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ is not monotone. Its monotonic property can only be demonstrated if the conditional entropy of $\widehat{y}_{t+1}$ satisfies a particular condition which is given in **Lemma 5.6**.

Lemma 5.6. With respect to Gaussian random variables $\widehat{y}_{t+1}$ given $y_{1:t}$, the set function of the conditional entropy $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ is monotone if

$$H\left(\widehat{y}_{t+1}|y_{1:t}\right) \in [\log(2\pi e)^{-N}, +\infty),$$

where N is the cardinality of $\widehat{y}_{t+1}$.

Proof: For notational simplicity, assume that $\mathcal{A} := \widehat{y}_{t+1,1:k}|y_{1:t}$ denotes a subset of $\widehat{y}_{t+1}|y_{1:t}$, and let $\Sigma_{\mathcal{A}\mathcal{A}}$ denotes the covariance matrix of $\widehat{y}_{t+1,1:k}$ conditioned on $y_{1:t}$. Therefore, the conditional differential entropy of A is given

$$H(\mathcal{A}) = \frac{1}{2}\log\{(2\pi e)^k \det(\Sigma_{\mathcal{A}\mathcal{A}})\}.$$

Let $\mathcal{B} = \widehat{y}_{t+1,k+1:r}|y_{1:t}$ be another subset of $\widehat{y}_{t+1}|y_{1:t}$ with cardinality $|\mathcal{B}| = d$, and $\Sigma_{\mathcal{B}\mathcal{B}}$ denote the covariance matrix of $\mathcal{B}$.

In order to prove the monotonicity of the differential entropy, it is required to show that $H(\mathcal{A} \cup \mathcal{B}) \geq H(\mathcal{A})$, for all sets $\mathcal{B}$. It can be clearly seen that

$$\triangle H = H(\mathcal{A} \cup \mathcal{B}) - H(\mathcal{A}) = \frac{1}{2}\log\{(2\pi e)^{k+d} \det(\Sigma_{\mathcal{A}\cup\mathcal{B}})\} - \frac{1}{2}\log\{(2\pi e)^k \det(\Sigma_{\mathcal{A}\mathcal{A}})\},$$

where $\Sigma_{\mathcal{A}\cup\mathcal{B}}$ denotes the covariance of the collective set of $\mathcal{A}$ and $\mathcal{B}$. Hence, it needs to illustrate that

$$\triangle H = \log(2\pi e)^d + \log\det(\Sigma_{\mathcal{A}\cup\mathcal{B}}) - \log\det(\Sigma_{\mathcal{A}\mathcal{A}}) \geq 0.$$

In probability theory, $\mathcal{A}$ and $\mathcal{B}$ are jointly Gaussian distributed as

$$\begin{bmatrix} \mathcal{A} \\ \mathcal{B} \end{bmatrix} \sim \mathcal{N}\left(\cdot, \Sigma_{\mathcal{A}\cup\mathcal{B}}\right) = \mathcal{N}\left(\cdot, \begin{bmatrix} \Sigma_{\mathcal{A}\mathcal{A}} & \Sigma_{\mathcal{A}\mathcal{B}} \\ \Sigma_{\mathcal{B}\mathcal{A}} & \Sigma_{\mathcal{B}\mathcal{B}} \end{bmatrix}\right),$$

where $\Sigma_{\mathcal{A}\mathcal{B}} (= \Sigma_{\mathcal{B}\mathcal{A}}^T)$, are cross-covariance matrices between $\mathcal{A}$ and $\mathcal{B}$. Here, the mean will not be exploited and hence not shown. Since $\Sigma_{\mathcal{A}\mathcal{A}}$ is invertible, it can be seen from the Leibniz formula that

$$\begin{bmatrix} \Sigma_{\mathcal{A}\mathcal{A}} & \Sigma_{\mathcal{A}\mathcal{B}} \\ \Sigma_{\mathcal{B}\mathcal{A}} & \Sigma_{\mathcal{B}\mathcal{B}} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathcal{A}\mathcal{A}} & \mathbf{0} \\ \Sigma_{\mathcal{B}\mathcal{A}} & I \end{bmatrix} \begin{bmatrix} I & \Sigma_{\mathcal{A}\mathcal{A}}^{-1}\Sigma_{\mathcal{A}\mathcal{B}} \\ \mathbf{0} & \Sigma_{\mathcal{B}\mathcal{B}} - \Sigma_{\mathcal{B}\mathcal{A}}\Sigma_{\mathcal{A}\mathcal{A}}^{-1}\Sigma_{\mathcal{A}\mathcal{B}} \end{bmatrix}.$$

which leads to

$$\det(\Sigma_{\mathcal{A}\cup\mathcal{B}}) = \det(\Sigma_{\mathcal{A}\mathcal{A}})\det(\Sigma_{\mathcal{B}\mathcal{B}} - \Sigma_{\mathcal{B}\mathcal{A}}\Sigma_{\mathcal{A}\mathcal{A}}^{-1}\Sigma_{\mathcal{A}\mathcal{B}}). \tag{5.34}$$

Using (5.34), it can be written

$$\log(2\pi e)^d + \log\det(\Sigma_{\mathcal{B}\mathcal{B}} - \Sigma_{\mathcal{B}\mathcal{A}}\Sigma_{\mathcal{A}\mathcal{A}}^{-1}\Sigma_{\mathcal{A}\mathcal{B}}) \geq 0. \tag{5.35}$$

The matrix in the second term of (5.35) is the covariance of the conditional distribution $p(\mathcal{B}|\mathcal{A})$ [154]. Therefore, (5.35) is equivalent to

$$\log(2\pi e)^d + \log\det(\Sigma_{\mathcal{B}|\mathcal{A}}) \geq 0.$$

Use of the expression of the conditional entropy (5.31) completes the proof. ■

Following the fundamental work by [23], this lower bound can be shown to guarantee the quality of the near-optimal solutions of the (5.30) as follows. If the conditional entropy of $\widehat{y^*}_{t+1}$ at resulting sensor locations s^*_{t+1}, which is obtained by the Algorithm 4, satisfies $H\left(\widehat{y^*}_{t+1}|y_{1:t}\right) \in [\log(2\pi e)^{-N}, +\infty)$, it is guaranteed by a lower bound that is within $1 - \frac{1}{e}$ of the optimal solution $H\left(\widehat{y^{opt}}_{t+1}|y_{1:t}\right)$, where $\widehat{y^{opt}}_{t+1}$ is the predicted measurements at s^{opt}_{t+1}.

It is to be noted that the lower bound on the conditional entropy in the **Lemma 5.6** is dependent on the number of mobile sensors, N. Therefore, lower bound and hence the stopping criteria of the algorithm can be varied by choosing an appropriate number of mobile sensors. On the other hand, as discussed in the Section 5.2, the quality of the prediction is a function of measurements that are results of a sampling strategy. As a consequence, by manipulating the number of the mobile sensors can ameliorate the prediction results to reach a user-defined accuracy while the sampling paths for the mobile sensors are simultaneously guaranteed within $1-\frac{1}{e}$ of the optimal strategy. Practical feasibility of the stopping criterion based on the lower bound will be discussed in Section 5.5.

5.4 Distributed Sensor Placement

In this section, the distributed learning and Gaussian inference algorithms for a MR-WSN in which every robotic wireless sensor employs only local information from local neighbouring agents are derived. By utilizing distributed computation tools consisting of Jacobi over-relaxation (JOR) and discrete-time average consensus (DAC) algorithms, it is proposed to compute centralized schemes in a distributed manner. In this proposition, at every time step a mobile wireless sensor not only senses the spatial field at its current location but also receives its neighbours' measurements. Upon receiving collective measurements, each mobile sensor itself learns its own model parameters and hyperparameters. Then, by exchanging information with its neighbours, each agent recursively updates its own estimation until it converges to global values of the parameters and the hyperparameters. In distributed prediction algorithms, each robotic sensor estimates the spatial field at unobserved locations based upon its collective measurements from the beginning to current time and global model parameters and hyperparameters before exchanging its predictions with its neighbours to compute globally predicted means and variances. A distributed sampling strategy is also presented at the end to drive mobile robotic sensors to the most uncertain locations in order to improve the quality of prediction.

5.4.1 Distributed computation tools

Here, two effective distributed computational approaches are briefly introduced.

5.4.1.1 Jacobi over-relaxation approach

The Jacobi Over-Relaxation approach [155] is an iterative algorithm widely utilized in distributed calculation in order to solve a linear system $\Sigma x = y$, where $\Sigma \in \mathbb{R}^{N\times N}$ is an invertible matrix and $y \in \mathbb{R}^N$ is a vector. Consider a network of a mobile agent system, it is defined that in each row $row_i(\Sigma) \in \mathbb{R}^N$, $\epsilon_{ij} = (\Sigma)_{ij} \neq 0$ if agent i and agent j are neighbours; that is, agent i can access to $row_i(\Sigma)$. In addition, if agent i knows y_i, individual element of the solution $x = \Sigma^{-1}y$ can be computed by the following recursion

$$x_i(k+1) = (1-\alpha)x_i(k) + \frac{\alpha}{\epsilon_{ii}}\left(y_i - \sum_{j\neq i}\epsilon_{ij}x_j(k)\right), \tag{5.36}$$

where $k \in \mathbb{Z}_{>0}$, $i \in \{1, ..., N\}$, and $\alpha \in (0, 1)$ is a constant that controls the speed of convergence.

In fact, the convergence properties of the JOR approach were comprehensively considered in [155] in terms of the eigenvalues of the matrix Σ. In [105], Cortés shows that if $\alpha < \frac{2}{N}$ and Σ is a symmetric and positive-definite matrix, then the JOR algorithm converges to the solution $x = \Sigma^{-1}y$ from any initial points $x(0)$. Therefore, by exchanging information with the neighbours, agent i can obtain the i-th element of x.

Remark 5.7. The significant computational complexity in GP regression, particularly as the number of the measurements is growing in MRWSNs, is to compute the inverse of the covariance matrix, which is cubic in the dimension [111]. Nonetheless, it can be seen that the JOR approach can resolve the linear system $\Sigma x = y$ without requiring to calculate Σ^{-1}. Hence, the JOR algorithm brings about particularly computational benefits for the MRWSNs in a distributed manner. ■

5.4.1.2 Discrete-time average consensus approach

The Discrete-time Average Consensus algorithm [156] is frequently utilized in a distributed mobile network in order to enable each agent to trace the average value of whole network signals. Consider a MRWSN in which every mobile sensor i has its own signal g_i. It is aimed that sensor i needs to compute the arithmetic mean of elements in the vector $g = (g_1, g_2, ..., g_N)^T \in \mathbb{R}^N$. It is assumed that each robotic sensor location is a vertex of a graph. If this graph is connected, then the arithmetic mean of g can be obtained by the following iterative procedure

$$x_i(k+1) = x_i(k) + \gamma \sum_{j \neq i} a_{ij} \left(x_j(k) - x_i(k) \right) \tag{5.37}$$

with initial condition $x(0) = g$, where $a_{ij} = 1$ iff i-th sensor and j-th sensor are neighbours and 0 otherwise, and γ is the step size. Olfati-Saber *et al.* [156] demonstrated that if γ holds

$$0 < \gamma < \frac{1}{\Delta},$$

the DAC algorithm converges to the solution, where $\Delta = \max_i \left(\sum_{j \neq i} a_{ij} \right)$ is the maximum degree of the graph. At the end of the iterations of the algorithm, every mobile sensor obtains the average value of the vector g that is equivalent to $\sum_{i=1}^{N} \frac{g_i}{N}$.

5.4.2 Distributed Gaussian inference

In this subsection, the distributed algorithm that each robotic wireless sensor can employ to predict the spatial field at unmeasured locations by exchanging local information within its neighbours in a communication range of R is proposed.

Consider the spatial model discussed in the Section 4.2, the measurements $y(\mathbf{s})$ taken at locations $\mathbf{s} = \{s_1, s_2, ..., s_N\} \subset \mathcal{Q} \subset \mathbb{R}^d$ has a Gaussian distribution $y(\mathbf{s}) \sim \mathcal{N}\left(X(\mathbf{s})\beta, \Sigma_{\mathbf{ss}} + \tau^2 I\right)$, where $\Sigma_{\mathbf{ss}}$ is a covariance matrix whose elements are computed by the kernel function $cov(s_i, s_j)$. It is supposed that the kernel function has a finite spatial range $f \in \mathbb{R}_{>0}$, and $cov(s_i, s_j) = 0$ if $\parallel s_i - s_j \parallel > f$. Let u_* denote an unobserved location in space $\mathcal{Q}$. The random variable $y(u_*)$ can be predicted in a centralized fashion as follows

$$y(u_*)|y(\mathbf{s}) \sim \mathcal{N}(m_{u_*|\mathbf{s}}, \Sigma_{u_*|\mathbf{s}}), \tag{5.38}$$

where

$$m_{u_*|\mathbf{s}} = X(u_*)\beta + \Sigma_{u_*\mathbf{s}}(\Sigma_{\mathbf{ss}} + \tau^2 I)^{-1}(y(\mathbf{s}) - X(\mathbf{s})\beta), \tag{5.39}$$

$$\Sigma_{u_*|\mathbf{s}} = \Sigma_{u_*u_*} - \Sigma_{u_*\mathbf{s}}(\Sigma_{\mathbf{ss}} + \tau^2 I)^{-1}\Sigma_{\mathbf{s}u_*}. \tag{5.40}$$

Here $m_{u_*|\mathbf{s}}$ and $\Sigma_{u_*|\mathbf{s}}$ are the mean vector and the covariance matrix of $y(u_*)$ given $y(\mathbf{s})$. $\Sigma_{\mathbf{ss}}$ and $\Sigma_{u_*u_*}$ are the covariance matrices of $y(\mathbf{s})$ and $y(u_*)$, respectively, and $\Sigma_{\mathbf{s}u_*} (= \Sigma_{u_*\mathbf{s}}^T)$ is the cross-covariance matrix between $y(\mathbf{s})$ and $y(u_*)$.

However, by employing two useful distributed tools described above, the distributed approaches to implement the equations (5.39) and (5.40) in distributed spatial estimators are presented. Given the measurements $y(\mathbf{s})$, by exchanging information with the neighbours, every robotic sensor can predict the field at u_* as in the following propositions.

Proposition 5.8. Given N measurements collected by N mobile robotic wireless sensors and the spatial field model parameters, i.e., $\beta, \sigma^2, \phi, \tau^2$, the posterior mean

at u_* can be computed by every mobile agent i as follows

$$m_{u_*|\mathbf{s}} = X(u_*)\beta + N\mu_i, \tag{5.41}$$

where μ_i is the average consensus value of the random field at u_* that is calculated by the mobile sensor i. Here,

$$\mu_i = \frac{1}{N}\left(\Sigma_{u_*\mathbf{s}}\left(\Sigma_{\mathbf{ss}} + \tau^2 I\right)^{-1}\left(y(\mathbf{s}) - X(\mathbf{s})\beta\right)\right) \tag{5.42}$$

in which $\Sigma_{u_*\mathbf{s}}$ is a cross-covariance vector between $y(u_*)$ and $y(\mathbf{s})$.

Proof: It is defined that

$$\varphi = \left(\Sigma_{\mathbf{ss}} + \tau^2 I\right)^{-1}\left(y(\mathbf{s}) - X(\mathbf{s})\beta\right). \tag{5.43}$$

Since the robotic sensor i knows the i-th row of the matrix $(\Sigma_{\mathbf{ss}} + \tau^2 I)$ and i-th element of $(y(\mathbf{s}) - X(\mathbf{s})\beta)$, utilizing the JOR algorithm allows it to solve the iterative solution of (5.43) within the i-th element of φ in each iteration specified by

$$\varphi_i(k+1) = (1-\alpha)\varphi_i(k) + \frac{\alpha}{\sigma^2 + \tau^2}\left(\left(y(s_i) - X(s_i)\beta\right) - \sum_{j\neq i}\epsilon_{ij}\varphi_i(k)\right),$$

where $\alpha < \frac{2}{N}$ and ϵ_{ij} is an element of the covariance matrix $(\Sigma_{\mathbf{ss}} + \tau^2 I)$. It is known that $\Sigma_{u_*\mathbf{s}} = (\epsilon_{u_*s_1}, \epsilon_{u_*s_2}, ..., \epsilon_{u_*s_N})$ is a row vector and φ is a column vector, thus

$$\mu = \Sigma_{u_*\mathbf{s}}\varphi = \sum_{i=1}^{N}\epsilon_{u_*s_i}\varphi_i \tag{5.44}$$

As a consequence, (5.44) can be effectively solved by the DAC approach in a distributed manner, and then at the end of the iterations of the DAC algorithm μ_i is

the arithmetic mean of a vector of $(\epsilon_{u_*s_1}\varphi_1, \epsilon_{u_*s_2}\varphi_2, ..., \epsilon_{u_*s_N}\varphi_N)^T$. This completes the proof. ■

Proposition 5.9. Let

$$\psi = \left(\Sigma_{\mathbf{ss}} + \tau^2 I\right)^{-1} \Sigma_{\mathbf{s}u_*}, \tag{5.45}$$

where $\Sigma_{\mathbf{s}u_*} = \Sigma_{u_*\mathbf{s}}^T$. Given initial conditions as in **Proposition 5.8**, the posterior variance at u_* can be computed by each sensor i as follows

$$\sigma^2_{u_*|\mathbf{s}} = \sigma^2 - N\omega_i, \tag{5.46}$$

where ω_i is the arithmetic mean of $(\epsilon_{u_*s_1}\psi_1, \epsilon_{u_*s_2}\psi_2, ..., \epsilon_{u_*s_N}\psi_N)^T$.

Proof: The proof is similar to that of the **Proposition 5.8**. ■

For the purpose of predicting the field at a set of the unobserved locations, similar algorithms can be derived.

5.4.3 Distributed estimators in MRWSNs over time

In this subsection, the distributed estimation approaches for every mobile sensor in the MRWSNs as they navigate through the environment and gather the measurements over time are presented. In the previous sections, it was shown that under the assumption of the communication range each robotic sensor can exchange its information with its neighbours. In equivalent words, at every time step each mobile wireless sensor not only senses the spatial field at its current location but also receives its neighbours' measurements. Therefore, it is assumed that every mobile agent can store all its own measurements and its neighbour sensors' readings that

have been transmitted to it from time 1 to t to estimate its own parameters and hyperparameters and to predict the spatial field.

Consider a MRWSN of N agents, the collectively environmental measurements of the i-th mobile wireless sensor at the time step t is given by

$$y_{i,1:t} = \left(\bigcup_{j=1}^{t} y(s_{j,i})\right) \bigcup \left(\bigcup_{j=1}^{t} \left(\bigcup_{k=1}^{N_{ij}} y(s_{j,k})\right)\right), \tag{5.47}$$

where N_{ij} is the number of the i-th sensor's neighbours at the time j. For the sake of simplicity, it is denoted $y_i := y_{i,1:t}$, $\mathbf{s}_i = \left(\bigcup_{j=1}^{t} s_{j,i}\right) \bigcup \left(\bigcup_{j=1}^{t} \left(\bigcup_{k=1}^{N_{ij}} s_{j,k}\right)\right)$.

Employing the spatial model as considered in the Section 4.2 when the measurements accumulate over time, it is supposed that the covariance function is defined as $\sigma_t^2 corr(\phi_t, h, t, t')$ with a separable correlation function $corr(\phi_t, h, t, t')$ proposed by 5.2. Here, σ_t^2 and ϕ_t are global hyperparameters learned at the time step t. Estimating these global hyperparameters in a distributed way is presented in the following.

5.4.3.1 Distributed parameters learning

The discussion followed here recovers the technique of learning parameters for the Gaussian field model, which is described in [5].

For each mobile agent i, it is straightforward to write down the marginal distribution of y_i as follows

$$y_i | \sigma_{t,i}^2, \phi_{t,i}, \tau_{t,i}^2, \beta_{t,i} \sim \mathcal{N}(X(\mathbf{s}_i)\beta_{t,i}, \Sigma_{t,i}), \tag{5.48}$$

where

- $\sigma_{t,i}^2, \phi_{t,i}, \tau_{t,i}^2, \beta_{t,i}$ are parameters that are estimated by the i-th sensor at time t from its collective measurements. The hyperparameters vector is defined as $\theta_{t,i} := (\sigma_{t,i}^2, \phi_{t,i}, \tau_{t,i}^2) \in \mathbb{R}_{\geq 0}^3$.

- $\Sigma_{t,i} = \Sigma_{\mathbf{s}_i\mathbf{s}_i} + \tau_{t,i}^2 I_i$. $\Sigma_{\mathbf{s}_i\mathbf{s}_i}$ is an $|\mathbf{s}_i| \times |\mathbf{s}_i|$ covariance matrix computed based on the i-th agent's own hyperparameters and I_i is an $|\mathbf{s}_i| \times |\mathbf{s}_i|$ identity matrix.

As a result, the log-likelihood function can be obtained by

$$\mathcal{L}(\theta_{t,i}, \beta_{t,i}) = -\frac{1}{2}\{(y_i - X(\mathbf{s}_i)\beta_{t,i})^T \Sigma_{t,i}^{-1}(y_i - X(\mathbf{s}_i)\beta_{t,i}) + \log\det(\Sigma_{t,i}) + |\mathbf{s}_i|\log(2\pi)\}. \tag{5.49}$$

If given $\theta_{t,i}$, $\beta_{t,i}$ will only be a variable. Thus, it is recognized that the log-likelihood function obtains the maximum value at

$$\widehat{\beta}_{t,i} = \left((X(\mathbf{s}_i)^T \Sigma_{t,i}^{-1} X(\mathbf{s}_i)\right)^{-1} X(\mathbf{s}_i)^T \Sigma_{t,i}^{-1} y_i. \tag{5.50}$$

However, in fact $\theta_{t,i}$ is not known, which can be estimated by utilizing the maximum likelihood approach [154]. Notice that the log-likelihood maximization procedure can be used recursively and for many hyperparameters.

