

BIO-INSPIRED ALGORITHMS FOR DATA FUSION IN
HAZARDOUS THREAT DETECTION

By
Champake Mendis

SUBMITTED IN TOTAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
THE UNIVERSITY OF MELBOURNE
MELBOURNE, AUSTRALIA
AUGUST 2014

Produced on archival quality paper.

© Copyright by Champake Mendis, 2014

Abstract

Bio-inspired systems focus on the design, development and understanding of systems composed of algorithms that mimic the behaviours or processes of biological entities. In order to achieve a better abstraction, these bio-inspired algorithms can be constructed on a multi-agent platform which comprises multiple interaction between autonomous agents. These components are equipped with cognitive and processing abilities and have access to information extracted from the environment. In our endeavour to explore the viability of using bio-inspired algorithms in chemical, biological, radiological and nuclear environments, we carried out four research studies.

First, we present a numerical verification of a simple population and physics-based epidemiological model for dynamic collaboration in a network of chemical sensors. The modelling approach is based on the known analogy between the information spread in a sensor network and the propagation of epidemics across a population. In this framework, we verify the derived analytical expressions, which relate the parameters to the network (e.g., number of sensors, their density, sensing time, etc.), with parameters of the external challenge (e.g., the chemical pollutant) and the environment (e.g., turbulence). Using numerical simulations of wireless sensor networks with random, line and circle topologies, we show that simulated and analytical results agree.

Secondly, we apply the epidemiology based protocol to a wireless chemical sensor network in a chemical environment with spatial characteristics. The chemical tracers dispersed by turbulent motion in the environment display rather complex and even chaotic properties. Meanwhile, chemical tracer detecting sensors with air sampling

to consume significant energy, hazardous chemical releases are rare events which will not require continuity in detection. If all sensors in a wireless chemical sensor network (WCSN) are left in the active state continuously, it would result in significant power consumption. Therefore, dynamic sensor activation is crucial for the longevity of WCSNs. Moreover, the statistical characteristics of chemical tracers to be detected (temporal and spatial correlations, etc.) and placement of chemical sensors can also become the key parameters that influence the WCSN design and performance. In this research study, we investigate the effect of the spatial correlation of a chemical tracer field, and also the effect of network topology, on the performance of a WCSN that employs an epidemiology-based dynamic sensor activation protocol. We present a simulation framework that comprises models of the spatially correlated tracer field, individual chemical sensor nodes, and the sensor network. After validating this simulation framework against an analytical model, we perform simulation experiments to evaluate the effect of spatial correlation and network topology on selected performance metrics: response time, level of sensor activation, and network scalability. Our simulations show that the spatial correlation of the chemical tracer field has a detrimental effect on the performance of a WCSN that uses an epidemiological activation protocol. These results also suggest that a WCSN with random network topology has poor performance compared to one with a regular grid topology in this application.

Thirdly, we apply a gossip based protocol to the wireless chemical sensor network to overcome the detrimental effect of the epidemiological protocol. In this study, we investigate the performance of a variant of epidemiology based protocols, the gossip-based sensor activation protocol of a WCSN in a chemical tracer field. The simulation framework with the gossip protocol is validated against an analytical model. We then perform simulation experiments to evaluate the performance of the gossip-based sensor activation protocol on selected performance metrics: the sensor activation and chemical tracer detection. We show by simulations that, by varying the communication radii of sensors; we can achieve better energy conservation while maintaining better performance of a WCSN with a gossip-based activation protocol which was the

drawbacks of the epidemiological protocol in a turbulent environment.

Fourthly, we explore the possibility of localising a radiological source using bio-inspired genetic algorithms. We consider localisation of point sources of gamma radiation using dose rate measurements. As bio-inspired genetic algorithms, we use binary and continuous genetic algorithms (GA). They are used to implement maximum likelihood estimation (MLE) of the position and strength of point radiation sources. MLE was achieved by minimising the objective function which computes the negative log likelihood. Real experimental data collected during a field trial was used to test and verify the performance of the algorithms. The performance of the GA-based implementation was compared to an implantation that used gradient descent optimisation. Source parameters estimated by the algorithms were also compared to the theoretical bounds obtained using a Cramer-Rao bound (CRB) analysis, which quantifies the accuracy with which it is possible to localise the source and estimate its strength. All three implementations localised a single-point source well, nearly approaching the CRB. Reasonable position estimates were achieved for two and three source cases, but the source strength estimates were found to have much larger root mean square (RMS) errors than that predicted by CRB. While the GA-based implementations took longer to converge compared to the gradient descent algorithm, they encountered fewer divergent runs than the latter algorithm. Also we examine the data collection geometries that influence the topography of objective function surfaces in radiological source localisation. It is shown that data gathered along the circumference of a circle around a point a radiation source has an associated mirror image source of different strength that exists outside the circle. It appears that data acquired along an irregular path generated using the random walk algorithm to eliminate the image sources making source parameter estimation easier.

Given these considerations, it is evident that Bio-inspired algorithms are viable solutions for CBRN data fusion.

Declaration

This is to certify that

- (i) the thesis comprises only my original work,
- (ii) due acknowledgment has been made in the text to all other material used,
- (iii) the thesis is less than 100,000 words in length, exclusive of table, maps, bibliographies, appendices and footnotes.

Signature:.....

Date:.....

Melbourne, Australia
August 15, 2014

Acknowledgments

I would like to thank Associate Professor Shanika Karunasekera and Dr Alex Skvortsov, my supervisors, for their encouragement, advice suggestions and constant support during this research. I am especially thankful to Dr Ajith Gunatilaka being a member of the advisory committee and his guidance throughout the years and in chaos and confusion during the course of the study.

The chair of the advisory committee Associate Professor Adrian Pearce expressing his interest in my work and supplied me with the valuable suggestions, which gave me a better perspective on my own research study. He shared with me his knowledge on sensor node failures, risks in the progression and his friendly encouragement. Professor Justin Zobel and Associate Professor James Bailey for their formal guidance in research through the Research Methods course.

It was a pleasure to work with Defence Science and Technology Organisation (DSTO) staff of Fishermans Bend for a period of three years and then DSTO-Sydney, where I had the opportunity to get first hand industrial research experience. They are wonderful people, and their support makes research like this practical and useful. The *Australian Postgraduate Award*, which was awarded to me for the period 2009–2012, was crucial to the successful completion of this project without a financial burden. I should also mention that my graduate studies in Australia were supported in part by the DSTO.

Further I acknowledge support provided by the colleagues in the University of Sydney, where I had the privilege of applying the knowledge, skills and expertise gathered in the research study to maintenance of multitude of distributed databases.

Of course, I am grateful to my wife Kaumudi and my parents for their patience and *love*. Without them, this work would never have come into existence (literally).

Champake Mendis, Melbourne, Australia
August 15, 2014.

Preface

Publications arising from this thesis

The content of this thesis comprises only my original work which, was conducted solely during my PhD candidature and has not been submitted for any other qualifications or degrees. Parts of this thesis have been extracted and published in various forums and publications in collaboration with my supervisors Associate Professor Shanika Karunasekera and Dr Alex Skvortsov, and colleagues, Dr. Ajith Gunatilaka and Dr. Branko Ristic.

In the publications, I developed most of the ideas, conducted all the described experiments, and was responsible for the writing of each paper with the support from co-authors and in some instances received the proof reading support from my friend, Jason Foster and Dr.Sudanthi Wijewickrema.

The following works were published prior to the PhD candidature as part of my research for Masters of Engineering Science in Mechatronics at the University of Melbourne, for which First Class Honours were awarded,

- C. Mendis, S. M. Guru, S. Halgamuge, and S. Fernando, ‘Optimized sink node path using particle swarm optimisation’,in Proc. AINA 2006: 20th International Conference on Advanced Information Networking and Applications, Vienna, Austria, 2006.
- C. Mendis, ‘Optimised Sink Trajectories for Sensor Networks’, Germany: LAP Publishing, 2009.

External presentations

- C. Mendis, 'Bio-inspired algorithms in engineering applications: In retrospect', Edinburgh Napier University, Edinburgh, UK, 2011.
- C. Mendis, 'Modelling a Wireless Chemical Sensor Network', DSTO-Fishermans Bend, Melbourne, Australia, 2011.

Document preparation, tools and data

The thesis was prepared with the LaTeX document formatting language, using the algorithm, algpseudocode, amsfonts, amsmath, amsopn, booktabs, caption, float, graphicx, listings, multirow, natbib, rotating, times, url, and xcolor packages. Statistical analysis, simulation experiments, and general programming tasks were performed in the MATLAB language with Signal Processing and Statistical Toolboxes. Some simulations were done in C++ with OpenGL. The special software tools Rationale, Microsoft Visio and Freemind for documenting concepts, Microsoft Project for Project Management, Microsoft Sourcesafe for version control and StarUML for software design were used during the research study.

A feasibility of using CUDA programming and high performance computing were explored but did not use due to time constraints as beyond the necessity for the research study.

Nomenclature

R_0	Basic reproduction rate
R_{corr}	Correlation radius
C^*	Contamination detection threshold of a sensor
r^*	Communication radius of a sensor
f	Correlation coefficient (R_{corr} / r^*)
τ_*	Fan out ratio of gossip algorithm
α	Infectious period
L	interaction or infection rate
C_0	Length of a side of the square contamination field
m	Mean concentration of contamination environment
N_+	Number of messages transacted
D_+	Number of sensors active
γ	Number of sensors detecting contamination
P	Parameter to be compliant with Kolmogorov turbulence model
Ψ	Probability of detection of contamination
τ	Unitized risk of detection (Reliability of detection)
a	Rise time/Network detection time
N	Spatial fraction of contamination environment
ν	Total number of sensors
δ	Tracer intermittency of contamination environment
	Variance of contaminant slice tracers

Chapter 1

Introduction

How can we use the inherent characteristics of our environment and the nature of human senses to inspire computational algorithms that can detect and localise threats and save energy while sending time sensitive data to save many lives?

This is the motivation for *bio-inspired algorithms for hazardous threat detection* which is the focus of this thesis. Population growth has increased the risks of political outbreaks, and new technological trends have enabled political radicals to create highly complex hazardous attacks against those with opposing views. However, some cases outbreaks could be innocently created due to carelessness and misconceptions.

In a hazardous substance release, such as the Sarin nerve gas attack in Tokyo in 1995, rapid detection of the event is important for minimising the damage, including casualties. Any immediate damage is sometimes imminent and may need detection and data collection to occur in the short period of time it takes for the effects to be manifest in the victims [36]. Furthermore, radiological and atomic sources can be lost or stolen [79], and in case of their release, it is of the utmost importance to identify the source of the atomic emissions.

The detection of complex hazardous material has therefore become the platform for a new-generation of research. Bio-inspired techniques have added more value to the arena by providing methods that have been stable for over a life-time. The popularity of these methods are the result of greater attention from the research community.

However, the importance and value of these methods has increased relation to the growing number of targets for a wide variety of attacks and the proliferation of threats to the security of the environment. For example, the number of reported chemical, biological, radiological and nuclear (CBRN) security incidents has increased from a single event in Tokyo in 1995 to regular use in civil wars [36].

In recent years, attackers have shown increasing complexity in their ability to launch attacks that target or utilize a large of number of locations that are spread over a wide geographical area [36]. For example, attackers can scan large numbers of locations simultaneously by using the remote activation of dangerous devices; they can use various contraptions such as dirty bombs, which contain lethal radioactive material; and they can also make use of the monitoring sensor network systems by compromising sensor nodes (i.e. distributed denial-of-service). We refer to these types of attacks as large-scale coordinated attacks.

There are roughly 12,000 incidents reported in the ITERATE database, spanning a 24-year period from 1968 to 2001. Importantly, however, only transnational terrorist events are considered. ITERATE defines an event as transnational using several criteria. These include the nationalities of the perpetrators and the location of the attack, as well as the nature of the target [68].

The environment resulting from a CBRN attack is very complex, as the constituents mix and spread due to environmental turbulences and therefore the evidence of the attack is spread across multiple domains. For example, consider the illustration of a CBRN attack in Figure 1.1, which depicts a wireless sensor network in a field under hazardous chemical threat.

In this figure, we show a typical data fusion system which can be connected to the internet, a grid or a cloud in the latest state-of-the-art technology. A CBRN attack is released at a point and dispersed through the environment. If a serious panic event is happening at a location, we can accept that it will be felt instantly by the victims. A sensor system with collaborated detection can be deployed to identify its location, direction, speed and extent of damage. However, if there are high winds and the

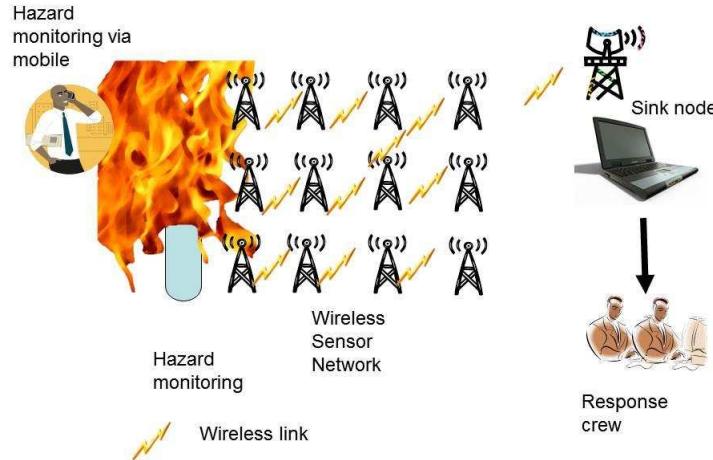


Figure 1.1: Hazard monitoring

CBRN constituents are carried away, less residue will be left for the sensors to trigger a detection. Similar scenarios occur during the early stages of a chemical outbreak, when an attack host starts to infect target systems with a virus, and when the master carrier infects the others to create an epidemic outbreak.

In order to detect these types of CBRN attacks, we need the ability to combine the evidence of suspicious network activity from several geographically-distributed sensor nodes. Rather than each sensor working in isolation, we can form a collaborative system, that analyses evidence from the several locations simultaneously. Although coordinated attacks may be easier to detect at later stages when the volume of attack is large, the benefit of hazard detection would be diminished because by that stage, the damage may have already been done. By combining evidence from multiple sensor nodes, we can detect such coordinated attacks at an early stage, before they have had a significant impact on the environment. The problem we address in this thesis is how to realize such a collaborative hazard detection system (CHDS). Given the large number of locations that can be affected by a coordinated attack or a distributed hazardous material, it is essential that a CHDS be highly scalable, while supporting

fully expressive bio-inspired algorithms to detect a wide variety of attacks in an accurate and computationally efficient manner. In this thesis, we propose several novel approaches to the problem of collaborative sensor activation for hazard detection, and evaluate their accuracy and scalability on a variety of different coordinated attacks in a realistic global-scale deployment on an area. The latest technology is smart-phone integrated chemical detection as developed by the Homeland Security Centre in the USA, known as CELL-ALL. The algorithms need to support data extraction from a large number of detecting sensor nodes.

Wireless sensor networks (WSNs) are of significant importance in the hazard detection process. The purpose of such WSNs is to detect, identify, and characterise the dynamic threats associated with tracer fields, such as aerosol, gas, moisture, etc., by using the advantages of a collaborative performance of sensors over those of individual nodes. Some recent initiatives (for example, by the USA Department of Homeland Security, [70]) are aimed at developing tiny chemical sensors that can be embedded in a conventional mobile phone and will make possible the construction of large-scale wireless chemical sensor networks (WCSNs) with massive numbers (thousands or even millions) of nodes. However, as the number of sensors increases, the design and management of such systems become a challenging task that requires mathematical modelling, computer simulation and parameter optimisation. In our study, we estimate the contribution of spatial correlations to the positive detection of hazardous threats by simulating different types of WCSNs in tracer fields with different correlation structures (short and long correlated; see Figure 1.2). We also briefly examine the effect of network topology (node connectivity) on the WCSNs performance.

The application of *WSN* in various topologies has been a topic of intensive research in areas spanning from technological and environmental monitoring to defence and security systems [34, 72]. Of particular interest is the optimisation of WSN to ensure the longevity and sustainability of the network system in the long run. This can be achieved by a simple reduction of the active time of individual sensors (scheduling) and

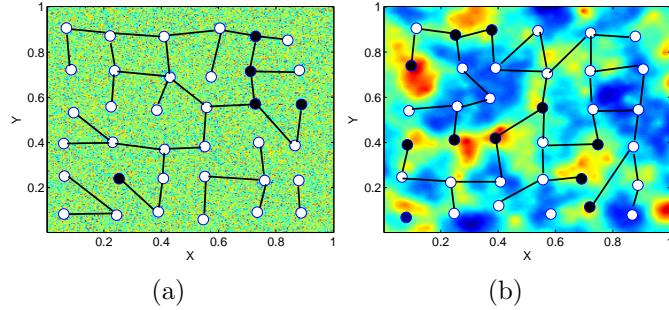


Figure 1.2: Graphical illustration of WCSNs operating in simulated chemical tracer fields that are: (a) spatially non-correlated; and (b) spatially correlated. The tracer fields are depicted as 2D xy -planes, with different colours representing different tracer concentrations. The empty white circles represent active sensors, and the solid black circles represent inactive sensors.

also by using various network topologies [27]. Because wireless networks, unlike their wired counterparts, are powered by batteries, the associated optimization problem sometimes becomes even more important as it has to meet energy supply constraints as well.

For a wireless network of sensors, the energy constraints often become the primary consideration because its design is driven by a trade-off between two opposite directions in developing operational protocols: prolonging the lifetime of the network and maximizing the reliability of event detection. The first direction, driven by the need to conserve energy, aims at reducing the active (sampling) time of all sensors in the network to save energy and thus prolong the network lifetime. Most modern chemical sensors contain an air sampling unit (fan), which turns on when the sensor is active and uses most of the battery. Therefore it is imperative that an effective protocol is devised to minimize the active time of the sensors. The second direction is ‘information driven’, which necessitates increasing the sensor sampling time and sampling volume to facilitate reliable detection of intermittently distributed chemical pollutants [43]. The latter, evidently, increases energy consumption.

‘Data fusion’ encompasses the process of association, correlation, and combination of data and information gathered from a single source or multiple sources to achieve

refined position and identity estimates, complete and timely assessments of the situation and the threat, and the evaluations of significance of the threat (see Fig. 1.3). This process is characterised by the continuous refinement of its estimates and assessments, and the evaluation of the need for additional sources, or the modification of the process itself, to achieve improved results [46]. To this end, intelligent techniques are integrated to the data fusion system. Numerous mathematical techniques, such as valuation networks, belief networks, random sets and evidential networks [8], are available for this purpose. However, these mathematical techniques are extremely complicated, computationally intensive and require higher resource overheads.

Data fusion from wireless networks of CBRN sensors has been tested and accepted as a viable option for detecting chemical, biological and radiological warfare agents. Since hazardous agents could be dispersed in a wider area within a short time, a network of sensors collaborating with each other can be used to greater effect over individual sensors to detect threats and ascertain the extent of dispersion.

However conserving energy is the main concern when sensors are left to operate without human intervention. This needs to be weighed against the need to operate the sensor network in the most efficient manner for information fusion. Energy is mainly conserved by reducing the active time of all sensors in the network and thus lengthening the network's lifetime. CBRN sensors and other accessories, such as air sampling devices, drain most of the battery power. Efficient data fusion demand requires enough sensor sampling time and sampling volume to support a reliable detection of an intermittently distributed chemical contaminant [43]. This inevitably increases energy consumption.

In this introductory chapter, the focus of the thesis in Section 1.1 is described in terms of the background to collaborative sensor activation and detection and the scope of the problem that is addressed in Section 1.2 outlines the structure of the proposed solution to the problem. Then in Sections 1.3 and 1.4 summarise the organisation of the thesis and the contributions of each chapter. Section 1.5 concludes with a list of the publications that have arisen from the research presented in this thesis.

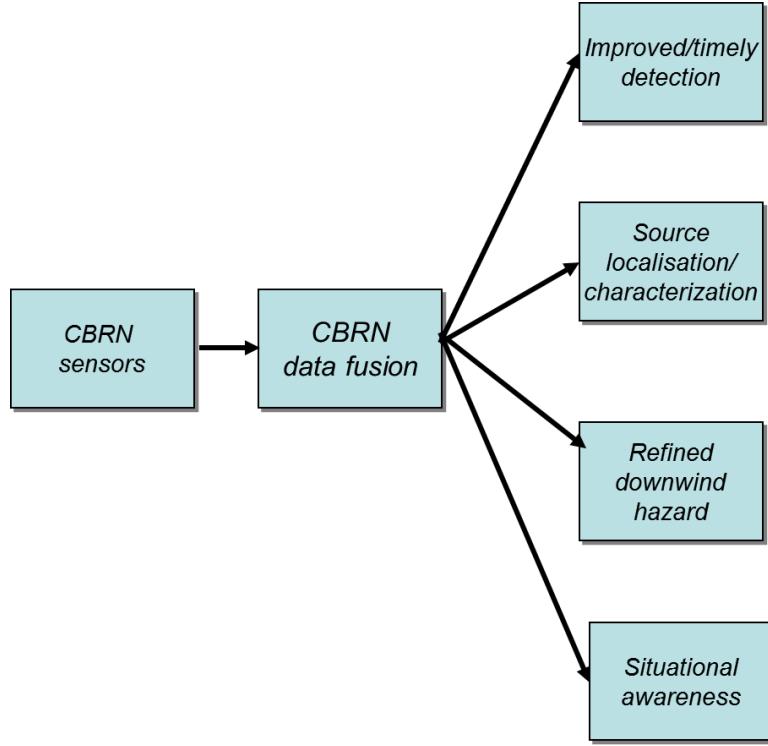


Figure 1.3: Data fusion process

1.1 Focus of the Thesis

In presenting the focus of the thesis, we begin by providing an introduction to the background behind collaborative sensor activation for hazard detection (CHD) and then outline the challenges in CHD that are addressed in this thesis.

1.2 Background

In recent years, special interest has been shown in using cross-disciplinary techniques in the networking arena. Bio-inspired and physics-inspired algorithms have been known in computer communication for a long time (e.g., genetic and gossip algorithms, quorum sensing, epidemiological models, graph and percolation theory, and self-organisation). They have recently become topics of intensive research as means of improving wireless sensor networks.

The message-flooding approach in wireless sensor networks is the easiest way of activating sensors. However Khelil *et al.* [58] show by experimentation that flooding-based approaches perform badly in comparison to selecting nodes to propagate messages. Therefore it can be accepted that detection and message passing protocols play a dominant role in efficient and reliable data fusion in wireless sensor networks.

Bio-inspired epidemiological models follow a similar approach to selective flooding, but sensors follow a localized way to inform other sensors. Therefore we can expect that an epidemiological model based on susceptible, infectious and susceptible (SIS) [73] is better than a flooding protocol which consumes a vast amount of energy by messaging unwanted sensors.

The analysis performed by a CHDS can typically be characterized as either misuse detection or anomaly detection: misuse detection searches for the known signature of malicious activity, namely a set of events that match a pre-defined signature of the traffic generated by a hazardous attack. In contrast, anomaly detection aims to model ‘normal’ detection in a network, and then detect any abnormal activity that does not match this model of normal behaviour.

While CHDS play an important role in providing an extra layer of defence to the security of the environment, there are still a number of ongoing challenges that need to be addressed by hazard detection in practice:

False alarms : CHDS platforms run the risk of overloading network management staff if they produce too many false alarms. Previous work from IBM Research observed that over 90% of alarms generated by a CHDS were caused by a few dozen persistent root causes. Such a flood of false alarms makes it very difficult to identify the cause of true alarms and discourages network managers from monitoring the output of a CHDS.

Irrelevant alerts : These are alarms generated by real attacks, which are not a threat because of the configuration of the target systems. For example, when a worm that is designed for the Windows operating system attacks a Linux platform, alerts will be generated by a CHDS that monitors the network. However, these alerts are

considered irrelevant since there is no threat to the target system. (e.g., aerosol sprays, perfumes, etc.)

1.3 Overview of the Problem

In this thesis, we address the problem of detecting CBRN hazards. These type of hazards involve attacks originating at multiple locations, or targeting multiple target destinations. In order to address the problem of detecting large-scale distributed hazardous attacks, we propose an approach based on a collaborative detection by a system comprising distributed sensor nodes (a CHDS). A CHDS consists of a set of individual hazard detectors (IDs) each monitoring a separate administrative geographical locations. These IDs cooperate to detect coordinated attacks by sharing evidence of locally observed attacks. Each ID reports any alerts of suspicious behaviour that it has collected from its local monitored network. The CHDS then correlates these alerts to identify coordinated attacks that affect multiple sub-networks.

In general, the traffic flow from hazardous attacks can originate either from outside the monitored sensor network environment (i.e., the monitored environment is being attacked), or from inside the monitored environment (i.e., the monitored environment is the source of the attack). While the CHDS approach that we propose is generally applicable to both cases. We focus on studying the underlying correlation properties of the hazards for the following reasons : (1) in the early stages of many coordinated hazard attacks, the high intensity source is the entity that is attacking the monitored environment, such as an external attacker scanning for vulnerable weakly monitored locations; (2) the study of correlation characteristics based on the hazard environment is a common practice among the collaborative intrusion detection community;

In this thesis, we address two open research challenges that are raised by such CHDS.

1. Information reliability: The aim of studying correlation in a hazard environment is to find its effect on the activation of distributed sensors. These correlation patterns can be based on the environmental factors, such as temperature, air pressure

and wind. In general, there is a combinatorial number of possible patterns that can be found by a CHDS. In practice, we need to constrain the reliability of information. An important research problem that we address in this thesis is how to design CHDSs that can support increasing levels of efficient information dissemination in their collaborative detection algorithm, while still being computationally efficient.

It is interesting to explore the events associated with past CBRN releases. When Sarin gas was released in the Tokyo railway station in Japan, thousands of people were affected because there were no proper systems in place. In Melbourne comparable events happened (without high fatalities) without the chemicals being deliberately released. One instance was in Cranbourne and another in Port Melbourne, where odours were detected around the area manually and their source, direction of travel and speed were determined subsequently.

In the biological world, the insects such as bees can detect the fragrance from a flower and identify the source using the biological sensors. Their sensory faculties are very stable after being refined through evolution. When we consider humans, there are five senses (i.e., vision, hearing, smell, touch, taste), and all discoveries and assumptions are built on these senses as were identified by ancient wisdom. In Buddhism, these sensing phenomenon are well explained (i.e., ‘Rupa’ or images, ‘Wedana’ or sensing, ‘Sandgha’ or signals, ‘Sankara’ or signal patterns, ‘Widjana’ or knowledge based on identified signal patterns). When we visualize something, we feel it and a signal is generated, and the brain identifies the visual object. In the modern world, multi-sensor data fusion behaves as per the principals identified thousands of years ago.

In any CHDS, a key issue is the trade-off between the detection rate for attacks versus the false alarm rate. In the context of finding a trade off between the expressiveness and scalability of a CHDS, we also need to take into consideration the extent to which different models of distributed correlation restrict the sharing of evidence within the CHDS, and hence how this affects the overall accuracy of the CHDS. In this thesis, we address the problem of how to achieve a high level of accuracy in

a CHDS, while limiting the communication and computational overhead within the system.

2. Energy conservation: Given that large-scale coordinated attacks can potentially span thousands of independently monitored sub-networks, we need to ensure that our CHDS can address the problem of scalability, because a centralized CHDS architecture may only be able to support limited use of correlation features for detection of hazardous attacks.