Therefore, every robotic wireless sensor can itself learn its own parameters and hyperparameters; nevertheless, in order to predict the spatial field at unmeasured locations, these sensors need to know the global parameters.

Proposition 5.10. If a MRWSN is connected, each mobile sensor i can calculate the global parameters and hyperparameters by exchanging local information with

its neighbours as given

$$\xi(k+1) = \xi(k) + \gamma \sum_{j \neq i} a_{ij}(\xi_j(k) - \xi_i(k)), \tag{5.51}$$

where $\xi_i = (\theta_{t,i}, \beta_{t,i})$.

Proof: The results are derived directly by the DAC approach. ■

5.4.3.2 Predicted means and variances

As discussed in Section 5.4.3.1, at the time instance t when the DAC algorithm converges, every robotic sensor can obtain the global hyperparameters vector $\theta_t = (\sigma_t^2, \phi_t, \tau_t^2)$ and the global mean parameters vector β_t. Thus, the conditional posterior mean and variance at unobserved location u_* at time t, given the collective measurements of this sensor, can individually be computed by the i-th agent as specified

$$m_{u_*|\mathbf{s}_i} = X(u_*)\beta_t + \Sigma_{u_*\mathbf{s}_i}\left(\Sigma_{gt,i} + \tau_t^2 I_i\right)^{-1}(y_i - X(\mathbf{s}_i)\beta_t) \tag{5.52}$$

and

$$\sigma^2_{u_*|\mathbf{s}_i} = \sigma_t^2 - \Sigma_{u_*\mathbf{s}_i}\left(\Sigma_{gt,i} + \tau_t^2 I_i\right)^{-1}\Sigma_{\mathbf{s}_i u_*} \tag{5.53}$$

where $\Sigma_{u_*\mathbf{s}_i} (= \Sigma^T_{\mathbf{s}_i u_*})$ is a cross-covariance vector between $y(u_*)$ and y_i, $\Sigma_{gt,i}$ is covariance matrix of y_i computed based upon the global hyperparameters. Eventually, the predicted mean and variance at u_* given all the measurements gathered from time 1 to t by the MRWSN can be obtained through the DAC algorithm.

5.4.4 Distributed sampling paths

The distributed algorithms have been designed in the previous section to predict the spatial field at time t. In this subsection, an efficient sampling strategy that drives the mobile robotic wireless sensors to the most uncertain locations at time $t+1$ is proposed to effectively improve the quality of prediction. It is shown that each mobile agent always looks for a next location at time instance $t+1$. In order to address this problem, the one-step-ahead forecast is utilized to predict the latent spatial values at time $t+1$, given the collective observations up to time t.

Given y_i, the agent i can forecast the latent spatial process at time step $t+1$ as follows

$$\hat{z}(s_{t+1,i})|y_i = z(s_{t,i})|y_i + \eta_{t+1,i}, \tag{5.54}$$

where $\eta_{t+1,i} \sim \mathcal{N}(0, \sigma^2_{sys})$, σ^2_{sys} is a system error that describes the changes in the elements of the parameters vector between times t and $t+1$. Then, the posterior variance of the posterior distribution of $\hat{z}(s_{t+1,i})$ conditioned on y_i and $\hat{y}(s_{t+1,i})$ is given by

$$\hat{\sigma}^2_{s_{t+1,i}|\mathbf{s}_i} = \sigma^2_t - \Sigma_{s_{t+1,i}\mathbf{s}_i}\left(\Sigma_{gt,i} + (\tau^2_t + \sigma^2_{sys})I_i\right)^{-1}\Sigma_{\mathbf{s}_i s_{t+1,i}}. \tag{5.55}$$

Since the sum of the spatial range and the movement range of each agent is smaller than the communication range R, the posterior variance of $\hat{z}(s_{t+1,i})$ given all measurements $y_{1:t}$ and $\hat{y}(s_{t+1})$ is only computed by the mobile sensors whose distances to $s_{t+1,i}$ are less than R. Note that the DAC method is utilized for this distributed calculation.

Finally, it is decided at time $t+1$ every robotic wireless sensor moves to one of the unmeasured locations inside its allowable movement region that has a maximum posterior variance forecasted.

5.5 Experimental Results

In this section, the effective centralized and distributed approaches for the sensor placement schemes in modelling, estimating and predicting the spatial fields are validated by the utilization of real-world data sets. Moreover, the sampling strategies that are designed by the base station from all the observations (centralized method) or the individual agent from its neighbours' information (distributed method) in order to move the mobile robotic sensors to the optimal sensing locations are efficiently demonstrated. The centralized approach was implemented in a small-scale environment, where all the sensors could transmit their observations to the sink, whereas the distributed approach was conducted in a large-scale deployed network. Note that all the experiments were implemented in a two dimensional environment.

5.5.1 Centralized sampling scheme

The efficiency of the centralized Gaussian inference for the spatial prediction is illustrated in the following experiment using a real-world data set from the Intel Berkeley Research Lab [135]. More importantly, the discussion of the centralized sampling navigation primarily concentrates on the proposed approach based on the conditional entropy; nevertheless, the analysed results are also compared with those of the locational optimization based method.

A spatial field of the indoor temperature was reconstructed as shown in Fig. 5.3. There were N MRWSs used with constrained individual displacements of a maximum radius of $v_{max} = 4$ m in every time step. The desired buffer width was set to $\omega = 0.3$ m. v_{max} and ω are usually chosen based on the dimensions and abilities of commercial robots. The spatial field was divided into a 100×100 regular grid,

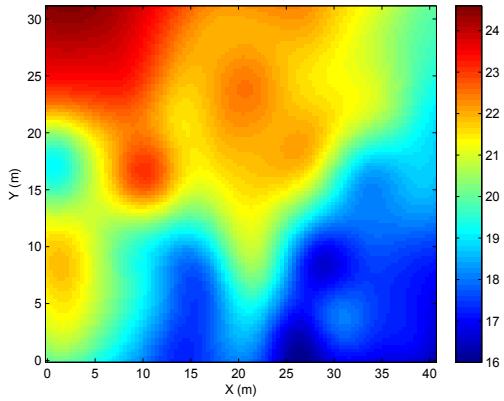


Figure 5.3: True field of spatial temperature. Colour bar shows range of temperature.

which makes $Q = 10000$. The simulations results were appraised in terms of the number of the wireless robotic sensors by identifying the period of sampling time in which the computed entropy is guaranteed by a lower bound and by computing the predicted field of interest at the time step of stopping.

For instance, it was considered that three scenarios of $N = 5$, $N = 10$, and $N = 20$ MRWSs were used to monitor the temperature field. In each case, all MRWSs were started from pre-defined locations. The near-optimal values of the conditional entropy $H\left(\widehat{y^*}_{t+1}|y_{1:t}\right)$ obtained by the proposed algorithm are demonstrated in Fig. 5.4. As expected, the uncertainty of prediction in each iteration reduces significantly as the number of the MRWSs increases. It can be noted that the conditions imposed on the conditional entropy as introduced in the **Lemma 5.6** for the scenarios $N = 5$, $N = 10$, and $N = 20$ MRWSs are $[\log(2\pi e)^{-5} = -14.1894, +\infty)$, $[\log(2\pi e)^{-10} = -28.3788, +\infty)$, and $[\log(2\pi e)^{-20} = -56.7575, +\infty)$, respectively. Thus, the conditional entropy solutions of the (5.30) were not guaranteed by a lower bound after time instants $t = 19$, $t = 12$, and $t = 6$ in the corresponding scenarios. Therefore, the wireless networks with $N = 5$, $N = 10$, and $N = 20$ mobile sensors

stop sampling at time steps $t = 19$, $t = 12$, and $t = 6$ respectively. The upper bounds are also shown in the Fig. 5.4 for completeness.

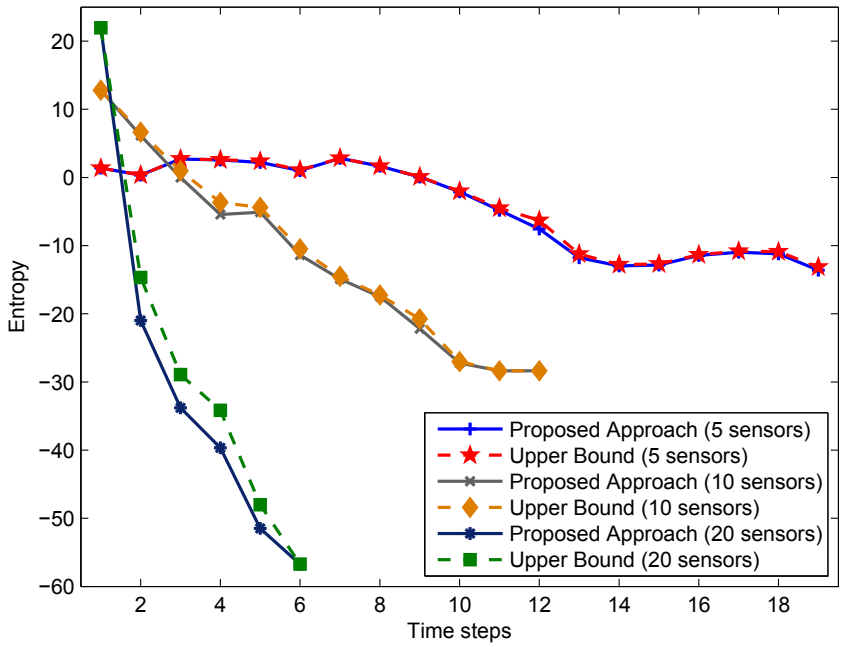
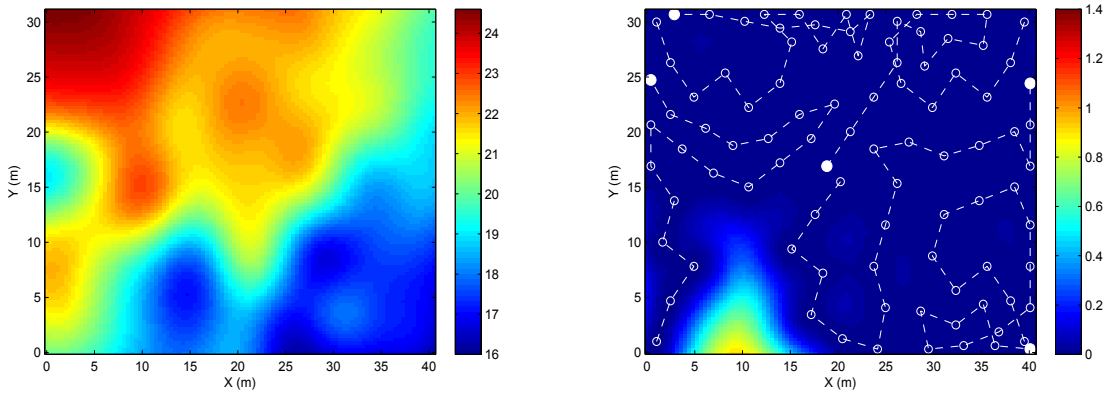


Figure 5.4: Conditional entropies $H\left(\widehat{y}_{t+1}|y_{1:t}\right)$ in the networks of 5, 10, 20 sensors, respectively.

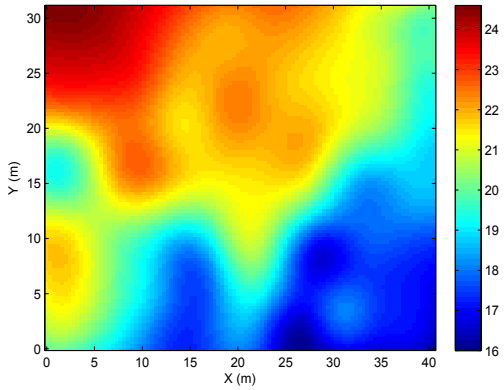
Let us consider the prediction accuaracies of above results. Figures 5.5a to 5.5e show the the predicted temperature fields by using $N = 5, 10$, and 20 mobile sensor networks at times $t = 19, 12$, and 6 respectively. It can be qualitatively seen that the predicted means are very much similar to the true field given in Fig. 5.3. White circles in Fig. 5.5b, 5.5d, and 5.5f show the trajectories of the near-optimal paths of the MRWSs up to time instants 19, 12, and 6 for three scenarios in this illustrative example whereas the colour map represents the prediction error variances. Note that the white solid dots in these figures correspond to the current sensor locations.

The final root mean square errors (RMSE) of the predicted field at Q spatial sites of all the scenarios have also been calculated. Here, the root mean square error

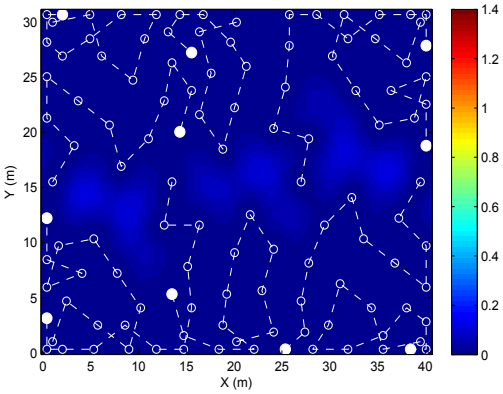


(a)

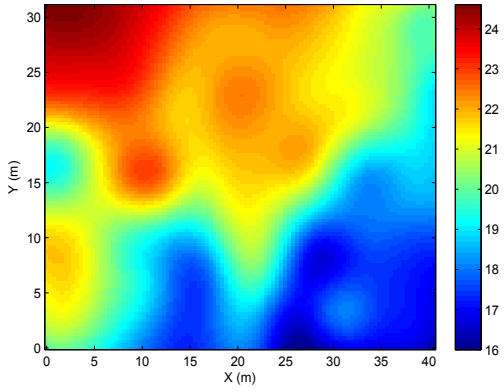
(b)



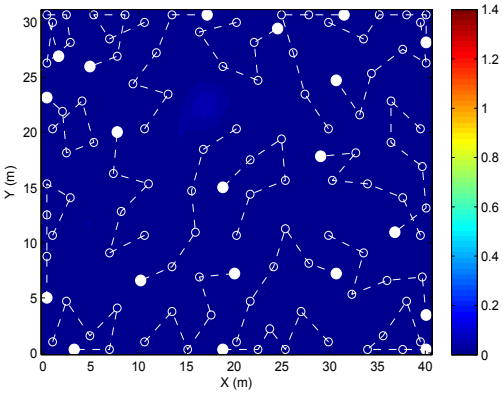
(c)



(d)



(e)



(f)

Figure 5.5: The predicted field (a) at time $t = 19$ using $N = 5$ sensors, (c) at time $t = 12$ using $N = 10$ sensors, and (e) at time $t = 6$ using $N = 20$ sensors. The predicted error variances and the sampling paths (b) at time $t = 19$ using $N = 5$ sensors, (d) at time $t = 12$ using $N = 10$ sensors, and (f) at time $t = 6$ using $N = 20$ sensors. Colour bars in left-wise column represent range of temperature, and colour bars in right-wise column represent range of errors.

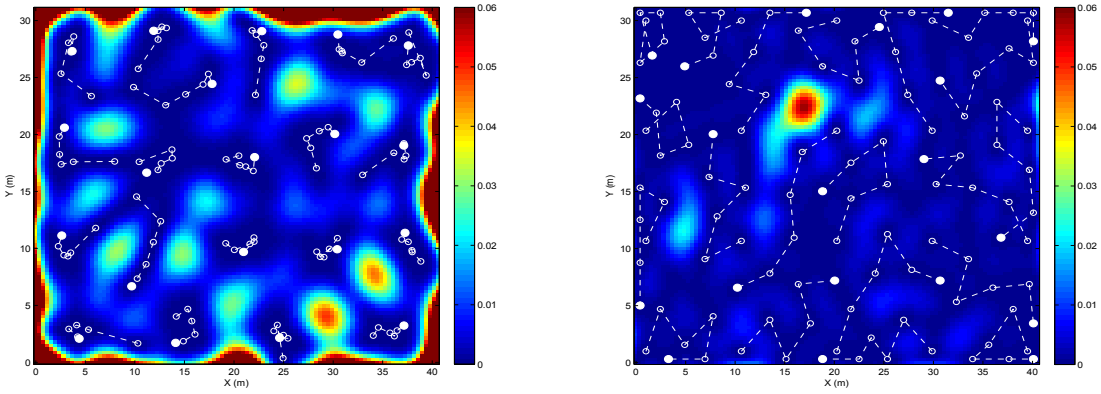


Figure 5.6: Comparison of the predicted error variances and the sampling paths between (a) locational optimization and (b) entropy based approaches. Range of errors is shown in colour bars.

is defined by

$$\text{RMSE} = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q}\left(\mu_{z_i|y_{1:t}} - z_i\right)^2}. \tag{5.56}$$

The computations show that the RMSE obtained by the network of 5 sensors at $t = 19$ is 0.2665^oC, the RMSE obtained by the network of 10 sensors at $t = 12$ is 0.1559^oC, and the RMSE obtained by the network of 20 sensors at $t = 6$ is 0.1347^oC. The results are reasonable for practical applications such as the one given in [157] where the prediction errors of daily temperature are set from 1.4^oC to 2.3^oC. However, if the requirement is tighter, it is still possible to achieve by increasing the number of mobile sensors which eventually contribute to a reduced RMSE.

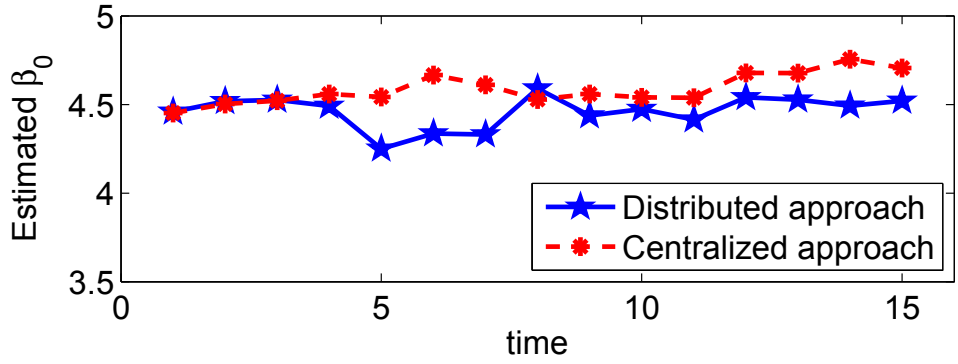
Another important contribution of this work is the reduced computational complexity. In this particular illustrative example of $v_{max} = 4$ m and $N = 5$, at $t = 19$ the proposed algorithm took approximately 15 seconds to address the problem (5.30), whereas the solution to (5.29) required approximately 25 hours run time implemented on R V3.0 with a PC of 3.1GHz Intel Core i5-2400 Processor.

More interestingly, the Figure 5.6 demonstrates the predicted error variances in the field and the sensing paths for $N = 20$ robotic sensors to discover the spatial phenomenon at the time steps from $t = 1$ to $t = 6$. The comparison has shown that the conditional entropy based approach tends towards the locations along the borders of the sensing areas at each time iteration; on the contrary, the locational optimization based method is inclined to pick the locations around centroids of the Voronoi partitions of the sensors' current locations. The former illustrates an ability to gain a better coverage, which results in very small values of the predicted errors in the whole field, while the later results in the high uncertainty in the borders. Note that due to very complicated shape of the density function, in some scenarios the locational optimization based method can only find the locally optimal locations for the mobile sensors.

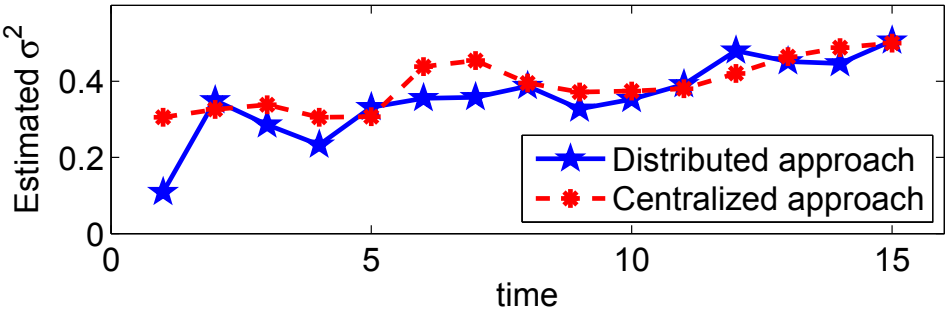
5.5.2 Distributed sampling scheme

In this subsection, experiments based on the published data set that was collected from a real-world field in Benton county, Indiana, USA [158] were conducted in order to validate the effectiveness of the distributed approaches proposed in the Section 5.4 for the sensor placement. More particularly, by utilizing the information from the local neighbours, individual mobile sensor can model the spatial field, estimate the model parameters, predict the field at locations of interest, and automatically deploy itself.