The interaction between these two challenges of efficiency of information dissemination and energy conservation provide the focus for our research on collaborative hazard detection in this thesis. Note that there are other issues raised by the design of a CHDS, such as expressiveness, scalability, trust, security and privacy among participant sensor nodes. These issues have already been the subject of research in the distributed computing community. In this thesis, we consider the case of a CHDS in which the participant sensor nodes are under the control of reliable network operators on a secure platform. This is often the case within a carrier's network or between network carriers. The problem of how to operate a CHDS when these assumptions of trust and security are relaxed, i.e., when a CHDS may be under the control of an attacker, are outside the scope of this thesis. In the next section, we provide an outline of the approach that we have taken to address these research problems.

1.4 Outline of Proposed Solution

The balance between energy consumption and detection capabilities is vital in designing sensor networks for CBRN threats. The statistical characteristics of CBRN tracers and the placement of sensors may also influence their detection reliability. Our aim was to develop bio-inspired algorithms for reliable, efficient, and scalable source localisation, detection, and information extraction while minimising energy consumption.

As such, we investigated bio-inspired data fusion techniques to enhance situational

awareness of emerging threats in a CBRN environment, such as on a military battlefield or in civilian counter-terrorism applications. This may involve utilising data from a CBRN-specific sensor network to enable the early warning of threats, source term estimation and refine hazard prediction.

To establish the foundation of our research study, we explored several areas: data fusion in relation to radiological, biological and chemical agents [8], radiological source localisation, data collection, and networks of chemical sensors. We investigated different chemical contaminant models, which we refer to as non-correlated and correlated, based on the dispersion characteristics. We explored static source localisation in radiological fields, more complex chemical tracer fields, and time variant chemical tracer fields with special tracer distribution characteristics.

As the solution to the research problem we have selected to address, we considered a simple synthetic chemical environment and a more complex time varying synthetic chemical environment with spatial correlated chemical tracer distributions. Protocols were designed to enhance the information fusion which facilitated the efficient detection of correlated chemical contaminants while saving energy. As a step towards a viable solution, we designed a sensor network model with an energy efficient sensor activation protocol based on epidemiology. We then used the model to evaluate its response to a chemical threat. For experimentation, we oversaw the problem in three aspects: chemical contaminant characteristics, wireless sensor network modelling and a protocol based on epidemiology and gossip. We extended the model in [94], which uses non-correlated chemical pollutants, to a model with correlated chemical contaminants. We compared the performance of WSN dynamic collaborative properties with that of the non-correlated model. The simulation was done using a network with a ‘grid’ topology. Later we selected radiological sources which we assumed to have static locations. Then We identified a reliable bio-inspired algorithm to identify the possible radiological threat locations.

We identified the information dissemination strategies and information utility measures introduced in [107]. Epidemiological DSC-based approaches were well dis-

cussed in [34, 72, 23, 90] and [108]. Zou *et al.* [108] used this approach to model a worm propagation on the internet. Khelil *et al.* [58] and Skvortsov *et al.* [94] analytically show that DSC-based algorithms involve continuous estimation of the state of each sensor in the network and require computationally intensive simulations. In modelling, they use simple differential equation-based models. Graph theory-based modelling approaches are introduced in [29] and [25] of the papers we reviewed.

The papers [11], [34] and [94] served as a good foundation for our study. In our research, we adopted a simple network model with a suitable energy efficient detection protocol that was an extension of the model introduced in [94]. Our challenge was to adapt the environmental model from [11] to a realistic external challenge for initial simulations. Later we adapted a time variant chemical model.

The basic approach that we take in this thesis to the problem of collaborative hazard detection is bio-inspired algorithms for collaboration. Correlating CHDS data across multiple network domains is the common practice in both industry and the academic research community. Publisher-subscriber models are widely used throughout the literature for tasks such as event notification, mobility support services and in the network message service. In our context, when a participant sensor node detects a possible attack in its monitored sub-network, it generates an alert, which is reported to the CHDS. This is known as *subscription*, i.e., the individual sensor node is registering its interest to the CHDS to confirm whether the alert is part of a large-scale coordinated attack. The role of the CHDS is to correlate alerts that are subscribed to participating individual sensor nodes. If enough subscribed alerts are received to confirm an attack, then the CHDS publishes a notification of a confirmed attack to the participating sensor nodes that subscribed to the attack. In this way, the local sensor node can take local network conditions into account in identifying suspicious flows from the raw network traffic, while the CHDs provides an additional analysis layer to identify patterns of alerts across multiple sub-networks.

In this thesis, we consider three main types of algorithms for hazard detection. We first consider the case of a non-correlated hazardous environment, where alerts

are not correlated on the basis of single intensity blurs. In this case, we present an algorithm to optimise the detection parameters for the non-correlated environment to maximize detection accuracy. We then consider the more general case of multi-dimensional correlated sources, where patterns of hazardous blurs can be found based on multiple environmental features. Finally we present algorithms to detect point hazardous sources.

We can classify the different approaches that we take to distributing computation in a CHDS into three collaboration models. The first approach is centralized collaboration, in which all correlations are performed at a centralized node. Alerts are subscribed to the centralized node by participating sensor nodes. All alerts are correlated at the centralized node, which notifies the relevant sensor nodes of any confirmed attacks. This centralized collaboration model should have the highest overall accuracy, as all information is available at a single location. However, the centralized correlation node can limit the scalability of a large CHDS, especially if we consider more computationally intensive correlation algorithms that support more expressive patterns of alerts.

Our second approach is the use of a collaborative hazardous environment. In this approach, some computation can be performed locally by the participating sensor nodes, so that not all alerts need to be subscribed to the central node. This can reduce the computational load of the centralized node, in order to support more sophisticated algorithms that can be used to find more expressive pattern alerts. However, there is a potential trade-off in terms of accuracy, since some alerts may be prematurely filtered at the local correlation stage.

Our third approach is to eliminate the need for a centralized collaborative node, so that the collaborating work can be distributed between the participating sensor nodes in a decentralized manner. In particular, we support a peer-to-peer (P2P) communication scheme in this approach. For this to work in a scalable manner, we require a method of routing subscribed alerts automatically to the responsible peer for responsive action, so that peer's do not need to keep track of which powers

are responsible for which attack instances. We achieve this by using P2P content-based routing overlay network between the participating sensor nodes. This raises the question of how to map the alert correlation task into this P2P overlay. While the elimination of a central node for collaborative sensor activation enables greater scalability in terms of distributing the computation load, the content based routing overlay network introduces a routing delay to the system. A key issue is whether the reduction in correlation delay outweighs the increase in communication and routing delay on wide-area networks in practice.

1.5 Organisation of the Thesis

The key concerns of this thesis are detection in chemical environments, hazard localisation in radiological environments and using bio-inspired algorithms. In detail, the chapters in the thesis are organised as follows.

Chapter 2 - A Survey of CBRN Attacks and Bio-inspired Algorithm-based Collaborative Detection, analyses current practices of data fusion and bio-inspired algorithms.

Chapter 3 - Epidemiology based dynamic sensor collaboration protocol, presents a numerical verification of a simple population and physics based model for dynamic collaboration in a network of chemical sensors.

Chapter 4 - Effect of chemical tracer distribution properties on epidemiology based sensor activation protocol, shows the statistical characteristics of chemical tracers (temporal and spatial correlations, etc.) and the placement of chemical sensors and their importance in WCSN design and performance.

Chapter 5 - Gossip inspired sensor activation protocol for a wireless chemical sensor network, evaluates the behaviour of gossip based sensor activation protocols in a chemical tracer environment.

Chapter 6 - Bio-inspired algorithms for radioactive source localisation, describes source localisation in radiological environments. This chapter considers localisation of point sources of gamma radiation using dose rate measurements. Binary

and continuous genetic algorithms (GA) were used to implement maximum likelihood estimation (MLE) of the position and strength of point radiation sources.

Chapter 7 - Conclusions, draws conclusions and suggest future work.

1.6 Contributions of the Thesis

This thesis makes several contributions towards identifying reliable bio-inspired algorithms to detect and localise hazardous threats in three areas: radiological source localisation, detection of chemical attacks with an epidemiological-based sensor activation, and using gossip-based protocols to improve the detection performance. The major contributions are:

1.6.1 Chapter 2

1. We conduct an extensive survey of CBRN Attacks and Bio-inspired Algorithm-based Collaborative Detection
2. We categorise the existing CHDS approaches.
3. We highlight the limitations of existing CHDS.

1.6.2 Chapter 3

1. we present a numerical verification of a simple population and physics based model for dynamic collaboration in a network of chemical sensors.
2. Development of an object-oriented multi-agent [37] simulation framework comprising objects for environment, sensors and network.
3. Further improvement of the wireless sensor network model for adaptability and realistic behaviour, optimising behaviour using evolutionary techniques we used in our previous studies [65], and adopting other chemical contaminant models introduced in [11].

1.6.3 Chapter 4

1. We investigate the effect of chemical tracer field properties on wireless sensor network performance
2. We evaluate the effect of statistical characteristics of chemical tracers and placement of sensors on WCSN design and performance. We introduce a metric measure spatial correlation of a chemical tracer field.
3. We perform simulations using different WCSN topologies and spatially correlated chemical tracer fields to evaluate the WCSN performance with the epidemiology based dynamic sensor activation protocol.

1.6.4 Chapter 5

1. We introduce a gossip based sensor activation protocol to overcome effects encountered with the epidemiology based protocol.
2. We evaluate the performance of the gossip-based sensor activation protocol on selected performance metrics: the sensor activation and chemical tracer detection.
3. We use simulations to show the benefits of the gossip based protocol in overcoming the effect of spatial correlation of the chemical tracer fields on the WCSN performance.

1.6.5 Chapter 6

1. We present research involving bio-inspired genetic algorithm to localise of radiological point sources.
2. We compare the genetic algorithm with the gradient descent method which is an unconstrained nonlinear optimization technique. We implement maximum likelihood estimation (MLE) of position and strength of point radiation sources.

3. We use real experimental data collected during a DSTO-conducted field trial to test the performance of the algorithms.

1.7 Publications leading to the thesis

The following publications were generated during the course of the research study and incorporated into the thesis.

1.7.1 Chapter 3

- S. Karunasekera, A. Skvortsov, A. Gunatilaka and C. Mendis, ‘Decentralized Dynamic Sensor Activation Protocol for Chemical Sensor Networks’, in Proceedings of the 9th IEEE International Symposium on Network Computing and Applications (NCA), 2010. IEEE Comp. Soc., July 2010, pp. 218-223.
- C. Mendis, ‘Wireless chemical sensor network simulation : A software framework’, in Proceedings of the IEEE Australia New Zealand Student Congress, ser. ANZCON 2010.

1.7.2 Chapter 4

- C. Mendis, A. Gunatilaka, A. Skvortsov and S. Karunasekera, ‘The effect of correlation of chemical tracers on chemical sensor network performance’, in Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Brisbane, Australia, 2010.
- C. Mendis, A. Skvortsov, A. Gunatilaka and S. Karunasekera, ‘Performance of wireless chemical sensor network with dynamic collaboration’, in Sensors Journal, IEEE, vol. 12, no. 8, pp. 2630-2637, Aug. 2012.

1.7.3 Chapter 5

- C. Mendis, ‘Gossip inspired sensor activation protocol for a correlated chemical environment’, in Bio-Inspired Models of Networks, Information, and Computing Systems of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 103, pp. 163-170, 2012.

1.7.4 Chapter 6

- C. Mendis, A. Gunatilaka, B. Ristic, S. Karunasekera and A. Skvortsov, ‘Experimental verification of evolutionary estimation algorithms for radioactive source localisation’, in Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 2009.
- A. Gunatilaka, B. Ristic, C. Mendis, S. Karunasekera and A. Skvortsov, ‘The effect of data collection geometry on radiological source localisation’, in Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 2009.

Chapter 2

A Survey of CBRN Attacks and Bio-inspired Algorithm-based Collaborative Detection

2.1 Survey of CBRN Hazardous Attacks

2.1.1 Chemical hazard detection

The external challenge or the chemical contaminant model in our case is a very important issue. The protocol should be adapted to the environment to reap maximum benefit from the detection network. Borgas *et al.* in [11] and [9] present models for chemical fume propagation and behaviour. These models are substantiated with experimental results. In our study we accept it as the foundation for chemical modelling. Interestingly, Bisignanesi and Borgas propose a multi-agent based pest-tracking model. Gunatilaka *et al.* [44] use a similar model to estimate the strength (emission rate) and the location of a chemical source. We accepted Borgas *et al.* in [11] as a foundation study on external challenges and adapted the model for our initial simulations.

2.1.2 Radiological hazard detection

Numerous incidents involving loss or theft of radioactive sources have been reported [79].

There is major concern that such sources could be acquired by terrorist networks and be used to build radiological dispersion devices. Therefore, it is of interest to be able to locate hidden radioactive sources. Gamma rays, which are highly penetrating electromagnetic radiations emitted by some radioactive material, can be detected at long stand-off distances, and measurements collected using gamma detectors can be used to detect sources of radiation. Assuming that the existence of a gamma radiation source has been established, we considered the problem of estimating the location and strength of the radiation source using dose rate measurements.

The detection and localisation of point-sources of gamma radiation has been studied recently by several authors [76] [96] [13] [82]. For example, the problem of detecting radiation sources carried in a vehicle by using a distributed sensor network placed along a section of a road is considered in [76, 96] and [13]. The maximum likelihood estimation (MLE) implemented using the *fminsearch* routine in MATLAB[®], extended Kalman filter (EKF) and unscented Kalman filter (UKF) were applied to the problem of localising point radiation sources in [42]. Sheng and Hu had used MLE in a wireless sensor network using acoustic energy measurements [91]. The MLE provided the best source estimates, asymptotically approaching the theoretical bound predicted by the Cramer Rao bound. UKF results were worse than those of MLE and the EKF failed to produce acceptable estimates. Rao *et al.* [82] propose localisation methods, mean of estimators (MOE) and mean of measurements, and compare them using simulations. Chin *et al.* [20] propose iterative pruning (ITP) to localise a single point radiation source, which is claimed to be more accurate than the MOE and more efficient than

the MLE. Although a *fminsearch*-based implementation of the MLE worked well for localising one or two sources, it was found to be unsatisfactory for locating three or more sources [71]. A progressive correction-based Bayesian approach is also used to localise multiple point sources in [71]. This research study explores the use of genetic algorithms (GA) for the localisation of multiple radiological point sources.

The networks have been modelled using simpler mathematical models and also using graph theory based approaches [30] [29]. The percolation of closed systems has been explained in [95] while self organising wireless sensor networks were introduced in [67]

2.2 Hazard Detection Systems

The Sensor Network Area Protection System (SNAPS) presented in [51] can monitor and predict the pattern of CBRN materials in a 10 mile radius. This surveillance takes place from a mobile command trailer where law enforcement personnel can track suspicious material during events with thousands of people in attendance. This counter-terrorism device is revolutionizing the way law enforcement and first responders deal with potential critical and mass disasters [51].

The question regarding CBRN wireless sensor network capability is whether this technology is suitable, reliable, user-friendly and quickly deployable. Furthermore, will this technology provide critical early-warning, detection and subsequent notification in real time? The goal of this thesis is to determine CBRN wireless sensor detection capability in terms of reliability, deployment, early-warning and notification. The objective is to outline a concept of operations document providing the need-structure for incorporating wireless sensor detection capability into public-safety

operations [75].

The aim of the response guidelines is to establish procedural guidelines for mid-level strategic/tactical planners responsible for CBRN preparedness and response. The response guidelines provide generic advice and guidance on procedures, capabilities and equipment required to implement an effective response. They are designed to improve multi-agency interoperability in first response to a CBRN incident and provide guidance on when regional, national or international assistance may be required. The guidelines have been prepared to help planners in EAPC nations determine their own level of capability through self-assessment. They serve as a checklist. Implementation of the guidelines is entirely optional [74].

CBRN are weaponized or non-weaponized Chemical, Biological, Radiological and Nuclear materials that can cause great harm and pose significant threats in the hands of terrorists. Weaponized materials can be delivered using conventional bombs (e.g., pipe bombs), improvised explosive materials (e.g., fuel oil-fertilizer mixtures) and enhanced blast weapons (e.g., dirty bombs). Non-weaponized materials are traditionally referred to as Dangerous Goods (DG) or Hazardous Materials (HAZMAT) and can include contaminated food, livestock and crops [17].

There are several existing databases that describe the incidence of terrorist activity. Perhaps the most well-known of these is the International Terrorism: Attributes of Terrorist Events (ITERATE) database. This data was originally compiled by Mickolus (1982) and spanned a period from 1968 to 1977. Included in the data are the date of attack, the type of event and the number of casualties (deaths or injuries), as well as various other characteristics of incidents. There are roughly 12,000 incidents reported in the ITERATE database, spanning a 24-year period from 1968 to 2001.

Importantly, however, only transnational terrorist events are considered. ITERATE defines an event as transnational along several criteria. These include the nationalities of the perpetrators, the location of the attack and the nature of the target [69].

Wireless sensor networks for hazard detection

There are many attributes that affect the detection of chemical tracers by a network of chemical sensors, such as the detection threshold of sensors, the communication range, the number of sensors, the sensor density, the connectivity, the sensor placement or topology and the sensor scheduling.

As mentioned in the previous section, for a wireless network of chemical sensors, the energy constraints often become the primary consideration [3], because the network design [2] and the development of operational protocols is driven by a trade-off between two opposing imperatives; to conserve energy and to facilitate reliable event detection. The first imperative is to reduce the active (sampling) time of all sensors in the network in order to save energy and thus prolong the network lifetime. Most modern chemical sensors contain an air sampling unit (fan), which turns on when the sensor is active, thereby using most of the battery. Therefore the need to activate the sensors only when necessary is a major concern. The other imperative is 'information' driven, which necessitates the incremental increase of the sensor sampling time and sampling volume to facilitate reliable detection of any intermittently distributed chemical contaminant [43] [5]. The latter, evidently, requires an increase in the energy supply.

In [33], Ertin *et al.* propose a dynamic sensor selection method for Bayesian filtering problems in sensor networks. This is an extension of the work done in [107]. The method is supported by proofs. As evidence of viability, they provide three cases

with empirical results. These are discrete observations, the Linear Gaussian model and acoustic array.

Dekker and Skvortsov [27] explore the use of different network topologies for a simple chemical contaminant model with non-spatial dynamics. A similar approach is presented in [72]. This information gain is an ideal metric to be evaluated against other performance criteria of the sensor system (e.g., energy constraints or detection time) to find optimal network parameters for a scenario.

In selecting a sensor network, the nature of sensors and their degree of connectivity is also important. In [83], the authors consider the connectivity of large-scale ad-hoc heterogeneous wireless networks. Secondary users in these networks solicit communication channels temporarily unused by primary users. A communication link exists between two secondary users that depends not only on the distance between two users, but also on the transmitting and receiving activities of nearby primary users. The authors establish a necessary condition and a sufficient condition for connectivity, which leads to an outer bound and an inner bound on the connectivity region.

Wireless sensor networks have been commissioned, but with a very small number of sensor nodes due to financial and physical constraints. Further research using physical networks seems to be limited [16]. Newer compact chemical sensor technologies, such as GUSTO, have high accuracy and throughput and are cheaper than conventional monitoring stations. Richards *et al.* in [84] point out if melioration of forecasting accuracy is notable with the application of spatiotemporal models, it will make deployment of many sensors economically justifiable. As the number of sensors increases, however, the design and management of such systems become challenging tasks that require mathematical modelling, computer simulation and parameter

optimisation. There are various approaches in designing wireless sensor networks. Khelil *et al.* [58] follow an analytical mathematical model-based approach and Paellegini *et al.* [25] follow a random graph theory-based approach. Akkaya and Younis [1] also analyse protocols used in WSN. Khelil *et al.* [58] show by experimentation that flooding-based approaches display poor performance in comparison to selecting nodes to propagate the messages. Therefore it can be accepted that detection and message-passing protocols play a dominant role in efficient and reliable wireless sensor networks.

The chemical pollutant in this scenario is assumed to be a static distributed source, but in reality it is dynamic and three-dimensional. The current field is considered to be a flat terrain without obstacles, which is a highly simplified version of reality. We are hoping to incorporate more realistic models in future simulation frameworks. The sensor network should have a dynamically changing composition, where new sensors should be introduced and malfunctioning sensors removed. Real-world epidemic sensor clusters can be considered with similar properties. Further sensors can incorporate buffers to store information until communication. Energy is ignored at this stage, but will be incorporated in future models. Communication contentions, frequency and other properties of a communication network need to be considered as well. In this study we considered epidemiological modelling, assuming nodes to be homogeneous, which is a simplified representation of the actual scenario: in reality networks divide into ‘clusters’ or ‘communities’. These clusters are tightly-knit groups with dense connections among nodes. The other phenomenon is ‘assortative mixing’: vertices in the network having various properties associated with them. Another explanation is that a network is said to show assortative mixing if the nodes in

the network that have many connections tend to be connected to other nodes with many connections. The important feature of network research is to identify not only the similarities between networks, but also their most important differences. In social networks, individuals can find short paths to others, despite having only very limited information about the structure of the network. This property of network ‘searchability’ depends on the ability of individuals to categorize themselves and each other into socially meaningful groups. This is called ‘social distance’ [77].

Optimization in wireless sensor network design

In general, optimisation of WSN is aimed at ensuring the longevity and sustainability of the network system in the long run. This can be achieved by the simple reduction of the active time of individual sensors (scheduling). Since wireless networks, unlike the wired, are powered by batteries, the associated optimisation problem sometimes becomes even more important to meet the constraints of the available energy supply.

In [19], the authors say that many new biological and chemical sensors have been or are continuously being developed for infrastructure and environmental protection (e.g., for protecting the quality of water and indoor and outdoor air). However, there is a lack of fundamental system-level research leading to the development of sensor networks that both maximize protection and minimize the system cost for indoor air protection. Four key parameters are usually used to evaluate sensor performance: sensor sensitivity, probability of correct detection, false positive rate and response time. The optimal design of a sensor system is affected by the above sensor performance parameters. This research study describes a preliminary study to:

1. identify simplified simulation and optimization strategies that can be used for sensor system design;

2. examine the relationships between sensor location, sensitivity and quantity, and
3. use both detection time and total occupant exposure as optimization objective functions for sensor system design.

Common building-attack scenarios, using a typical chemical and biological warfare (CBW) agent, are simulated for a small commercial building. A genetic algorithm (GA) is then applied to optimize the sensor sensitivity, location and quantity, thus achieving the best system behaviour while also reducing the total system cost. Assuming that each attack scenario has the same probability of occurrence, optimal system designs that account for the simulated possible attack scenarios are obtained.

Mendis *et al.* [65] use a moving sink to gather information from sensors and devise an optimum path for the sink using an evolutionary computational method based on particle swarm optimization(PSO). In our studies, we investigated the issues in the placement of sensors for the localisation of sources using optimisation techniques.

2.3 Bio-inspired algorithm-based Collaborative Detection Systems (BICDS)

We see many innovations in science and technology where we try to subdue natural processes to our advantages, but the natural world around us has variety, adaptability and sophistication which we have not yet explored [10]. Computational scientists continue to create faster more efficient algorithms and hardware that exhibit centralized control or to place less emphasis on speed and efficiency than on robustness, adaptability and emergent organisation from the interaction of many loosely coupled processes, which we call ‘bio-inspired computing’. Figure 2.1 shows one such instance of the complexity of nature.

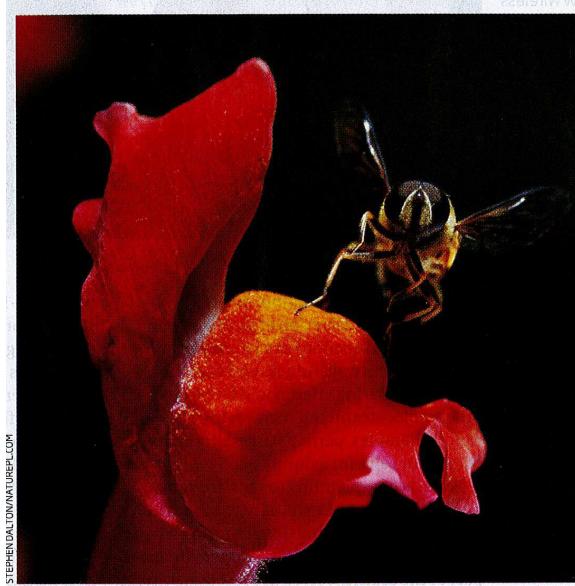


Figure 2.1: Mutant snapdragon flowers revealed the secrets of gene expression (Courtesy New Scientist)

Nature has many complex processes where it converges to a stable-state interaction with the environment. The fascination is the evolution of biological entities with time. Several scientists (such as Darwin and Mendel) have tried to unravel the mysteries of nature by postulating theories whose work can be summarised as follows,

- Evolution operates on chromosomes and does not operate on organisms. Chromosomes are the organic tools by means of which the structure of a certain living being is encoded; that is, the features of a living being are defined by the decoding of a collection of chromosomes. These chromosomes (or more precisely, the information they contain) pass from one generation to another through reproduction.
- The evolutionary process takes place precisely during reproduction. Nature

exhibits a plethora of reproductive strategies. The most essential ones are mutation (that introduces variability in to the gene pool) and recombination (that introduces the exchange of genetic information among individuals).

- Natural selection is the mechanism that relates chromosomes with the adequacy of the entities they represent, favouring the proliferation of effective, environment-adapted organisms, and conversely causing the extinction of less effective, non-adapted organisms.

These principles are comprised within the most orthodox theory of evolution, the Synthetic Theory. Although alternative scenarios that introduce some variety in this description have been proposed (for example, the Neutral Theory and, very remarkably, the Theory of Punctuated Equilibria) it is worth considering the former basic model. Despite the apparent simplicity of the principles upon which it rests, Synthetic Theory reveals the unparalleled power of nature in developing and expanding new life forms [78].

The brain is considered to consist of a grid-like lay-out. Similarly, applications in a computer grid are considered to be an ideal topology for monitoring where sensors are placed uniformly covering a field. In regard to the geometric placement issues, we say random placement is considered to avoid mirror-type detection of sources. Turin and Neumann had tried to compare capacity of real artificial intelligence to mimic the processes in the brain, but hitherto the exercise has failed. In our attempt to use bio-inspired processes in engineering applications, we feel that we have achieved to a greater extent the proposition of reliable bio-inspired techniques. Using source localisation, sensor placements, sensor activation using a bio-inspired epidemiological protocol [85] [24] [97] and then a gossip protocol, we show that our results converge

to a reliable value.

WSNs are networks of distributed autonomous devices that can sense or monitor physical or environmental conditions cooperatively. WSNs face many challenges, mainly caused by communication failures, storage and computational constraints and limited power supplies. Paradigms of computational intelligence (CI) have been successfully used in recent years to address various challenges, such as data aggregation and fusion, energy aware routing, task scheduling, security, optimal deployment and localization. CI provides adaptive mechanisms that exhibit intelligent behavior in complex and dynamic environments like WSNs. CI brings about flexibility, autonomous behavior, and robustness against topology changes, communication failures and scenario changes. However, WSN developers are usually unaware or insufficiently aware of the potential that CI algorithms offer. On the other side, CI researchers are not familiar with all real problems and subtle requirements of WSNs. This mismatch makes collaboration and development difficult. This paper intends to close this gap and foster collaboration by offering a detailed introduction to WSNs and their properties. An extensive survey of CI applications to various problems in WSNs from various research areas and publication venues is presented in the paper. A discussion on advantages and disadvantages of CI algorithms over traditional WSN solutions is also offered. In addition, a general evaluation of CI algorithms is presented, which will serve as a guide for using CI algorithms for WSNs [61]. In [109], Zungeru *et al.* also survey several swarm intelligence based routing protocols for WSN.

A generalization of the sampling method introduced by Metropolis *et al.* [66] is presented, along with an exposition of the relevant theory, techniques of application

and methods and difficulties of assessing the error in Monte Carlo estimates. Examples of the methods, including the generation of random orthogonal matrices and potential applications of the methods to numerical problems arising in statistics, are discussed [47].