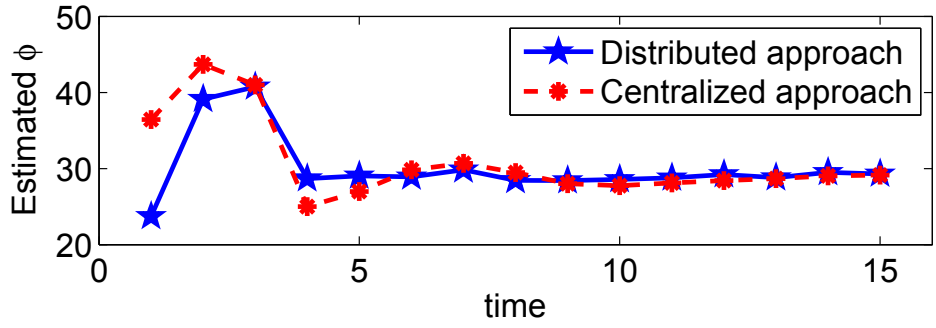
Nitrogen fertilizer applications are essential for the production of crops in the agricultural field. Since nitrogen levels, supplied by mineralization of soil organic matter, can vary considerably across the field, soil sampling is crucial for the accurate description of spatial patterns in soil fertility. Therefore, in this study, it has been



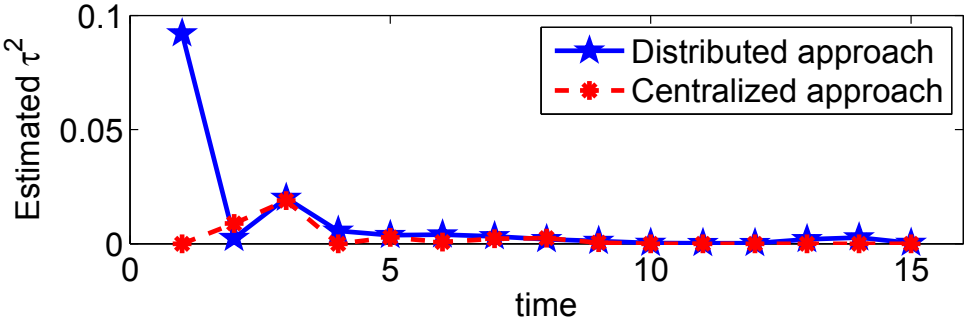
(a)



(b)



(c)



(d)

Figure 5.7: The globally estimated: (a) mean parameter β_0 (b) marginal variance σ^2, (c) bandwidth of the kernel ϕ, and (d) noise variance τ^2.

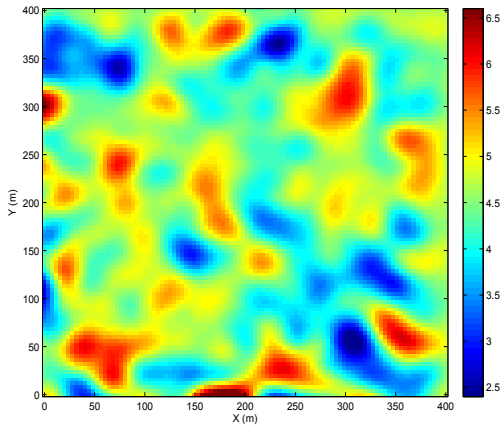
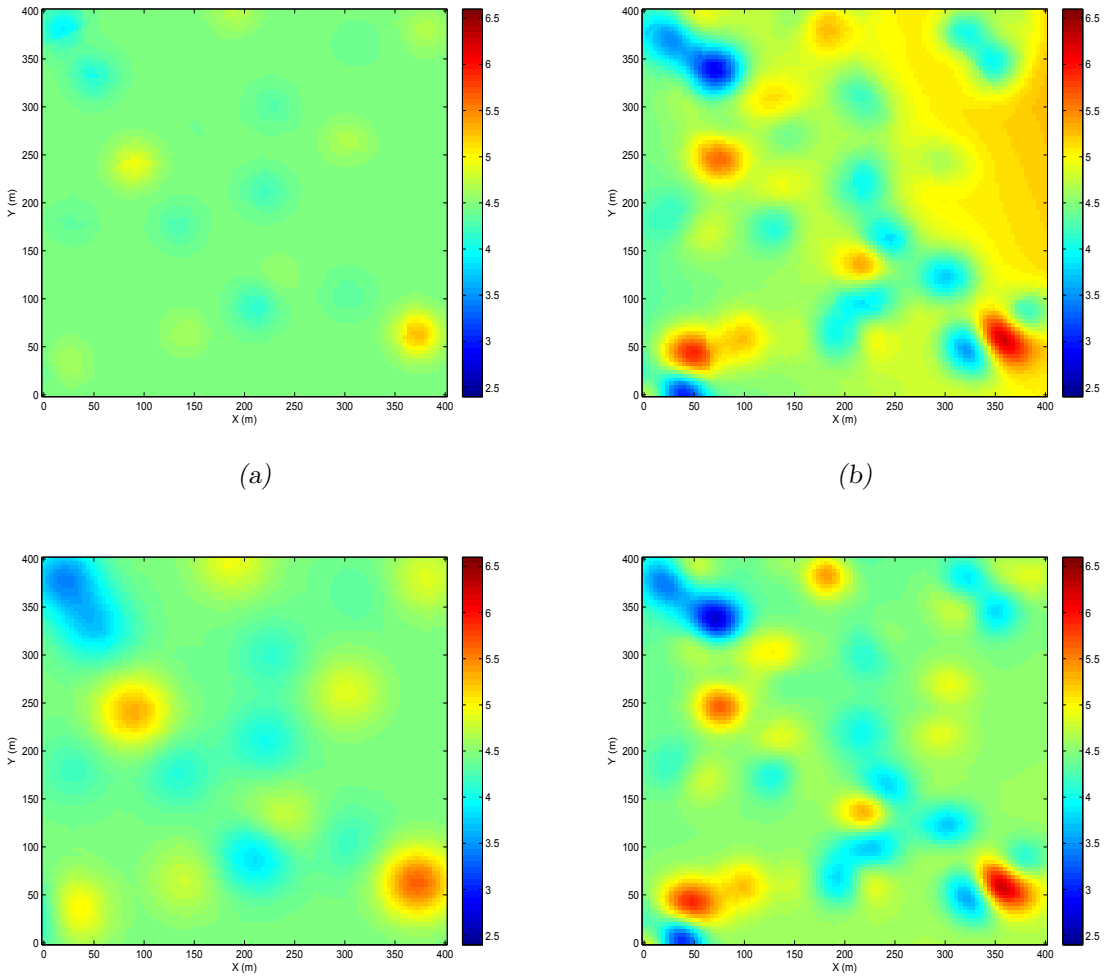


Figure 5.8: The true field of the soil organic matter. Percentage of the soil organic matter is shown in color bar.

proposed to employ the MRWSNs to construct the soil organic matter distribution that eventually can be used to recommend a strategy for application of the nitrogen fertilizer. The soil organic matter field was regenerated in [159] from the data of 1375 soil measurements. The size of this field is 400 m ×400 m. The true levels of the soil organic matter are illustrated in Fig. 5.8. In this experiment, at the beginning, $N = 20$ mobile robotic wireless sensors were randomly deployed in the farm as demonstrated in Fig. 5.10a in which white dots are sensor locations. The maximum distance that each mobile agent can travel between time steps was set to 10 m. The spatial range f and the communication range R were set to 70 m and 100 m, respectively. It was assumed that σ^2_{sys} was known and set to 0.05.

For comparison purposes, the centralized GP method was also implemented in the reference field. It is to be noticed that due to using all measurements for computations, the centralized algorithm shows the best performance in all cases. Fig. 5.7 demonstrates globally estimated model parameter and hyperparameters. It can be clearly seen that the distributed parameters computed by the proposed algorithms are highly comparable to those of the centralized method. In Fig. 5.9,

the predicted field is illustrated at particular times by the distributed algorithm and the standard GP regression method. The column-wise pairs of Fig. 5.9a and Fig. 5.9c, Fig. 5.9b and Fig. 5.9d, Fig. 5.9e and Fig. 5.9g, Fig. 5.9f and Fig. 5.9h are distributed and centralized images of the spatial field of interest estimated at the times $t = 1$, 5, 10, and 15, respectively. Fig. 5.10 shows not only the trajectories of the sampling paths of the mobile robotic wireless sensors in white circles but also



(a) *(b)*

(c) *(d)*

Figure 5.9: The column-wise pairs compare the predicted fields obtained by distributed and centralized approaches at times (a) and (c) $t = 1$, (b) and (d) $t = 5$. Percentage of the soil organic matter is shown in color bar.

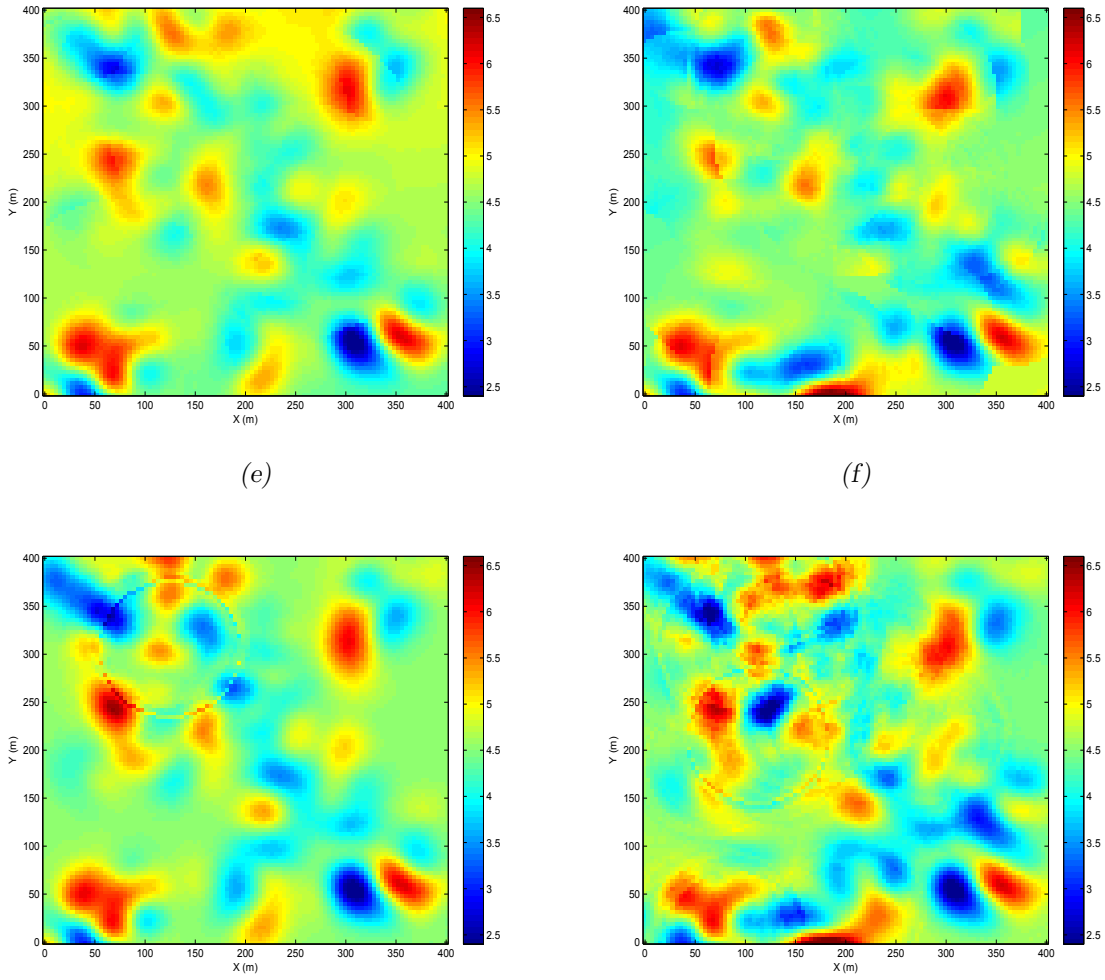


Figure 5.9: The column-wise pairs compare the predicted fields obtained by distributed and centralized approaches at times (e) and (g) $t = 10$, (f) and (h) $t = 15$. Percentage of the soil organic matter is shown in color bar.

the prediction error variances of interested points, corresponding to $1^{st}, 5^{th}, 10^{th}$, and 15^{th} instants. The white dots in these figures correspond to the current mobile sensor locations.

Fig. 5.11 shows the root mean square errors (RMSEs) of the predicted field, which is computed via (5.56). As expected, Fig. 5.11 shows a gradual reduction of RMSEs with increased number of observations or with more exploration. The

results of the proposed algorithm are reasonably close to those of the standard GP technique with centralized approach which is appealing.

The convergences of the DAC employed in the algorithm are shown in Fig. 5.12. These instances are considered at time step 10, in which the model parameters and hyperparameters at each mobile wireless sensor are globally updated until they converge to global solutions. For the purpose of stochastic evaluation, these parameters were taken at the different sensors.

5.6 Summary

This chapter has presented the sensor placement for the MRWSNs in predicting the spatial fields. It has been proposed to efficiently and effectively model the spatial phenomena by the Gaussian processes; the centralized and distributed inference approaches are statistically derived.

In the context of the centralized schemes, the sampling strategy has been considered in both the continuous and discrete spaces. The optimal sampling paths for the MRWSs have been addressed by the use of the locational optimization technique, where the density function has been computed based on the estimated variances. However, a near-optimal solution to the motion planning problem for a network of mobile, wireless sensors for monitoring spatial phenomena in polynomial time has been presented in the discretized space. The computational complexity of the near-optimal but efficient proposition has been shown to be practically feasible for the resource-constrained MRWSNs. The conditional entropy solutions of the adaptive sampling optimization problem have been proven to be bounded by an upper bound and a lower bound. Particularly, the lower bound has been proposed as a stopping

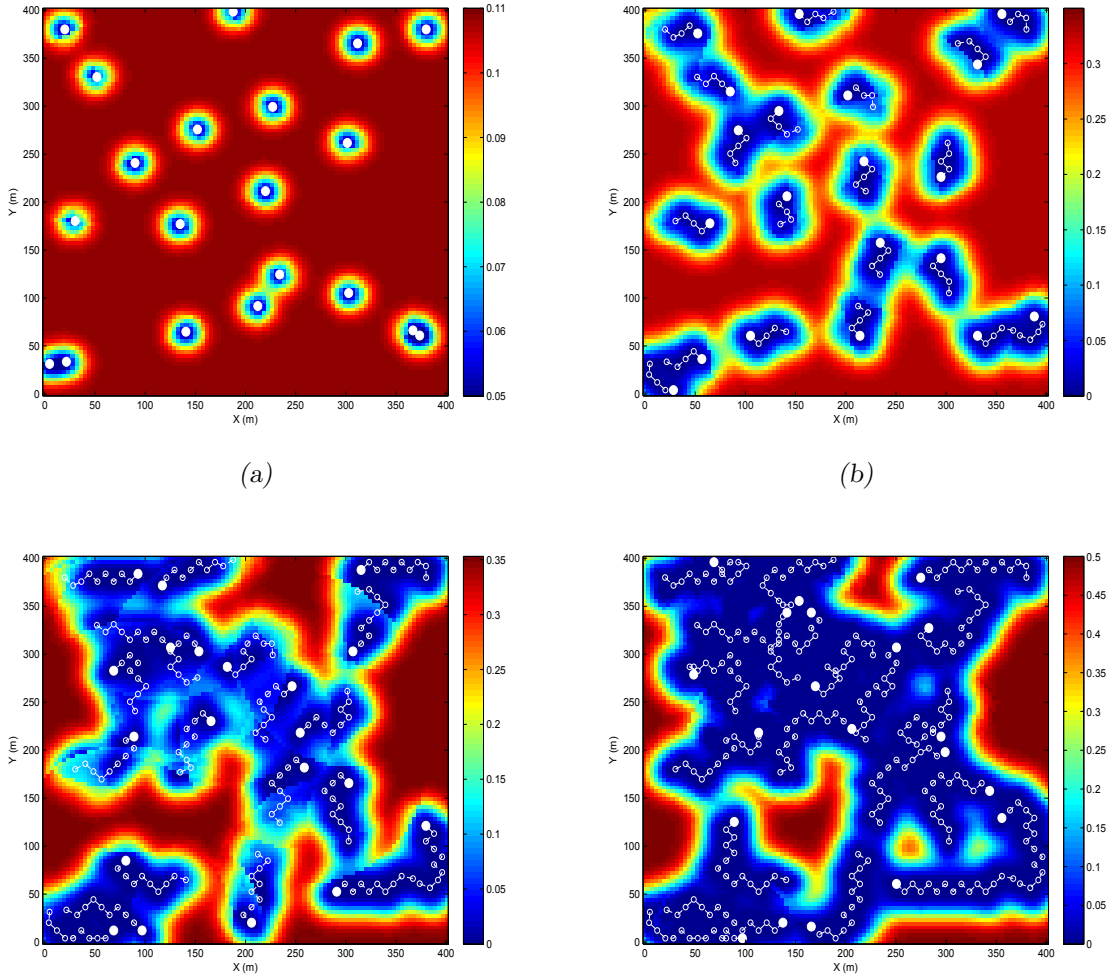


Figure 5.10: The globally predicted variances and the sampling paths at times (a) $t = 1$, (b) $t = 5$, (c) $t = 10$, and (d) $t = 15$. Range of errors is illustrated in colour bars.

criterion for the sampling algorithm, which guarantees the prediction results to be within a user-defined accuracy.

Furthermore, a distributed spatial prediction method to efficiently compute the centralized architecture of the sophisticated statistical inference for a resource-constrained MRWSN performing spatial estimation tasks has been proposed. The algorithm has been proposed based upon two useful computation tools of the JOR

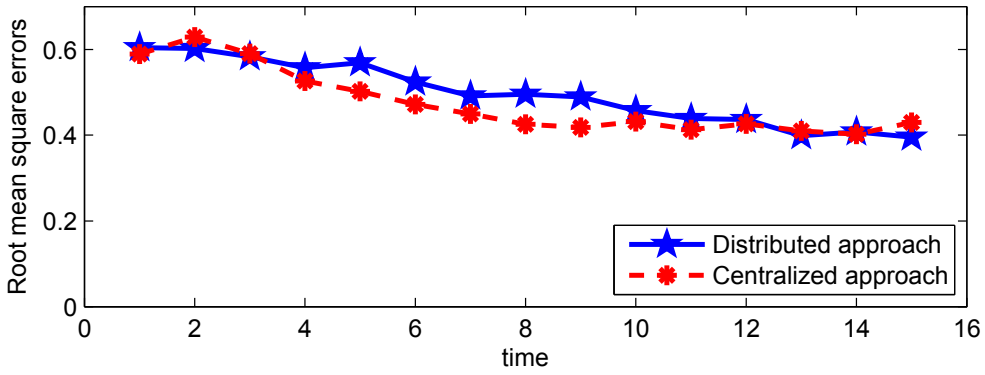
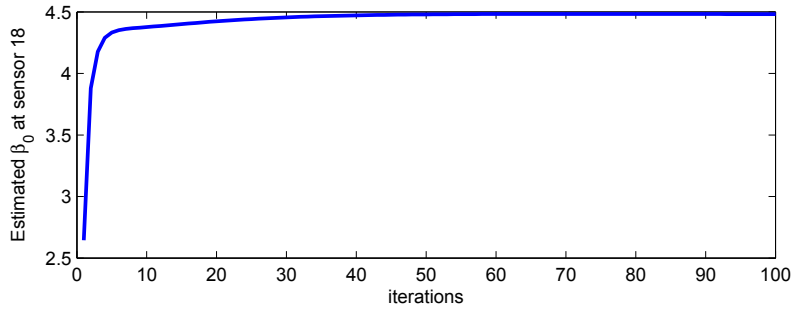


Figure 5.11: Root mean square errors.

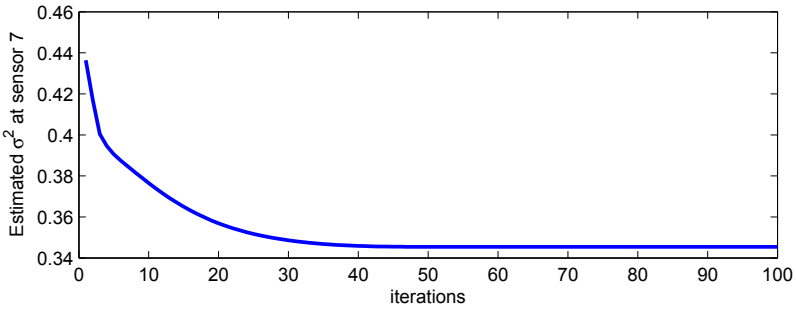
and the DAC and has been executed by each mobile sensor using only local information from its neighbours. The proposed approach allows a MRWSN with limited wireless communication range, memory and computational power to compute Gaussian inference without requiring the inverse of the covariance matrix whose complexity is cubic in dimension. A locally acting mobile sensor can predict globally the spatial field at unmeasured locations of interest by exchanging the information in the neighbourhood structure. Sampling paths for mobile robotic sensors in a distributed fashion have been designed using the most uncertain locations so that the quality of parameter estimation and spatial inference is improved.

Finally, the proposed approaches have been extensively evaluated using the experiments of the real-life pre-published data sets, which has been then demonstrated with promising results.

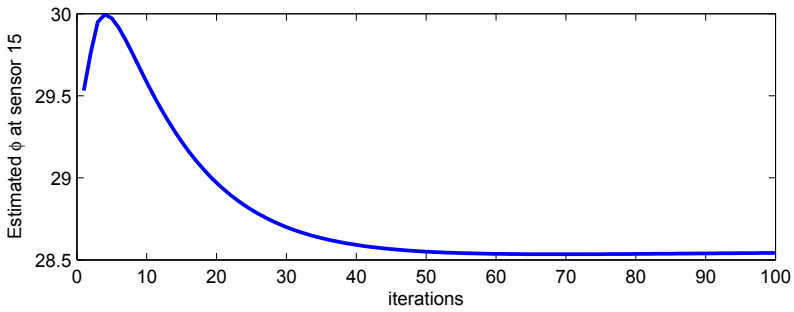
Nonetheless, the prime disadvantage of the GP is in computing the covariance matrix. Since factorizing this dense matrix is cubic in the dimension, computation in the GP regression faces more difficulty if the number of measurements is increasing. Consequently, in the next chapter, a new model based on the GMRF will be proposed to replace the GP in modelling the spatial field, which eventually will lead to considerable benefits in computation for the resource-constrained MRWSNs.



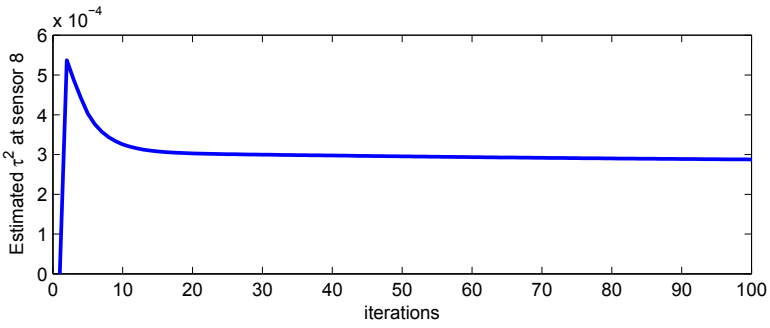
(a)



(b)



(c)



(d)

Figure 5.12: Convergences of the DAC algorithms of different sensors at time step 10: (a) estimated mean parameter β_0 of sensor 18, (b) estimated marginal variance σ^2 of sensor 7, (c) estimated bandwidth of the kernel ϕ of sensor 15, and (d) estimated noise variance of sensor 8.

Chapter 6

Spatial Estimation using Gaussian Markov Random Field

6.1 Introduction

The advantages of utilizing the GP to model the spatial field in the spatial prediction problem have been investigated to solve the sensor placement problem in chapter 5. The GP delineates the continuous interpretation of the spatial field very well; nevertheless, the size of the covariance matrix is proportional to the number of sensor measurements. That is, the level of computational complexity in the GP regression as well as the sampling optimization problem increases as the MRWSs navigate throughout the environment over time. There have been some pre-published works addressing this issue in different ways. For instance, in their work [112], Williams *et al.* proposed to employ the Nyström method, which is used to numerically address the eigenproblems, to approximately compute the eigendecomposition of the Gram matrix. The proposition enables the kernel-based predictors to be carried out on

a smaller system compared with the original one of all the observations. Besides, [113] introduced a simple sparse greedy approach to calculate the approximation of the maximum posteriori estimation of the GP, which diminishes the amount of computation required to find the solution scales. In an effort to overcome the computational complexity in the GP regression, the authors in [114] introduced a framework for sparse GP techniques. The method first considered randomized greedy selections based on information-theoretic principles, then alternatively represented the GP model that facilitates generalizations to regression. In [109], the authors proposed to truncate the sensor measurements in order to deal with the complexity in computing the covariance matrix. The proposition was assumed that the model parameters were known a priori. However, in the present study, it is proposed to estimate the mean parameters and hyperparameters online at each time iteration as observations are gathered, which requires all the measurements to be presented. Recently, the challenge regarding the covariance matrix in the GP has been dealt with by replacing the GP by a computationally efficient Gaussian Markov random field (GMRF) [15, 101, 115–118]. The GMRF approximately represents the spatial field on a lattice. It is to be noted that in the extreme cases of very small neighbours in the lattice, the GMRF can approximate the GP surprisingly well [115]. The computational benefits of the GMRF come from the sparsity of the precision matrix, an inverse of a dense covariance matrix, whose zero elements relate directly to conditional independence assumptions, introduced in section 2.3.2. Due to computational advancement and scalability, the GMRF has received much attention for resource-constrained MRWSNs as compared to the standard GP [101, 119–122]. In [122], Xu *et al.* proposed a new class of GP built on a GMRF for the spatial prediction problem. Nonetheless, this proposition requires the precision matrix to be given a priori. Although [119–121] provided algorithms to derive the sampling locations for

the MRWSNs, their solutions were restricted to a regular lattice. Fitting the GMRF to the GP on the regular lattice, as discussed in [118], has to be precomputed for a discrete set of parameter values. As a consequence, the contribution in this work is to employ the GMRF model described on an irregular lattice so as to study a spatial estimation approach in the resource-constrained MRWSNs.

More importantly, designing optimally adaptive sampling paths that enable the MRWSs to be driven to the most informative locations to observe the physical environment, as proved in [160], is NP-hard but can be approximately computed by a greedy heuristic algorithm. However, the complexity of resolving sampling strategies by the greedy approach that requires factorizing of covariance or precision matrices has not received much attention from the existing works. Particularly, the algorithm often becomes intractable due to the need of finding the optimal sampling points in a large number of unmeasured locations of interest. Hence, the present work in the first stage concentrates on modelling the spatial field by the GMRF representation on an irregular discrete space in order to derive the near-optimal navigation schemes with a reasonable running time.

The remainder of this chapter is organized as follows: In Section 6.2, it is proposed to utilize the GMRF that is represented on an irregular lattice of a triangulation by using a specific SPDE approach to model the physical spatial field. Centralized and distributed prediction schemes from the GMRF standpoint for the physical process are derived in Section 6.3. By analysing a technique of the one-step-ahead forecast to predict the future measurements, an adaptive sampling optimization problem in terms of information-theoretic criteria by the use of the GMRF models is first formulated in Section 6.4. It is demonstrated that the existing approximate greedy algorithms to address the combinatorial and NP-hard problem

of optimizing the sampling paths for the MRWSNs are intractable if the networks search for a large number of unobserved locations in the field. Then, an efficient novel optimality criterion for the adaptive sampling problem is proposed so that the running time of a greedy algorithm is deterministic even though either the number of measurements goes up or the unmeasured locations are dense on a very large mesh. For the case where each mobile sensor has a short wireless communication range, a distributed sampling scheme for each robotic agent is also designed in this section. The distributed approach enables each robotic sensor to communicate only with its neighbouring sensors but to approximately reach to global estimation and prediction values. Efficiency of the proposed approaches is extensively evaluated using two pre-published data sets in Section 6.5. Finally, the summaries are provided in Section 6.6.

6.2 Spatial Field Model via GMRF

In this section, the basic concepts and results on the spatial field model that is used in this chapter are introduced. The discussion here follows the model represented in Section 4.2; yet the component of the random field is specified by a GMRF as introduced in the Section 2.3.2.

Let $\mathcal{Q} \subset \mathbb{R}^d$ be considered as the spatial field of interest that is discretized into n Voronoi partitions 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. 6.1 and is created by exploiting the stochastic partial differential equations (SPDE) that is discussed in Section 6.3.1. In a real application, a much denser mesh will be considered. Let spatial locations at the vertices of the mesh

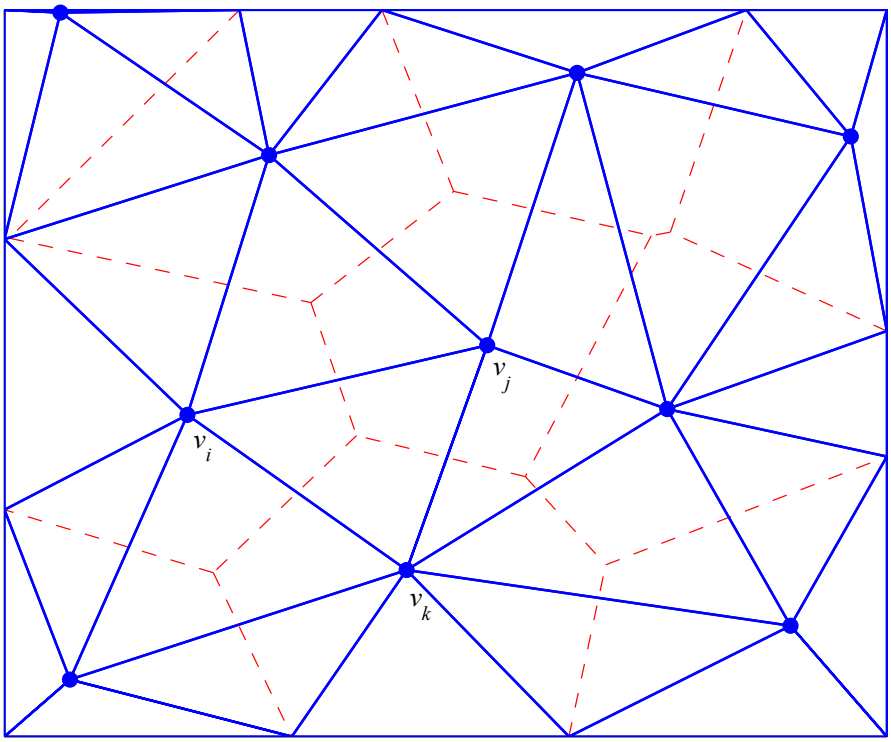


Figure 6.1: The spatial field $\mathcal{Q} \subset \mathbb{R}^2$. Voronoi cells are partitioned by dash red lines and a mesh of a triangulation is created by solid blue lines.

denote as $\mathbf{v} := (v_1^T, v_2^T, ..., v_n^T)^T$ and the random vector $y'(\mathbf{v})$ denoted by $y'(\mathbf{v}) := (y'(v_1), y'(v_2), ..., y'(v_n))^T$ describes a vector of latent variables of the field. In this study, the latent variables are modelled as a summation of a large scale component and a random field. The model is defined by

$$y'(\mathbf{v}) = X(\mathbf{v})\beta + z(\mathbf{v}), \tag{6.1}$$

where

- $X(\mathbf{v})\beta = \mu(\mathbf{v})$ is the expectation of $y'(\mathbf{v})$. $X(\mathbf{v})$ denotes a matrix of spatially referenced non-random variables (known as covariates) at location $\mathbf{v}$ and β is

a vector of mean parameters;

- z is a GMRF with a zero mean vector and a 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 6.3

As discussed in [118], 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 in Section 4.2

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

where $h := \| v_i - v_j \|$, 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.

When the model of a network of mobile wireless sensors is considered, as generally analysed in Section 2.6.1, agents are equipped with identical sensors and take measurements at times $t \in \mathbb{Z}_{>0}$. At time t, agent i takes a noisy measurement at its current location $s_{t,i} \in \mathcal{Q}$, specified by

$$y_{t,i} = X(s_{t,i})\beta_t + z(s_{t,i}) + \varepsilon(s_{t,i}), \tag{6.2}$$

where β_t is the mean parameter vector estimated at time t, $\varepsilon(s_{t,i})$ is a noise with a zero mean and a known variance σ^2_{ε}.

6.3 Spatial Predictive Inference via GMRF

In this section, the spatially centralized prediction schemes via the GMRF using the SPDE approach are first introduced. The spatially distributed implementation of the centralized schemes is then proposed. In these schemes, the unobserved locations of interest are on the vertices of the mesh as defined in Section 6.2. In the first subsection, how the GMRF is constructed by the use of the SPDE is presented.

6.3.1 The SPDE approach

When dealing with the spatial field, the SPDE approach proposed by Lindgren *et al.* [118] 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, for example 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" [111]. In fact, the "big n problem" in the GP defines a computational burden while working with algebra operations that are with $n \times n$ dense covariance matrices.

As shown in [118], the SPDE approach utilizes the Finite Element method [161] 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 L spatial locations $s_1, s_2, ..., s_L$; 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

$$z(s) = \sum_{k=1}^{n} f_k(s) w_k, \tag{6.3}$$

where $\{f_k(s)\}$ represents the basis function, and $\{w_k\}$ represent Gaussian distributed weight. The functions $\{f_k(s)\}$ are piece-wise linear on each triangle, which denotes that $f_k(s)$ is 1 at the k^{th} vertex of the mesh and 0 at all other vertices. w_k is the value of the spatial field at each triangle vertex k. Therefore, in the perspective of the continuous domain, the GP can be developed by the joint distribution of the weights $\{w_k\}$. It can be clearly seen that the finite element representation (6.3) 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 funtion to the elements of the precision matrix Q.

The precision matrix Q with the size of $n \times n$ is calculated by, as shown in [118],

$$Q = \tau^2(\kappa^4 C + 2\kappa^2 G + G C^{-1} G), \tag{6.4}$$

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 and is computed by

$$\tau = \sqrt{\frac{\Gamma(\nu)}{\Gamma(\nu+1) 4\pi \sigma_{\varepsilon}^2 \kappa^{2\nu}}}.$$

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, let $\theta := (\log(\tau), \log(\kappa))$ define a hyperparameter vector.

It can be seen that the GMRF representation constructed by the linear basis functions holds the sparse property of the precision matrix.

6.3.2 Centralized prediction schemes

In this subsection, it is proposed to consider the spatial field model as described in Section 6.2 and impose unmeasured locations upon the vertices on the mesh. Suppose that the latent random field at the n vertices of the triangulation at time t is distributed as

$$z_t|\theta_t \sim \mathcal{N}\left(\mathbf{0}, Q_t^{-1}\right), \tag{6.5}$$

where θ_t is the hyperparameter vector learned at time step t, $Q_t = Q(\theta_t)$.

Furthermore, a concept of a projector matrix, A, is introduced that projects the latent random field modeled at the triangulation vertices to the data locations. This is equivalent to saying that this sparse weight matrix is needed to map between the basis function representation responded at n locations and the random field with a dimension of tN. At the time instant t, the size of the projector matrix, A_t, is $tN \times n$. The projector matrix A_t is built in such a way that each row has only one non-zero element of value 1. This non-zero element is indexed by the correspondingly observed location in the vertices set of the triangulation.

Under the spatial field model, it is given

$$y_{1:t}|z_t, \theta_t, \beta_t, \sigma_\varepsilon^2, A_t \sim \mathcal{N}\left(X(s_{1:t})\beta_t + A_t z_t, \sigma_\varepsilon^2 I\right), \tag{6.6}$$

where $X(s_{1:t})$ denotes a matrix of covariates, β_t, θ_t are parameters estimated by using all the available observations from time 1 to t, and I denotes the $tN \times tN$ identity matrix. If all model parameters are learned, the joint distribution that is computed by employing the technique in [154] is specified by

$$z_t, y_{1:t}|\theta_t, \beta_t, \sigma_\varepsilon^2, A_t \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ X(s_{1:t})\beta_t \end{bmatrix}, \begin{bmatrix} Q_t^{-1} & Q_t^{-1}A_t^T \\ A_t^T Q_t^{-1} & \sigma_\varepsilon^2 I + A_t Q_t^{-1} A_t^T \end{bmatrix}\right). \tag{6.7}$$

One can also represent the covariance matrix in (6.7) in terms of the corresponding partitioned precision matrix. To do this, it is proposed to utilize the block-wise inversion approach [162] and the Schur complement. It is given by

$$\begin{bmatrix} Q_t^{-1} & Q_t^{-1}A_t^T \\ A_t^T Q_t^{-1} & \sigma_\varepsilon^2 I + A_t Q_t^{-1} A_t^T \end{bmatrix}^{-1} = \begin{bmatrix} Q_t + A_t^T(\sigma_\varepsilon^2 I)^{-1}A_t & -A_t^T(\sigma_\varepsilon^2 I)^{-1} \\ -(\sigma_\varepsilon^2 I)^{-1}A_t & (\sigma_\varepsilon^2 I)^{-1} \end{bmatrix}. \tag{6.8}$$

In probabilistic terms, the full conditional distribution of z_t given $y_{1:t}$ is also Gaussian, derived from (6.7) and (6.8) as follows

$$z_t|y_{1:t}, \theta_t, \beta_t, \sigma_\varepsilon^2, A_t \sim \mathcal{N}\left(\mu_{z_t|y_{1:t}}, Q_{z_t|y_{1:t}}^{-1}\right), \tag{6.9}$$

where

$$\mu_{z_t|y_{1:t}} = Q_{z_t|y_{1:t}}^{-1} A_t^T (\sigma_\varepsilon^2 I)^{-1}(y_{1:t} - X(s_{1:t})\beta_t), \tag{6.10}$$

$$Q_{z_t|y_{1:t}} = Q_t + A_t^T(\sigma_\varepsilon^2 I)^{-1}A_t. \tag{6.11}$$

Note that the full conditional distribution represents the posterior distribution of the random field given the observations. Nonetheless, in order to compute the prediction mean of the predicted field of interest, the large scale component needs to be added to $\mu_{z_t|y_{1:t}}$. For instance, the predicted mean at an unobserved vertex v_i at time t is $X(v_i)\beta_t + \mu_{z_t|y_{1:t}}[i]$. Furthermore, the predicted variance at v_i at instant t is $Q^{-1}_{z_t|y_{1:t}}[i,i]$.

Remark 6.1. The primary computation of (6.10) is to factorize the sparse matrix $Q_{z_t|y_{1:t}}$, which can be obtained in a short time. On the other hand, since $Q_{z_t|y_{1:t}}$ is not dependent on collection of measurements, $\mu_{z_t|y_{1:t}}$ can be computed in $\mathcal{O}(1)$. If the standard GP model [13] is utilized, where the computational complexity of the prediction scheme is cubic in the dimension of the dataset, the GP-based computational issues become intractable with the increase in the number of observations over time. ■

6.3.3 Distributed prediction schemes

Employing the useful distributed discrete-time average consensus (DAC) technique described in Section 5.4.1.2, the distributed estimation approach is presented for every mobile sensor in the MRWSNs as they navigate through the environment and gather the measurements over time. It is shown that under the assumption of the communication range of R, each robotic sensor can exchange its information with its neighbours. In equivalent words, at every time step the mobile wireless sensor not only senses the spatial field at its current location but also receives its neighbours' measurements. Therefore, it is assumed that every mobile agent can store all its own measurements and its neighbour sensors' readings that have been transmitted

to it from time 1 to t to estimate its own parameters and hyperparameters and to predict the spatial field.

Consider a MRWSN of N agents; it is supposed that every robotic sensor knows the mesh of the triangulation and is located at an arbitrary vertex of the mesh. However, each sensor does not know the locations of any other sensors. At the time step t, the collectively environmental measurements of the i^{th} mobile wireless sensor are given by

$$y_{i,1:t} = \left(\bigcup_{k=1}^{t} y(s_{k,i})\right) \bigcup \left(\bigcup_{k=1}^{t} \left(\bigcup_{g=1}^{N_{ik}} y(s_{k,g})\right)\right), \tag{6.12}$$

where N_{ik} is the number of the i^{th} sensor's neighbours at the time k. For the sake of simplicity, denote $y_i = y_{i,1:t}$. The i^{th} mobile sensor can also know the sampling locations corresponding to $y_{i,1:t}$ as follows

$$\mathbf{s}_i = \left(\bigcup_{k=1}^{t} s_{k,i}\right) \bigcup \left(\bigcup_{k=1}^{t} \left(\bigcup_{g=1}^{N_{ik}} s_{k,g}\right)\right). \tag{6.13}$$

It is supposed that by exchanging information with the neighbours, every mobile agent can only predict the field at an unobserved vertex v_j that is inside the sphere of radius $\frac{R}{2}$ centered at the sensor. Hence, the global posterior prediction at any unmeasured locations can be obtained as the following proposition.

Proposition 6.2. Given the measurements y_i collected by the mobile robotic wireless sensor i from time 1 to t and the global parameters of the spatial field model, that is β_t, θ_t, every mobile agent i can approximately compute the global posterior prediction at v_j, where $\| v_i - v_j \| \leq \frac{R}{2}$, as follows.

(1) Mean:

$$m_{v_j|s_{1:t}} = X(v_j)\beta_t + \frac{1}{N_j}\sum_{i=1}^{N_j} \varphi_i[j], \tag{6.14}$$

where

$$\varphi_i = Q^{-1}_{z_{t,i}|y_{1:t}} A^T_{t,i} (\sigma^2_\varepsilon I_i)^{-1} (y_i - X(\mathbf{s}_i)\beta_t),$$

and

- N_j is the number of sensors whose distances to v_j are less than or equal to $\frac{R}{2}$;
- $A_{t,i}$ is the projector matrix that projects the random field modelled at the triangulation vertices to $\mathbf{s}_i$;
- I_i is a $|\mathbf{s}_i| \times |\mathbf{s}_i|$ identity matrix;
- $Q_{z_{t,i}|y_{1:t}} = Q_t + A^T_{t,i}(\sigma^2_\varepsilon I_i)^{-1} A_{t,i}$.

(2) Variance:

$$\sigma^2_{v_j|s_{1:t}} = \frac{1}{N_j} \sum_{i=1}^{N_j} Q^{-1}_{z_t|y_i}[j,j]. \tag{6.15}$$

Proof: In the centralized schemes, it can be seen that $\phi = A^T_t(\sigma^2_\varepsilon I)^{-1}(y_{1:t} - X(s_{1:t})\beta_t)$ is a n column vector with non-zero elements corresponding to sensor locations on the mesh. Production of the j^{th} row of $Q^{-1}_{z_t|y_{1:t}}$ and ϕ is the prediction mean of the random field at the vertex v_j on the triangulation.