In this they present the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is one of essential equations or, more generally, one of integral-differential equations that occur in various branches of the natural sciences [66].

A major limitation on the more widespread implementation of Bayesian approaches is that obtaining the posterior distribution often requires the integration of high-dimensional functions. This can be computationally very difficult, but several approaches short of direct integration have been proposed (reviewed by Smith 1991, Evans and Swartz 1995, Tanner 1996). We focus here on Markov Chain Monte Carlo (MCMC) methods, which attempt to simulate direct draws from some complex distribution of interest. MCMC approaches are so-named because one uses the previous sample values to randomly generate the next sample value, generating a Markov chain (as the transition probabilities between sample values are only a function of the most recent sample value) [103].

There is a deep and useful connection between statistical mechanics (the behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature) and multivariate or combinatorial optimization (finding the minimum of a given function depending on many parameters). A detailed analogy with annealing in solids provides a framework for optimization of the properties of very large and complex

systems. This connection to statistical mechanics exposes new information and provides an unfamiliar perspective on traditional optimization problems and methods [59].

Epidemiology

Epidemiology-based information driven dynamic sensor collaboration(DSC) provides a flexible approach that often allows the accommodation of these opposing criteria of saving energy and information gain, and the provision of an optimal protocol for a given scenario.

Eubank *et al.* [34] provide a comparison between epidemiology and wireless communication. The introduced approach is based on the known analogy between the information spread in a sensor network and the propagation of the epidemic across a population. They then compare and contrast epidemiology and wireless communication. They introduce a critical value R_o , known as the reproductive number. The parameter is important for predicting the reaction of networks.

Eubank *et al.* [34] further suggest that the relationship between disease transmission and message transmission is a metaphor rather than an analogy. Interactions among people that lead to disease transmission are accepted as being homogeneous. In the real world, more complicated correlation structures, such as demographically determined compliance rates, can be represented just as easily, as long as the data is available. This may be as simple as a finite state machine distinguishing the susceptible state from the infectious state.

The primary goal of algorithms/protocols for communication networks is rapid information diffusion based solely on local information. Their effectiveness can be measured by the total amount of information that is moved into the network to

accomplish the task, the time it takes for updates to be complete, the kind of global information assumed and the environmental constraints.

Eubank *et al.* [34] raise questions regarding the epidemic nature of a disease:

1. How many hosts will have been infected when the outbreak dies out ?
2. How many hosts are infected at any given time ?
3. Which hosts are vulnerable ?
4. Which hosts are critical ?

They suggest compartmental models as the answer to the first and second issues, but they are unable to address the third and fourth. The authors further raise issues on the number of infectious hosts, the number of hosts infected at a given time, the number of hosts vulnerable and number of hosts critical.

Two models, the Barabasi-Albert (BA) and Watts-Stogatz (WS), are considered in [72]. BA is a highly heterogeneous network and WS is a homogeneous network. They define a simple epidemic data dissemination model, which is a modified version of the Daley and Kendall model [23].

Zhao *et al.* [107] present a DSC-based approach, in which a sensor in the network is activated only when there is an information gain associated with the measurements in close proximity to it. This method is presented with theoretical background and experimental results and has attracted a large number of citations.

In [94], a physics-based model of a wireless network of randomly-located chemical sensors is introduced, in which dynamic sensor collaboration (DSC) is driven by the concentration of a contaminant (the external challenge) on each individual sensor. This simple model can describe an information epidemic (IE) in a sensor network.

It can also express the relationships between the parameters of the network (e.g. number of sensors, their density, sensing time), the network performance parameters (probability of detection, response time of the network) and the parameters of the external challenge (environment, contaminant).

In [90], a model for epidemics on an adaptive network is considered. The study considers patterns where nodes follow a susceptible-infective-recovered-susceptible (SIRS) model. It claims that the introduction of rewiring affects both the network structure and the epidemic dynamics. Further more, the authors claims that the rewiring leads to regions of bi-stability where either an endemic or a disease-free steady state can exist. When fluctuations around the endemic state and the lifetime of the endemic state are considered, the fluctuations have been found to exhibit power law behaviour.

These physics-based models are used to describe the network resilience, to estimate the coverage area, or to simulate ‘peer-to-peer’ communication over the internet [72]. These models are commonly used to describe worm propagation in computer networks [108], connectivity properties of ‘virtual’ (random) networks [58] and similar issues. Some publications contain rigorous derivations of these models [29], but very often they are successfully applied based on a semi-empirical or purely heuristic approach [108] by using the visual similarity of ‘breakthrough’ curves responsible for the observed variables (number of infected nodes, fraction of coverage area, transmission radius, etc.). Similarly several studies have been done in evaluating these techniques. In [67], Mills also present a review of self-organisation in WSNs.

Localisation and detection of hazardous threats from chemical, biological, radiological and nuclear sources are very complex tasks and involve NP-based computation.

We have put forth a systematic survey of the state-of-the-art threat detection research. While progress has been made on the problem of threat detection and localisation, there still remain a range of open issues that need to be addressed.

2.4 Hazard detection and Environmental properties

The effect of the correlation structure of the environment on WSN architecture and its operational protocol has been actively investigated, and the results are well-documented in the literature (see [98], [102], and [105] and references therein). Many of these studies are concerned with data redundancy, aggregation, and energy consumption of WSNs. The investigation of these considerations in WCSNs has been started only recently; the research study by Ma *et al.* [62] is one such instance. The reason for this relative lack of studies of WCSNs is two-fold. Firstly, a reliable chemical sensor is a complex technical device (which often contains an air-sampling unit), so development of WCSNs with an appreciable number of sensors has been considered as impractical until recently. Secondly, high fidelity statistical modelling of tracer distribution in turbulent flow (consistent with the results of the theory of turbulent mixing and dispersion) is a challenging task and still an area of active research [92]. A full-scale implementation of such models still requires expert knowledge in computational fluid dynamics and advanced computer resources. As for simplified models, there are surprisingly few that can provide consistent temporal and spatial statistics, and this is still an area of active research as well [9, 7, 87].

Some of the first examples of simulating a WCSN operating in a turbulent environment are presented by Dekker and Skvortsov [27] and then by Skvortsov *et al.*

in [94]. There, the authors proposed a simple mathematical model for WCSN with an epidemiology-based activation protocol and a turbulent chemical tracer environment modelled by a random time series. Dekker and Skvortsov [27] investigated the effect of network topology on WCSN performance. Of six networks examined, they found a square grid and a network with short range random links to have the best performance. Skvortsov *et al.* [94] derived analytical expressions that relate the parameters of the network (e.g., number of sensors, their density, sensing time) with the parameters of network performance (probability of detection, response time of network) and the parameters of the external challenge (the chemical pollutant and environment). In physical implementations, Vito *et al.* [26] present polymer-carbon chemical sensors with inbuilt intelligence for distributed sensing and censoring transmission to reduce power consumption. This substantiates the feasibility of using an epidemiology-based protocol in intelligent sensor activation. While the analogy between epidemiology and a wireless communication operational protocol has become a topic of interest in intensive sensor network research and development activity in the recent past (see [34, 58] and references therein), due to aforementioned reasons, there is a notable lack of studies focused on WCSN.

2.5 Gossip Algorithm

A gossip algorithm was initially introduced by Demers *et al.* in [28] for replicating databases. Eugster *et al.* [35] have studied factors such as membership management in message transmission associated with the gossip based algorithm application. Rao *et al.* [81] have applied a gossip algorithm to peer-to-peer networking to extract aggregated information from the network. The studies of Haas *et al.* [45] and Boyd *et*

al. [12] are instances for the application of the gossip based protocols to wireless sensor networks. Kashyap *et al.* [54] propose a gossip-based data aggregation scheme but uses only one communication partner node. Wuhib *et al.* [104] investigate the use of gossip-based protocols for the continuous monitoring of network-wide aggregates under crash failures. They evaluate a gossip-based monitoring system against a tree-based monitoring protocol. The authors also consider sensor failures, but restrict their study to non-contiguous failures. They claim that the global average across all nodes is independent of the network size. In the literature, the use of the gossip-based algorithms is grouped into three types: information spread, aggregate computation, overlay management. The gossip based protocols are expressed in complex mathematical forms in the literature and deriving optimal parameters is not trivial.

In physical implementations, Gamm *et al.* [38] present a sensor with two power modes (sleeping and receiving transmission). This substantiates the feasibility of using a gossip-based protocol in sensor activation with existing sensor technology. We did studies on correlated and non-correlated chemical tracer fields, random and regular grid topologies.

The gossip-based algorithms have been applied to information dissemination in large-scale systems due to its simplicity in deployment, robustness and resilience to failures. It mimics the spread of contagious diseases in a way similar to the epidemiological protocol. In health related research, they are more focused on stopping epidemics rather than encouraging them. When applied to communication networks, the stress moves towards the better spread of information. The gossip-based protocols generally have the following properties: buffer capacity, relay count, fan-out

ratio. When gossip-based algorithms are employed, network overlay, message filtering and sensor node membership maintenance become an issue [6]. The gossip-based protocols are robust and have no centralised control. They are also the solutions for scalability and sensor change issues.

Chapter 3

Epidemiology based dynamic sensor collaboration protocol

In this Chapter¹, we present a numerical verification of a simple population and physics based model for dynamic collaboration in a network of chemical sensors. The modelling approach is based on the known analogy between the information spread in a sensor network and the propagation of epidemics across a population. In this framework we verify the derived analytical expressions which relate the parameters of the network (e.g., number of sensors, their density, sensing time, etc.), with the network performance parameters (probability of detection, response time of network) the parameters of the external challenge (the chemical pollutant) and the environment(turbulence). Using numerical simulations of wireless sensor networks with random, line, and circle topologies, we show that simulated and analytical results agree.

¹Presentations/Publications:

- S. Karunasekera, A. Skvortsov, A. Gunatilaka and C. Mendis, ‘Decentralized Dynamic Sensor Activation Protocol for Chemical Sensor Networks’,in Proceedings of the 9th IEEE International Symposium on Network Computing and Applications (NCA), 2010. IEEE Comp. Soc., July 2010, pp. 218-223.
- C. Mendis, ‘Wireless chemical sensor network simulation : A software framework’, in Proceedings of the IEEE Australia New Zealand Student Congress, ser. ANZCON 2010.

3.1 Introduction

Information-driven dynamic sensor collaboration (*DSC*) [33, 107, 65] provides a flexible approach that often allows to accommodate these opposite criteria and to provide an optimal protocol for a given scenario. In the DSC approach, a sensor in the network is activated only when there is an information gain associated with the measurements in its close proximity[107]. A similar approach is presented in [72]. This information gain is an ideal metric to be evaluated against other performance criteria of the sensor system (e.g., energy constraints or detection time) to find optimal network parameters. Usually, DSC-based algorithms involve continuous estimation of the state of each sensor in the network and require computationally intensive simulations [33, 107, 58].

In [94], a physics-based model of a wireless network of randomly-located chemical sensors is introduced, in which DSC is driven by the concentration of a pollutant (the external challenge) on each individual sensor. The introduced approach is based on the known analogy between the information spread in a sensor network and the propagation of epidemics across a population [34, 72, 108]. This simple model allows to describe an information epidemic (*IE*) in a sensor network and expresses the relationships between the parameters of the network (e.g., number of sensors, their density, sensing time, etc.), the network performance parameters (probability of detection, response time of network) and the parameters of the external challenge (environment, pollutant).

Initially, we verify the model of [94] by extensive numerical simulations and extend it to two other topologies (line and circle), which are derivatives of previously used random topology (see Fig. 3.1).

This Chapter is organised as follows. Section 3.2 describes the system modelling

and simulation framework. The simulation approaches are stated in 3.3. The results from simulations are presented in Section 3.4. Section 3.5 draws conclusions and suggests future work.

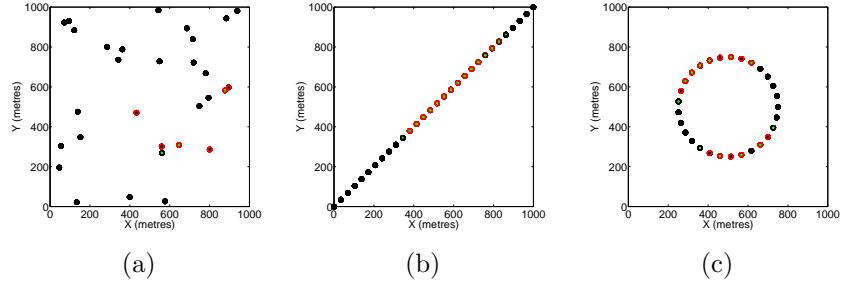


Figure 3.1: The three network topologies considered in our simulations: (a) random (b) line (c) circle. Black, green and red circles denote passive sensors, active sensors, and sensors that are contaminated, respectively.

3.2 System modelling and simulation

For the purpose of DSC system simulation, we developed a simulation framework comprising three independent models representing the environment (the spatiotemporal realisation of the concentration field of the chemical tracers caused by turbulent mixing), an individual chemical sensor node, and the whole wireless chemical sensor network (WCSN). This modular design of the simulation framework allows gradual improvements in any of these models. We validated our simulation model by comparing simulated results against an analytical model. The three component models of our simulation framework as well as the analytical model are discussed below.

3.2.1 The model of environment

The external challenges are modelled by random time series which mimic the turbulent fluctuations of concentration at each sensor of the network. In this approach the fluctuations in concentration C are described by its probability density function $\rho(C)$, with mean concentration of pollutant in the area, C_0 , as a parameter [9, 11]:

$$\begin{aligned}\rho(C) &= (1 - u)\delta(C) \\ &\quad + \frac{u^2(\gamma - 1)}{C_0(\gamma - 2)} \left(1 + \frac{u}{(\gamma - 2)} \frac{C}{C_0}\right)^{-\gamma}.\end{aligned}\quad (3.2.1)$$

Here the value $\gamma = 26/3$ has been chosen to make $\rho(C)$ compliant with the theory of tracer dispersion in Kolmogorov turbulence (see [9]), but it may vary with the meteorological conditions. The parameter u , which models the tracer intermittency in the turbulent flow, is in the range $[0, 1]$, with $u = 1$ corresponding to the non-intermittent case.

The measured concentration time series can be simulated by drawing random samples from the probability density function $\rho(C)$ at each time step. The random number generator is implemented using the *inverse transform* method based on the following steps [43] :

1. Draw a sample u from the standard uniform distribution: $u \sim U[0, 1]$.
2. Compute the value of C that satisfies $F(C) = u$, where $F(\cdot)$ is the cumulative distribution function (*CDF*) of the distribution of (3.2.1).

We ensure that the generated concentration values are non-negative by applying

the following condition:

$$C = \begin{cases} C_0 \left(\frac{\gamma-2}{u} \right) \left[\left(\frac{1-u}{u} \right)^{-\frac{1}{\gamma-1}} - 1 \right], & u \geq 1 - u \\ 0, & u < 1 - u. \end{cases} \quad (3.2.2)$$

This simple formula is the only one required to implement a reasonably accurate model of the contaminated environment (i.e., to generate the concentration realisation at each sensor over time). The parameters γ and u are typically extracted from meteorological data and will be assumed known.

An example concentration realisation generated using this algorithm is presented in Figure 3.2, where different colours are used to represent different concentrations of contaminant. In this chapter, we will be concentrating on non-correlated case represented by Figure 3.2(a).

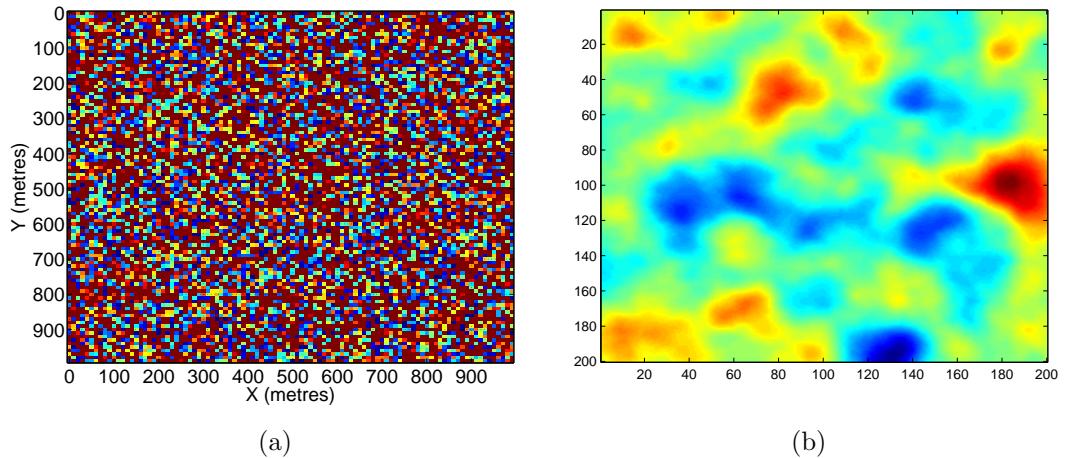


Figure 3.2: An example of random concentration realisation where different concentrations are represented by different colours: (a) static (b) dynamic

3.2.2 The model of a chemical sensor node

In a typical chemical sensor, a set volume of air is drawn first, ionized molecules when sent through an electric field, are accelerated and with different arrival times of different ion types at the ion collector, bar readings can be created and classified by agent type [86]. Here, we adopt a simple binary (or ‘threshold’) model of a sensor with air sampling. In this model, the sensor reading V is given by:

$$V = \begin{cases} 1, & C \geq C_* \\ 0, & C < C_*, \end{cases} \quad (3.2.3)$$

where C is the tracer concentration at a sensor location, and C_* is the chemical tracer detection threshold, an internal characteristic of sensors, unrelated to the mean concentration C_0 . The parameter C_* enables continuous estimation of probability of the tracer concentration to exceed sensor threshold: $p = 1 - F(C_*)$, where $F(\cdot)$ is the cumulative distribution function of tracer concentration for a given form of $p(c)$. A sensor is also characterised by its detection period τ_* ; i.e., the time taken for the sensor to sample the air and make a determination whether a tracer concentration above its detection threshold is present.

3.2.3 The model of chemical sensor network

We consider a WCSN consisting of N identical chemical sensor nodes and covering the full area of the chemical tracer field. Each sensor node can be in one of two states: *passive* or *active* behaviour as per the DSC protocol as shown in Figure 3.3. Apart from the controlled activation of a small number of sensor nodes (four in our simulations) at the start of each simulation for bootstrapping the WCSN, a passive sensor can become active only when it receives a message from an active neighbour. If

an active sensor detects a chemical tracer concentration above the detection threshold, it remains active and broadcasts an ‘activation’ message to its neighbours that are within its broadcast range r_* (we assume ideal, error-free communication and avoid considerations of interferences, contentions, etc.); otherwise, it becomes passive and remains so until ‘woken up’ again by an active neighbour. In other words, only those active sensors that make a detection are retained in the active state. This constitutes a single cycle or a time step in the WCSN life cycle. In the present study, we consider three WCSN topologies: uniform random grid, line and circle.

3.2.4 The analytical model

The behaviour of the WCSN described in the previous section is similar to the epidemic SIS (susceptible-infected-susceptible) model in [73], in which an individual in a community is in one of two states: susceptible to an identified disease; or infected by the disease. In this context, the passive sensors are analogous to susceptible individuals, and active sensors are analogous to infected individuals. For simplicity, we accept all inactive or passive sensors as susceptible unlike in a real world scenario, where susceptible individuals are the ones who are in contact with the infected individuals. Based on this intimate analogy with the process of an epidemic, an analytical model in the form of a closed system of ODE can be derived to describe the dynamics of the WCSN [94]:

$$\frac{dN_+}{dt} = \alpha N_+ N_- - \frac{N_+}{\tau_*}, \quad (3.2.4)$$

$$\frac{dN_-}{dt} = -\alpha N_+ N_- + \frac{N_+}{\tau_*}, \quad (3.2.5)$$

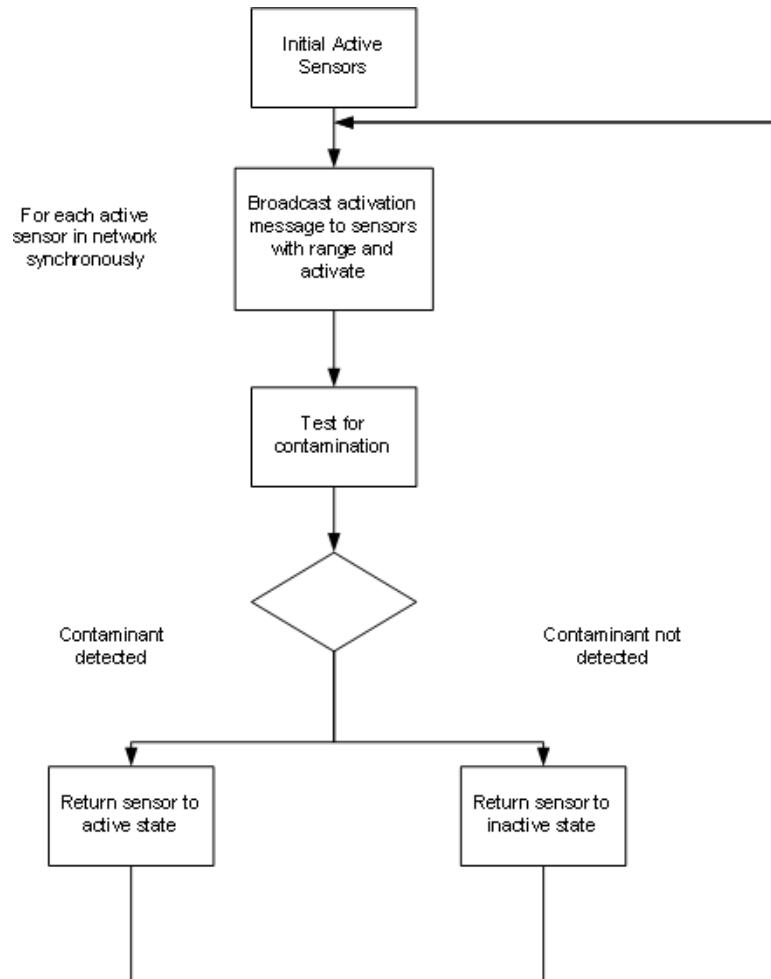


Figure 3.3: Flow chart for the DSC protocol of WSN.

where N_+ and N_- denote the number of active and passive sensors, respectively. The population size (i.e., the total number of sensors) is conserved, that is $N_+ + N_- = N = \text{const}$. The non-linear terms on the RHS of Eq. (3.2.4) and Eq. (3.2.5) are responsible for the interaction between individuals (i.e., sensors), with the parameter α being a measure of this interaction. Based on the epidemiological analogy, the following analytical expression for the interaction parameter α is proposed by Skvortsov *et al.* [94]:

$$\alpha = G \frac{\pi r_*^2}{\tau_* S} p, \quad (3.2.6)$$

where S is the square area covered by a WCSN, and G is the calibration factor, being of order unity, and parameter p , as above, is expressed in terms of the cumulative distribution function: $p = 1 - F(C_*)$; parameter p is a function of the sensor detection threshold C_* .

An expression for parameter α in Eq. 3.2.4) and Eq. 3.2.5 can be derived by invoking the epidemiological analogy [94]. The values of α derived for random, line and circle topologies, which we denote by α_r , α_l , and α_c , respectively, are:

$$\alpha_r = G \frac{\pi r_*^2}{\tau_* S} p, \quad \alpha_l = G \frac{2r_*}{\tau_* L} p, \quad \alpha_c = G \frac{r_*}{\pi \tau_* R} p, \quad (3.2.7)$$

where S is the area of the rectangular sensor field, R is the radius of the circular sensor field, and L is the length of linear sensor field; p is the probability of detection, and G is a constant calibration factor, being of order unity.

The SIS protocol, represented by Eq. (3.2.4) and Eq. (3.2.5), can be easily treated analytically and allows us to model a variety of responses and perform various optimisation studies. This analytical form also provides us with a valuable performance check for our simulations. Skvortsov *et al.* [94] found the analytical solution of the

system to be given by:

$$z(t) = \frac{z_0}{(1 - z_0) \exp(-bt) + z_0}, \quad (3.2.8)$$

where $z(t) = N_+/N$, $z_0 = z(0)$, and

$$b = \alpha N - \frac{1}{\tau_*}. \quad (3.2.9)$$

It can be seen that $z(t)$ increases only if $b > 0$ (otherwise $z(t) \rightarrow 0$). This condition corresponds to an *information epidemic* (IE) [94, 93] and can be re-written as

$$\alpha N \tau_* > 1, \quad (3.2.10)$$

or by substituting from Eq. 3.2.6, expressed as

$$GNp \left(\frac{\pi r_*^2}{S} \right) > 1. \quad (3.2.11)$$

For given values of other parameters, this condition can always be met by increasing the number of sensors N . From our computer simulations, we can estimate the fraction of active sensors as $z(t) = \frac{N_+}{N} = 1 - \theta$. Then the interaction rate α can be computed from:

$$\alpha = \frac{1}{\theta \tau_* N}. \quad (3.2.12)$$

The parameter θ determines the fraction of active sensors N_+/N_- at this new state and can be used as a measure of the network (positive) response to the event of chemical contamination. From Eq. 5.2.2, it is clear that the time constant for the network to reach the new state (the rise-time) is given by:

$$\tau \approx \frac{1}{b} = \frac{\tau_*}{\alpha \tau_* N - 1}. \quad (3.2.13)$$

From Eq. 3.2.13, it is evident that an “epidemic threshold” for the sensor network is simply,

$$R_0 = \alpha\tau_*N > 1. \quad (3.2.14)$$

The parameter R_0 is well-known in epidemiology where it has the meaning of a *basic reproductive number* [73]. Observe that sensor sampling time τ_* will disappear from the expression for R_0 for any topology (Eq. 3.2.7). This means that it is possible to create an IE (i.e., detect a chemical pollutant) for any value of τ_* , provided this time is long enough for a sensor to detect the chemical tracer.

From Eq. 3.2.12 and Eq. 3.2.13, we can deduce a few important expressions that will be later validated by our computer simulation:

$$\frac{N_+}{N_-} = \frac{\tau_*}{\tau} = \frac{1-\theta}{\theta}, \quad \theta = \frac{1}{\alpha\tau_*N}, \quad \alpha = \frac{1}{\theta\tau_*N}. \quad (3.2.15)$$

3.3 Simulation approaches

We experimented with four simulation approaches,

1. ‘Random’ network, WCSN for each scenario comprising N, r^*, C^* , R_{corr} random values. In each iteration, a new set of initial sensors was chosen.(new scenario - new network)
2. ‘Random’ network, WCSN for each scenario comprising random N, r^* but having the same network for R_{corr}/a value. In each iteration, a new set of initial sensors is chosen.(Each R_{corr} scenario - new network)
3. ‘Random’ network, WCSN for each iteration in scenarios comprising N, r^*, C^* , R_{corr} random values, with a new set of initial sensors chosen, but with identical

networks for different R_{corr} iterations. (Each scenario iteration - new network, but each with a similar pattern of change)

4. ‘Grid’ network, with the same WCSN for each scenario comprising N, r^*, C^* , R_{corr} random values, and a set of initial sensors. Only one iteration is performed. (Each scenario - same network)

Approach (1) is considered realistic, but chaotic. It is not valid for the case of different sensor detection threshold values, as the same set of points is not chosen for the network and values of parameter p are not the same.

The simulation parameters were $a = 0.2$, $N = 150$, $r^* = 20$, varied C^* and number of iterations = 50 for the first three approaches and no iterations for the fourth.

We used a more chaotic overall simulation approach and later diverged into investigating more specific characteristics of the WCSN in detail. We removed the stochastic nature of the WCSN by having a ‘grid’ topology instead of a ‘random’ topology.