In the distributed schemes, due to lack of information of all sensor locations, the mobile agent i cannot compute the projector matrix A_t. This leads to (6.10) not being calculated in a distributed way. However, look into (6.10), suppose that if a sensor location v_k and v_j are distant, then covariance $Q^{-1}_{z_t|y_{1:t}}[j,k]$ between the random variables at v_k and v_j is trivial. In other words, the predicted field at v_j would not change much if the robotic sensor at v_k does not participate in this prediction computation. Therefore, in this consideration, it is proposed that each mobile agent i employs its own measurements and its neighbour sensors' readings

y_i but uses the global parameters (β_t, θ_t) to predict the random field φ_i at the triangulation vertices. It can seen that φ_i is a n column vector but only the element $\varphi_i[j]$ that is estimated by the mobile sensor i is used to calculate the global random field at v_j.

Notice that the assumption $\parallel v_i - v_j \parallel \leq \frac{R}{2}$ is required for making communications among robotic sensors that can estimate the field at v_j, which holds that $\frac{1}{N_j}\sum_{i=1}^{N_j}\varphi_i[j]$ and the variance can be obtained by the group of these agents as average consensuses. ■

Remark 6.3. Each mobile agent in the v_j-estimated group predicts the spatial field at v_j in a distributed way; yet its own predicted mean and variance are exchanged with its neighbouring sensors in the group to compute the global prediction mean and variance at this predicted location. The distributed algorithm for a robotic sensor is summarised in Algorithm 5. ■

Algorithm 5 Distributed prediction algorithm for each mobile agent i at time step t

Input: $\bigcup_{k=1}^{t} y(s_{k,i})$, $\bigcup_{k=1}^{t}\left(\bigcup_{g=1}^{N_{ik}} y(s_{k,g})\right)$, $\bigcup_{k=1}^{t} s_{k,i}$, $\bigcup_{k=1}^{t}\left(\bigcup_{g=1}^{N_{ik}} s_{k,g}\right)$, β_t, θ_t, v_j

Output: Predicted field at unobserved location v_j, including mean $m_{v_j|s_{1:t}}$ and variance $\sigma^2_{v_j|s_{1:t}}$

At time $t \in \mathbb{Z}_{>0}$, do

1: Compute Q_t based on θ_t
2: Compute $A_{t,i}$
3: Compute $Q_{z_{t,i}|y_{1:t}} = Q_t + A_{t,i}^T(\sigma_\varepsilon^2 I_i)^{-1}A_{t,i}$
4: Compute $\varphi_i = Q^{-1}_{z_{t,i}|y_{1:t}} A_{t,i}^T(\sigma_\varepsilon^2 I_i)^{-1}(y_i - X(\mathbf{s}_i)\beta_t)$
5: Compute average random value $\frac{1}{N_j}\sum_{i=1}^{N_j}\varphi_i[j]$ via DAC approach, where j is the index of the location v_j on triangulation mesh
6: Compute mean $m_{v_j|s_{1:t}} = X(v_j)\beta_t + \frac{1}{N_j}\sum_{i=1}^{N_j}\varphi_i[j]$
7: Compute variance $\sigma^2_{v_j|s_{1:t}} = \frac{1}{N_j}\sum_{i=1}^{N_j} Q^{-1}_{z_{t,i}|y_{1:t}}[j,j]$ via DAC approach

6.3.4 Parameters estimation

The discussion followed here recovers the technique of learning parameters for the Gaussian field model, which is described in [5].

6.3.4.1 Centralized learning

From (6.7), it is straightforward to write down the marginal distribution of $y_{1:t}$ as

$$y_{1:t}|\theta_t, \beta_t, \sigma_\varepsilon^2, A_t \sim \mathcal{N}\left(X(s_{1:t})\beta_t, \sigma_\varepsilon^2 I + A_t Q_t^{-1} A_t^T\right). \tag{6.16}$$

Let $\Sigma_t = \sigma_\varepsilon^2 I + A_t Q_t^{-1} A_t^T$, the log-likelihood function is

$$f(\theta_t, \beta_t) = -\frac{1}{2}\{(y_{1:t} - X(s_{1:t})\beta_t)^T \Sigma_t^{-1}(y_{1:t} - X(s_{1:t})\beta_t) + \log\det(\Sigma_t) + tN\log(2\pi)\}. \tag{6.17}$$

The recursive algorithm to estimate the model parameters θ_t, β_t is referred to in Section 5.3.1.

Remark 6.4. The complexity of the evaluation of the log-likelihood function is considerably dependent on inverting the covariance matrix Σ_t. The complexity of inverting this matrix is $\mathcal{O}(n^3)$. Since n is vary large, it is worth considering how to efficiently compute the logarithm of the determinant of the covariance matrix Σ_t and its inversion, that is the precision matrix Σ_t^{-1}.

By the matrix determinant lemma, the logarithm of the determinant of Σ_t can be given by

$$\log\det(\Sigma_t) = \log\det(Q_t + \frac{1}{\sigma_\varepsilon^2} A_t^T A_t) - \log\det(Q_t) + 2tN\log(\sigma_\varepsilon).$$

Similarly, the precision matrix Σ_t^{-1} can be obtained by employing the Woodbury maxtrix identity as follows

$$\Sigma_t^{-1} = \frac{1}{\sigma_\varepsilon^2}\left(I - A_t(Q_t + A_t^T A_t \frac{1}{\sigma_\varepsilon^2})^{-1} A_t^T\right).$$

As a result, these formulae ameliorate the calculations very considerably in the MRWSNs with limited resources. ■

6.3.4.2 Distributed learning

Consider a distributed MRWSN where it is assumed that a graph with vertices of mobile sensor locations is connected over time. In a distributed way, every robotic sensor i can itself learn its own parameters and hyperparameters by utilizing its collective measurements y_i. Each mobile agent can employ the recursive algorithm introduced in the previous section to estimate its own model parameters. Nevertheless, in order to predict the physical field at any unobserved locations, the GMRF regression technique, as described in (6.14) and (6.15), requires the global parameters that can be obtained by the following proposition.

Proposition 6.5. If a MRWSN is connected, each mobile sensor i can calculate the global parameters and hyperparameters by exchanging local information with its neighbours as given

$$\xi_i[k+1] = \xi_i[k] + \gamma \sum_{j \neq i} a_{ij}(\xi_j[k] - \xi_i[k]), \tag{6.18}$$

where $\xi_i = (\theta_{t,i}, \beta_{t,i})^T$, and $\beta_{t,i}$ and $\theta_{t,i}$ are a mean parameter vector and a hyperparameter vector estimated at the time instant t by the mobile sensor i.

Proof: Since, it can be clearly seen that (6.18) is similar to (5.37), the results are directly derived by the DAC approach that has been described in Section 5.4.1.2. ■

6.4 Adaptive Sampling Approach

The objective in the present study is to estimate and predict efficiently and effectively the spatial field by using a limited number of the MRWSs. Therefore, there is a need to design optimal adaptive sampling paths that enable to drive the MRWSs to the most informative locations to observe the physical environment. It is shown that each MRWS always looks for a next location at time instant $t+1$ given all measurements up to time t. Note that the robotic sensors collect measurements of the field at the vertices of the triangulation. In the following, the one-step-ahead spatial prediction technique before presenting the adaptive sampling approach is considered.

6.4.1 One-step-ahead spatial inference

It can be clearly seen that looking for the most information locations s_{t+1} at time instant $t+1$ without y_{t+1} is intractable. Hence, in order to address this problem, the one-step-ahead forecast is utilized to predict the latent spatial values at time $t+1$, given the collective observations up to time t. The readers are referred to [163] for more detailed discussion.

The spatial field model at time t can be written as

$$y_{1:t} = X(s_{1:t})\beta_t + A_t z_t + \varepsilon(s_{1:t}), \tag{6.19}$$

where $\varepsilon(s_{1:t}) \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 I)$.

Given $z_t|y_{1:t}$ (6.9), the latent random process at time step $t+1$ given observations $y_{1:t}$ can be forecasted by

$$\widehat{z}_{t+1}|y_{1:t} = z_t|y_{1:t} + \eta_{t+1}, \tag{6.20}$$

where $\eta_{t+1} \sim \mathcal{N}(\mathbf{0}, \sigma_{sys}^2 I_\eta)$, I_η is a $n \times n$ identity matrix, and σ_{sys}^2 is a system error that describes the changes in the elements of the parameters vector between times t and $t+1$.

Let $Q_{\eta,t+1}^{-1} = \sigma_{sys}^2 I_\eta$ and suppose that σ_{sys}^2 is to be known, the prior conditional distribution of $\widehat{z}_{t+1}$ given $y_{1:t}$ can be obtained by

$$\widehat{z}_{t+1}|y_{1:t} \sim \mathcal{N}(\mu_{z_t|y_{1:t}}, Q_{z_t|y_{1:t}}^{-1} + Q_{\eta,t+1}^{-1}). \tag{6.21}$$

Therefore, the measurements forecasted at time $t+1$ is

$$\widehat{y}_{t+1}|y_{1:t} = X(s_{t+1})\beta_t + A_t(\widehat{z}_{t+1}|y_{1:t}) + \varepsilon(s_{t+1}), \tag{6.22}$$

where $\varepsilon(s_{t+1}) \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 I_{t+1})$ and I_{t+1} is a $N \times N$ identity matrix.

Moreover, if VT is a set of the vertices on the triangulation, let $b_{t+1} \in \mathbb{R}^n$ define a column vector at time $t+1$ as

$$b_{t+1}[i] = \begin{cases} 1, & \text{if } VT(i) = s_{t+1,j} \text{ and } j = 1, ..., N \\ 0, & \text{otherwise.} \end{cases}$$

Then, the posterior precision matrix of the posterior distribution of $\widehat{z}_{t+1}$ conditioned on $y_{1:t}$ and $\widehat{y}_{t+1}$ is specified by

$$\widehat{Q}_{z_{t+1}|y_{1:t},\widehat{y}_{t+1}} = (Q^{-1}_{z_t|y_{1:t}} + Q^{-1}_{\eta,t+1})^{-1} + \frac{1}{\sigma^2_\varepsilon} b_{t+1} b^T_{t+1}. \tag{6.23}$$

For the sake of simplicity, from here it will be used $\widehat{Q}_{z_{t+1}} := \widehat{Q}_{z_{t+1}|y_{1:t},\widehat{y}_{t+1}}$, $Q_{z_t} := Q_{z_t|y_{1:t}}$, and $Q_\eta := Q_{\eta,t+1}$.

Note that these precision matrices are very sparse ones that will be utilized for fast computation in the following sections.

6.4.2 Centralized sampling strategy

In recent related works, the approaches mostly used to find the most informative sampling locations at the next time step $t+1$ are derived from information-theoretic criteria such as entropy [133]. These works proposed to optimize the conditional entropy on the GP models [97], [107], or on the GMRF models restricted to a regular lattice [119–122]. In this subsection, the conditional entropy based adaptive sampling method is extended on the GMRF models that are described on an irregular lattice. Moreover, the complexity of these existing methods is demonstrated and a novel approach is proposed, which is computationally practical for the resource-constrained MRWSNs.

The crucial idea of the proposed technique is to drive the MRWSs to new sampling positions in such a way that the overall uncertainty at the n spatial sites is minimized. In terms of the GP technique, let $\widehat{\Sigma}_{z_{t+1}}$ define a very dense covariance matrix of $\widehat{z}_{t+1}|y_{1:t}, \widehat{y}_{t+1}$. $\widehat{\Sigma}_{z_{t+1}}$ can be obtained by $\widehat{\Sigma}_{z_{t+1}} = \widehat{Q}^{-1}_{z_{t+1}}$. The conditional

entropy of $\widehat{z}_{t+1}|y_{1:t}, \widehat{y}_{t+1}$ is computed in terms of the closed form expression (6.23) as

$$H(\widehat{z}_{t+1}|y_{1:t}, \widehat{y}_{t+1}) = \frac{1}{2}\{\log\det(\widehat{\Sigma}_{z_{t+1}}) + n\log(2\pi) + n\}.$$

In a simplified form, it is denoted

$$H = \log\det(\widehat{\Sigma}_{z_{t+1}}) \tag{6.24}$$

as a conditional entropy function. As a result, the next optimal sampling locations at time instant $t+1$, s_{t+1}, can be obtained by minimizing the uncertainty as follows

$$s_{t+1}^{opt} = \underset{s_{k+1,i} \in \Omega_{VT}^{(t)}(i)}{\textbf{argmin}} \log\det(\widehat{\Sigma}_{z_{t+1}}), \tag{6.25}$$

where $\Omega_{VT}^{(t)}(i)$ is a set of the spatial sites inside VT but restricted by $\Omega_i^{(t)}$ that is the allowable movement region of the MRWS i at time t, explained in Section 2.6.2. This optimization problem can be rewritten in terms of the precision matrix. One has

$$s_{t+1}^{opt} = \underset{s_{k+1,i} \in \Omega_{VT}^{(t)}(i)}{\textbf{argmax}} \log\det(\widehat{Q}_{z_{t+1}}). \tag{6.26}$$

It can be clearly seen that the above combinatorial optimization problems are NP-hard [53]. Usually, these issues can be addressed by employing a greedy algorithm that sequentially finds a set of the next near-optimal sampling locations at time $t+1$, s_{t+1}^{*}. Nonetheless, the costs of these algorithms need to be considered in the resource-constrained MRWSNs. The computational complexity of the algorithms can be described as below.

Lemma 6.6. Given a GP model, the greedy algorithm can resolve the optimization problem (6.25) in time $\mathcal{O}(Nn^4)$.

Proof: Since the size of $\widehat{\Sigma}_{z_{t+1}}$ is $n \times n$, exact computations of $\log\det(\widehat{\Sigma}_{z_{t+1}})$ require $\mathcal{O}(n^3)$ operations [152]. In addition, VT has n elements, so there are $\mathcal{O}(n)$ such logarithms of the determinants to be calculated.

Moreover, in the adaptive sampling strategy, it needs to find N next MRWS locations at every time step. Thus, the greedy algorithm has to run N iterations in time $\mathcal{O}(N)$. So, the optimization problem (6.25) can be finalized in running time $\mathcal{O}(Nn^4)$. ■

Remark 6.7. In general, n is very large; despite being under a greedy approach, the problem (6.25) is still intractable in the resource-constrained MRWSNs. ■

Lemma 6.8. Given a GMRF model, the greedy algorithm can address the optimization problem (6.26) in time $\mathcal{O}(Nn^{\frac{5}{2}})$.

Proof: Note that the size of $\widehat{Q}_{z_{t+1}}$ is also $n \times n$. As discussed in [164],

$$\log\det(\widehat{Q}_{z_{t+1}}) = 2\sum_{i=1}^{n}\log(L_{ii}),$$

where $\widehat{Q}_{z_{t+1}} = LL^T$. This Cholesky decomposition is calculated in time $\mathcal{O}(n^{\frac{3}{2}})$ in two dimensional space.

The rest of this proof is similar and refers to the proof of the **Lemma 6.6**. Therefore, the computational complexity of (6.26) is $\mathcal{O}(Nn^{\frac{5}{2}})$. ■

Remark 6.9. The GMRF models enable the greedy algorithm to very considerably reduce the computational time of finding the near-optimal sampling locations for

the MRWSNs. Nevertheless, this approach is clearly impractical. In the following, an efficient novel optimality criterion for the adaptive sampling problem in the MRWSNs is proposed. ■

Proposition 6.10. Given Q_{z_t} computed at time t by (6.11) and σ^2_{sys}, the next sampling positions for the MRWSs at the time step $t+1$ is the optimal solution of the following problem.

$$s^{opt}_{t+1} = \underset{s_{t+1,i} \in \Omega^{(t)}_{VT}(i)}{\textbf{argmax}} \quad b^T_{t+1} \left(Q^{-1}_{z_t} + Q^{-1}_{\eta}\right) b_{t+1}. \tag{6.27}$$

Proof: Let $\Psi = \left(Q^{-1}_{z_t} + Q^{-1}_{\eta}\right)^{-1}$. From (6.26) and (6.23), and by utilizing the matrix determinant lemma, one has

$$\begin{aligned} H &= -\log\det\{\Psi + \frac{1}{\sigma^2_{\varepsilon}} b_{t+1} b^T_{t+1}\} \\ H &= -\log\left\{\left(1 + \frac{1}{\sigma^2_{\varepsilon}} b^T_{t+1}\Psi^{-1} b_{t+1}\right)\det(\Psi)\right\} \\ H &= -\log\left(1 + \frac{1}{\sigma^2_{\varepsilon}} b^T_{t+1}\Psi^{-1} b_{t+1}\right) - \log\det(\Psi) \end{aligned}$$

It is stated that Ψ is constant at each time instant, Ψ^{-1} is symmetric positive definite, and the logarithm function is monotonic. For these reasons, it can be in turn defined

$$H = b^T_{t+1}\Psi^{-1} b_{t+1}$$

to complete the proof. ■

The optimization problem in the **Proposition 6.10** is still combinatorial and NP-hard. Nonetheless, the greedy algorithm can cope with this maximization issue in a reasonable computational time as follows.

Theorem 6.11. *Given a GMRF model, the near-optimal solution of (6.27) can be obtained in time* $\mathcal{O}(Nn)$.

Proof: Due to the sparse structure of Q_{z_t}, $Q_{z_t}^{-1}$ can be computed in a very short time at time t and constant at time $t+1$. Therefore, the computation of $b_{t+1}^T \left(Q_{z_t}^{-1} + Q_\eta^{-1}\right) b_{t+1}$ requires $\mathcal{O}(1)$ operations. As shown in the **Lemma 6.6**, the total time to resolve the optimization problem (6.27) is $\mathcal{O}(Nn)$. ■

Remark 6.12. It is apparent that the proposed approach can be efficiently implemented in the resource-limited MRWSNs. It is also to be noted that the computational complexity remains approximately constant even with an increasing number of measurements. ■

6.4.3 Distributed sampling strategy

The distributed algorithms have been designed in the previous section to predict the spatial field at time t. In this subsection, an efficient navigation strategy is developed to implement the equation (6.27) in a distributed fashion. Under the assumption of the allowable measurement region in Section 2.6.2, each mobile agent i can find the most informative location inside $\Omega_{VT}^{(t)}(i)$ as below.

Proposition 6.13. Given Q_t obtained at time t, the next sampling location for the robotic sensor i at the time step $t+1$ is $v_k \in \Omega_{VT}^{(t)}(i)$ that satisfies

$$s_{t+1,i}^{opt} = \underset{v_k \in \Omega_{VT}^{(t)}(i)}{\textbf{argmin}} \quad Q_t[k,k]. \tag{6.28}$$

Proof: At the mobile sensor i, a column vector $b_{t+1,i} \in \mathbb{R}^n$ is denoted as

$$b_{t+1,i}[j] = \begin{cases} 1, & \text{if } VT(j) = s_{t+1,i} \\ 0, & \text{otherwise.} \end{cases}$$

It can be seen that $b_{t+1,i}$ is a vector with one non-zero element. Hence, from (6.27), σ_ε^2 will be excluded from the result as the noise corresponds to measurements. The constant-diagonal matrix Q_η^{-1} can also be eliminated. Eventually, the result of the right hand side of (6.27) obtained by the sensor i is $Q_t^{-1}[k,k]$ for computing iteration of v_k. ■

Obviously, the computational complexity of (6.28) is $\mathcal{O}(N_i)$, where N_i is the number of the vertices inside $\Omega_{VT}^{(t)}(i)$. This computational time is very encouraging for each resource-constrained robotic sensor to find an optimal sampling path.

6.5 Experimental Results

In this section, in order to illustrate the efficiency of the proposed approaches, experiments were conducted using published data sets described as follows. The aim was to compute the predicted field and the prediction error variances in the whole environment as well as the root mean square errors. More particularly, the computational time for finding the efficient sampling paths for MRWSs was also calculated. Simulated results are demonstrated in two following subsections: *centralized scheme* and *distributed scheme*. Usually, the centralized approach is utilized in a small network, where all sensors can easily transmit all measurements to the sink. Hence, the centralized algorithm was carried out on the numerically generated data set [119], where a small network of 5 mobile sensors was utilized to observe the physical field.

In contrast, the distributed method is often used in a large-scale network. In this experiment, the distributed algorithm was implemented in a network of 30 mobile sensing agents that were deployed in a simulated agricultural field. Data set for this experiment was derived from a real-world application [158] and regenerated by [159]. Note that all experiments were implemented in two dimensional environments.

6.5.1 Centralized navigation scheme

In this discussion, one considers a situation where, at time step t, all mobile sensors make new observations and transmit them to the sink via a specific routing tree. Then the base station computes the centralized sampling strategy and sends control commands back to each robotic sensor. All prediction activities are conducted by the sink.