The advantage of the ‘grid’ model is that all properties except c^* , r^* and R_{corr} are fixed. Therefore a better correlation between varying variables can be identified. This model is also less cumbersome compared to the previous model.

3.4 Numerical Simulations

The goal of our numerical simulations was to validate the modeling framework with the DSC protocol, proposed in this study. The modeling framework (model of environment, model of sensor, and model of network) was implemented in MATLAB. Numerical simulations were carried out for various values of parameters r_* , τ_* , p , and α , each

distinct combination of parameter values defining a different scenario; over 500 scenarios were considered in the study. Twenty five Monte Carlo runs were performed on each scenario to suppress fluctuations and produce meaningful “ensemble-averages”. Then these averages were compared with the predictions of the deterministic model (Eq. 3.2.4).

The aims of our study were:

1. To simulate conditions for IE and observe this effect in the network of chemical sensors. Identify the influence of topology and the main parameters (r_* , τ_* , C_* , C_0 , and N) on the IE;
2. To compare saturation limits and activation times with the values predicted by the models represented by Eq 3.2.13), Eq. 3.2.15;
3. To validate some universal relationships (scaling laws) predicted by the model represented by Eq. 3.2.15;
4. To derive the value of parameter α from computer simulations (i.e., from the solution of Eq. 3.2.12 and compare it with physics - based estimates (Eq. 3.2.7);
5. To calibrate our model for different topologies;
6. To find a universal fitting function for the activation phase of IE.

Figures 3.1(a), 3.1(b) and 3.1(c) show the different wireless sensor network topologies, namely, random, line, and circle, simulated in the study. In these figures, black, green, and red circles show inactive, active, and contaminated sensors, respectively. Four randomly-selected sensors were activated at $t = 0$ in each simulation. We observed that when the condition for an IE was satisfied (see Eq.(3.2.14)), the network

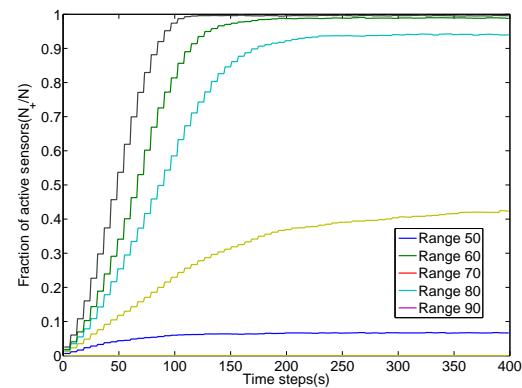
supported the IEs; otherwise it “died out”. The results of our simulations are shown in Figures 3.4 to 3.8.

Figures 3.4(a) and 3.4(b) show how the communication radius r_* and the detection threshold C_* affect the onset of an IE in the sensor network. The basic reproductive number R_0 in model represented by Eq. 3.2.14 computed for each of these simulations shows that an IE occurs only for $R_0 > 1$.

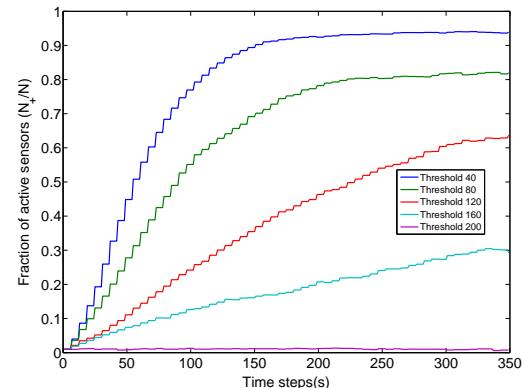
In general, results presented in Figure 3.4, support our intuitive conclusion that the number of active sensors involved in an IE (i.e., network response) increases with increasing communication radius and decreasing detection threshold of individual sensors. It comes as a surprise, however, that sometimes these trends may be non-monotonic. These anomalies require further investigation.

Our simulation results also allowed us to estimate calibration constants G in Eq. 3.2.7: $G = 1$ for the random topology, $G = 2.4$ for the line topology, and $G = 5.4$ for the circle topology.

Our next step was to validate some general properties of the deduced theoretical expressions in Eq.3.2.15 and to verify the theoretical approach for estimating the activation time of the network. We used the relationship $\frac{N_+}{N_-} = \frac{\tau_*}{\tau}$ (see Eq.3.2.15) as it does not contain any assumption about α . The activation time τ was estimated as the time for N_+ to reliably reach its saturation value. Initially, we assumed a general power-law relationship $\frac{N_+}{N_-} = \eta \left(\frac{\tau_*}{\tau} \right)^\kappa$, with constant η and κ . Figure 3.5 shows a plot of $\frac{N_+}{N_-}$ versus $\frac{\tau_*}{\tau}$ in a log-log scale. From the best fit line, shown in red, we estimated $\eta \approx 330$ and $\kappa \approx 1.4$. Because the estimated value of κ happens to be close to its theoretical value (i.e., $\kappa \approx 1$), we concluded that our theoretical model (Eq.3.2.4) provides an adequate description of network behaviour in the case of IE.



(a)



(b)

Figure 3.4: The effect of (a) communication radius r_* and (b) detection threshold C_* on the onset of IEs in a network of chemical sensors

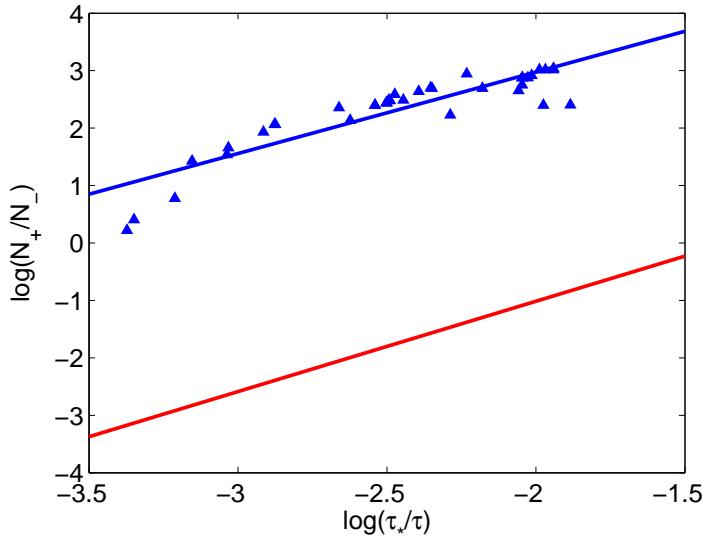


Figure 3.5: Comparison of theoretical (Eq.3.2.15) and simulated relationship between τ_*/τ and N_+/N_-

In order to validate analytical values of parameter α_t provided by expressions (Eq.3.2.7) we compare them with the results derived from our simulations. For a given saturation value of N_+ a value of α_d was estimated based on one of the Equations in .3.2.15. depending on the network topology. The results of our study is presented in Figure 3.6. The red line corresponds to the best fit $\alpha_d = \mu\alpha_t$ with $\mu \approx 1.6$. The agreement seems to be reasonable for the simple model behind estimate (Eq.3.2.7) and the rather narrow range of simulated values of α . This also means that expression (Eq.3.2.7) consistently overestimates the value of α . The overestimation can be explained based on the fact that model (Eq.3.2.7) does not take into account the accurate balance of ‘redundant messages’ (i.e., multiple activated messages to the same sensor during its passive time) and hence overestimates the effective ‘infectious’ rate.

Motivated by the analytical solution (Eq.5.2.2), we proposed a universal fitting

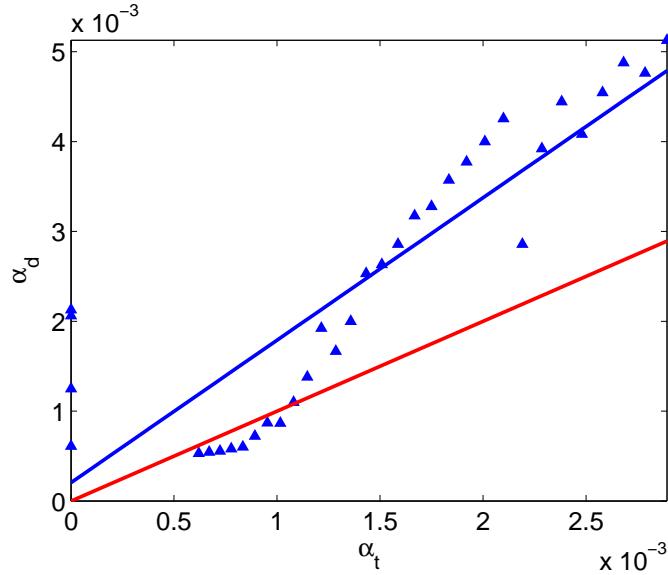


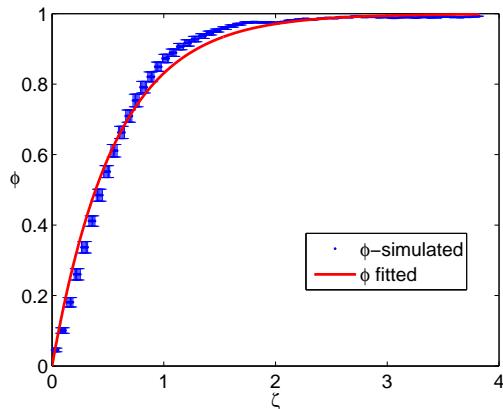
Figure 3.6: Comparison of derived and theoretical values of parameter α (3.2.7)

function to describe the IE in a network of chemical sensors in a wide range of operational scenarios (i.e., different combinations of topology, p , N , C_* , γ , and u):

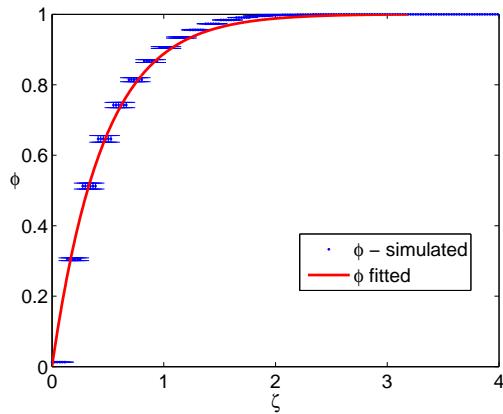
$$\phi(\zeta) = \phi_0(1 - \exp(-q\zeta)), \quad \zeta = t/\tau, \quad (3.4.1)$$

where $\phi = N_+/N$, τ is the activation time defined by (3.2.13), and $\phi_0 = b/(\alpha N)$ is the saturation value of N_+/N as $t \rightarrow \infty$; the non-dimensional parameter q is introduced to capture a flexibility in definition of the activation time and, hence, it is expected to have a universal value.

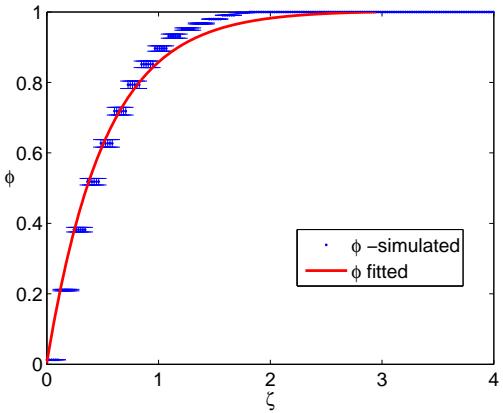
Normalised IE onset data (blue) and the corresponding fits (red) using a function of the form of (3.4.1) for random, line, and circle topologies are shown in Figures 3.7(a), 3.7(b), and 3.7(c), respectively. The value for parameter q estimated from these fits for each topology is found to be close to 2 ($q = 1.7810$, $q = 2.1792$, and $q = 1.9255$ for random, line, and circle topology, respectively).



(a)



(b)



(c)

Figure 3.7: Logistic curves (red) obtained by fitting Eq.(3.4.1) to normalised IE onset data (blue) for (a) random (b) line and (c) circle topologies. The blue error bars indicate variance over all simulations within each topology.

The fitting curves for all network topologies are presented in Figure 3.8 and provide quite a reasonable agreement. The visual deviation towards large values of ζ noticeable in Figure 3.8 has no practical implications since it corresponds to a relative error of the order of 10^{-5} in the estimated fraction of active sensors. This can be compared to similar situations obtained in [40].

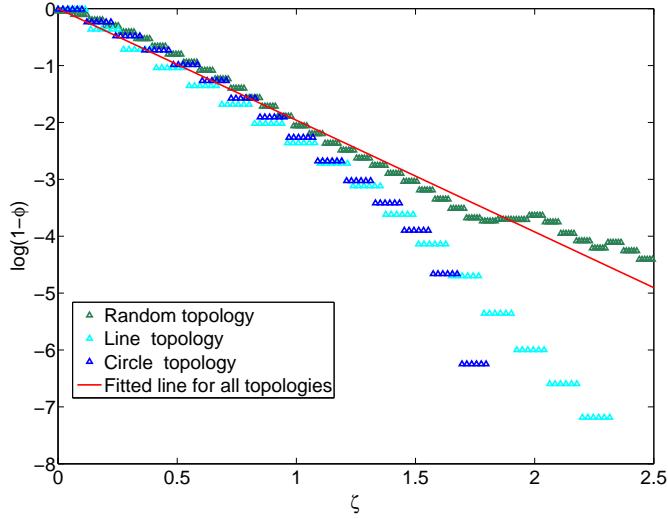


Figure 3.8: Universal logistic fit (3.4.1) for all network topologies

3.5 Summary

We evaluated a “bio-inspired” model for a network of chemical sensors with dynamic collaboration for the purpose of energy conservation and information gain. The model leverages on the existing theoretical discoveries from epidemiology resulting in a simple analytical model for the analysis of network dynamics. The model formulated

analytically the conditions for the network performance. Thus, we found that matching the optimal configuration with the analytical model , within the underlying assumptions, yields a balance between the number of sensors, detected concentration, the sampling time and the communication range. The findings are well affirmed by numerical simulations. Further work is required to address the model refinements and generalisations.

The chemical pollutant in this scenario is accepted to be a static distributed source but in reality it is a dynamic and three dimensionally distributed. The current field considered a flat terrain without obstacles, which is a highly simplified version of the more complex and challenging problem in reality. More realistic models are used in the in the following work. The sensor network should have dynamically changing composition, where there should be introduction of new sensors and removal of malfunctioning sensors.

While the model presented in this Chapter is specific to a network of chemical sensors, the underlying analytical approach can be easily adapted to other applications and other types of networks by a simple change of the model of external challenge which we consider in the next Chapter.

Chapter 4

Effect of chemical tracer distribution properties on epidemiology based sensor activation protocol

In this Chapter¹, we investigate the effect of chemical tracer field properties on wireless sensor network performance. The chemical tracers dispersed by turbulent motion in the environment display rather complex and even chaotic properties. Meanwhile, chemical tracer detecting sensors with air sampling consume significant energy. Hazardous chemical releases are rare events. If all sensors in a wireless chemical sensor network (WCSN) are left in the active state continuously, it would result in significant power consumption. Therefore, dynamic sensor activation is crucial for the longevity

¹Presentations/Publications:

- C. Mendis, A. Gunatilaka, A. Skvortsov and S. Karunasekera, ‘The effect of correlation of chemical tracers on chemical sensor network performance’, in Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Brisbane, Australia, 2010.
- C. Mendis, A. Skvortsov, A. Gunatilaka and S. Karunasekera, ‘Performance of wireless chemical sensor network with dynamic collaboration’, in Sensors Journal, IEEE, vol. 12, no. 8, pp. 2630-2637, Aug. 2012.

of WCSNs. Moreover, the statistical characteristics of chemical tracers to be detected (temporal and spatial correlations, etc.) and the placement of chemical sensors could be the key parameters that influence the WCSN design and performance. In this Chapter, we investigate the effect of spatial correlation of a chemical tracer field, and also the effect of network topology, on the performance of a WCSN that employs an epidemiology-based dynamic sensor activation protocol. We present a simulation framework that comprises models of the spatially correlated tracer field, individual chemical sensor nodes, and the sensor network. After validating this simulation framework against an analytical model, we perform simulation experiments to evaluate the effect of spatial correlation and network topology on selected performance metrics: response time, level of sensor activation, and network scalability. Our simulations show that spatial correlation of a chemical tracer field has a detrimental effect on the performance of a WCSN that uses an epidemiological activation protocol. The results also suggest that a WCSN with random network topology has poorer performance compared to one with a regular grid topology in this application.

4.1 Introduction

It is well known that chemical tracers (throughout this Chapter, we use the terms ‘chemical tracers’, ‘chemical contamination’ and ‘threat’ interchangeably) dispersed by turbulent motion in the environment display rather complex and even chaotic properties (so-called phenomenon of scalar turbulence [92]). The underlying complexity of the tracer field imposes additional challenges on the WCSN architecture. Moreover, the statistical characteristics of chemical tracers to be detected (temporal and spatial correlations, etc.) can become the key parameters that influence the

WCSN design and performance. In application to other types of WSNs (sonar, radio, optical, etc.), the studies of these effects have a long history, and they are well documented. The current study deals with the correlation effects that are specific to a tracer advected by turbulent flow in the ambient environment (i.e., low velocity of tracer propagation, intermittent distributions and power-law tails of probability density function (PDF)). In this context, we estimate the contribution of spatial correlations to the positive detection of hazardous threats by simulating different types of WCSNs in tracer fields with different correlation structures (short and long correlated; see Fig. 1.2(a) and Fig. 1.2(b)). We also briefly examine the effect of network topology (node connectivity) on the WCSN’s performance.

Apart from the aforementioned characteristics of the underlying chemical tracer field, chemical sensors themselves also have unique characteristics that differentiate them from other types of sensors that are typically used in wireless sensor networks. Particularly, unlike other types of sensors, such as optical or acoustic sensors, typical chemical sensors with air sampling consume significant energy for sensing activity compared to that consumed for communication. Because hazardous chemical releases are rare events, if all sensors in a WCSN are left in the active state continuously, it would result in significant power consumption. Therefore, dynamic sensor activation is crucial for the longevity of WCSNs. Dynamic sensor activation for chemical sensor networks using an ‘epidemiological’ sensor activation protocol has been proposed in previous work [94]. In this protocol, the information on tracer presence is spread amongst WCSN nodes similar to a disease epidemic in a community. This protocol provides a well developed analytical framework, rich physics-based analogies

(epidemiology, percolation, network models, etc.), and simple computer implementation in the form of coupled ordinary differential equations (ODE). The significant simplifications associated with employing the ‘epidemiological’ approach enables the conducting of some optimisation studies which would be otherwise difficult, if not unfeasible. Several performance aspects such as the effect of topology [27] and energy consumption of WCSNs with epidemiological sensor activation have been studied previously assuming the underlying tracer fields to be spatially uncorrelated.

In this Chapter, we present the findings of a simulation study performed to evaluate the performance of a WCSN with an epidemiology-based sensor activation protocol in the presence of non-correlated and correlated chemical tracer fields. Specifically, we make the following contributions:

- We evaluate the effect of spatial correlation of tracer fields on some performance metrics of the WCSN (namely, fraction of active sensors, response time, interaction rate, and network scalability)
- We assess the impact of two sensor network topologies (namely, regular grid and uniform random grid) on the same four metrics (i.e., fraction of active sensors, response time, interaction rate, and network scalability).

This Chapter is organised as follows. Section 4.2 describes the System model. Section 4.3 explores the validation of the model. The results from simulations are presented in Section 4.4. Section 4.5 draws conclusions and suggests future work.

4.2 System modelling and simulation

For the purpose of system simulation using a correlated chemical tracer field, we replaced the synthetic model used in Section 3.2.1 and used it in the simulation framework comprising three independent models representing the environment (the spatiotemporal realisation of the concentration field of the chemical tracers caused by turbulent mixing), an individual chemical sensor node, and the whole WCSN. This modular design of the simulation framework allows gradual improvements in any of these models. We validated our simulation model by comparing simulated results against an analytical model. The three component models of our simulation framework as well as the analytical model are discussed below.

4.2.1 The model of the environment

We consider a chemical tracer field covering a square region of side length L and area S ($S = L^2$). We assume that all sensor nodes of our system are deployed at the same height; so, we are interested in 2D tracer concentrations. In Chapter 3, we used a chemical tracer model that had no significant spatial characteristics as it was not a necessity for our research study. The chemical tracer concentration was extracted from inverse CDF. In this chapter, we generate chemical tracers from a filtering based method.

Our modelling framework allows generation of two dimensional concentration slices for a given functional form of PDF, $p(c)$; for details see [43]. The chemical tracer field is modelled as a sequence of time-varying random data slices which mimics the turbulent fluctuations of concentration at each point in the 2D xy -plane of interest where the sensor nodes are deployed. Because the main focus in this study is

on the effect of spatial correlation of chemical tracers on WCSN performance, in our simulations we use concentration slices that have spatial correlation but are temporally uncorrelated. This mimics the case where the characteristic time of the turbulent fluctuations of tracers is greater than the typical sampling time of the chemical sensor (usually, a few seconds to tens of seconds).

As a benchmark for comparing the effect of spatial correlation on WCSN performance, we use the case where the tracer is spatially non-correlated. To simulate non-correlated tracer fields, we use the simple model of a clipped Gaussian distribution and draw random data that represent uncorrelated concentration fluctuations. For simulating the spatially correlated concentration fields, we first generate non-correlated concentration data as above and filter the data using a two-dimensional exponential filter. The amount of spatial correlation is controlled by changing the sharpness of the exponential filter kernel. The mean concentration of the tracer field is C_0 while the variance of the tracer distribution depends on the level of spatial correlation. We normalise the correlation radius R_{corr} by tracer field size L and express the degree of correlation by a non-dimensional parameter ω as follows:

$$\omega = \frac{R_{corr}}{L}. \quad (4.2.1)$$

Low values of ω correspond to underlying tracer fields with low spatial correlations and vice versa; the non-correlated case, in which the correlation radius R_{corr} approaches zero, corresponds to $\omega = 0$. Further details on correlated and non-correlated chemical tracers are available in Appendix B.

4.2.2 The model of a chemical sensor node

As explained in section 3.2.2, we adopt a simple binary (or ‘threshold’) model of a sensor with air sampling. The environmental model [43] display spatial characteristics. Figure 4.1 shows the influence of C_* on the probability of detection p of chemical tracers by a sensor for fields with different correlation structures.

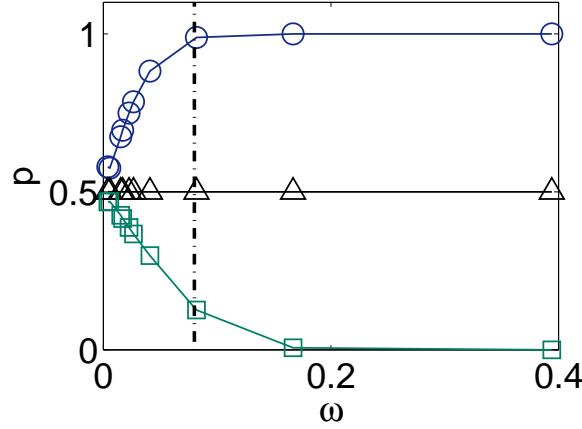


Figure 4.1: Variation of probability of detection (p) of a sensor as a function of the non-dimensional correlation parameter (ω), defined in Eq. (4.2.1), for different value ranges of the sensor detection threshold (C_*). $C_* < C_0$ (\circ), $C_* = C_0$ (\triangle), and $C_* > C_0$ (\square). C_0 is the mean concentration of the chemical tracer field. The dashed line corresponds to $\omega = 0.08$.

It can be seen that when $C_* < C_0$, p increases with ω , and when $C_* > C_0$, p decreases with ω ; when $C_* = C_0$, p is independent of ω , and it is an ideal value to keep the p value at a constant level for experimentation.

We consider each sensor node to be capable of communicating in an omni-directional fashion with its neighbouring nodes that are within a communication range r_* . By normalising the communication radius r_* by tracer field size L , we express it as a non-dimensional parameter δ as follows:

$$\delta = \frac{r_*}{L}. \quad (4.2.2)$$

We also define another non-dimensional parameter ϵ to describe the correlation of a chemical tracer field as a factor of communication radius r_* as follows:

$$\epsilon = \frac{R_{corr}}{r_*}. \quad (4.2.3)$$

When ϵ is less than one, communication occurs beyond a correlated tracer island.

4.2.3 The model of chemical sensor network

The model of WCSN is similar to one described in Section 3.2.3. Here consider two WCSN topologies: regular grid and uniform random grid.

4.3 Validation of WCSN model

To validate our WCSN simulation model against the analytical model of Eq. (3.2.8), simulation experiments were performed using WCSNs of both the regular grid topology and the uniform random topology, with non-correlated ($\omega = 0$) and correlated ($\omega = 0.02$ to 0.39) chemical tracer fields, we show results for ($\omega = 0.08$) here, as it is considered as an ideal value with diverse properties as shown in Figure 4.1. The following simulation parameter values were used in these validation experiments: $N = 400$ sensors, $L = 250$ units, the sensor detection threshold was held at $C_* = C_0$ and $r_* = 20$ units (equivalently, $\delta = 0.08$ in non-dimensional form). All results in the Chapter will be presented in terms of non-dimensional parameters to make it possible to scale the results without being tied to particular measurement units.

Figure 4.2 shows the results of validation experiments for the two network topologies in the presence of non-correlated and correlated tracer fields; in this figure, elapsed time t is shown as a non-dimensional parameter $\kappa = t/\tau_*$. The time evolution of the

fraction of active sensors (N_+/N) obtained from simulations show sigmoidal patterns similar to that obtained by Eq. (3.2.8). We first empirically estimated b by fitting curves of the analytical form of Eq. (3.2.8) to simulated data and estimated the corresponding calibration constant G using Eq. (3.2.6) and Eq. (3.2.9). For all cases, the calibration parameter G was found to be of the order of unity.

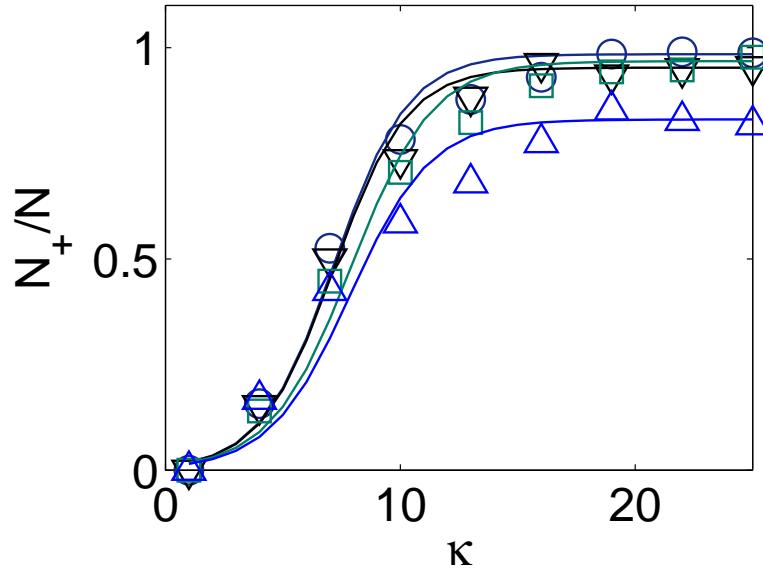


Figure 4.2: Comparison of the analytical model defined Eq. (3.2.8) against the simulation models of a WCSN: $\omega = 0.00$, Regular grid topology (\circ), $\omega = 0.00$, Random topology (∇), $\omega = 0.08$, Regular grid topology (\square), and $\omega = 0.08$, Random topology (\triangle). The solid lines show the analytically fitted curves. Here, N_+ is the number active sensors, and N is the total number of sensors in the WCSN. $\kappa = t/\tau_*$ is the normalised time where τ_* is the detection period of a sensor.

As shown in Figure 4.2, the good agreement between analytical and simulated results validates our simulation model. The experiments to evaluate the performance are presented in the Section 4.4.

4.4 Numerical results

In this section, we describe and present the results of three numerical experiments performed using our simulation framework to evaluate the performance of WCSNs. In these experiments, simulations were carried out using WCSN models of both regular grid and uniform random topologies in the presence of simulated tracer fields with several different levels of spatial correlation, including the non-correlated limit.

4.4.1 Experiment 1: Effect of sensor collaboration protocol

The first experiment, carried out to determine the fraction of sensors detecting chemical contamination, was run without any sensor collaboration protocol but with all sensors always active (i.e., $N_+ = N$). This is similar to the case where sensors operate independently and only report their positive detections of chemical pollution to the base station. The time evolution of D_+ (the number of sensors detecting contamination) was obtained for chemical tracer fields of different levels of spatial correlation (i.e., different R_{corr} values). In this experiment, the number of sensors was fixed at $N = N_+ = 400$, and the detection threshold of each sensor satisfied the condition: $C_*/C_0 = 0.93$. Because the collaboration protocol was not used in this experiment, parameter r_* was not relevant. The average number of sensors with detections was expected to be $\bar{D}_+ = N \times P$ [94] and the simulation results in Figure 4.3 are in good agreement with this value. For any other scenario, if the network uses a DSC protocol, we may expect the number of active sensors N_+ to be higher than the D_+ we obtained in this experiment. Then the increased ratio of D_+/N_+ may be considered as a benefit of DSC.

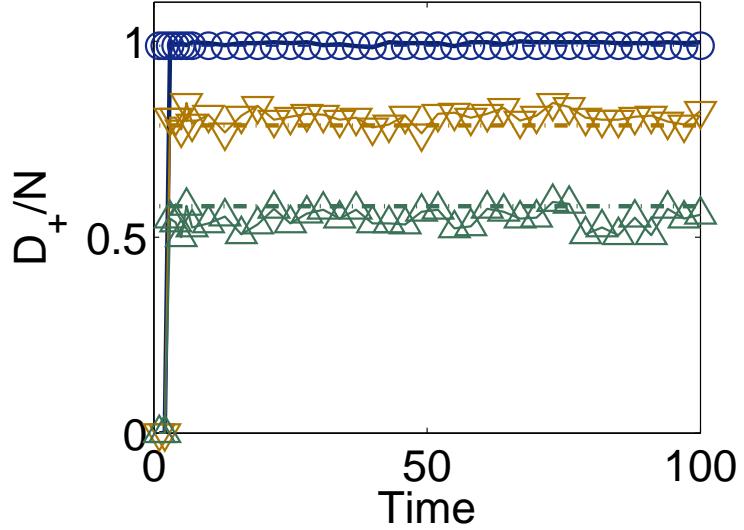


Figure 4.3: The number of sensors detecting chemical tracers (D_+/N) in evolution of time without an activation protocol and with different R_{corr} values, $R_{corr}/L = 0.08$ (\circ), $R_{corr}/L = 0.03$ (\triangledown), and $R_{corr}/L = 0$ (\triangle) or non-correlated tracers.

4.4.2 Experiment 2: Effect of sensor detection threshold

The second experiment was carried out to investigate the effect of different detection thresholds (C_*) on the behaviour of a sensor network with a DSC protocol in the presence of both correlated and non-correlated distributions of chemical tracers. In this experiment, the sensor detection threshold C_* was varied while the number of sensors was held at $N = 400$ and the communication radius was fixed at $r_* = 20$. Figure 4.4 shows example results corresponding to the three cases: $C_* < C_0$, $C_* = C_0$, and $C_* > C_0$. We see a reduction of N_+ and an increase of τ in the correlated case in comparison to those in the non-correlated case.

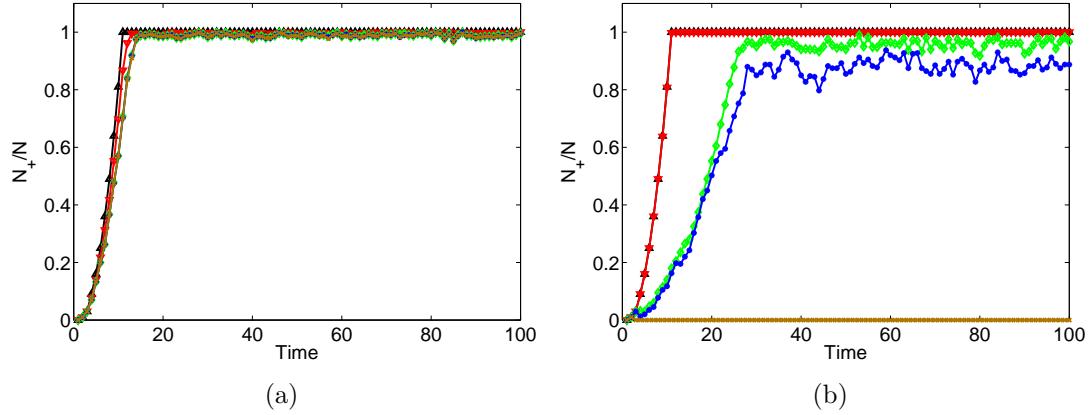


Figure 4.4: The effect of C_* on the onset of IE (N_+/N) in a WCSN: (a) non-correlated tracers and (b) correlated tracers, $C_*/C_0 = 0.267$ (black upward triangle), $C_*/C_0 = 0.933$ (red downward triangle), $C_*/C_0 = 1.0$ (green diamond), $C_*/C_0 = 1.007$ (blue star) and $C_*/C_0 = 1.013$ (brown x-mark).

4.4.3 Experiment 3: Sensor interaction

The fourth experiment, aimed at verifying the analytical model of [94] with our simulated data, was carried out by varying r_* while keeping $N = 400$, $C_*/C_0 = 1$, and $R_{corr}/L = 0.08$. The interaction rate α was computed analytically (using Eq. (3.2.6)) and from simulation results (based on Eq. (3.2.15)). We denote the analytical value of α by α_a and the simulated value by α_s . Figure 4.5 shows plots of α_s vs. α_a for both correlated and non-correlated cases.

We see that an optimal communication radius r_* is visible at which there will be no improvements to WCSN performance. The optimal communication range was found to be 171 and α_s was 0.125 for both cases (saturation values in Fig. 4.5). We observe that, based on the results of our simulations, we can write a simple yet universal estimate for the interaction parameter α : $\alpha_a \leq \alpha \leq 0.125$ (and $\alpha = 0.125$ otherwise), where α_a is given by Eq. (3.2.6) and α_s is given by Eq. (3.2.15).

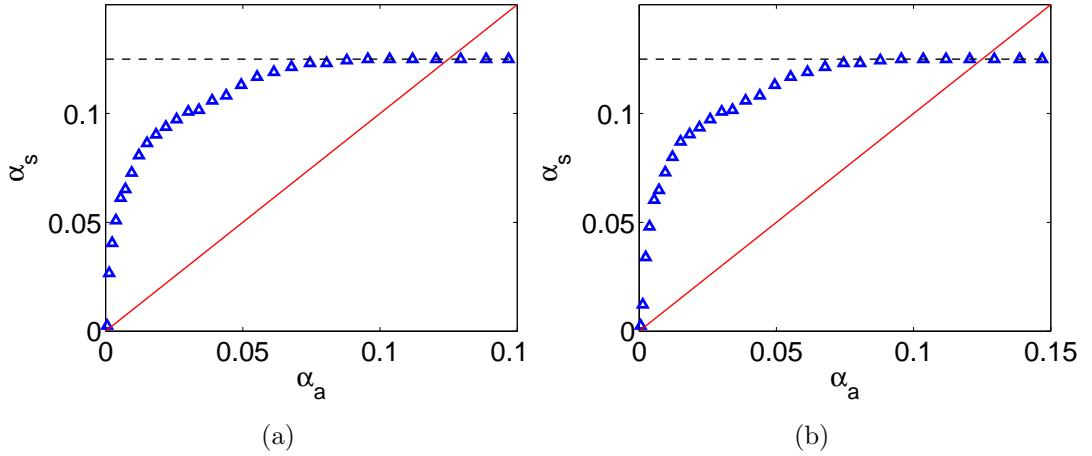


Figure 4.5: Comparison of simulated (α_s) and analytical (α_a) values of parameter α on (a) non-correlated tracers and (b) correlated tracers, with the red line showing the expected outcome, and black dashed line showing the saturation point.

Our simulation results also allowed us to estimate the calibration constant in the above equations: $G = 0.13$. Figure 4.5 shows that Eq. (3.2.15) for both non-correlated and correlated chemical tracers, consistently underestimates the value of α for lower r_* and overestimates for higher r_* . However the overall estimation is within an acceptable range.

4.4.4 Experiment 4: Effect of network topology

The fourth experiment, aimed at verifying the analytical model of [94] with our simulated data, was carried out by varying the WCSN topology while keeping $N = 400$, $C_*/C_0 = 1$, $r_* = 20$ and $R_{corr}/L = 0.08$. Figure 4.6 shows plots with different WCSN topologies.

We see a rise in the saturation level of the IE and a decrease of response time τ in the line topology compared to other topologies. The regular grid topology presents a

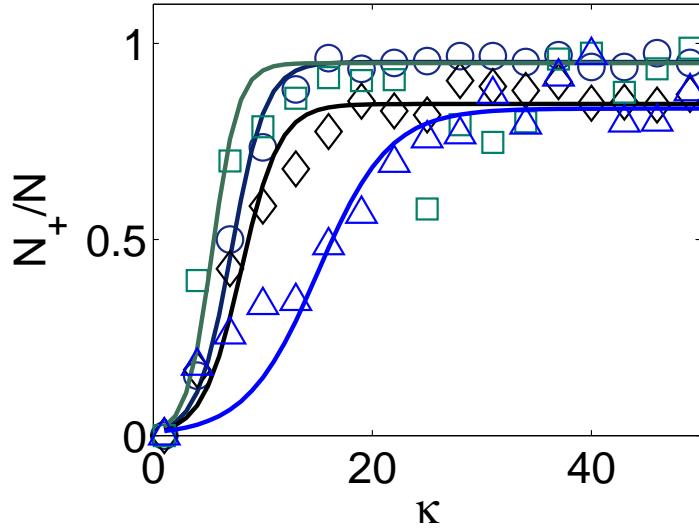


Figure 4.6: The effect of correlation characteristics of chemical tracer field (ϵ) on the onset of IEs (N_+/N) in a regular grid WCSN. The range of WCSN topologies used are: regular grid (\circ), random (\triangleright), line (\square), and circle (\triangle). $\kappa = t/\tau_*$ is the normalised time where τ_* is the detection period of a sensor.

stable activation which is preferable for reliability. The circle and random topologies show a degraded performance.

4.4.5 Experiment 5: Effect on sensor activation

The first experiment was carried out to determine the effect of network topology and chemical tracer field correlation on the sensor activation of WCSN. A fraction of active sensors N_+ at saturation state is a measure of the network (positive) response to the event of chemical contamination. In this experiment, WCSNs of both regular grid and uniform random topologies were simulated in the presence of a simulated chemical tracer field with no spatial correlation ($\omega = 0$) and one with correlation ($\omega = 0.08$). The number of sensors was again set at $N = 400$, and the sensor detection threshold

was held at $C_* = C_0$. Simulations were performed for several values of communication radius r_* (or equivalently, the non-dimensional parameter δ); specifically, the following δ values were used for plotting: $\delta = 0.052$, $\delta = 0.056$, $\delta = 0.080$, $\delta = 0.160$, and $\delta = 0.480$. Each scenario was simulated 25 times, varying the bootstrapping sensors in the case of the regular grid sensor network and varying the sensor locations themselves in the case of the random sensor network. Figures 4.7(a) and 4.7(b) show ensemble averaged results obtained for regular grid and Figures 4.7(c) and 4.7(d) show those for random WCSN.

We see a rise in the saturation level of the IE and a decrease in response time τ when the communication radius r_* (or δ) is increased or ϵ is decreased. We see that some low values of δ do not support an IE; i.e., the IE ‘dies out’, as the condition for an IE is not satisfied in these cases. In a regular grid topology, sensors are well distributed and capable of bearing the drastic changes in the environment, where in the case of random topology sensors may be not within a chemical tracer concentration for collaborative activation and IE to continue. We can say that blobs instead of well dispersed chemical tracers of high concentration and insufficient sensor connectivity tend to hinder the IE process. We observe that results are in compliance with Eq. (3.2.8), but with different values for calibration constant G .

Figure 4.8 shows ensemble averaged results obtained for several correlation values with regular grid topology.

4.4.6 Experiment 6: Empirical modelling of interaction rate

The aim of the second experiment was to derive an empirical model for the interaction rate α for a WCSN of regular grid topology in a correlated chemical environment,

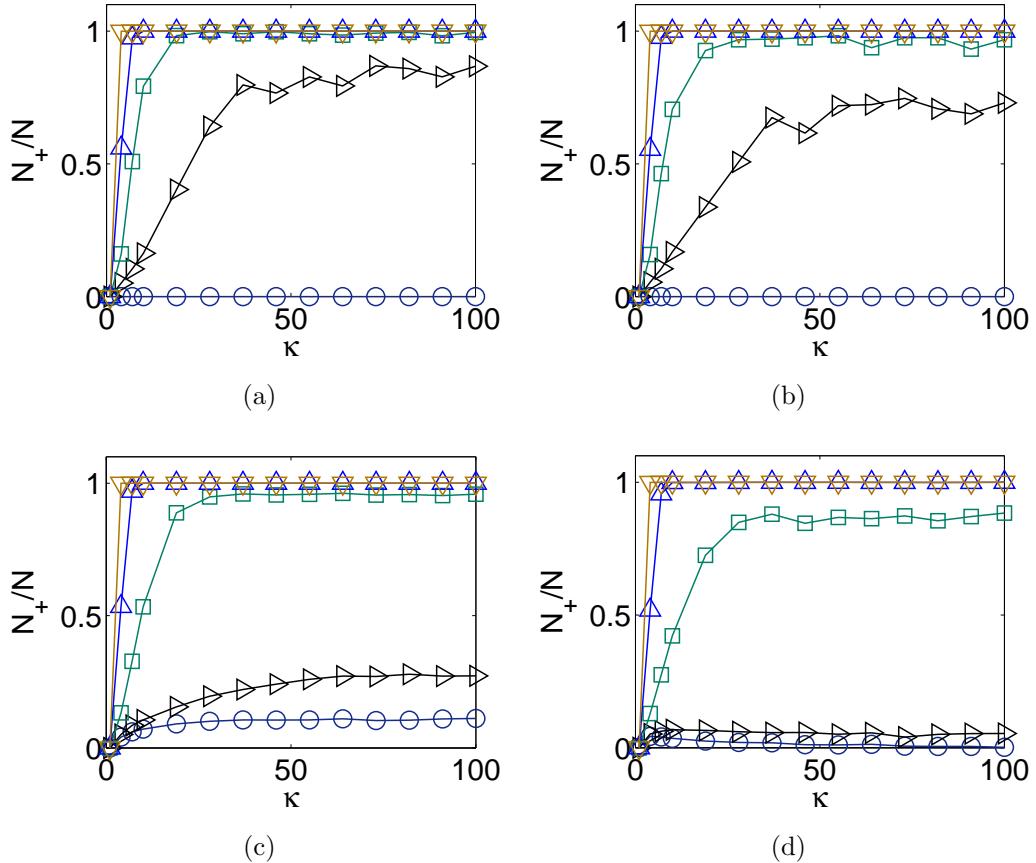


Figure 4.7: The effect of communication range (r_*) on the onset of IEs (N_+/N) in a WCSN: (a) regular grid topology in non-correlated tracer field; (b) regular grid topology in correlated tracer field; (c) random topology in non-correlated tracer field; (d) random topology in correlated tracer field. The range of values used for the non-dimensional communication radius parameter δ , defined in Eq. (4.2.2), are: $\delta = 0.052$ (\circ), $\delta = 0.056$ (\triangleright), $\delta = 0.080$ (\square), $\delta = 0.160$ (\triangle), $\delta = 0.480$ (∇). $\kappa = t/\tau_*$ is the normalised time where τ_* is the detection period of a sensor.

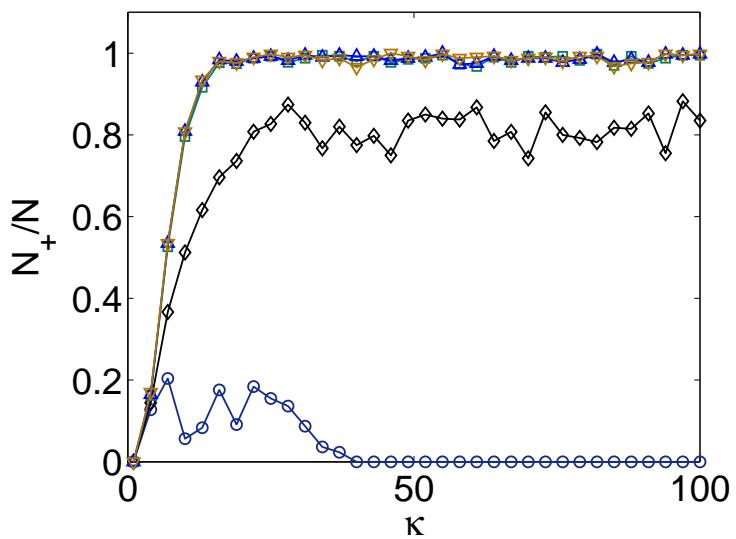


Figure 4.8: The effect of correlation characteristics of chemical tracer field (ϵ) on the onset of IEs (N_+/N) in a regular grid WCSN. The range of values used for the non-dimensional communication radius parameter ϵ , defined in Eq. (4.2.3), are: $\epsilon = 4.93$ (\circ), $\epsilon = 2.08$ (\diamond), $\epsilon = 0.51$ (\square), $\epsilon = 0.28$ (\triangle), $\epsilon = 0.19$ (\triangledown). $\kappa = t/\tau_*$ is the normalised time where τ_* is the detection period of a sensor.

based on our numerical simulation data. The experiment was performed by varying r_* (or δ) over a wide range. (from very small values to a radius large enough to communicate from one corner to the other in the field covered by WCSN). The sensor detection threshold was fixed at $C_* = C_0$, side length $L = 500$ units and calibration constant $G = 0.13$. In this, we used the best performing WCSN of regular grid topology in the chemical tracer field with spatial correlation.

As r_* was incremented, the interaction rate α was computed analytically (using Eq. (3.2.6)) and from simulation results (based on Eq. (3.2.12)). We denote the analytical value of α by α_a and the simulated value by α_s . Figure 4.9 shows the comparison of α_s and α_a . As r_* is increased, as the interaction rate estimated from simulations α_s increases gradually and reaches a saturation value, which we denote by α^* .

We did not find notable differences of α^* with different deployed areas, network topologies, and correlation structures. However, when the process was repeated using a different number of sensors, the saturation value α^* was found to change. Therefore, we repeated the simulations for a range of N (all square numbers between 400 and 3600) and obtained the simulated and analytical interaction rates.

Figure 4.9 shows that Eq. (3.2.6) consistently underestimates the value of α for lower r_* and overestimates for higher r_* . But the overall estimate is within an acceptable range, and we can write a simple but universal estimate for the interaction parameter for a given N as: $\alpha_a \leq \alpha \leq \alpha^*$, where α^* is the saturation level. We believe this discrepancy between the simulated and analytical values of the interaction rate is due to the effect of double counting of activation messages in overlapping areas of activation in the sensor network [94]. We can expect a better match by varying the

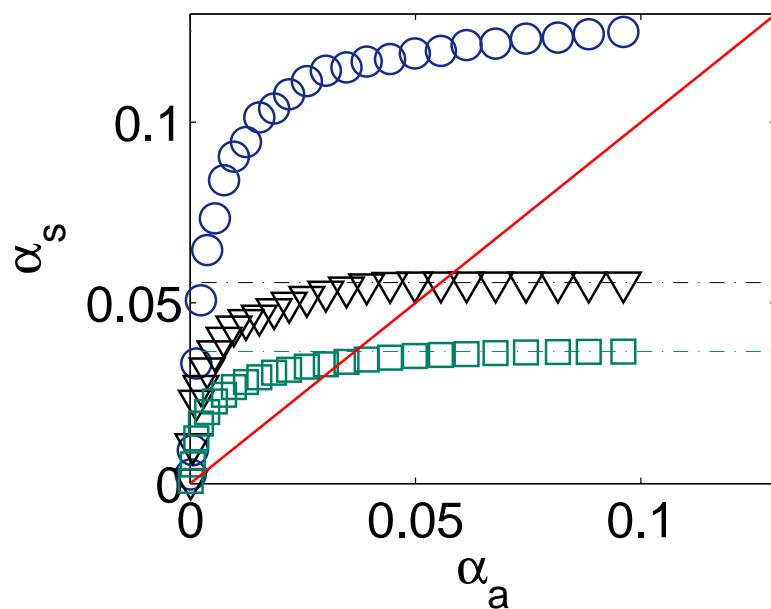


Figure 4.9: Comparison of simulated (α_s from Eq. (3.2.6)) and analytical (α_a from Eq. (3.2.12)) values of interaction rate α in a WCSN with regular grid topology in the presence of a correlated tracer field, with the number of sensors equal to: $N = 400$ (\circ), $N = 900$ (∇), $N = 1369$ (\square). Dash-dot lines show the saturation levels corresponding to the scenarios. The solid line shows the expected outcome if simulated and analytical values are in perfect agreement.

calibration parameter G . As a better alternative, we derive a robust empirical model for the interaction rate α . In compliance with our intuitive assumption, we see that the interaction rate α and saturation state α^* decrease when the number of sensors N increases. Based on the intuitive assumption that α^* and N have a power law relationship, we can write

$$\alpha^* \propto N^\beta, \quad (4.4.1)$$

where the power law exponent β is a fitting parameter. From the plot of α^* vs. N in log-log scale, shown in Figure 4.10, we estimate that $\beta \approx -1$.

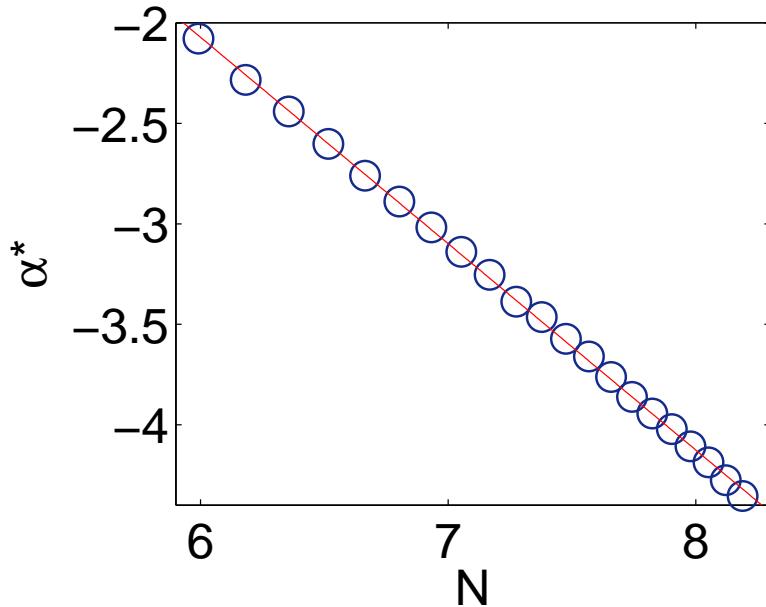


Figure 4.10: Saturation value of interaction rate (α^*) as a function of the number of sensors (N), plotted in log-log scale. The best linear fit is shown as a solid line.

Next, for different values of N , we plot α , normalised by the saturation value α^* , as a function of the normalised communication radius δ , as shown in Figure 4.11. From this figure, we see that the curves corresponding to all values of N collapse into

a single universal curve which fits a function of the form:

$$\frac{\alpha}{\alpha^*} = 1 - \exp(-a\delta). \quad (4.4.2)$$

The fitting parameter in Eq. (4.4.2) is estimated to be: $a = 5.8$.

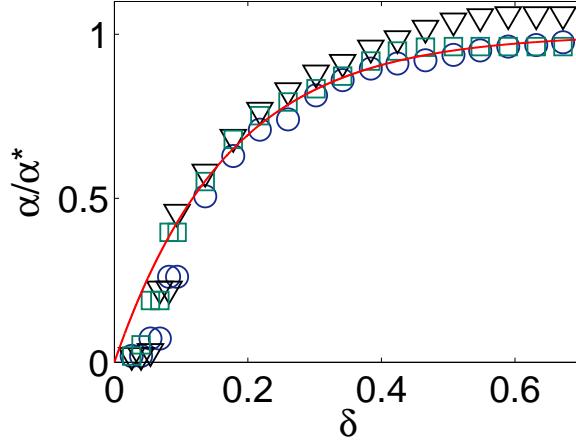


Figure 4.11: Interaction rate (α), normalised by its saturation value (α^*), as a function of the non-dimensional communication radius parameter (δ), defined in Eq. (4.2.2), for a WCSN operating in a correlated tracer field and having a regular grid topology with different number of sensors: $N = 400$ (\circ), $N = 900$ (∇), and $N = 1369$ (\square). The solid line shows a curve fitted by varying a of Eq. (4.4.2).

4.4.7 Experiment 7: Scaling laws of WCSN performance

The final experiment was performed to ascertain the scalability of the WCSN. The scalability of any sensor system is a parameter that usually describes the system performance as a function of the number of deployed sensors. This is one of the important parameters that is used to shape the network architecture. The proposed framework enables straightforward derivation of the scalability factor for the WCSN.

Let us characterise the WCSN performance as a ratio $\Psi = \sigma_{N_+}/\mu_{N_+} \propto \Xi^q$, where μ_{N_+} is the time-averaged mean of N_+ (i.e., around the saturation level), and σ_{N_+}

is the time-averaged fluctuations of N_+ and $\Xi = 1/N$. The scalability properties of the WCSN can then be characterised by the power exponent q in the functional relationship $\Psi \sim \Xi^q$ where N is the number of sensors, and the case $q \geq 0$ should be held for any sustainable network architecture.

In this experiment, we varied the number of sensors N while keeping the normalised communication radius fixed at $\delta = 0.08$ and sensor detection threshold at $C_* = C_0$. The lower bound of N was chosen as the minimum number of sensors that can propel an IE, and the upper bound was chosen subject to computer memory constraints. Simulations for this experiment were performed for networks with regular grid and random network topologies, operating in tracer fields with various levels of spatial correlation.

Plots of relative fluctuations Ψ versus Ξ , the reciprocal of the number of sensor nodes, are shown in Figures 4.12(a) and 4.12(b) for the regular grid WCSN and in Figures 4.12(c) and 4.12(d) for the random WCSN. For comparison, each figure also shows asymptotes in red which correspond to a scalability factor $q = 1/2$ that holds for any system of independent sensors with Gaussian noise, following from the standard estimate $\Psi \sim \sigma_{D_+}/\mu_{N_+} \sim \Xi^{1/2}$ [56].

From these plots, we can see an asymptotic relationship of ϵ and fluctuations of N_+ , demonstrating a power law behaviour, and we estimate the scalability factor q to be: $q \approx 0.1$ in all scenarios considered. We can see that the scalability of the WCSN with epidemic protocol follows correct trends (i.e., $q > 0$). Depending on a particular operational scenario, some other parameters (energy conservation, reliability of detection, activation time, budget, etc.) may become more important and would favour a particular technical solution.