A spatial field of the physical quantity was generated within 100 units×50 units as shown in Fig. 6.2. The range of the field varies from 15.00 to 28.30. There were five ($N = 5$) MRWSs used with constrained individual displacements of a maximum radius of 5 units in every time step. The desired buffer width was set

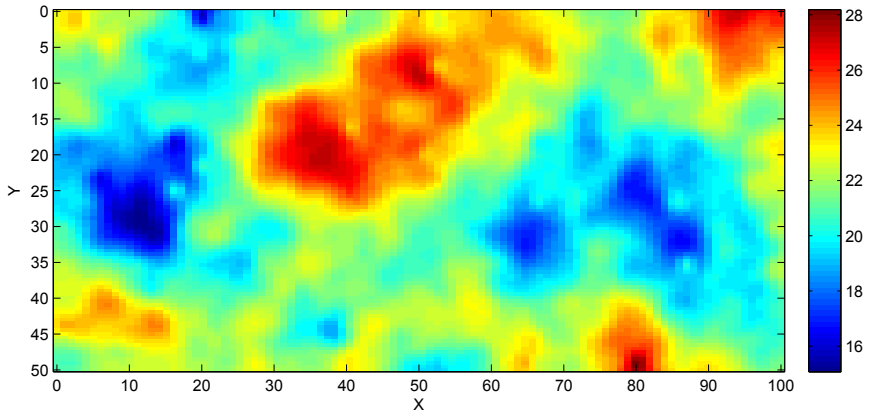
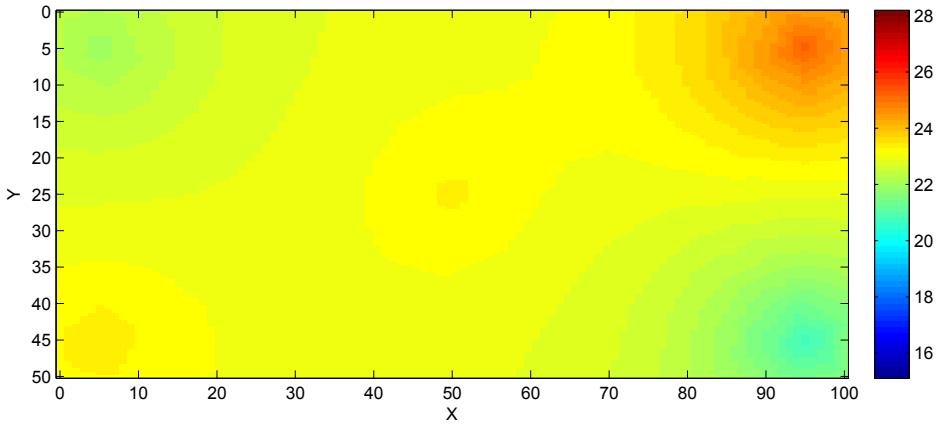
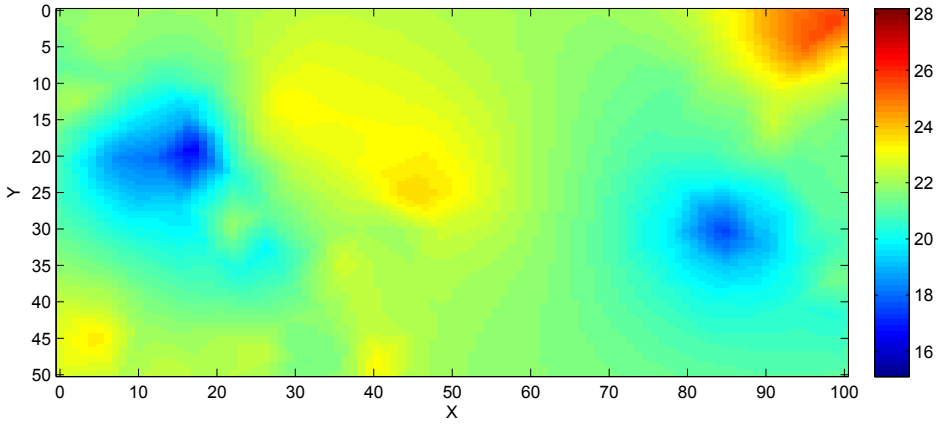


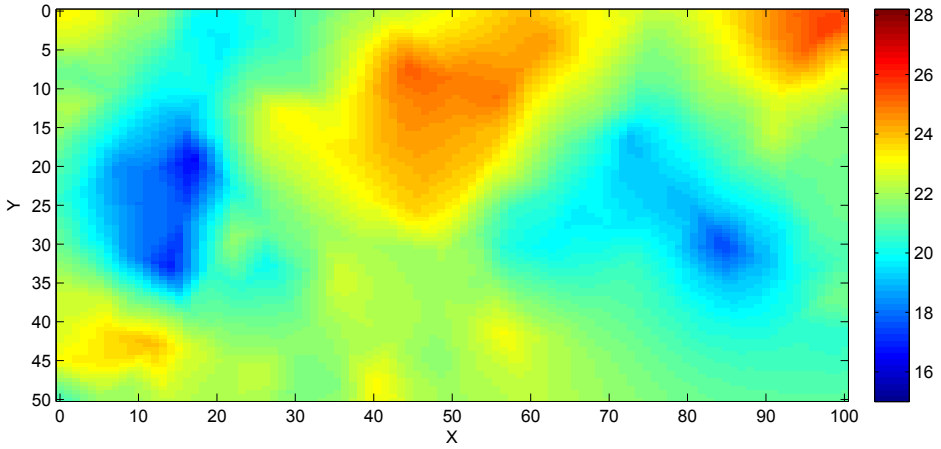
Figure 6.2: The true field of the numerically generated data set. Range of the fields is shown in color bar.



(a)



(b)



(c)

Figure 6.3: The results for numerically generated field in Case 1: The predicted fields at time (a) $t = 1$, (b) $t = 10$, and (c) $t = 20$. Range of the fields is illustrated in color bars.

to $\omega = 0.3$ units. In this evaluation, three cases are considered in three different starting conditions as illustrated in Case 1 Fig. 6.4a, Case 2 Fig. 6.5a and Case 3 Fig. 6.5b. In Case 1, for comparison purposes, all MRWSs were started from the pre-defined locations as shown in Fig. 6.4a (in white dots), similar to the work in [119]. It was assumed the knowledge of the measurement noise $\sigma_{\varepsilon}^2 = 0.2$ and the system error $\sigma_{sys}^2 = 0.25$. The spatial field was represented by approximately 5000 vertices of triangles which were considered as spatially interested sites. In this case, the percentage of non-zero elements in the sparse precision matrix is around 18%, whereas the covariance matrix is comprehensively dense.

Figures 6.3a, 6.3b, and 6.3c demonstrate the predicted fields at times $t = 1, 10$ and 20. It can be seen that the predicted means are approaching the true field as illustrated in Fig. 6.2 when the number of the observations increases. White circles in Fig. 6.4a, 6.4b, and 6.4c show the trajectories of the near-optimal paths of the MRWSs up to time instants 1, 10, and 20, respectively. Fig. 6.4 also shows the gradual reduction in the prediction error variances with the growth of time steps, corresponding to 1^{st}, 10^{th}, and 20^{th} instant respectively. The prediction errors are higher in the areas distant from the measured points. Nonetheless, the predicted variances in the whole space are trivial when 5 MRWSs collect 20 observations each.

Fig. 6.6 shows the root mean square errors (RMSEs) of the predicted field at n spatial sites, which are based on,

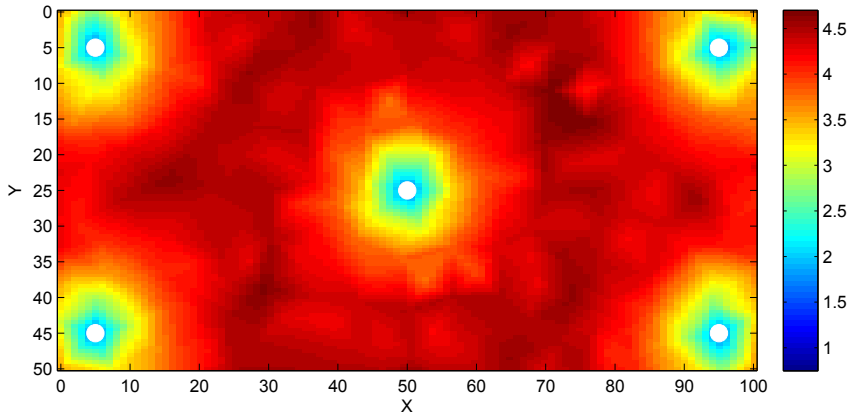
$$\text{RMSE}(t) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mu_{z_t|y_{1:t}}[i] - z_t[i]\right)^2},$$

where z_t is a vector of the values actually observed.

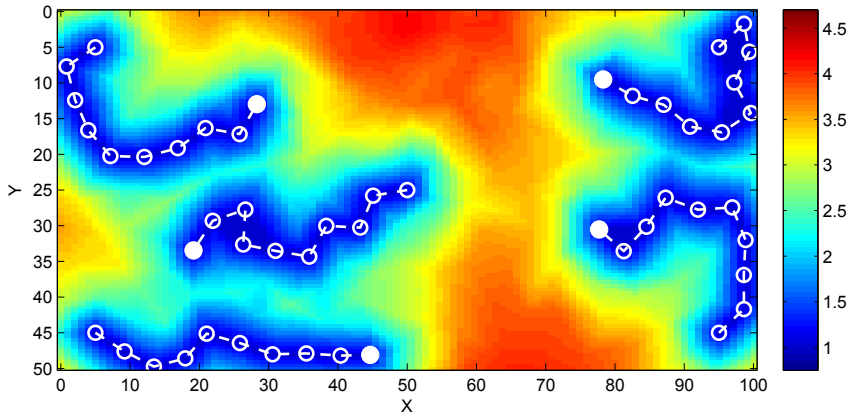
As expected, Fig. 6.6 shows a gradual reduction of RMSEs with increased number of observations or with more exploration.

For comparison purposes, a generic GP regression [13] for estimating and predicting the spatial phenomenon was implemented in Case 1. Note that since the standard GP continuously interpret the spatial field, its prediction results should illustrate the best performance in all scenarios studied. However, computing the dense covariance matrix is a main obstacle to addressing the spatial prediction issue, which has led this study to propose the GMRF in this chapter. Under starting conditions in Case 1, as can be seen in Fig. 6.6, results of the proposed approach are highly comparable to those of the standard GP based method. Furthermore, the implementation of the approach proposed by Xu *et al.* [119] was also conducted. It can be clearly seen that the proposed GMRF outperforms [119]. The rationale could be in the fact that [119] chooses the hyperparameters prior to measurements whereas the proposed GMRF updates these parameters online. In order to highlight the consistency of resulting performance, the proposed approach was then evaluated in Case 2 and Case 3, respectively. Results corresponding to the prediction quality as demonstrated in Fig. 6.6 in Cases 1, 2 and 3 show that the proposed algorithm always generates best solutions for all the cases studied in this work.

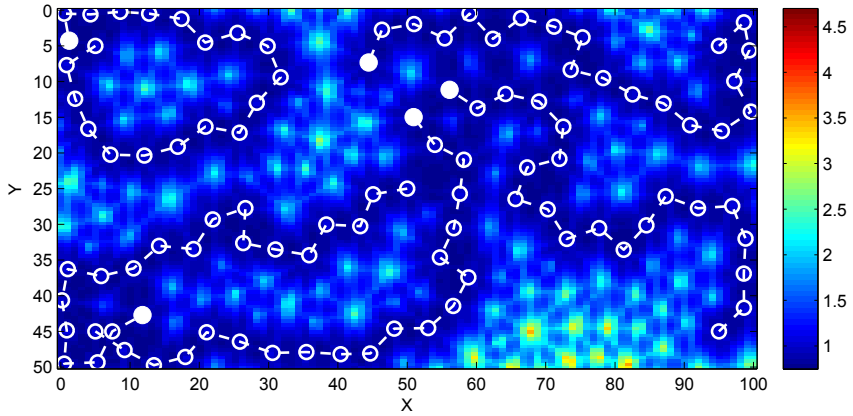
Another important aspect to compare is the computational complexity in finding the sampling paths for the MRWSs. In section 6.4.2, three strategies to optimize the informative navigations in the MRWSNs have been formulated. They all aim to minimize the uncertainty at unobserved locations of interest. The first strategy uses a GP model to describe the spatial field and then to find sensing locations by minimizing the logarithm of the determinant of the covariance matrix. The strategy is specified by (6.25). The second strategy represents the field by the GMRF and



(a)

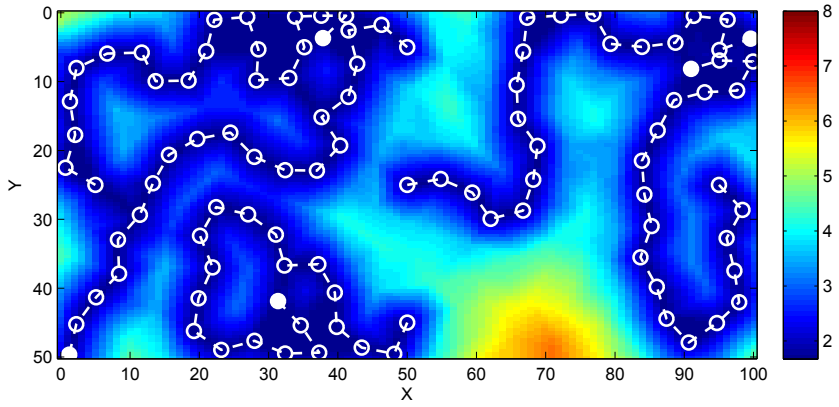


(b)

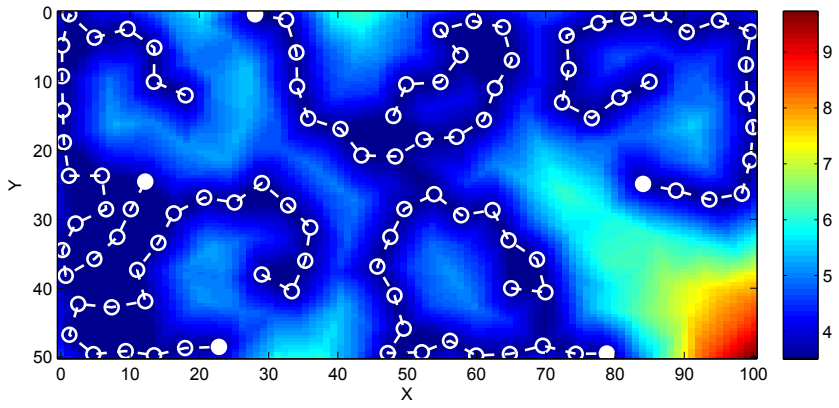


(c)

Figure 6.4: The results for numerically generated field in Case 1: The predicted variances in the field at time (a) $t = 1$, (b) $t = 10$, and (c) $t = 20$. Range of the errors is demonstrated in color bars. The sampling paths for MRWSs up to time steps 1, 10 and 20, respectively, are illustrated by white circles, where current mobile sensor locations are shown in white dots.



(a)



(b)

Figure 6.5: The sampling paths for MRWSs up to time step 20 are illustrated by white circles, where current mobile sensor locations are shown in white dots. (a) Case 2 and (b) Case 3.

solves the sampling problem by maximizing the logarithm of the determinant of the precision matrix. The issue is now shown in (6.26). In the third scheme, the GMRF is utilized to model the spatial field, but a new criterion for the sampling path issue has been proposed, which removes the need for computing the logarithm of the determinant. The proposed criterion is described by (6.27). These strategies are all resolved by the greedy algorithm. It was noted that the run time of the algorithms did not ascend with the rise of the number of observations. More importantly, in this

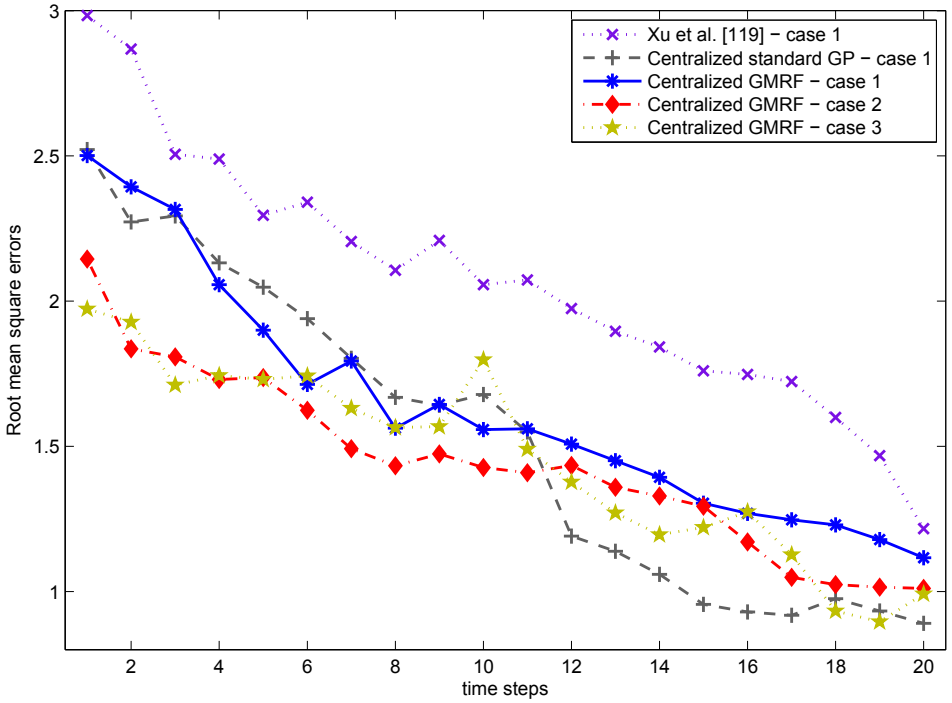
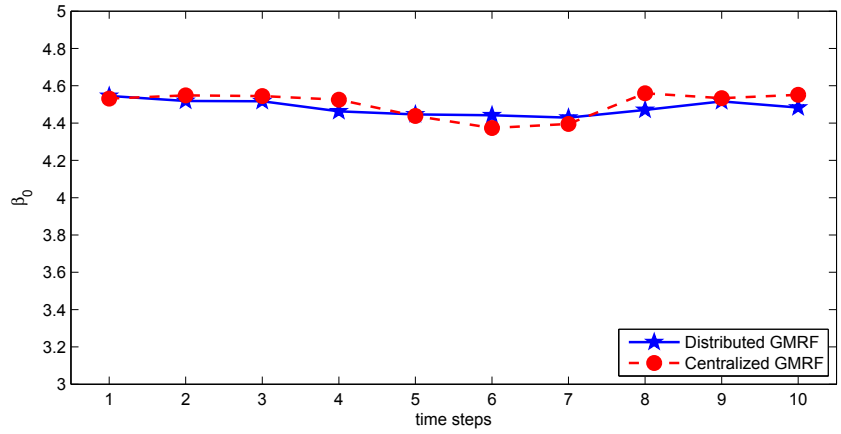


Figure 6.6: The root mean square errors for numerically generated field at different starting conditions.

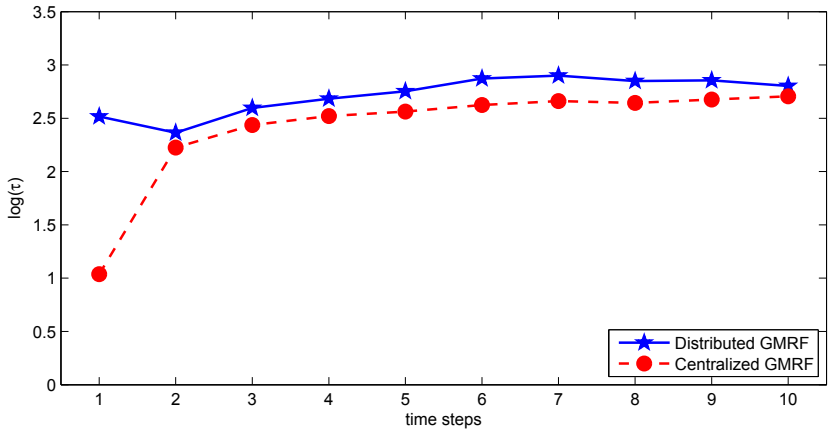
particular illustrative experiment, in each time instant, the algorithm took approximately 9 seconds to address the problem (6.27), whereas solutions of the problems (6.25) and (6.26) required approximately 7 hours and 4 hours run time respectively implemented on R V3.0 with a PC of 3.1GHz Intel Core i5-2400 Processor.

6.5.2 Distributed navigation scheme

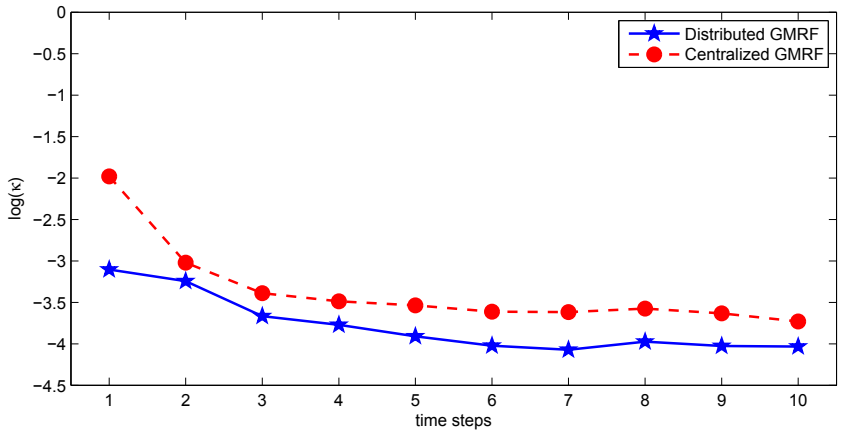
Consider a MRWSN in a distributed fashion in which at time instant t every individual mobile agent can transmit data to and receive information from its neighbouring sensors. There is no central station in this case; each robotic sensor itself computes prediction of the field of interest and finds the optimal sampling path by employing its own and its neighbours' measurements as well as exchanging information with



(a)



(b)



(c)

Figure 6.7: The results for agricultural field: The model parameters of GMRF are estimated by centralized and distributed approaches proposed. These parameters are (a) mean parameter β_0, (b) $\log(\tau)$, and (c) $\log(\kappa)$.

the neighbours. The data set in this consideration is derived from a real application [158]. It has been known that nitrogen fertilizer applications are essential for production of crops in the agricultural field. Since nitrogen levels, supplied by mineralization of soil organic matter, can vary considerably across the field, soil sampling is crucial for the accurate description of spatial patterns in soil fertility. It was proposed utilizing the MRWSNs to construct the percentage of the soil organic matter patterns that recommend a strategy of nitrogen fertilizing.

The percentage of the soil organic matter was regenerated in [159] from the data of 1375 soil measurements. The size of this field is 400m$\times$ 400m. The true levels of the percentage of the soil organic matter are illustrated in Fig. 6.8. In this experiment, at the beginning, $N = 30$ mobile robotic wireless sensors were randomly deployed in the field as demonstrated as white dots in Fig. 6.9b. The maximum distance that each mobile agent can travel between time steps was set to 15m. The communication range R was set to 100m. The minimum distance between any two robotic sensors at any time was ω =3m. It was assumed that σ_{ε}^2 was known and set to 0.01.

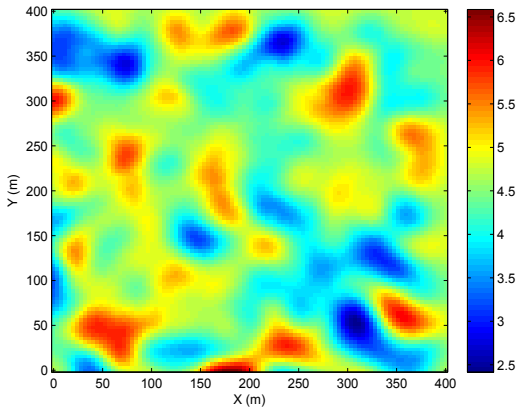


Figure 6.8: The true field of the agricultural data set. Percentage of the soil organic matter is shown in color bar.

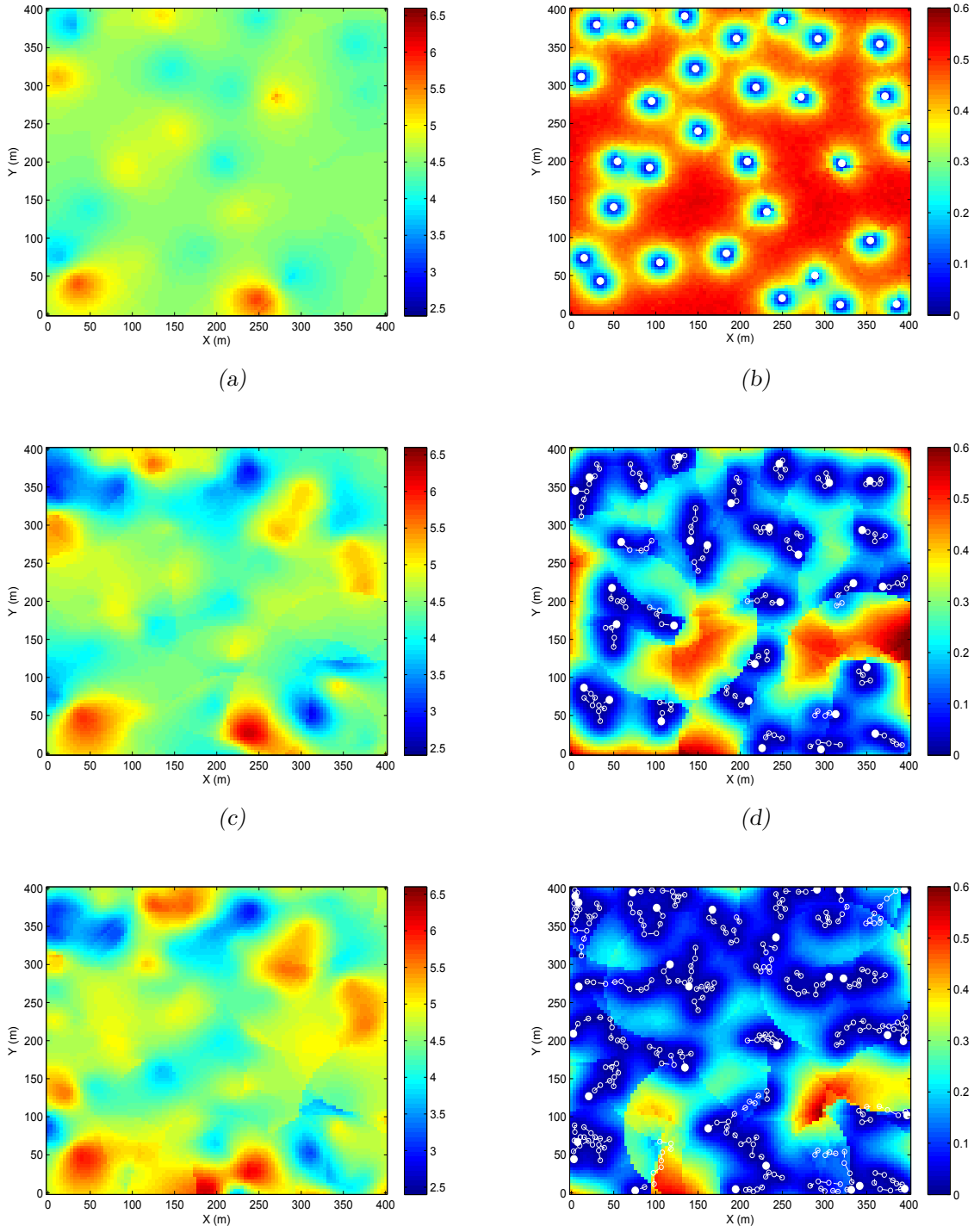


Figure 6.9: The results for agricultural field: (a), (c) and (e) are the predicted field at times $t = 1$, $t = 5$, and $t = 10$, respectively. The predicted variances in the field at times (b) $t = 1$, (d) $t = 5$, and (f) $t = 10$. The sampling paths for MRWSs up to time steps 1, 5 and 10, respectively, are shown by white circles, where current mobile sensor locations are demonstrated in white dots.

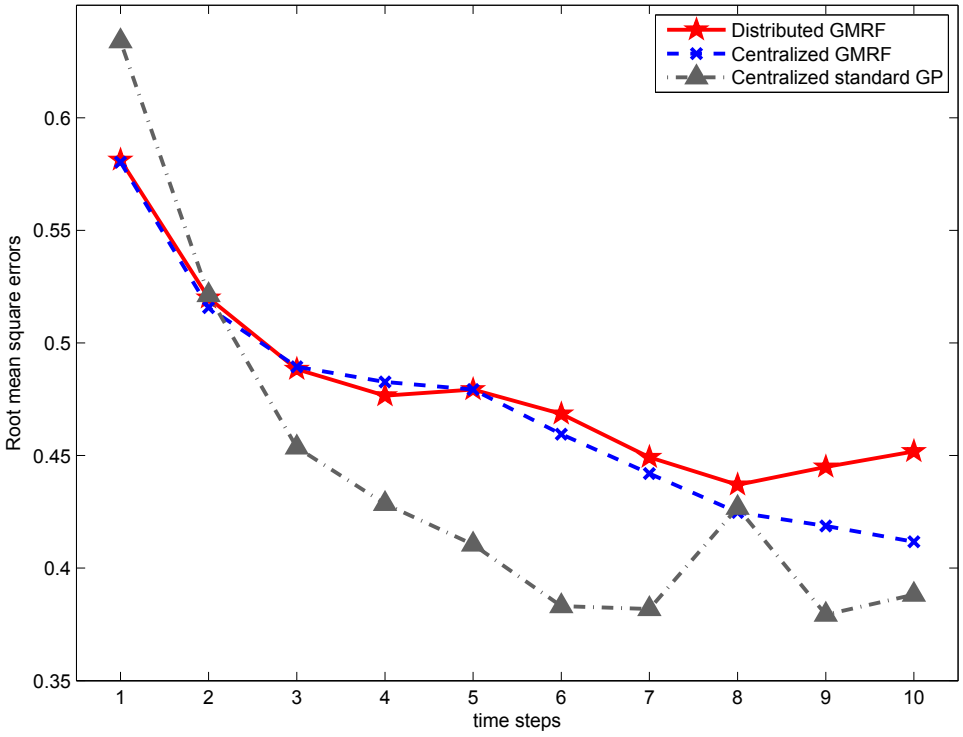


Figure 6.10: The root mean square errors for agricultural field.

In the proposed method, it is assumed that each mobile sensor can store all its own measurements from time step 1 to current time step t and transmit them to its neighbours when required. Therefore, at a particular time, every mobile sensing agent possesses not only its own measurements but also its neighbours' readings by only exchanging local information. Particularly, these collective measurements include the measurements gathered distantly from the sensor. This leads to the fact that the sensor can obtain measurements at locations even without navigating them. As a result, there are encouraging results demonstrated in Figures 6.7 and 6.9. For instance, in order to evaluate the distributed technique, the mean parameters and the hyperparameters of the GMRF model are estimated using both centralized and distributed methods and illustrated in Fig. 6.7. It can be clearly seen that the global model parameters obtained by the distributed method approach those obtained by

the centralized method. In the left column of Fig. 6.9, the predicted fields are illustrated at particular times $t = 1$, 5, and 10, respectively. The fields represent the percentage of the organic matter in the soil. When the MRWSs navigate throughout the environment, the predicted fields shown in Figures 6.9a, 6.9c, and 6.9e are gradually getting close to the true field shown in Fig. 6.8.

In order to find the new optimal sampling paths for the MRWSs, the centralized criterion demonstrated by (6.27) was solved in a distributed fashion. In other words, the centralized strategy (6.27) is replaced by the distributed scheme (6.28) so that each mobile sensor can itself find the next sensing location. Consequently, the trajectories of the sampling paths of the mobile robotic wireless sensors at time steps 1, 5 and 10 are demonstrated by white circles in Figures 6.9b, 6.9d, and 6.9f, where the current sensor locations are shown in white dots. Although each mobile sensing agent cannot communicate to other sensors outside of its communication range, the mobile sensors adjust themselves according to the proposed strategy so that their measured locations cover the whole space as time accumulates. Moreover, the prediction error variances in the whole environment are also described in the right column of Fig. 6.9, where Figures 6.9b, 6.9b, and 6.9f correspond to $1^{st}, 5^{th}$, and 10^{th} sampling instants. In accordance with the color bars, in the vicinity of the measured locations, the predicted variances are small compared with those at distant points. Particularly, it can be seen that the prediction errors in the whole space gradually reduce when more observations are collected.

In this experiment, for comparison purposes, the centralized standard GP and the centralized GMRF methods were also implemented in the reference field to compute the RMSEs. Notice that due to using all measurements to continuously model the physical field, the centralized GP algorithm should show the best performance

in all cases. The results of the RMSEs obtained by the centralized and distributed GMRF approaches and the centralized standard GP method are shown in Fig. 6.10. Generally speaking, the centralized GMRF approach outperforms the distributed GMRF method. In addition, the RMSEs obtained by the proposed approaches are not much different from those obtained by the centralized standard GP. More importantly, in regards to computational time of addressing the sampling path problem (6.28) to find the next optimal sampling location, each mobile robotic sensor can successfully solve (6.28) in approximately one second.

6.6 Summary

This chapter has proposed to utilize the MRWSNs to observe and predict the physical spatial field of interest using the GMRF model. A method based on the SPDE approach projects the approximate GMRF on a triangulation of an irregular lattice, which not only allows the parameters and hyperparameters to be updated online but also contributes to significant reduction in computational time owing to the sparse structure of the precision matrix. The optimal sampling strategy proposed in this chapter also forecasts future measurements using a one-step-ahead prediction. This combinatorial and NP-hard problem was effectively solved by a greedy algorithm, which led to run time computational complexity of $\mathcal{O}(Nn)$. A distributed algorithm was also proposed for the MRWSNs with a limited communication range to address the limitation of the centralized prediction and navigation schemes. The proposed algorithms were demonstrated with convincing results in terms of both accuracy of the spatial prediction and computational savings of finding near-optimal sampling paths.

Chapter 7

Conclusions and Future Research

In this chapter, the summary of the research and the crucial findings performed in this thesis are first summarized. Then, recommendations for promising future research carrying on from what has been done in this thesis are also delineated.

7.1 Thesis Conclusions

In this section, the key contributions in the field of efficiently and effectively sampling the spatial phenomena using the WSNs presented in this thesis are summarized as follows.

In the first part of Chapter 4, selecting the best subset of the environmental wireless sensors has been considered in an unconstrained environment. In order to deal with this issue, the selective criterion, ARMSE, has been proposed so that the average uncertainty at each unmeasured but predicted location is minimized. The NP-hard and combinatorial minimization has then been resolved by the simulated annealing based algorithm, where the replacement is conducted between a selected

sensor location and its neighbours at each iteration. Experiment results carried out in many indoor and outdoor data sets have indicated that the proposed algorithm outperforms the other conventional approaches. In the second part of this chapter, the selection of sensor observations has been conducted in a constrained sewage system in which the phase of the H_2S is monitored by the gaseous sensors and modelled by a modified GP. Due to the boundaries of the sewer pipes, correlations among H_2S contents cannot be computed as those of random variables in a free environment. Hence, an improved covariance function has been developed so that the covariances are ultimately calculated based upon the lengths along the sewage pipes. Additionally, an efficient algorithm based on a novel mutual information criterion has been introduced in order to choose the best H_2S sensor locations in the sewage system. Results obtained by the proposed approach on the H_2S data sets have been also verified by a very complex model developed by Sydney Water.

Chapter 5 has presented new approaches to cope with the sensor placement problem in the MRWSNs by using the GP. After introducing a centralized Gaussian predictive inference approach over time, the chapter has described new algorithms to find the most informative sampling trajectories for the MRWSs to take measurements at each time step. These algorithms have been developed based on an information-theoretic basis, where uncertainty at all unmeasured locations of interest aims to be minimized. First of all, a continuous density function of posterior variances has been proposed for a locational optimization based approach in order to resolve the optimal sensing agent locations in a continuous space. It has been then proven that the next sensing location is the centroid of the Voronoi partition of the current location. More interestingly, the environment has been considered in a discretized stage from which the sampling issue in the MRWSNs has been formulated into a combinatorial and NP-hard optimization by the use of conditional entropy.

The minimization has been proven to be intractable in a resource-constrained MRWSN, then a near-optimal but very efficient criterion for this problem is proposed. With this new proposition, the resource-constrained MRWSNs can address the entropy based sampling optimization in practically feasible time by a greedy algorithm. Particularly, solution of this maximization is guaranteed by bounds; the lower bound within $1-\frac{1}{e}$ of the optimal solution is also utilized to stop the MRWSs from sampling the environmental field to preserve the restricted energy. At the same time, a class of new methods has been developed to address the centrally-computed approaches in a distributed manner. In this scheme, an individual mobile sensor exchanges local information with its neighbours by the use of distributed computation tools of the JOR and the DAC. By using its own collective measurements and its neighbours' collective observations, each mobile agent first calculates its own model parameters and then computes global values via distributed techniques. Predicting procedure can also be obtained in the same way with the global parameters. Distributed navigation strategy has been additionally given to drive the MRWSs to the most uncertain points in the field. Testing performance on pre-published data sets has shown the proposed methods to be very encouraging.

In Chapter 6, in an effort to diminish the computational complexity of factoring the covariance matrix in the GP regression, the GMRF has been utilized to model the spatial field. Sparsity of the precision matrix, which is created by the fact that the GMRF represents the spatial phenomenon on an irregular lattice of triangulation, enables the GMRF spatial estimation and prediction approach to be computed quickly. A GMRF centralized adaptive navigation scheme for the MRWSNS has been formulated, yet this entropy based minimization is unfeasible to be solved by the resource-constrained MRWSNs in polynomial time. Therefore, a novel and efficient criterion for finding the optimal sampling paths for the mobile sensing agents

has then been developed, where the running time of a greedy algorithm used to solve this problem has been proven to be deterministic. Furthermore, a distributed approach for GMRF spatial inference and a distributed navigation strategy for each robotic sensor have been given in the case where wireless communication range of each sensor is short. Again testing results on pre-published and realistic data sets have proven the significance of the proposed approaches.

7.2 Future Research

This section describes the future research directions.

- The sensor selection and the sensor placement problems in the WSNs probably need to be considered under more network related constraints such as energy, communication range, routing structure.
- Moreover, the optimal sampling schemes for mobile wireless sensors will have to be dealt with under more robot related constraints such as nonholonomy, obstacles and localization error.
- In addition, the more general spatio-temporal field models will be utilized to expand the work in this thesis.
- In the context of the sensor selection, the SWSNs may be addressed in a distributed fashion, or considered under a condition in which the spatial field is modelled by the GMRF.
- Algorithmic implementation and large scale testing will be conducted.

Appendix A

Mathematical Background

This appendix presents mathematical results that are effectively utilized in linear algebra conversions.

A.1 Block-wise Inversion of Matrix

Matrices can be inverted block-wise by using the following analytic inversion formula:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}, \tag{A.1}$$

where A, B, C and D are matrix sub-blocks of arbitrary size. A and D must be square; and A and $D - CA^{-1}B$ must be non-singular.

A.2 Matrix Determinant Lemma

Assume that A is an invertible $n \times n$ square matrix; and u and v are column vectors. Then the matrix determinant lemma states that

$$\det(A + uv^T) = (1 + v^T A^{-1} u) \det(A), \tag{A.2}$$

where uv^T is the outer product of two vectors u and v.

In generalization, the lemma can also be expressed as

$$\det(A + UWV^T) = \det(W^{-1} + V^T A^{-1} U) \det(W) \det(A), \tag{A.3}$$

where U, V are $n \times m$ matrices, and W is an invertible $m \times m$ matrix.

A.3 Woodbury Matrix Identity

The Woodbury matrix identity is

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}, \tag{A.4}$$

where A, B, C and D are $n \times n$, $n \times k$, $k \times k$ and $k \times n$ matrices, respectively.

A.4 Cholesky Decomposition

Suppose that A is a symmetric, positive definite matrix, then the Cholesky decomposition of A decomposes A into a product of a unique lower triangular matrix L

and its transpose

$$A = LL^T, \tag{A.5}$$

where L is referred to as the Cholesky factor.

Specifically, the determinant of A can be efficiently computed by

$$|A| = \prod_{i=1}^{n} L_{ii}^2, \tag{A.6}$$

or

$$\log |A| = 2 \sum_{i=1}^{n} \log L_{ii}. \tag{A.7}$$

Appendix B

Information Theory

The following results are really useful for calculating entropy.

B.1 Chain Rule for Entropy

Let $X_1, X_2, ..., X_n$ be random variables. Then, the joint entropy $H(X_1, X_2, ..., X_n)$ can be computed as

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, ..., X_1), \tag{B.1}$$

where $H(X_i|X_{i-1}, ..., X_1)$ is the conditional entropy of X_i conditioned on $X_{i-1}, ..., X_1$, and $i = 2, ..., n$.

Bibliography

[1] M. P. Pata, T. Kneib, C. Cadarso-Suárez, V. Lustres-Pérez, and E. Fernández-Pulpeiro. Categorical structured additive regression for assessing habitat suitability in the spatial distribution of mussel seed abundance. *Environmetrics*, 23:75–84, 2012.

[2] M. Francisco-Fernández, A. Quintela del Río, and R. Fernández-Casal. Nonparametric methods for spatial regression. An application to seismic events. *Environmetrics*, 23:85–93, 2012.

[3] Z. Zhu and M. L. Stein. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:24–44, 2006.

[4] T. Harmon, C. Kvien, D. Mulla, G. Hoggenboom, J. Judy, and J. Hook. Precision agriculture scenario. In *Proc. NSF Workshop on Sensors for Environmental Observatories*, Baltimore, MD, USA, 2005.

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

[6] A. Garcia-Sanchez, F. Garcia-Sanchez, and J. Garcia-Haro. Wireless sensor network deployment for integrating video-surveillance and data-monitoring in

precision agriculture over distributed crops. *Computers and Electronics in Agriculture*, 75:288–303, 2011.

[7] J. Lindström, A. A. Szpiro, P. D. Sampson, A. P. Oron, M. Richards, T. V. Larson, and L. Sheppard. A flexible spatio-temporal model for air pollution with spatial and spatio-temporal covariates. *Environmental and Ecological Statistics*, 21:411–433, 2013.

[8] M. Jadaliha and J. Choi. Environmental monitoring using autonomous aquatic robots: Sampling algorithms and experiments. *IEEE Transactions on Control Systems Technology*, 21:899–905, 2013.

[9] N. D. Le and J. V. Zidek. *Statistical analysis of environmental space-time processes*. Springer, New York, 2006.

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

[11] R. Graham and J. Cortes. Spatial statistics and distributed estimation by robotic sensor network. In *Proc. IEEE American Control Conference*, pages 2422–2427, Baltimore, MD, USA, 2010.

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

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

[14] 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*, 9:235–284, 2008.

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

[16] G. Matheron. The intrinsic random functions and their application. *Advances in Applied Probability*, 5:439–468, 1973.

[17] J. P. Chilès and P. Delfiner. *Geostatistics: Modelling spatial uncertainty*. Wiley, 1999.

[18] C. K. I. Williams and C. E. Rasmussen. Gaussian processes for regression. *Advances in Neural Information Processing Systems*, 8:514–520, 1996.

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

[20] R. Graham and J. Cortes. A cooperative deployment strategy for optimal sampling in spatiotemporal estimation. In *Proc. IEEE Conference on Decision and Control*, pages 2432–2437, Cancun, Mexico, 2008.

[21] S. Fujishige. *Submodular functions and optimization*. Elsevier, 2005.

[22] L. Lovasz. Submodular functions and convexity. *Mathematical Programming: The State of the Art*, pages 235–257, 1983.

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

[24] W. W. Eric. Monotonic function, 2014. URL `http://mathworld.wolfram.com/MonotonicFunction.html`.