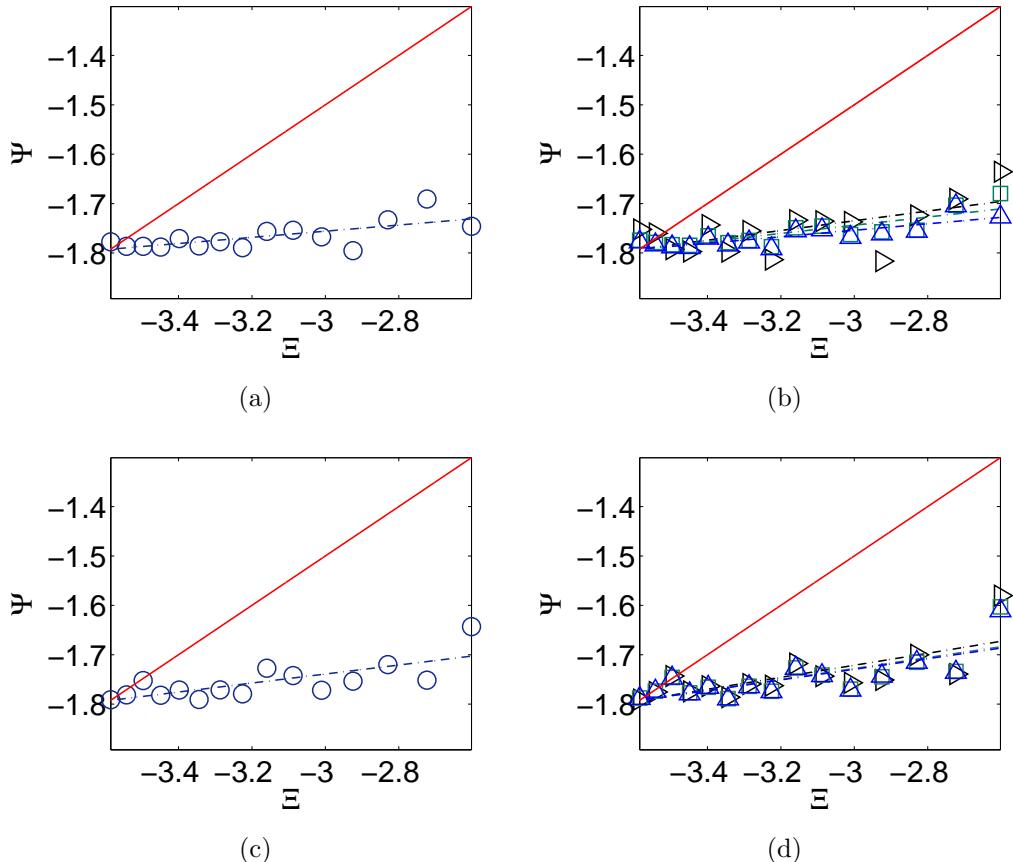


Figure 4.12: Relative fluctuations $\Psi = \sigma_{N_+}/\mu_{N_+} \propto \Xi^q$, where $\Xi = 1/N$, plotted in log-log scale for a WCSN: (a) regular grid topology in non-correlated tracer field, (b) regular grid topology in correlated tracer field; (c) random topology in non-correlated tracer field, (d) random topology in correlated tracer field. The correlation of tracer fields is characterised by $\epsilon = 0.00$ (\circ), $\epsilon = 1.02$ (\triangleright), $\epsilon = 0.33$ (\square), and $\epsilon = 0.28$ (\triangle). The dash-dot lines show lines fitted by varying q . The solid lines show the outcome for a WCSN of independent sensors with Gaussian noise.

4.5 Summary

In this Chapter, we investigated the effect of spatial correlation of a chemical tracer field, and also the effect of network topology, on the detection of a chemical threat using a WCSN. To perform this study, we constructed a simple integrated simulation framework comprising models of the environment, individual chemical sensors, and the sensor network. Our WCSN used an epidemiology-based sensor activation protocol. Using a ‘bio-inspired’ analytical model for a WCSN, initially introduced by Skvortsov *et al.* [94], we verified the simulation framework for WCSNs of regular grid and uniform random network topologies operating in the presence of simulated tracer fields with and without spatial correlation. Based on simulation results, we evaluated some important performance metrics (namely, detection time, fraction of active sensors, sensor interaction rate, and scalability factor).

In general, we found random network topology and spatial correlation of tracer distribution (which is typical for most operational scenarios) to have a negative effect on the performance of a WCSN with the ‘epidemiology’ based sensor activation protocol. It manifests in longer network detection time and a lower fraction of active sensors, which is detrimental to good detectability [106]. This result contrasts with what is observed more commonly in other types of WSNs. The other possible reason for the different behaviour in our simulation is the nature of the chemical tracer field itself, involving blobs of high concentration, which is a direct consequence of mass conservation. This structure is different to the more coherent variations typical of phenomena such as optical or acoustic fields. The work reported in this Chapter is extended to overcome the negative effects posed by the DSC protocol which is presented in the following chapter.

Chapter 5

Gossip inspired sensor activation protocol for a wireless chemical sensor network

In this chapter¹, we evaluate the behaviour of gossip based sensor activation protocol in a chemical tracer environment. Energy conservation in chemical sensor networks is crucial as chemical sensors with air sampling consume significant energy for sensing activity compared to that used for communication unlike other types of sensors, such as optical or acoustic. When considering the threat environment, the chemical tracers dispersed by turbulent motion in the environment display rather complex and even chaotic properties. Hazardous chemical releases are rare events. If all sensors in a wireless chemical sensor network (WCSN) are left in the active state continuously, it will result in significant power consumption. Therefore, dynamic sensor activation is essential for the durability of WCSNs. Dynamic sensor activation for chemical sensor

¹Presentation/Publication:

- C. Mendis, ‘Gossip inspired sensor activation protocol for a correlated chemical environment’, in Bio-Inspired Models of Networks, Information, and Computing Systems of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 103, pp. 163-170, 2012.

networks using an epidemiology-based sensor activation protocol has been proposed in the literature. In this chapter, we investigate the performance of a variant of epidemiology based protocols, the gossip-based sensor activation protocol of a WCSN in a chemical tracer field. The simulation framework with the gossip protocol is validated against an analytical model. We then perform simulation experiments to evaluate the performance of the gossip-based sensor activation protocol on selected performance metrics: the sensor activation and chemical tracer detection. We show by simulations that, by varying the communication radii of sensors, we can achieve better energy conservation while maintaining better performance of a WCSN with a gossip-based activation protocol.

5.1 Introduction

A chemical detection system in a environment prone to a possible threat from hazardous chemicals, is comprised of individual sensors or a network of sensors (WCSN). The sensors detect and immediately inform the base station or aggregate and transmit the data in later stage [31]. The chemical sensors have unique characteristics that distinguish them from other types of sensors that are typically used in wireless sensor networks. Particularly, unlike other types of sensors, such as optical or acoustic sensors, typical chemical sensors with air sampling consume significant energy for sensing activity compared to what is consumed for communication. When considering the threat environment, the chemical tracers dispersed by turbulent motion in the environment display rather complex and even chaotic properties (the phenomenon of so-called scalar turbulence [92]). These chemical tracers have statistical characteristics such as temporal and spatial correlations. These parameters influence the WCSN

design and performance.

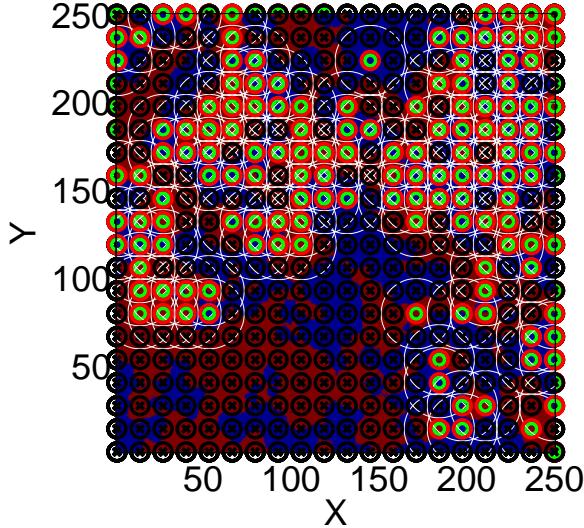


Figure 5.1: Graphical illustration of a WCSN operating in simulated chemical tracer fields with spatial correlation. The tracer field at a given time instance is depicted as 2D xy -plane, with different colours representing different tracer concentrations. The gray circles with larger white circle represent active sensors with their communication range, and the black circles represent inactive sensors.

Minimising sensor activation to improve the lifespan of WCSNs is a subject which has received special attention in the literature in the recent past [80], [4], [15]. Research has focused on the development of optimal strategies for sleep-wake scheduling based on a detection decision performed at a central fusion center. In [80], the authors present similar strategies for the closed-loop case where it is dynamic and based on a detection decision that is fed back from the base station. They also consider an open-loop case where the scheduling is optimised for a particular attack pattern. In both cases, the optimisation is based on a Markov decision process for detection. Dynamic sensor activation for chemical sensor networks using an ‘epidemiological’ sensor activation protocol has been proposed in [93] and [53]. We have evaluated the effect

of spatial correlation of tracer fields and assessed the impact of two sensor network topologies (namely, regular grid and uniform random grid) on the three metrics (i.e., fraction of active sensors, response time, interaction rate, and network scalability). They have found that random network topology and the spatial correlation of tracer distribution (which is typical for most operational scenarios of chemical environments) have a negative effect on the performance of a WCSN with an epidemiology-based sensor activation protocol.

Our motivation being a solution to the negative effect on the performance of the WCSN in the previous study, we focus on a dynamic, decentralised, sleep-wake strategy without feedback from a centralised base station using a ‘gossip’ based protocol. Our goal is to minimise the number of active sensors in the network in the absence of a chemical release and to rapidly increase the number when a chemical release happens, also we strive to improve the WCSN performance in a chemical field with spatial correlation of tracer distribution. We propose to achieve our goals through a simple sensor activation protocol: sensors that detect an event wake up a selected number of neighbouring sensors. In this study, we use a stochastic approach as in the case of [35]. In compliance with our intuitive assumption that the gossip-based protocol will help to improve the sensor activation inefficiency in a correlated chemical environment by extending the communication radius while reducing the number of active sensors, furthermore, in order to retain the simplicity of expressing its behaviour by an analytical expression, we introduce an additional ‘gossip-factor’ (throughout this chapter, we use the terms ‘gossip-ratio’, ‘gossiping probability’ and ‘fanout ratio’ interchangeably) to the existing epidemiology-based WCSN model.

We use a synthetic chemical threat environment to study the behaviour of WCSN.

When the gossip-factor reaches unity, we show that our protocol behaves similarly to an epidemiology-based protocol. When the gossip-factor is zero, the network becomes dysfunctional. We perform simulation experiments to study the behaviour of the gossip-based protocol on a chemical environment using selected performance metrics: the network activation and the chemical detection (see Section 5.4), which are well-established metrics for other types of sensor networks (see [106]). In this chapter, we make the following specific contributions:

- We evaluate the behaviour of a WCSN in complex chemical tracer fields based on certain performance metrics of the WCSN (namely, the number of active sensors and the unitised risk of detection) to obtain optimal network sensor configurations.

We believe that these contributions sufficiently illustrate the viability and optimal use of gossip based algorithms, as well as establish the novelty of the present study.

The Chapter is organised as follows. Section 5.3 describes Gossip based sensor activation protocol which we used to overcome the short-comings in we saw in the previous chapter. Section 5.3 describes the System model. The results from simulations are presented in Section 5.5. Section 5.6 draws conclusions and suggests future work.

5.2 Gossip-based sensor activation protocol and analytical model

5.2.1 The activation protocol

In our research studies, we use a variant of the epidemiological model presented in Chapter 4, we are capable of validating a WCSN model with a gossip-based protocol using the same analytical model. Unlike typical research carried out so far, we use a chemical tracer field and evaluate the performance of the network based on a gossip-based protocol. In our model, we accept that a node communicates with more than one node and activates all nodes contacted based on a predefined contacts table. When communicating with the contacted node, sensors exchange their chemical concentration measurements with each other with time stamping and node identification. We employ a push approach. The goal of the research is to activate the efficient number of sensors and achieve good information dissemination while conserving energy. Peers selected for communicating with are chosen randomly based on the fanout ratio f . We also accept that the process occurs in synchronisation across the network. We currently use a stochastic approach in selection, but have the option of using a deterministic approach later, which we do not consider due to time constraints. We use a mesh or regular grid topology and assume that the sensor will not change the degree of connectivity during the operation of the network. With uniformly placed sensors covering the area, we assume that it will be the best solution to overcome coverage issues. The sensor nodes communicate in an undirected and duplex manner. With wireless communication, although the sensor node's activation message is broadcasted, we accept that only a sensor pre-selected to activate, will wake up. The

sensor node can be clustered, based on data collection, but here we are concerned only with network activation. We assume that there are neither interferences nor node or transmission failures.

Though sensors communicate with all neighbouring sensors, sensor activates only f selected number of sensors. If an active sensor detects a chemical tracer concentration above the detection threshold, it remains active and broadcasts an ‘activation’ message fraction f sensor nodes of its neighbours that are within its broadcast range r_* (We assume ideal, error-free communication and avoid considerations of interferences, contentions, etc.); otherwise, it becomes passive and remains so until ‘woken up’ again by an active neighbour. In other words, only those active sensors that make a detection are retained in the active state. This constitutes a single epoch, cycle or a time step in the WCSN life cycle.

5.2.2 The analytical model

The overall behaviour of the WCSN considered by us, is analogous to extension of an epidemic SIS (susceptible-infected-susceptible) model [73]. Based on this close analogy with the process of an epidemic, an analytical model in the form of a closed system of ODE can be derived to describe the dynamics of the WCSN [93]. We can express the interaction rate of a gossip-based WCSN α_G as a function of α given in Eq. (3.2.6):

$$\alpha_G = \alpha f, \quad (5.2.1)$$

where f is the *fanout* ratio [35] or gossiping probability [45] of the gossip-based sensor activation protocol. The epidemiological protocol, represented by Eq. (3.2.4) and Eq. (3.2.5), can be easily treated analytically as the gossip protocol with Eq. (5.2.1)

and allows us to model a variety of responses and perform various optimisation studies.

Skvortsov *et al.* [93] found the analytical solution of the system to be given by:

$$z(t) = \frac{z_0}{(1 - z_0) \exp(-bt) + z_0}, \quad (5.2.2)$$

where $z(t) = N_+/N$, $z_0 = z(0)$, and $b = \alpha_G N - \frac{1}{\tau^*}$. This analytical solution provides us with a valuable performance check for our simulations. When $f = 1$, the WCSN overall behaviour becomes similar to epidemiological-based model, and when $f = 0$, the WCSN becomes dysfunctional.

5.3 WCSN system model

For the purpose of system simulation using a correlated chemical tracer field, we replaced the synthetic model used in the Section 5.3 of Chapter 3 and used with the simulation framework comprising three independent models representing the environment (the spatiotemporal realisation of the concentration field of the chemical tracers caused by turbulent mixing), an individual chemical sensor node, and the whole WCSN. This modular design of the simulation framework allows gradual improvements in any of these models. We adopt a non-binary (or ‘threshold’) model of a sensor with air sampling unlike sensors used in studies presented in Chapter 4. In this model, the sensor reading V is given by:

$$V = \begin{cases} C, & C \geq C_* \\ 0, & C < C_*, \end{cases} \quad (5.3.1)$$

where C is the tracer concentration at a sensor location, and C_* is the chemical tracer detection threshold, an internal characteristic of sensors, unrelated to the mean concentration C_0 . The parameter C_* enables continuous estimation of probability of the

tracer concentration to exceed sensor threshold: $p = 1 - F(C_*)$, where $F(\cdot)$ the cumulative distribution function of tracer concentration for a given form of $p(c)$. A sensor is also characterised by its detection period τ_* ; i.e., the time taken for the sensor to sample the air and make a determination whether a tracer concentration above its detection threshold is present. We validated our simulation model by comparing simulated results against an analytical model. The three component models of our simulation framework as well as the analytical model are discussed below.

The environment module allows generation of two dimensional concentration slices for a given functional form of PDF, $p(c)$; for details see [43]. The chemical tracer field is modelled as a sequence of time-varying random data slices which mimics the turbulent fluctuations of concentration at each point in the 2D xy -plane of interest where the sensor nodes are deployed. Because the main focus in this study is to find the means to overcome the effect of spatial correlation of chemical tracers on WCSN performance, in our simulations we use concentration slices that have spatial correlation but are temporally uncorrelated. This mimics the case when the characteristic time of the turbulent fluctuations of tracers is greater than the typical sampling time of the chemical sensor (usually, a few seconds to tens of seconds).

We use the cases where the tracer are spatially correlated and non-correlated. To simulate non-correlated tracer fields, we use the simple model of a clipped Gaussian distribution and draw random data that represent uncorrelated concentration fluctuations. For simulating the spatially correlated concentration fields, we first generate non-correlated concentration data and filter the data using a two-dimensional exponential filter. The amount of spatial correlation is controlled by changing the sharpness of the exponential filter kernel.

In our simulations, we predefine the activation pattern of all sensors-based on f , where we deviate from the highly decentralised sensor activation approach proposed in Chapter 4. For implementation, we use a vectorised form to select sensors synchronously for activation to improve the simulation speed. In real physical implementation, this can be sent as a ‘black list’, which contains the references to the sensors which are not intended to be activated. This can be embedded in a message sent by a sensor to activate its neighbour after detecting a chemical tracer concentration. We consider simulation scenarios each having a distinct r and f . Based on measurements presented in [38], we say that the energy consumed when a sensor is activated is greater than the energy consumed by extending the communication radius of a sensor.

We can illustrate the gossip-base sensor activation as shown in Figure 5.2.

5.3.1 Validation of WCSN simulation model

To validate our WCSN simulation model with selective activation protocol against the analytical model of Skvortsov *et al.* [93] described in Section 5.2, simulation experiments were performed using a WCSN of a regular grid topology. The following simulation parameter values were used for these validation experiments: $N = 400$ sensors, $L = 250$ units, and $r^* = 20$ units (equivalently, $\delta = 0.08$ in the non-dimensional form as stated in Eq. (4.2.2)) and $f = 0$, $f = 0.4$, $f = 0.6$ and $f = 1$. All results in the chapter will be presented in terms of non-dimensional parameters to make it possible to scale the results without being tied to particular measurement units.

Figure 5.3 shows the results of validation experiments for the WCSN model. In this figure, the onset of an *information epidemic* (IE), characterised by the fraction of

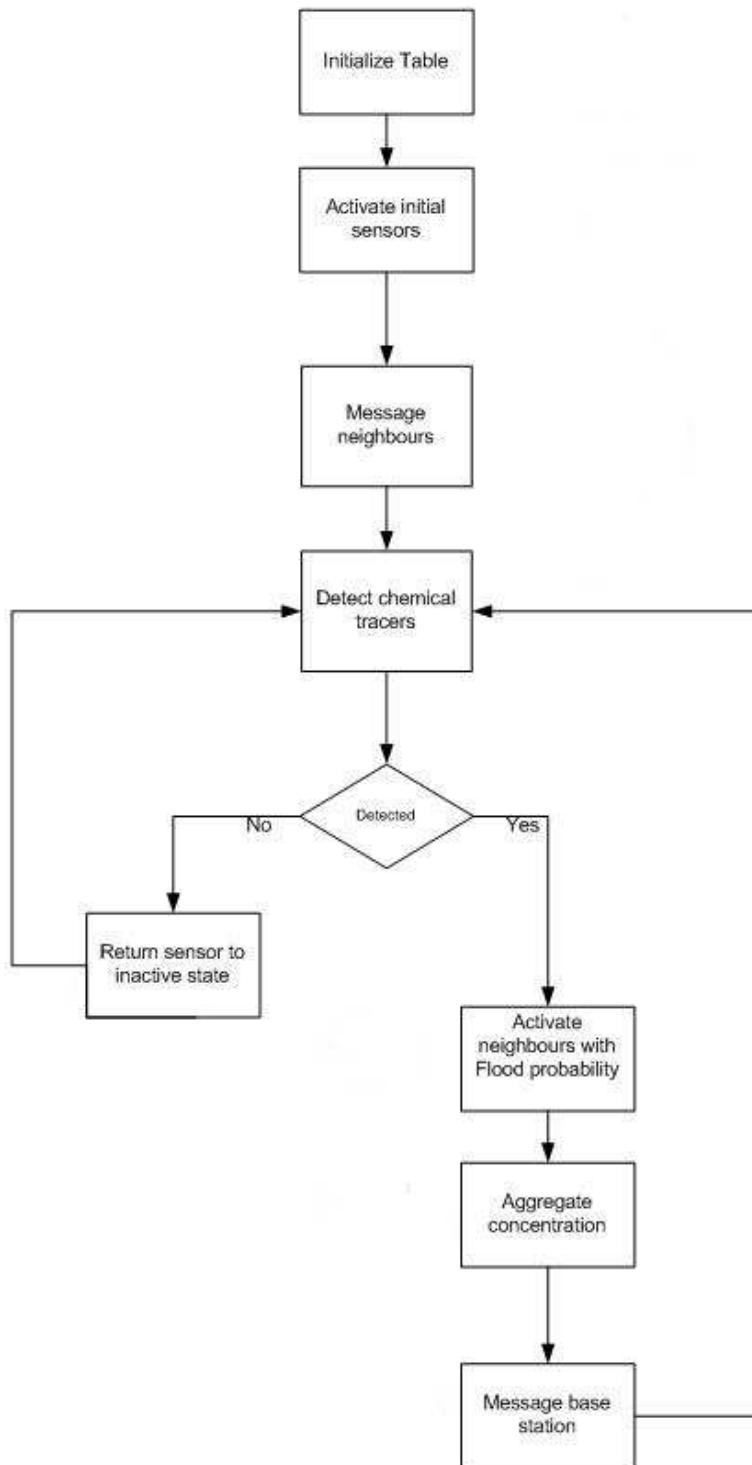


Figure 5.2: *Gossip-base sensor activation protocol*

active sensors N_+/N is plotted against normalised elapsed time $\kappa = t/\tau^*$. The time evolution of the fraction of active sensors (N_+/N) obtained from simulations shows sigmoidal patterns similar to those obtained by Eq. (5.2.2). We empirically estimated the corresponding calibration constant G by fitting simulated data to the analytical equation. Here we measure the onset of an IE with a fraction of active sensors N_+/N .

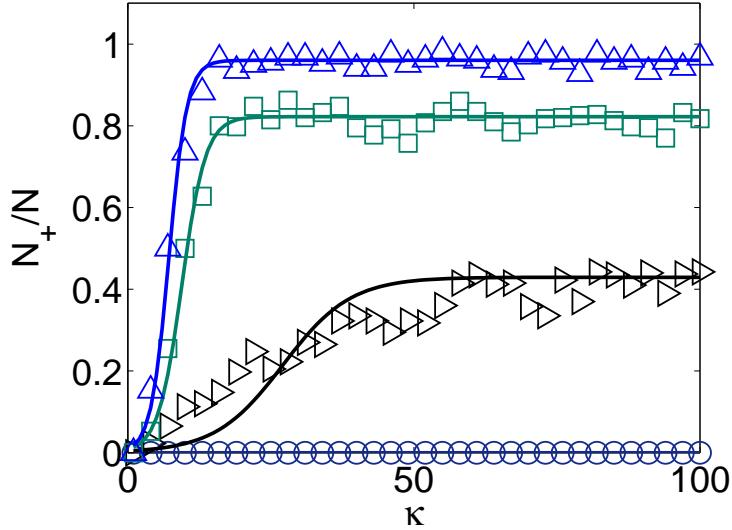


Figure 5.3: The effect of fanout ratio (f) on the onset of IEs (N_+/N) in a WCSN with a regular grid topology in the presence of a chemical tracer field. The range of values used for the f are: $f = 0.0$ (\circ), $f = 0.4$ (\triangleright), $f = 0.6$ (\square) and $f = 1.0$ (\triangle). The analytical fitted curves are shown in solid lines. $\kappa = t/\tau^*$ is the normalised time where τ^* is the detection period of a sensor.

As shown in Figure 5.3, the good agreement between analytical and simulated results validates our simulation model. We see a rise in the saturation level of the IE and a decrease of response time τ when fanout ratio f is increased.

5.4 Network performance metrics

The network performance objectives defined in [14] are low Power consumption, low cost, worldwide availability, network type, security, data throughput, message latency, mobility. Because hazardous chemical releases are rare events, if all sensors in a WCSN are left in the active state continuously, it would result in significant power consumption. Therefore, dynamic sensor activation while preserving reliable detection capability is crucial for the longevity of WCSNs.

5.4.1 Network activation

The number of active sensors is a important characteristic of a WCSN performance. However there is no importance if sensors in the most required region is not activated or confined to limited area. The diversive sensor activation will enable to detection rate even though less sensors are activated. From The sensor interaction rate is also important and correlation plays a major role in faster activation of sensors.

5.4.2 Concentration detection

Most important necessity of the network is to monitor and inform the control centre of the hazard at the least possible time. It is valuable that WCSN provides stable measurements. We characterise the unitised risk of detection (this is referred as relative standard error in [60]) as the coefficient of variation of concentration detection C by active sensors $\Psi = \sigma_C/\mu_C$, where μ_C is the time-averaged mean of C (i.e., around the saturation level), and σ_C is the time-averaged fluctuations of C . Our intuitive assumption is that σ_C need to be lesser than the chemical tracer concentration

variation in overall field as only sensor nodes in the higher concentration regions are activated and measurements obtained. We characterise the error of detection e as $e = (C - \bar{C})/\bar{C}$, where \bar{C} is the mean field value of C [6].

5.5 Simulation results

Describing the WCSN behaviour with analytical expressions provides a valuable insight into maximising the energy saving, network information gain and network performance optimisation. The simulation allows us to mimic the behaviour of a real system design based on analytical derivations. In this section, we describe and present the results of five experiments performed to optimise the energy efficiency of WCSNs while maintaining a good network performance. In these experiments, simulations were carried out using WCSN models in the presence of simulated chemical tracer fields.

5.5.1 Experiment 1: Network activation

The effect of the spatial correlation of tracer distribution on the sensor network's performance with different correlation radius R_{corr} was evaluated in the first experiment. WCSNs of regular grid topologies were simulated in the presence of a simulated chemical tracer field. The number of sensors was set at $N = 400$, communication radius was fixed at $r^* = 20$ or $\delta = 0.08$, and the sensor detection threshold was held at $C^* = C_0$. Simulations were performed for several values of correlation radius R_{corr} (or equivalently, the non-dimensional parameter ϵ); specifically, 19 values of ω ranging from 0.0 to 0.2 however here we show only plots for four values for clarity. Each

scenario was simulated 50 times and the ensemble-averaged results of IE are shown in Figure 5.4.

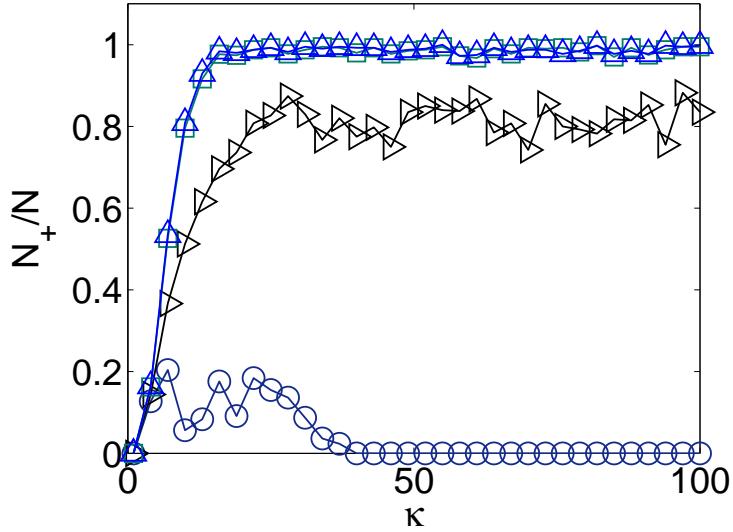


Figure 5.4: The effect of correlation characteristics of chemical tracer field (ϵ) on the onset of IEs (N_+/N) in a regular grid WCSN. The range of values used for the non-dimensional communication radius parameter ϵ , defined in Eq. (4.2.3), are: $\epsilon = 4.93$ (\circ), $\epsilon = 2.08$ (\triangleright), $\epsilon = 0.51$ (\square), $\epsilon = 0.28$ (\triangle). $\kappa = t/\tau_*$ is the normalised time where τ_* is the detection period of a sensor.

We see a fall of saturation level Z^* of the IE (N_+/N) in highly correlated chemical environments. This has been studied well in research work presented in Chapter 4. We observe that when ϵ is greater than one where communication radius of sensor is shorter than the correlation radius R_{corr} the IE tends to ‘die out’.

5.5.2 Experiment 2: Improving network activation

As presented in Chapter 4, we found random network topology and spatial correlation of tracer distribution (which is typical for most operational scenarios) to have a negative effect on the performance of a WCSN with the epidemiology-based sensor

activation protocol. It manifests in longer network detection time and a lower fraction of active sensors, which is detrimental to good detectability.

Based on our intuitive assumption that by extending the communication radius r^* and incorporating a gossip-ratio f , we should be able to achieve a better activation rate while saving energy by not activating sensors that are not necessary for detection. We examined the characteristic in the second experiment and the results are shown in Figure 5.5. WCSNs of regular grid topologies were simulated in the presence of a simulated chemical tracer field. The number of sensors was set at $N = 400$, the sensor detection threshold was held at $C^* = C_0$, and three values of communication radius $\delta = 0.056$, $\delta = 0.160$ and $\delta = 0.480$ were used. Simulations were performed for several values of correlation radius R_{corr} (or equivalently, the non-dimensional parameter ω); specifically, 19 values ranging from 0.0 to 0.2 and for fanout ratio $f = 0.5$ and $f = 1.0$. Each scenario was simulated 25 times and the IE saturation values Z^* derived from ensemble-averaged results are shown.

We see a drop of saturation level Z^* of the IE (N_+/N) in highly correlated chemical environments but we managed to overcome the issue by extending the communication radius beyond the correlation radius. We can say that for an IE to occur, following criteria need to be satisfied. If there are two sensors s_1 and s_2 , chemical contaminant plume location p_1 and at the next time step it is at location p_2 .

1. $d_{s_1, s_2} \leq r^*$.
2. $C(s_1) \geq C^*, C(s_2) \geq C^*$.
3. for (2) to be true, $d_{p_1, s_1} \leq R_{corr}, d_{p_2, s_2} \leq R_{corr}$

Since shortest distance between two points is the straight line connecting them, d_{s_1, s_2}

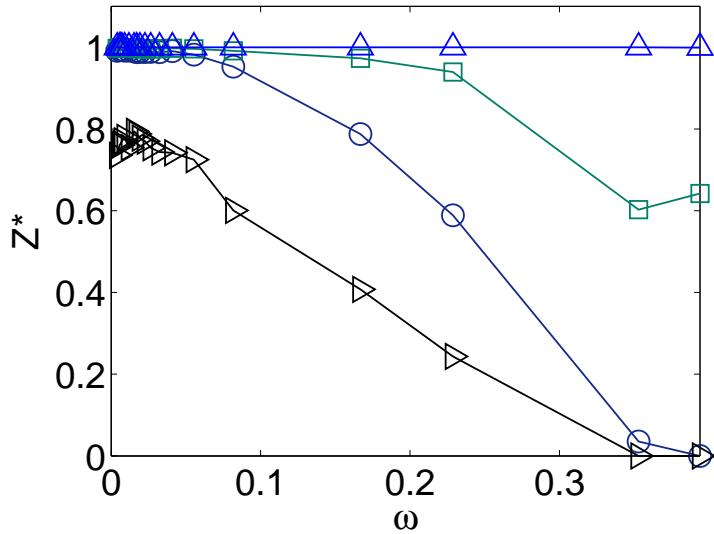


Figure 5.5: The effect of correlation characteristics of chemical tracer field (ϵ) on the saturation value Z^* of (N_+/N) with different fanout ratio f in a regular grid WCSN. The values used for the non-dimensional communication radius parameter δ , defined in Eq. (4.2.2) and fanout ratio f , are: $\delta = 0.056$, $f = 1.0$ (\circ), $\delta = 0.056$, $f = 0.5$ (\triangleright), $\delta = 0.160$, $f = 0.5$ (\square) and $\delta = 0.480$, $f = 0.5$ (\triangle).

is the straight line connecting two sensors s_1 and s_2 . Also $d_{s_1,s_2} \leq 2R_{corr} + d_{p_1,p_2}$, where d_{p_1,p_2} is the temporal correlation distance, $C(s_1)$ is the chemical concentration detected at sensor location s_1 , $C(s_2)$ is the chemical concentration detected at sensor location s_2 .

5.5.3 Experiment 3: Optimal fanout ratio

Further extending facts on our intuitive assumption, the effect of the fanout ratio on the sensor network's performance with different sensor communication ranges r^* or normalised parameter δ was examined in the third experiment. WCSNs of regular grid topologies were simulated in the presence of a simulated chemical tracer field with a correlation coefficient of $\omega = 0.08$. The number of sensors was set at $N = 400$, and the sensor detection threshold was held at $C^* = C_0$. Simulations were performed for several values of communication radius r^* (or equivalently, the non-dimensional parameter δ); specifically, the following δ values were used: $\delta = 0.056$ ($\epsilon = 1.43$), $\delta = 0.080$ ($\epsilon = 1.00$), $\delta = 0.160$ ($\epsilon = 0.50$), and $\delta = 0.480$ ($\epsilon = 1.43$) ($\epsilon = 0.17$) with 26 values of f ranging from 0 to 1.0. Each scenario was simulated 100 times and the IE saturation values Z^* derived from ensemble-averaged results for sensor activation are shown in Figure 5.6.

We see a rise in the saturation level Z^* of the IE (N_+/N) when the communication radius r^* (or δ) is increased or when the fanout ratio (f) is increased. This can be explained with the analytical expression obtained in [93] modified and described by Eq. (5.2.2). We also observe that some low values of δ and f do not support an IE; i.e., the IE ‘dies out’, as the condition for an IE is not satisfied in these cases.

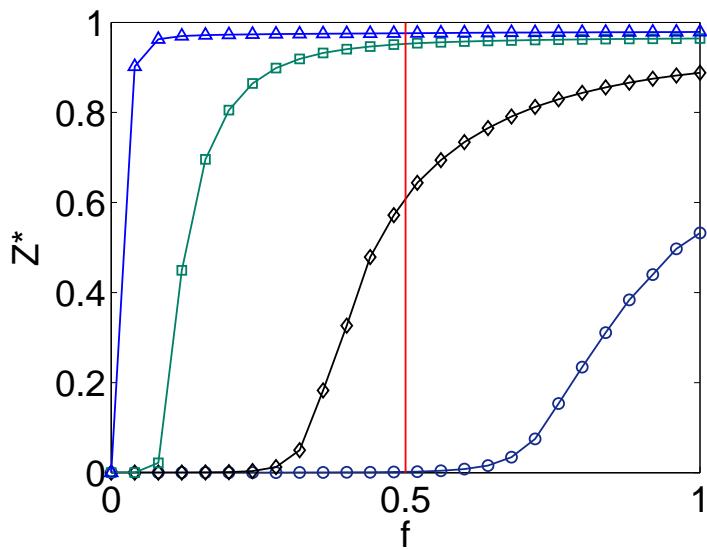


Figure 5.6: The effect of the fanout ratio (f) on the saturation value Z^* of (N_+/N) with different communication ranges (δ) in a WCSN with a regular grid topology in the presence of a chemical tracer field. The range of values used for the non-dimensional communication radius parameter δ , defined in Eq. (4.2.2), are: $\delta = 0.056$ (\circ), $\delta = 0.080$ (\diamond), $\delta = 0.160$ (\square), $\delta = 0.480$ (\triangle). The solid line shows $f = 0.5$.

5.5.4 Experiment 4: Chemical detection

In the fourth experiment, we evaluate the unitised risk of detection for the results obtained by running the same experiment as stated in the Section 5.5.1 ; the results being shown in Figure 5.7. Plots of Ψ versus f corresponding to several values of communication ranges δ are shown.

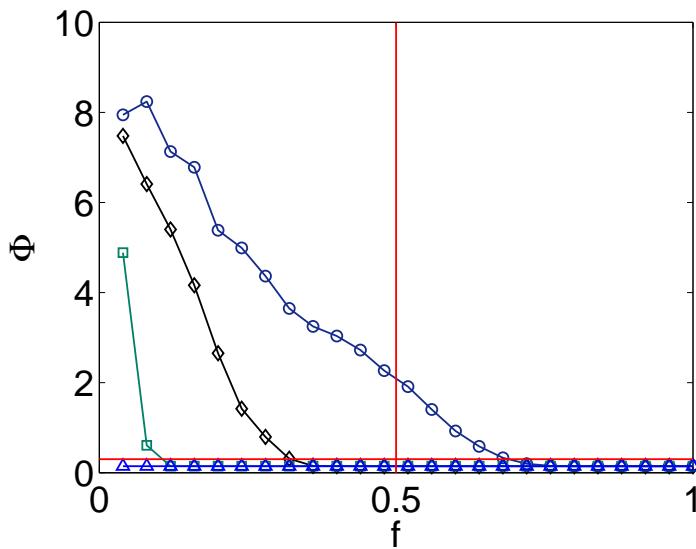


Figure 5.7: The effect of the fanout ratio (f) and communication radius (δ) on the standard error of detection (Ψ) of chemical tracer detection by a WCSN with a regular grid topology in the presence of a chemical tracer field. The values used for the non-dimensional communication radius parameter δ , defined in Eq. (4.2.2), are: $\delta = 0.056$ (\circ), $\delta = 0.080$ (\diamond), $\delta = 0.160$ (\square), $\delta = 0.480$ (\triangle). The solid line shows $f = 0.5$ and $\Phi = 0.3$.

It is necessary that the detecting WCSN should provide values closer to the mean, here we explore error encountered in detection. Plots of e versus f corresponding to several values of communication ranges δ are shown in Figure 5.8 for the WCSN.

From these plots, we can see that even when we decrease f by increasing r^* , we can obtain reliable readings from the WCSN, and we can estimate the optimal

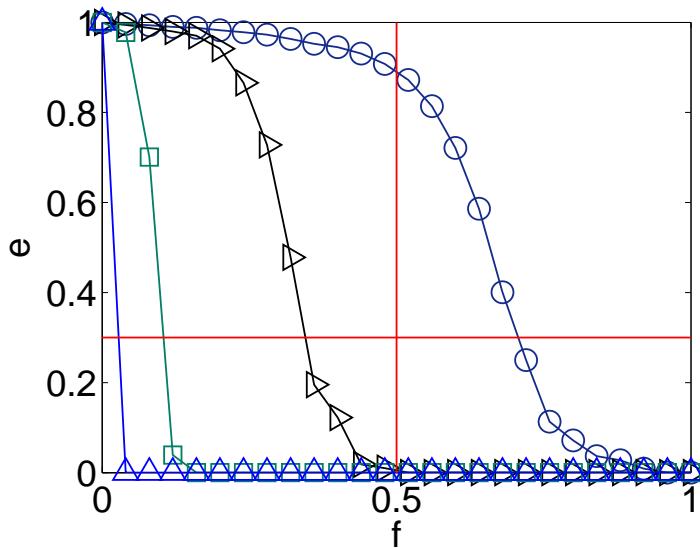


Figure 5.8: The effect of the fanout ratio (f) and communication radius (δ) on the error of detection (e) by a WCSN with a regular grid topology in the presence of a chemical tracer field. The values used for the non-dimensional communication radius parameter δ , defined in Eq. (4.2.2), are: $\delta = 0.056$ (\circ), $\delta = 0.080$ (\triangleright), $\delta = 0.160$ (\square), $\delta = 0.480$ (\triangle). The solid line shows $f = 0.5$ and $e = 0.3$.

points where Ψ reaches zero or no fluctuations which evident in error of detection e as well. Depending on a particular operational scenario, some other parameters (energy conservation, WCSN activation time, sensor interaction rate, budget, etc.) may become more important and would favour a particular technical solution.

5.5.5 Experiment 5: Optimal network parameters

The fifth experiment was performed to derive a cost function to optimise energy efficiency in a WCSN. A WCSN of a regular grid topology was simulated in the presence of a simulated chemical tracer field with a correlation ($\omega = 0.08$). The number of sensors was held at $N = 400$ and the detection threshold was fixed at $C^* = C_0$. Simulations were done with the communication radius r^* varied from $r^* = 15$ ($\delta = 0.06$) to $r^* = 315$ ($\delta = 1.26$) and the fanout ratio f was varied from 0 to 1. Each scenario was simulated 25 times, varying the initially active sensors used for bootstrapping of IE in the regular grid WCSN. Plots of Ψ versus f corresponding values of communication ranges δ are shown in Figure 5.9 for the WCSN.

In the experiment, we observed that better performance is achieved when the sensor communication radius r^* or fanout ratio f is increased. We see that R^2 values for both cases are below the acceptable level 0.80 and complexity in fitting a surface to derive an empirical equation. With the fitted surface, $\Phi = b_1 b_2 f + b_3 \delta + b_4 f \delta + b_5 f^2 + b_6 \delta^2$, where $b_1 = 1.79$, $b_2 = -2.19$, $b_3 = -2.87$, $b_4 = 1.60$, $b_5 = 0.71$, $b_6 = 1.25$. We obtained $f = 0.86$, $\delta = 0.60$ for the minima using genetic algorithm optimisation (GA) function in Matlab®.

The another experiment was performed to derive the minima for the WCSN. A WCSN of a regular grid topology was simulated in the presence of a simulated

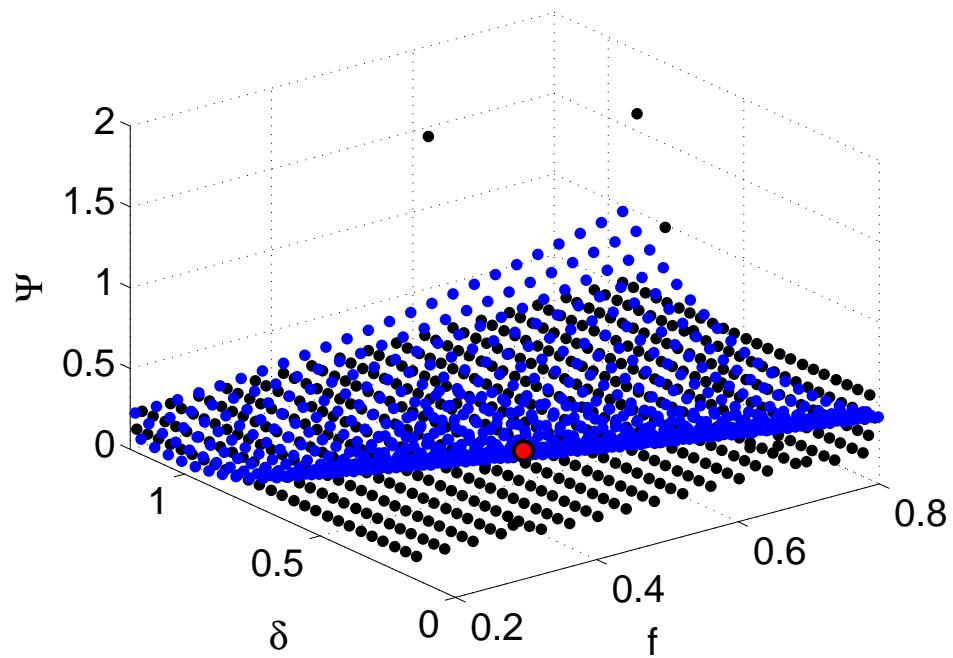


Figure 5.9: The effect of the fanout ratio (f) and non-dimensional communication parameter (δ) on the relative standard error of detection (Ψ) of a WCSN with a regular grid topology and in the presence of a chemical tracer field shown as black diamond (\diamond), a fitted surface shown as blue box (\square) and minima as a red filled circle (\circ).

chemical tracer field with a correlation ($\omega = 0.08$). The number of sensors was held at $N = 400$ and the detection threshold was fixed at $C^* = C_0$. Simulations were done with the communication radius r^* varied from $r^* = 14$ ($\delta = 0.06$) to $r^* = 350$ ($\delta = 1.26$) and the fanout ratio f was varied from 0.3 to 1. We used the Matlab based GA integrated with our WCSN simulation framework. For four initial active sensors were used for bootstrapping the network. The points explored by GA is shown in Figure 5.10 for the WCSN. We obtained $f = 0.80$, $\delta = 0.29$ as the coordinates for the minima.

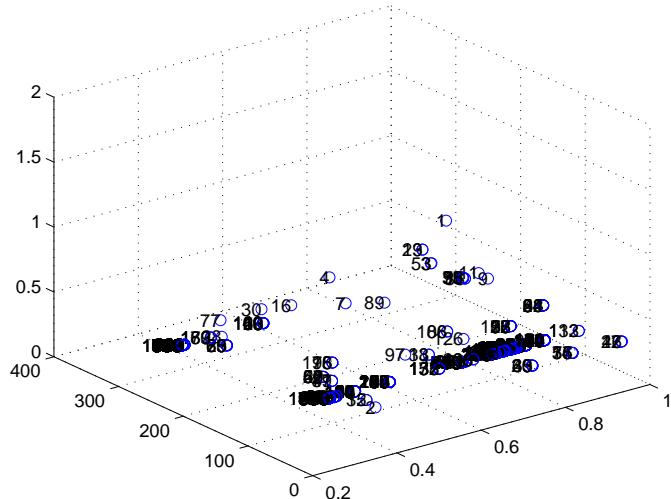


Figure 5.10: Locating the optimal parameters for the WCSN with a regular grid topology and in the presence of a chemical tracer field.

5.5.6 Experiment 6: Decentralised aggregation

In this study, we try to examine the performance of aggregation of detection by the WCSN. Consider a network of N nodes, with node i having a neighbour table, K_i with θ vectors mentioned in previous section, of size $k_i = |K_i|$. Node $j \in K_i$ if node

j is contained in node i 's neighbor table. The network is therefore a un-directed graph and in our work we consider the case where a self loop is not present. Also, $K'_i = \{u | i \in K_u\}$ is the set of nodes for which i is a neighbour. In our case as we consider undirected communication, $K_i = K'_i$. Let C_i be a value at activated node i . Assuming the network is stable (no nodes join or leave) and the node values remain unchanged during the computation period, the system average computed over all nodes is:

$$\bar{\theta}_i = \frac{1}{r} \sum_{j=1}^r \theta_j, \quad (5.5.1)$$

where r denotes the number of concentration values collected by the node i .

$$\bar{C}_t = \frac{1}{N_i} \sum_{i=1}^{N_i} \bar{\theta}_i, \quad (5.5.2)$$

\bar{C}_t denotes the mean concentration value at an instance of an epoch t , N_i denotes the number of active sensor detecting chemical tracer concentration in an epoch.

$$\mu_C = \frac{1}{t_{max}} \sum_{t=1}^{t_{max}} \bar{C}_t, \quad (5.5.3)$$

μ_C denotes the mean concentration value of in a WCSN instance averaged over the time period.

We adjust the communication radius (r^*) and the fanout ratio (f). The experiment was performed to assess the impact of the gossip fanout ratio on centralised and decentralized data aggregation techniques. The WCSN with a regular grid topology was simulated in the presence of a simulated chemical tracer field with correlation ($\omega = 0.08$). The number of sensors was held at $N = 400$ and the detection threshold was fixed at $C^*/C_0 = 1$. Simulations were done with the communication radius

r^* set to following values: $r^* = 15$ ($\delta = 0.16$) to $r^* = 315$ ($\delta = 1.26$) and the fanout ratio f was varied from 0 to 1. Each scenario was simulated only once. In order to overcome the issue of the positioning of the initially active sensors, we run simulations multiple times, changing the positions each time. However, if we use multiple iteration, some aggregation values will have zeros and will give odd values for overall aggregation. Therefore the figures 5.14 and 5.12 are from a single iteration. In the centralised approach, chemical tracer concentration are collected from the initiation from all the active sensors, in the decentralised approach, each active sensor stores the historical chemical tracer concentration values and sends the aggregated value to the base station.

The Figure 5.14 show stable detection value from the centralised aggregation and lesser stability from decentralized aggregation. As presented in Chapter 4, we note a rise in the saturation level of the IE and a decrease of response time τ when the communication radius r^* (or δ) is increased or δ is decreased. We see that some low values of δ do not support an IE; i.e., the IE ‘dies out’, as the condition for an IE is not satisfied in these cases therefore detection will not happen.

In Figure 5.12, We can see a normally distributed detection values for $f = 0.30$ and $f = 0.60$. Unlike the centralised approach, in a decentralised approach, it will be necessary to activate the sensors in the high concentration regions. The two approaches gives similar results and shows feasibility of using the decentralised approach. This prevents from data losses from sensor detection anomalies and also prevents data input implosions at centralised base stations.

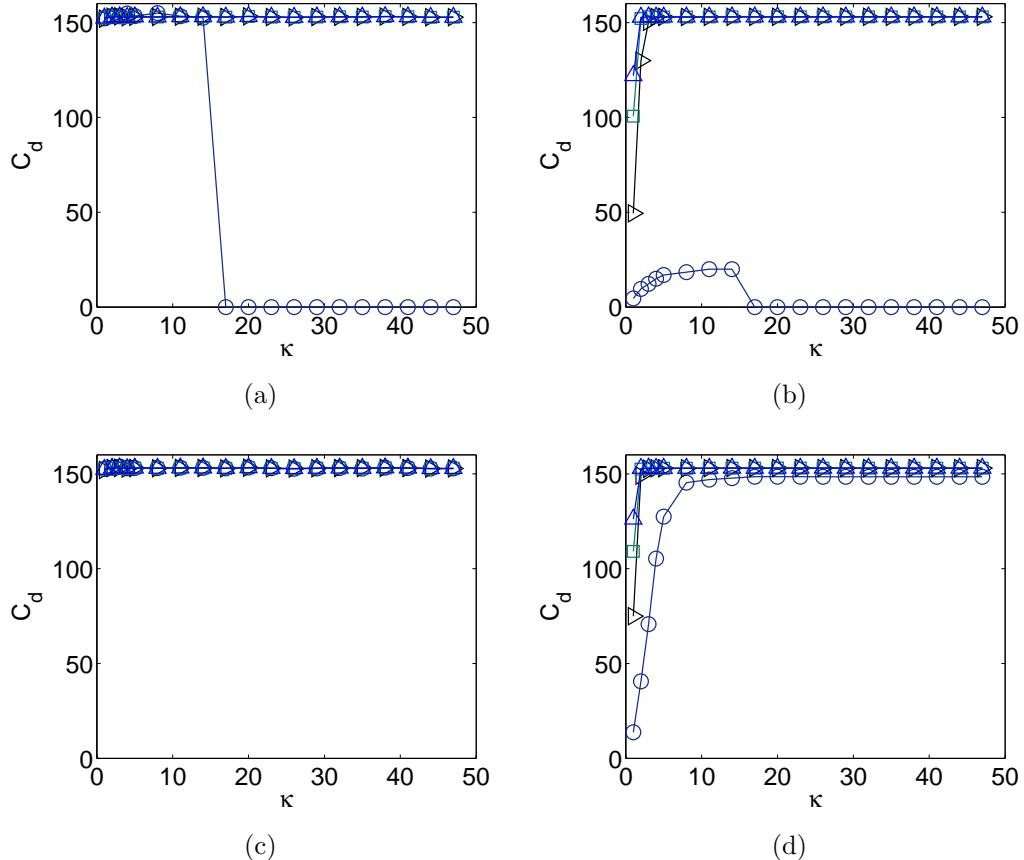


Figure 5.11: Chemical tracer detection value aggregation for a WCSN with regular grid topology in the presence of tracer field: and when : (a) Centralised aggregation, fanout ratio $f = 0.15$; (b) Decentralised aggregation, fanout ratio $f = 0.15$; (c) Centralised aggregation, $f = 0.20$; (d) Decentralised aggregation, $f = 0.20$. The chemical fields are characterised by $\delta = 0.06$ (\circ), $\delta = 0.32$ (\diamond), $\delta = 0.58$ (\square), and $\delta = 0.83$ (\triangle). $\kappa = t/\tau^*$ is the normalised time where τ^* is the detection period of a sensor.

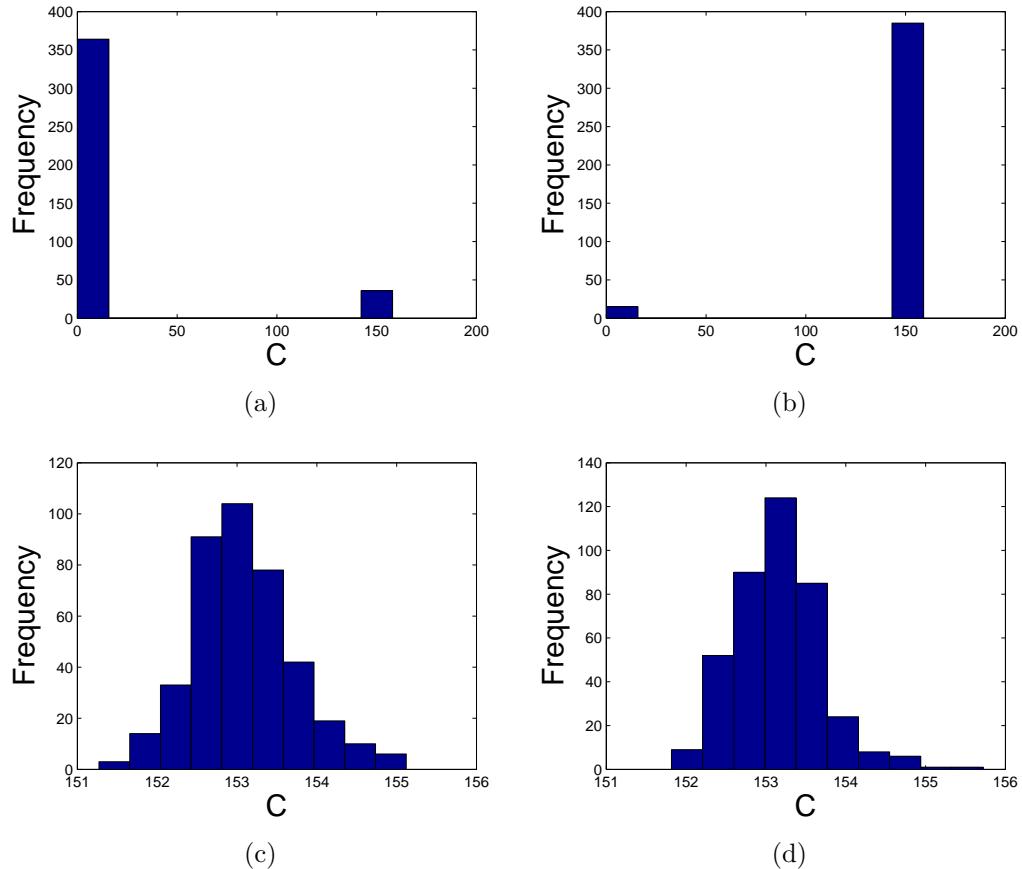


Figure 5.12: The distribution of Chemical tracer detection value aggregation for a WCSN with regular grid topology in the presence of tracer field. (a) $f = 0.20, \kappa = 0$, (b) $f = 0.20, \kappa = 10$ (c) $f = 0.30, \kappa = 10$, (d) $f = 0.60, \kappa = 7$.