[25] K. Shih, Y. Chen, C. Chiang, and B. Liu. A distributed active sensor selection scheme for wireless sensor networks. In *Proc. 11th IEEE Symposium on Computers and Communications*, pages 923–928, Sardinia, Italy, 2006.

[26] M. Cardei and D. Du. Improving wireless sensor network lifetime through power-aware organization. *ACM Wireless Networks*, 11:333–340, 2005.

[27] M. Naderan, M. Dehghan, and H. Pedram. Sensing task assignment via sensor selection for maximum target coverage in WSNs. *Journal of Network and Computer Applications*, 36(1):262–273, 2013.

[28] L. Kaplan. Global node selection for localization in a distributed sensor network. *IEEE Transactions on Aerospace and Electronic Systems*, 42:113–135, 2006.

[29] F. R. Armaghani, I. Gondal, and J. Kamruzzaman. Dynamic sensor selection for target tracking in wireless sensor networks. In *Proc. IEEE Vehicular Technology Conference*, pages 1–6, San Francisco, USA, September 2011.

[30] F. R. Armaghani, I. Gondal, J. Kamruzzaman, and D. G. Green. Sensor selection for tracking multiple groups of targets. *Journal of Network and Computer Applications*, 46:36–47, 2014.

[31] J. Szurley, A. Bertrand, P. Ruckebusch, I. Moerman, and M. Moonen. Greedy distributed node selection for node-specific signal estimation in wireless sensor networks. *Signal Processing*, 94:57–73, 2014.

[32] D. Bajović, B. Sinopoli, and J. Xavier. Sensor selection for event detection in wireless sensor networks. *IEEE Transactions on Signal Processing*, 59:4938–4953, 2011.

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

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

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

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

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

[38] S. Joshi and S. Boyd. Sensor selection via convex optimization. *IEEE Transaction on Signal Processing*, 57:451–462, 2009.

[39] F. Ghassemi and V. Krishnamurthy. Upper bounds for sensor subset selection problem. In *Proc. International Conference on Information Fusion*, pages 110–117, Seattle, WA, USA, July 2009.

[40] F. Ghassemi and V. Krishnamurthy. Separable approximation for solving the sensor subset selection problem. *IEEE Transaction on Aorospace and Electronic Systems*, 47:557–568, 2011.

[41] M. Naeem, U. Pareek, and D. C. Lee. Swarm intelligence for sensor selection problems. *IEEE Sensors Journal*, 12(8):2577–2585, 2012.

[42] H. Jamali-Rad, A. Simonetto, and G. Leus. Sparsity-aware sensor selection: Centralized and distributed algorithms. *IEEE Signal Processing Letters*, 21(2): 217–220, 2014.

[43] S. B. Crary and Y. Jeong. Bayesian optimal design of experiments for sensor calibration. In *Proc. International Conference on Solid-State Sensors and Actuators, and Eurosensors IX*, pages 48–51, Stockholm, Sweden, 1995.

[44] A. Krause and C. Guestrin. Optimal value of information in graphical models. *Journal of Machine Learning Research*, 35:557–591, 2009.

[45] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg. Robust sensor placement at informative and communication efficient locations. *ACM Transactions on Sensor Networks*, 7:1–33, 2011.

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

[47] M. Jadaliha, Y. Xu, J. Choi, N. S. Johnson, and W. Li. Gaussian process regression for sensor networks under localization uncertainty. *IEEE Transactions on Signal Processing*, 61:223–237, 2013.

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

[49] 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*, 3:1–13, 2014.

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

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

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

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

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

[55] S. Kirkpatric, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[56] V. Gupta, T. Chung, B. Hassibi, and R. Murray. On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. *Automatica*, 42(2):251–260, 2006.

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

[58] L. Paninski. Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17:1480–1507, 2005.

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

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

[61] B. Korte and J. Vygen. *Combinatorial optimization: Theory and algorithms*. Springer, 2012.

[62] L. Yao, W. A. Sethares, and D. C. Kammer. Sensor placement for on-orbit modal identification via a genetic algorithm. *AIAA*, 31:1922–1928, 1993.

[63] B. Suman and P. Kumar. A survey of simulated annealing as a tool for single and multiobjective optimization. *Operational Research Society*, 57:1143–1160, 2006.

[64] J. P. Sheu, K. Y. Hsieh, and P. W. Cheng. Design and implementation of mobile robot for nodes replacement in wireless sensor networks. *Journal of Information Science and Engineering*, 24:393–410, 2008.

[65] G. T. Sibley, M. H. Rahimi, and G. S. Sukhatme. Robomote: A tiny mobile roboti platform for large-scale ad-hoc sensor networks. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 1143–1148, Washington DC, USA, May 2002.

[66] S. Bergbreiter and K. S. J. Pister. CotsBots: An off-the-shelf platform for distributed robotics. In *Proc. IEEE/RSJ International Conferece on Intelligent Robots and Systems*, pages 1632–1637, Nevada, USA, October 2003.

[67] M. B. McMickell, B. Goodwine, and L. A. Montestruque. MICAbot: A robotic platform for large-scale distributed robotics. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 1600–1605, Taipei, Taiwan, September 2003.

[68] M. D. Francesco, S. K. Das, and G. Anastasi. Data collection in wireless sensor networks with mobile elements: A survey. *ACM Transactions on Sensor Networks*, 8(1):7:1–7:31, 2011.

[69] Q. Dong and W. Dargie. A survey on mobility and mobility-aware MAC protocols in wireless sensor networks. *IEEE Communications Surveys and Tutorials*, 15(1):88–100, 2013.

[70] R. C. Shah, S. Roy, S. Jain, and W. Brunette. Data MULEs: modeling a three-tier architecture for sparse sensor networks. In *Proc. IEEE International Workshop on Sensor Network Protocols and Applications*, pages 30–41, Anchorage, AK, USA, May 2003.

[71] C. H. Wu, W. Sheng, and Y. Zhang. Mobile sensor networks self localization based on multi-dimensional scaling. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 4038–4043, Rome, Italy, April 2007.

[72] S. Gao, H. Zhang, and S. Das. Efficient data collection in wireless sensor networks with path-constrained mobile sinks. In *Proc. IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks and Workshops*, pages 1–9, Kos, Greece, June 2009.

[73] O. Tekdas, V. Isler, J. H. Lim, and A. Terzis. Using mobile robots to harvest data from sensor fields. *IEEE Wireless Communications*, 16(1):22–28, 2009.

[74] T. C. Chen, T. S. Chen, and P. W. Wu. On data collection using mobile robot in wireless sensor networks. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems*, 41(6):1213–1224, 2011.

[75] L. He, J. Pan, and J. Xu. An on-demand data collection scheme for wireless sensor networks with mobile elements. In *Proc. IEEE International Conference on Communications*, pages 1–5, Kyoto, Japan, June 2011.

[76] G. Fletcher, L. Xu, A. Nayak, and I. Stojmenovic. Randomized robot-assisted relocation of sensors for coverage repair in wireless sensor networks. In *Proc. IEEE Vehicular Technology Conference*, pages 1–5, Ottawa, Canada, September 2010.

[77] L. Xu, R. Falcon, A. Nayak, and I. Stojmenovic. Servicing wireless sensor networks by mobile robots. *IEEE Communications Magazine*, 50(7):147–154, 2012.

[78] Z. Yong and W. Li. A particle filtering method for odor-source localization in wireless sensor network with mobile robot. In *Proc. 8th World Congress on Intelligent Control and Automation*, pages 7032–7036, Jinan, China, July 2010.

[79] A. Howard, M. J. Mataric, and G. S. Sukhatme. Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem. In *Proc. 6th International Symp. on Distributed Autonomous Robotics Systems*, pages 299–308, Fukuoka, Japan, June 2002.

[80] Y. C. H. Y. Mei, Y. H. Lu, and C. S. G. Lee. Reducing the number of mobile sensors for coverage tasks. In *Proc. IEEE/RSJ International Conferece on*

Intelligent Robots and Systems, pages 1426 – 1431, Alberta, Canada, August 2005.

[81] P. Corke, R. Peterson, , and D. Rus. Localization and navigation assisted by networked cooperating sensors and robots. *International Journal of Robotics Research*, 24(9):771–786, 2005.

[82] P. L. Pham, A. T. Mai, and S. Commuri. Mobile robots assisted target tracking in wireless sensor networks. In *Proc. IEEE Globecom Workshops*, pages 1574 – 1578, Anaheim, CA , USA, December 2012.

[83] Y. Mei, C. Xian, S. Das, Y. C. Hu, and Y. H. Lu. Replacing failed sensor nodes by mobile robots. In *Proc. IEEE International Conference on Distributed Computing Systems, Workshop on Wireless Ad hoc and Sensor Networks*, pages 1 – 6, Lisboa, Portugal, July 2006.

[84] Z. Zhang, Z. Li, D. Zhang, and J. Chen. Path planning and navigation for mobile robots in a hybrid sensor network without prior location information. *International Journal of Advanced Robotic Systems*, 10:1–12, 2013.

[85] C. Tang and P. McKinley. Energy optimization under informed mobility. *IEEE Transactions on Parallel and Distributed Systems*, 17(9):947 – 962, 2006.

[86] G. Dini, M. Pelagatti, and I. M. Savino. An algorithm for reconnecting wireless sensor network partitions. In *Proc. 5th European conference on Wireless Sensor Networks*, page 253–267, Bologna, Italy, January 2008.

[87] K. A. Luthy, E. Grant, and T. C. Henderson. Leveraging RSSI for robotic repair of disconnected wireless sensor networks. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 3659 – 3664, Rome, Italy, April 2007.

[88] J. Cortes, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20:243–254, 2004.

[89] N. E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. Davis. Collective motion, sensor networks and ocean sampling. *Proceedings of the IEEE*, 95(1):48–74, 2007.

[90] S. Martínez. Distributed interpolation schemes for field estimation by mobile sensor networks. *IEEE Transactions on Control Systems Technology*, 18:491–500, 2010.

[91] S. Huang and J. Tan. Adaptive sampling using mobile sensor networks. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 657–662, Minnesota, USA, May 2012.

[92] J. Binney, A. Krause, and G. S. Sukhatme. Optimizing waypoints for monitoring spatiotemporal phenomena. *The International of Robotics Research*, 32(8):873–888, 2013.

[93] R. Ouyang, K. H. Low, J. Chen, and P. Jaillet. Multi-robot active sensing of non-stationay Gaussian process-based environmental phanomena. In *Proc. International Foundation for Autonomous Agents and Multiagent Systems*, pages 573–580, Paris, France, 2014.

[94] H. M. La, W. Sheng, and J. Chen. Cooperative and active sensing in mobile sensor networks for scalar field mapping. In *Proc. IEEE International Conference on Automation Science and Engineering*, pages 831–836, WI, USA, 2013.

[95] J. Suh and S. Oh. Efficient environmental monitoring using cost-aware path planning. In *Proc. 13th International Conference on Control, Automation and Systems*, pages 1362–1365, Gwangju, Korea, 2013.

[96] R. Graham and J. Cortes. Distributed sampling of random fields with unknown covariance. In *Proc. American Control Conference*, pages 4543–4548, MI, USA, 2009.

[97] R. Graham and J. Cortes. Cooperative adaptive sampling via approximate entropy maximization. In *Proc. IEEE Conference on Decision and Control*, pages 7055–7060, Shanghai, China, 2009.

[98] D. O. Popa, M. F. Mysorewala, and F. L. Lewis. EKF-based adaptive sampling with mobile robotic sensor nodes. In *Proc. IEEE/RSJ International Conferece on Intelligent Robots and Systems*, pages 2451–2456, Beijing, China, October 2006.

[99] D. O. Popa, M. F. Mysorewala, and F. L. Lewis. Adaptive sampling using non-linear EKF with mobile robotic wireless sensor nodes. In *Proc. IEEE International Conference on Control, Automation, Robotics and Vision*, pages 1–6, Grand Hyatt, Singapore, December 2006.

[100] J. Choi, J. Lee, and S. Oh. Swarm intelligence for achieving the global maximum using spatio-temporal Gaussian processes. In *Proc. American Control Conference*, pages 135–140, Washington, USA, June 2008.

[101] J. L. Ny and G. J. Pappas. On trajectory optimization for active sensing in Gaussian process models. In *Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, pages 6286–6292, Shanghai, China, 2009.

[102] J. Euler, A. Horn, D. Haumann, J. Adamy, and O. V. Stryk. Cooperative n-boundary tracking in large scale environments. In *Proc. IEEE International Conference on Mobile Adhoc and Sensor Systems*, pages 1–6, Las Vegas, NV, USA, October 2012.

[103] W. Wu and F. Zhang. Robust cooperative exploration with a switching strategy. *IEEE Transactions on Robotics*, 28(4):828 – 839, 2012.

[104] C. H. Caicedo-Nú nez and N. E. Leonard. Symmetric coverage of dynamic mapping error for mobile sensor networks. In *Proc. IEEE American Control Conference*, pages 4661–4666, San Francisco, CA, USA, 2011.

[105] J. Cortés. Distributed Kriged Kalman filter for spatial estimation. *IEEE Transactions on Automatic Control*, 54:2816–2827, 2009.

[106] S. Oh and J. Choi. Distributed learning in mobile sensor networks using cross validation. In *Proc. 49th IEEE International Conference on Decision and Control*, pages 3845–3850, Atlanta, USA, 2010.

[107] Y. Xu and J. Choi. Adaptive sampling for learning Gaussian processes using mobile sensor networks. *Sensors*, 11:3051–3066, 2011.

[108] Y. Xu, J. Choi, S. Dass, and T. Maiti. Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks. *IEEE Transactions on Automatic Control*, 57:2078–2084, 2012.

[109] Y. Xu, J. Choi, and S. Oh. Mobile sensor network navigation using Gaussian processes with truncated observations. *IEEE Transactions on Robotics*, 27: 1118–1131, 2011.

[110] R. Marchant and F. Ramos. Bayesian optimization for intelligent environmental monitoring. In *Proc. IEEE/RSJ International Conferece on Intelligent Robots and Systems*, pages 2242–2249, Algarve, Portugal, October 2012.

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

[112] C. K. I. Williams and M. Seeger. Using the Nyström method to speed up kernel machines. *Advances in Neural Information Processing Systems*, 13, 2001.

[113] A. J. Smola and P. Bartlett. Sparse greedy Gaussian process regression. *Advances in Neural Information Processing Systems*, 13, 2001.

[114] N. Lawrence, M. Seeger, and R. Herbrich. Fast sparse Gaussian process methods: The informative vector machine. In *Proc. 16th Annual Conference on Neural Information Processing Systems*, pages 609–616, Vancouver, BC , USA, 2003.

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

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

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

[118] 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)*, 73: 423–498, 2011.

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

[120] M. Jadaliha and J. Choi. Fully Bayesian simultaneous localization and spatial prediction using Gaussian Markov random fields. In *Proc. American Control Conference*, pages 4599–4604, Washington DC, USA, June 2013.

[121] M. Jadaliha, Y. Xu, and J. Choi. Efficient spatial prediction using Gaussian Markov random fields under uncertain localization. In *Proc. ASME Dynamic Systems and Control Conference*, Florida, USA, October 2012.

[122] Y. Xu and J. Choi. Spatial prediction with mobile sensor networks using Gaussian processes with built-in Gaussian Markov random fields. *Automatica*, 48:1735–1740, 2012.

[123] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical methods for physics and engineering*. Cambridge University Press, 2006.

[124] S. Chen and B. L. Luk. Adaptive simulated annealing for optimization in signal processing applications. *Signal Processing*, 79:117–128, 1999.

[125] B. Suman. Study of simulated annealing based multiobjective algorithm for multiobjective optimization of a constrained problem. *Computer and Chemical Engineering*, 28:1849–1871, 2004.

[126] A. G. Boon and A. R. Lister. Formation of sulfide in rising main sewers and its prevetion by injection of oxygen. *Prog. Water Technology*, 7(2):289–300, 1975.

[127] P. Gostelow, S. A. Parsons, and R. M. Stuetz. Odour measurement for sewage treatment works. *Water Research*, 35(2):579–597, 2001.

[128] K. Sharma, D. W. de Haas, S. Corrie, K. OHalloran, J. Keller, and Z. Yuan. Predicting hydrogen sulfide formation in sewers: A new model. *Journal of the Australian Water Association*, pages 60–65, 2008.

[129] K. Sharma, Z. Yuan, D. W. de Haas, G. Hamilton, S. Corrie, and J. Keller. Dynamics and dynamic modelling of H_2S production in sewer systems. *Water Research*, pages 2527–2538, 2008.

[130] L. Sutherland-Stacey, S. Corrie, A. Neethling, I. Johnson, O. Gutierrez, R. Dexter, Z. Yuan, J. Keller, and G. Hamilton. Continuous measurement of dissolved sulfide in sewer systems. *Journal of Water Science and Technology*, pages 375–381, 2008.

[131] Y. C. Wang, N. Nobi, T. Nguyen, and L. Vorreiter. A dynamic ventilation model for gravity sewer networks. *Journal of Water Science and Technology*, pages 60–68, 2012.

[132] T. Hvitved-Jacobsen, J. Vollertsen, and A. H. Nielsen. *Sewer processes: Microbial and chemical process engineering of sewer networks*. CRC Press, Washington, DC, 2002.

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

[134] L. Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003.

[135] P. Bodik, C. Guestrin, W. Hong, S. Madden, M. Paskin, and R. Thibaux. Intel lab data, 2004. URL `http://db.csail.mit.edu/labdata/labdata.html`.

[136] P. J. Ribeiro and P. J. Diggle. geor: A package for geostatistical analysis. *R NEWS*, 1:15–18, 2001.

[137] L. F. Merino, J. R. Caballero, J. M. de Dios, and A. O. Ferruz. A cooperative perception system for multiple UAVs: Application to automatic detection of forest fires. *Journal of Field Robotics*, 23:165–184, 2006.

[138] C. E. Corrigan, G. C. Roberts, M. V. Ramana, and V. Ramanathan. Capturing vertical profiles of aerosols and black carbon over the indian ocean using autonomous unmanned aerial vehicles. *Atmospheric Chemistry and Physics Discussions*, 7:11429–11463, 2007.

[139] S. Susca, F. Bullo, and S. Martínez. Monitoring environmental boundaries with a robotic sensor network. *IEEE Transactions on Control Systems Technology*, 16:288–296, 2008.

[140] K. M. Lynch, I. B. Schwartz, P. Yang, and R. A. Freeman. Decentralized environmental modeling by mobile sensor networks. *IEEE Transactions on Robotics*, 24:710–724, 2008.

[141] J. Choi, S. Oh, and R. Horowitz. Distributed learning and cooperative control for multi-agent systems. *Automatica*, 45:2802–2814, 2009.

[142] M. A. Demetriou and I. I. Hussein. Estimation of spatially distributed processes using mobile spatially distributed sensor network. *SIAM Journal on Control and Optimization*, 48(1):266–291, 2009.

[143] G. M. Hoffmann and C. J. Tomlin. Mobile sensor network control using mutual information methods and particle filters. *IEEE Transactions on Automatic Control*, 55:32–47, 2010.

[144] M. Zhong and C. G. Cassandras. Distributed coverage control and data collection with mobile sensor networks. *IEEE Transactions on Automatic Control*, 56(10):2445–2455, 2011.

[145] D. Gu and H. Hu. Spatial Gaussian process regression with mobile sensor networks. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8): 1279–1290, 2012.

[146] M. A. Osborne, S. J. Roberts, A. Rogers, and N. R. Jennings. Real-time information processing of environmental sensor network data using Bayesian Gaussian processes. *ACM Transactions on Sensor Networks*, 9(1):1:1–1:32, 2012.

[147] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95:215–233, 2007.

[148] G. C. Calafiore, L. Carlone, and M. Wei. A distributed technique for location of agent formations from relative range measurements. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems*, 42(5):1065–1076, 2012.

[149] Z. Butler and D. Rus. Controlling mobile sensor for monitoring events with coverage constraints. In *Proc. IEEE International Conferece on Robotics and Automation*, pages 1563–1573, New Orleans, LA, USA, May 2004.

[150] R. C. Hibbeler. *Engineering mechanics: Statics and dynamics*. Prentice Hall, 1990.

[151] J. E. Goodman and J. O'Rourke. *Handbook of discrete and computational geometry*. CRC Press, 2004.

[152] M. L. Stein, Z. Chi, and L. J. Welty. Approximating likelihood for large spatial data sets. *Journal of the Royal Statistical Society (Series B)*, 66:275–296, 2004.

[153] S. Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39:55–72, 1978.

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

[155] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and distributed computation: Numerical methods*. Englewood Cliffs: Prentice Hall, 1999.

[156] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95:215–233, 2007.

[157] P. V. Bolstad, L. Swift, F. Collins, and J. Régnière. Measured and predicted air temperatures at basin to regional scales in the southern appalachian mountains. *Agricultural and Forest Meteorology*, 91:161–176, 1998.

[158] D. Mulla, M. Beatty, and A. Sekely. Evaluation of remote sensing and targeted soil sampling for variable rate application of nitrogen. In *Proc. 5th International Conference on Precision Agriculture*, America, 2001.

[159] P. Tokekar, J. V. Hook, D. Mulla, and V. Isler. Sensor planning for a symbiotic UAV and UGV system for precision agriculture. Technical Report 13-010, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455-0159, USA, March 2013.

[160] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser. Efficient informaive sensing using multiple robots. *Journal of Artifitial Intelligence Research*, 34: 707–755, 2009.

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

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

[163] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer, New York, 1999.

[164] 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)*, 71:319–392, 2009.