5.5.7 Experiment 7: Effect of false alarm

We used the CHEMSEN; epidemiological protocol [1] based chemical tracer detection simulation system. As system parameters we used $N = 400$, $C* = 150$, $r* = 14, 20, 40$, False negative alarms from active sensors $F_n = 0.05$ and False positive alarms from active sensors $F_p = 0.10$. A Correlation environment has chemical tracers with a correlation radius $R_{corr} = 20.44$ or $\omega = 0.08$ and non-correlated environment has tracers with $R_{corr} = 0.0$ or $\omega = 0.0$. The side length L of field is 250. The mean concentration Co of the chemical tracer environment is 150. We use intermittent tracers. Each scenario was simulated 100 times and ensembled averages were obtained. To reduce time we ignore further simulations when no active sensor can detect chemical tracers. When a fraction of sensors is active, and detects we remove F_p from number of detecting sensors and add F_n of sensors to the number of sensors detecting from number of active sensors. Figures 5.13 and 5.14 show the fraction of active sensors with different communication radii with and without false alarm correction.

We observe the saturation level of the IE to drop slightly and the response time (τ) to increase when faulty sensors are present. We see less effect of false alarm rate on the detection of chemical tracer concentration. Here we considered faulty sensors with the epidemiological algorithm. The false alarm positive and negative fractions we used $F_p = 0.1$, $F_n = 0.05$ do not affect the performance of the WCSN.

We can see a note worthy effect on the sensor activation by faulty sensors when communication radius is decreased. When communication radius is increased, the number of links from a sensor also increased resulting greater number sensors active.

We integrated the false alarm properties to WCSN model and we used a selected false alarm fractions to test our system. The performance of WCSN was lesser when

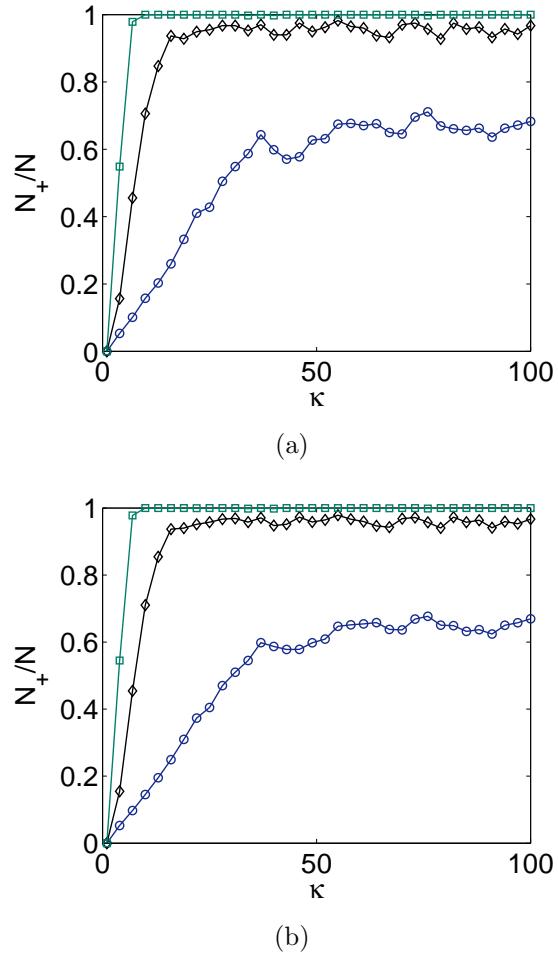


Figure 5.13: The effect of F_{out} on the onset of IEs (N_+/N) in a WCSN with regular grid topology in the presence of a correlated tracer field: (a) without false alarms (b) with false alarms. The range of values used for are: $\delta = 0.06$ (\circ), $\delta = 0.08$ (\diamond), $\delta = 0.16$ (\square). $\kappa = t/\tau^*$ is the normalised time where τ^* is the detection period of a sensor.

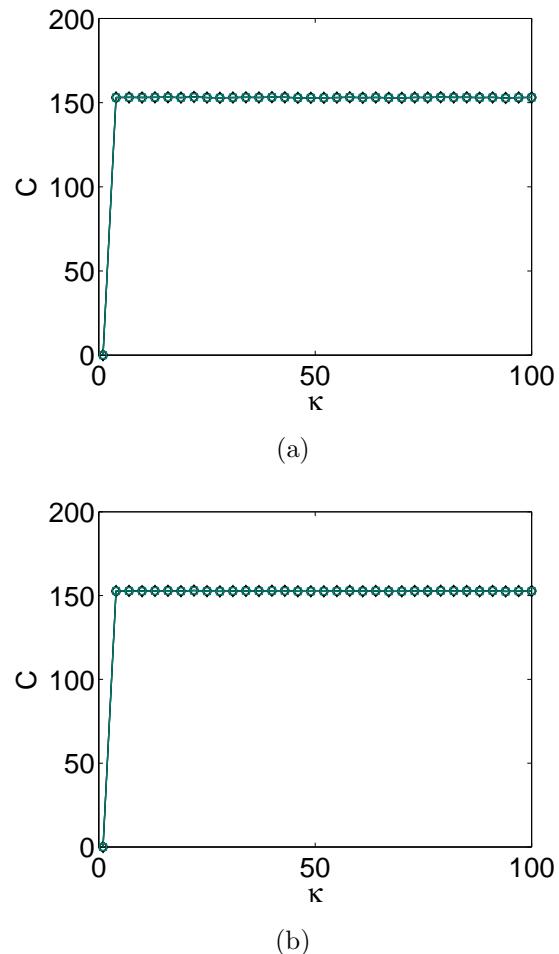


Figure 5.14: The effect of false alarms on the chemical concentration detected by active sensors in a WCSN with a regular grid topology in the presence of a correlated tracer field: (a) without false alarms (b) with false alarms. The range of values used for are: $\delta = 0.06$ (\circ), $\delta = 0.08$ (\diamond), $\delta = 0.16$ (\square). $\kappa = t/\tau^*$ is the normalised time where τ^* is the detection period of a sensor.

communication radius was decreased. We will be studying further to identify the threshold false alarm fraction with coefficient of variance [3].

5.6 Summary

In this chapter, we investigated the behaviour of a WCSN with a bio-inspired, gossip-based sensor activation protocol in a turbulent chemical tracer field, which is highly complex in nature. Using a bio-inspired analytical model for a WCSN, initially introduced by Skvortsov *et al.* [93], we verified the simulation framework for WCSNs (Fig. 5.3). Based on simulation results, we evaluated some important performance metrics: the number of active sensors, as shown in Figure 5.6, and the unitised risk of detection, as shown in Figure 5.7. This metric is ideal to avoid unnecessary sensor activation which are power greedy and not contributing to the overall detection. We tried to obtain the optimal parameters by fitting a surface to the plot. The strategy employed can be used to overcome nature of the chemical tracer field itself, involving blobs of high concentration, which is a direct consequence of mass conservation and hindrance to good sensor activation (see Fig. 5.1). We considered only a single chemical threat, but in reality there can be various chemical agents present and multiple activation protocols can be employed to address different scenarios. Based on our results, we can consider extending the communication range of sensors, rather than activating a greater number of neighbouring sensors, which offers a more energy-efficient strategy [38]. By using a deterministic gossip-based protocol, sensors can be equipped with intelligence when selecting neighbours for activation. Studies on gossip based protocols can also be applied to a real system where sensor communication links become dysfunctional over time due to aging. In this regard, here we

used a centralised aggregation of detected concentration values, we can incorporate a gossip-based aggregation protocol [81]. This approach also helps to store the localised chemical tracer concentration data and enables to map the concentrations within the affected area. In future studies we hope to use a bio-inspired optimisation technique to maximise the energy efficiency of a WCSN with network configuration parameters [53].

We integrated the false alarm properties to WCSN model, and we used selected false alarm fractions to test our system. The performance of WCSN was lesser when the communication radius was decreased. We will be studying further to identify the threshold false alarm fraction with the coefficient of variance. The research work presented so far explained the DSC protocol improved the detection capability while achieving good energy saving and information dissemination. We explored various methods of efficient hazardous detection with collaborative approaches, in the following chapter, we explore the possibility of detecting hazard source locations.

Chapter 6

Bio-inspired algorithms for radioactive source localisation

This chapter ¹ considers localisation of point sources of gamma radiation using dose rate measurements. Binary and continuous genetic algorithms (GA) were used to implement maximum likelihood estimation (MLE) of position and strength of point radiation sources. MLE was achieved by minimising the objective function which computes the negative log likelihood. Real experimental data collected during a DSTO-conducted field trial were used to test the performance of the algorithms. The performance of GA-based implementation was compared to an implementation that uses the MATLAB built-in routine *fminsearch*. Source parameters estimated by the algorithms were also compared to the theoretical bounds obtained using Cramer-Rao

¹Presentations/Publications:

- C. Mendis, A. Gunatilaka, B. Ristic, S. Karunasekera and A. Skvortsov, ‘Experimental verification of evolutionary estimation algorithms for radioactive source localisation’, in Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 2009.
- A. Gunatilaka, B. Ristic, C. Mendis, S. Karunasekera and A. Skvortsov, ‘The effect of data collection geometry on radiological source localisation’, in Fifth International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Melbourne, Australia, 2009.

bound (CRB) analysis which quantifies the accuracy with which it is possible to localise the source and estimate its strength. All three implementations localised a single point source well, nearly approaching the CRB. Reasonable position estimates were achieved for two and three source cases, but the source strength estimates were found to have much larger RMS errors than what is predicted by the CRB. While the GA-based implementations took longer to converge compared to the *fminsearch* algorithm, they encountered fewer divergent runs than the latter algorithm.

6.1 Introduction

Numerous incidents involving loss or theft of radioactive sources have been reported [79]. There is a major concern that such sources could end up in the hands of terrorist networks and be used for building radiological dispersion devices. Therefore, it is of interest to be able to locate hidden radioactive sources. Gamma rays, which are highly penetrating electromagnetic radiations emitted by some radioactive material, can be detected at large standoff distances, and measurements collected using gamma detectors can be used to localise the radiation sources. Assuming an area has been identified where sources of gamma radiation potentially exist, we consider the problem of estimating the locations and strengths of sources using dose rate measurements.

Detection and localisation of point-sources of gamma radiation have been studied recently by several authors [76, 96, 13, 82]. For example, the problem of detecting radiation sources carried in a vehicle by using a distributed sensor network placed along a section of a road is considered in [76, 96] and [13]. Maximum likelihood estimation (MLE) implemented using the *fminsearch* routine in MATLAB[®], extended Kalman filter (EKF) and unscented Kalman filter (UKF) were applied to the problem

of localising point radiation sources in [42]. The MLE provided the best source estimates, asymptotically approaching the theoretical bound predicted by the Cramer Rao bound. UKF results were worse than MLE and the EKF failed to produce acceptable estimates. Rao et al [82] proposed the mean of estimator methods (MOE). Chin et al [20] proposed iterative pruning (ITP) to localise a single point radiation source, which is claimed to be more accurate than MOE and more efficient than the MLE. Although a *fminsearch*-based implementation of MLE worked well for localising one or two sources, it was found to be unsatisfactory for locating three or more sources [71]. A progressive correction-based Bayesian approach is also used to localise multiple point sources in [71]. This chapter explores the use genetic algorithms (GA) for localisation of multiple radiological point sources.

Three sets of real data acquired during a field trial in Puckapunyal Military Area (Victoria, Australia) are used in the study. The data sets were collected using the DSTO-developed Low Cost Advanced Radiological Survey (LCAARS) system in the presence of one, two, and three radiological point sources of unequal strengths and well separated in space. The chapter presents the parameter estimation results obtained by implementing MLE using binary GA (GA-b) and continuous GA (GA-c) for all three data sets and compares them with corresponding Cramer-Rao bounds. We also compare the performance of the GA implementation of MLE algorithm to that of the *fminsearch*-based implementation.

The chapter is organised as follows. Section 6.1.1 formulates the problem and Section 6.2 describes the maximum likelihood estimation and Cramer-Rao bound. Section 6.4 describes how the MLE was implemented using the GA. Section 6.5 discusses the evaluation of algorithms, and results are presented and discussed in Section

6.6. Section 6.7 draws conclusions.

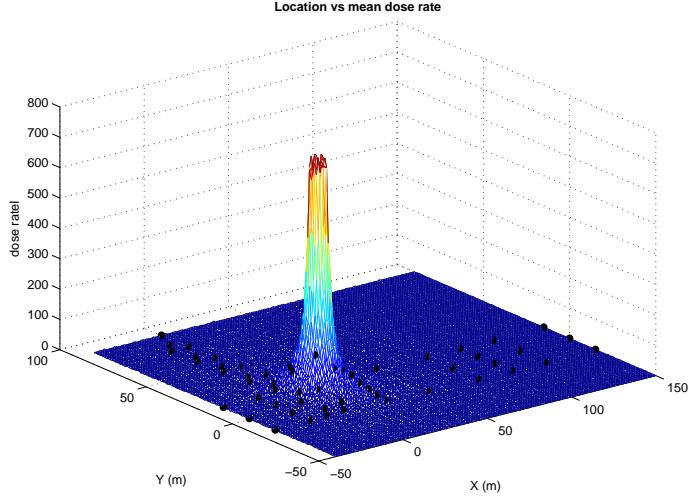


Figure 6.1: *Distribution of mean dose rate. Black dots indicates Measurements*

6.1.1 Problem formulation

Assume that $r \geq 0$ sources are present in the area of interest. In this chapter, we assume r to be known. Let $\boldsymbol{\theta}_i = [x_i \ y_i \ \alpha_i]^\top \in \mathbb{R}^2 \times \mathbb{R}^+$, $i = 1, \dots, r$, denote the parameter vector of the i th source where (x_i, y_i) is the source position in Cartesian coordinates and α_i is the source strength. The source parameter vectors are collected into a stacked vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^\top \ \dots \ \boldsymbol{\theta}_r^\top]^\top$.

For $j = 1, \dots, m$ measurements $z_j \in \mathbb{Z}^+$ of radiation dose are taken at locations $(\xi_j, \zeta_j) \in \mathbb{R}^2$.

The following assumptions are made:

- the gamma probe has a uniform directional response;
- radiation sources are point sources emitting uniformly in all directions;

- radiation measurements are independently distributed;
- each measurement is taken with equal exposure time.

The joint density of the measurement vector $z = [z_1 \dots z_m]^\top$ conditional on the parameter vector $\boldsymbol{\theta}$ and the knowledge that r sources are present, can then be written as stated in,

$$l(z|\boldsymbol{\theta}) = \prod_{j=1}^m P(z_j; \lambda_j(\boldsymbol{\theta})), \quad (6.1.1)$$

where $P(z; \lambda) = e^{-\lambda} \lambda^z / z!$ is the Poisson probability density function evaluated at $z \in \mathbb{Z}^+$ with parameter λ , and $\lambda_j(\boldsymbol{\theta})$ is the mean radiation count for the j th sensor location:

$$\lambda_j(\boldsymbol{\theta}) = \lambda_b + \sum_{i=1}^r \alpha_i / d_{j,i}, \quad (6.1.2)$$

with

$$d_{j,i} = (\xi_j - x_i)^2 + (\zeta_j - y_i)^2 \quad (6.1.3)$$

being the squared distance between the source i and the j th sensor. The constant λ_b in (6.1.2) represents the average count due to the background radiation only, and is assumed to be known.

The problem is to estimate the parameter vector $\boldsymbol{\theta}$ using the measurement vector z .

6.2 Estimation Algorithms

This section reviews the maximum likelihood (MLE) estimation algorithm and the theoretical estimation error bound.

6.2.1 Cramér-Rao bound

In many estimation problems optimal parameter estimates can be found only through the use of numerical techniques, the reliability of which cannot be guaranteed. Theoretical error performance bounds provide a way of assessing these numerical techniques even before any measurements are collected. The most popular performance bound is the Cramér-Rao bound (CRB) which places a lower bound on the variance of unbiased estimators. In particular, for an unbiased estimator $\hat{\boldsymbol{\theta}}$ of the deterministic parameter vector $\boldsymbol{\theta}$ the CRB for $\text{cov}(\hat{\boldsymbol{\theta}})$ is \mathbf{J}^{-1} where \mathbf{J} is the Fisher information matrix (FIM) [101, p.80],

$$\mathbf{J} = -\mathbb{E}[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \log l(z|\boldsymbol{\theta})] \quad (6.2.1)$$

with $\nabla_{\boldsymbol{\theta}}$ the gradient operator with respect to $\boldsymbol{\theta}$ and $l(z|\boldsymbol{\theta})$ the likelihood function of the measurement vector z given by (6.1.1). The CRB holds in the sense that $\text{cov}(\hat{\boldsymbol{\theta}}) - \mathbf{J}^{-1}$ is positive semidefinite.

It has been shown in [71] that the FIM can be written as:

$$\mathbf{J} = \sum_{j=1}^m \frac{\nabla_{\boldsymbol{\theta}} \lambda_j(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^T \lambda_j(\boldsymbol{\theta})}{\lambda_j(\boldsymbol{\theta})}, \quad (6.2.2)$$

where the partial derivatives in $\nabla_{\boldsymbol{\theta}} \lambda_j(\boldsymbol{\theta})$ can be found as

$$\frac{\partial \lambda_j(\boldsymbol{\theta})}{\partial x_i} = 2\alpha_i(\xi_j - x_i)/d_{j,i}^2, \quad (6.2.3)$$

$$\frac{\partial \lambda_j(\boldsymbol{\theta})}{\partial y_i} = 2\alpha_i(\zeta_j - y_i)/d_{j,i}^2, \quad (6.2.4)$$

$$\frac{\partial \lambda_j(\boldsymbol{\theta})}{\partial \alpha_i} = 1/d_{j,i}, \quad (6.2.5)$$

for $i = 1, \dots, r$ and $j = 1, \dots, m$.

The theoretical Cramér-Rao bounds are computed as diagonal elements of \mathbf{J}^{-1} for Tests 1, 2 and 3. The square-root values (which correspond to the lower bounds

of estimation error standard deviations) are shown in the first columns of Tables 6.3, 6.5 and 6.7. The resulting bounds indicate that it is possible in all three tests to accurately localise all sources (with standard deviation of the positional error in the order of a meter). The bound also shows that the standard deviation of source strength estimation error is relatively larger.

6.2.2 Maximum likelihood estimation

The MLE is widely used for parameter estimation because, if an asymptotically unbiased and minimum variance estimator exists for large sample sizes, it is guaranteed to be the MLE [101]. The MLE is determined as the vector $\boldsymbol{\theta}$ which maximises the likelihood function $l(z|\boldsymbol{\theta})$:

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg \max_{\boldsymbol{\theta}} l(z|\boldsymbol{\theta}). \quad (6.2.6)$$

In [42] and [71] MLE was implemented using the *fminsearch* routine in MATLAB®. Because there are three parameters per source, the dimensionality of the search space increases rapidly when multiple sources are to be localised, and it was difficult to estimate parameters well for more than two sources using this implementation.

Searching for minima in high dimensional spaces is often difficult because of the presence of local minima and search algorithms, particularly those based on gradient search, may find a local minimum rather than the global minimum depending on the initial value. Because evolutionary algorithms such as GA are known to be effective in searching in high dimensional spaces and less vulnerable to getting trapped in local minima, in this Chapter we explore the use of evolutionary algorithms to carry out MLE. Effectively, we performed maximum likelihood estimation by defining the

negative log likelihood function as the objective function and searching for the minimum of this objective function. The log likelihood function is obtained by taking the logarithm of (6.1.1) as shown below:

$$\log(l(z|\boldsymbol{\theta})) = \sum_{j=1}^m \log(P(z_j; \lambda_j(\boldsymbol{\theta}))), \quad (6.2.7)$$

$$= C + \sum_{j=1}^m [z_j \log \lambda_j(\boldsymbol{\theta}) - \lambda_j(\boldsymbol{\theta})], \quad (6.2.8)$$

where $C = \sum_{j=1}^m \log(z_j!)$ is a constant independent of $\boldsymbol{\theta}$.

6.3 Particle swarm optimisation

Particle swarm optimisation (PSO) is an evolutionary algorithm which is based on swarm of particles. It was first introduced by Kennedy and Eberhart in 1995 [57]. Similar to other population-based optimisation methods such as genetic algorithms, PSO starts with the random initialisation of a population of particles in the search space. It converges to the global best solution by simply adjusting the trajectory of each individual particle towards its own best location and towards the best particle of the entire swarm at each time step (generation). The position vector and the rate of position change (velocity) vector of the particle in the d-dimensional search space is represented as $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{id})$ and $V_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{id})$ respectively. According to a defined fitness function P for the required scenario, by assuming the best position of each particle (which corresponds to the best fitness value obtained by that particle at a selected time) is $P_i = (p_{i1}, p_{i2}, p_{i3}, \dots, p_{id})$, and the fittest particle

found at the selected time is $P_g = (p_{g1}, p_{g2}, \dots, p_{gd})$, the new velocities and the positions of the particles for the next fitness evaluation can be evaluated with the following two equations:

$$v_{id} = \omega \times v_{id} + c_1 \times \text{rand}() \times (p_{id} - x_{id}) + c_2 \times \text{Rand}() \times (P_{gd} - x_{id}), \quad (6.3.1)$$

$$x_{id} = x_{id} + v_{id}, \quad (6.3.2)$$

where ω is the inertia factor, c_1 and c_2 are constants known as acceleration coefficients, $\text{rand}()$ and $\text{Rand}()$ and are two separately generated uniformly distributed random numbers in the range [0,1].

The first part of (6.3.1) denotes the previous velocity, which provides the necessary momentum for particles to roam across the search space. The second part, known as the ‘cognitive’ component, denotes the personal thinking of each particle. The cognitive component instigates the particles to move toward their own best positions at the time. The third part is known as the ‘social’ component, which denotes the collaborative effect of the particles, in finding the global optimal solution. The social component drags the particles toward the global best particle at current time.

An initial population of particles is generated with random positions, and then random velocities are assigned to each particle. The fitness of each particle is then calculated based on the user defined objective function. At each iteration, the velocity of each particle is evaluated according to (6.3.1) and the position for the next function calculation is updated according to (6.3.2). Each time if a particle finds a better position compared to the previously found best position, its location is stored. Generally, a maximum velocity (V_{maxd}) for each of the particles (v_{id}) is defined in order to control excessive scattering of particles outside the user defined search space[3]. Numerous versions of PSO have been developed such as time varying acceleration

coefficients [65].

6.4 Genetic algorithm (GA)

In this section we briefly review the genetic algorithm and describe the two implementations of GA we used to perform the MLE.

The genetic algorithm is an evolutionary algorithm that mimics evolution of biological processes [52, 39]. It is a powerful optimisation tool that can be used to find maxima or minima in complex multidimensional spaces. Unlike algorithms such as Simulated Annealing, which explores the space using a single Markov Chain, the GA uses a population of random vectors to search the space. Evolutionary algorithms have been used to solve complex problems including inverse radiation problems [99], assimilating sensor data in dispersion [49, 50] and finding optimal placement of biological sensors in a CBR fusion context [100].

In GA, a randomly initialised population of solution strings is evolved by manipulating the strings using a set of operators such as crossover and mutation. At each iteration, the population evolves such that the *fitness* of the population improves in a manner similar to the survival-of-the-fittest principle in the evolution of biological systems. GA-b represents variables as encoded binary strings and works with the binary strings to minimise the cost, while GA-c works with continuous variables.

The two genetic algorithms we implemented in MATLAB used the negative log likelihood function in (6.2.8) as the fitness function. Selection of individuals for mating was based on a chromosome's rank ordering of fitness rather than the actual (or scaled) fitness value. Relative to proportional selection, ranking selection reduces selection pressure when fitness variance is high and increases selection pressure when

the variance is low. Ranking schemes can avoid premature convergence [48]. Because three parameters (x, y, α) are used to characterise a radiation source, the GA-c implementation used a chromosome consisting of $3r$ floating point numbers, where r is the number of sources. In GA-b, each parameter was represented using a 16 bit binary string and, therefore, a binary chromosome of length $48r$. Both implementations of GA used a population size of 12 and the maximum number of iterations was 4500.

6.5 Evaluation of algorithms

We evaluated the performance of the two GA and PSO implementations of the MLE using real experimental data collected during a DSTO-conducted radiation field trial. For the purpose of comparison, we also evaluated an implementation of MLE based on the MATLAB built in function, *fminsearch*, as described in [42] and [71]. In this section we first describe the field trial and the data and then discuss how the algorithm evaluation was carried out.

6.5.1 Experimental setup

A radiological field trial was conducted on a large, flat, and open area without any obstacles. The LCAARS survey system consists of an AN/PDR-77 radiation survey meter equipped with an RS232 interface module, a gamma probe (Geiger-Müller tube), and software written in Visual Basic and running on a laptop computer. When connected to the AN/PDR-77 radiation survey meter, the gamma probe is capable of measuring gamma radiation over a wide range of dose rates without saturating at high levels (although the sensitivity is poor at low dose rates close to background and

at low gamma energies). Dose rate data measured with this probe were recorded in microgray per hour ($\mu\text{Gy}/\text{h}$).

The measured directional response pattern of the gamma probe (mounted vertically) is shown in Figure 6.2. This plot shows that our assumption in Section 6.1.1 about the uniformity of the detector response is approximately satisfied.

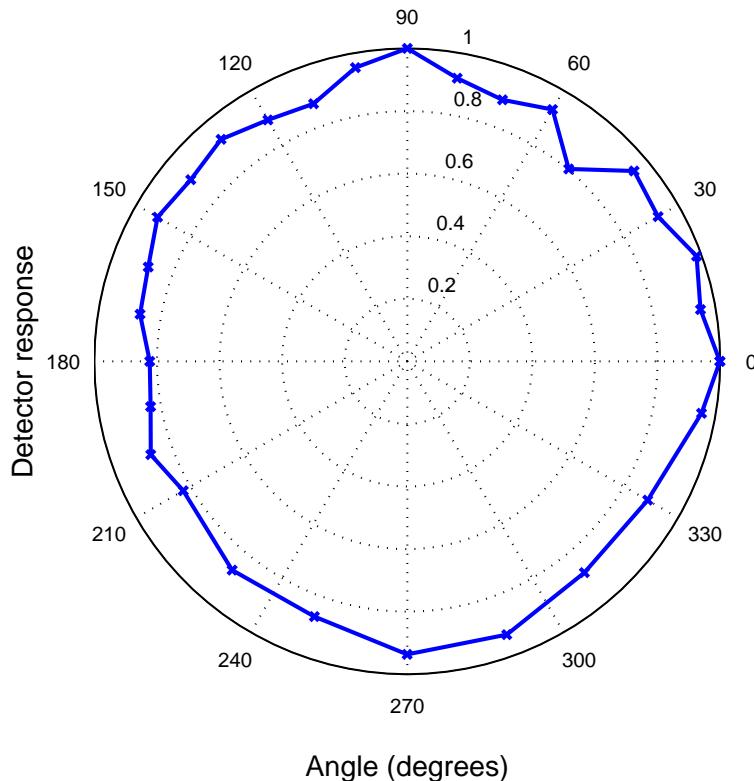


Figure 6.2: *Directional response pattern of the gamma probe obtained by normalising the mean unfiltered data.*

Three radiation sources were used in the field trial: two cesium sources (^{137}Cs) and one cobalt (^{60}Co) source (see Table 6.1). To ensure that radiation sources appear as isotropic, they were placed in a vertical configuration such that their handling rods were pointing up and did not distort the radiation field in the horizontal plane where

data were acquired.

Table 6.1: *Radiation sources used in the field trial*

Source	Type	Activity (GBq)
1	^{137}Cs	26
2	^{137}Cs	5
3	^{60}Co	0.2

Radiation dose measurements were collected in grid points that were carefully measured and marked beforehand in the local Cartesian coordinate system on the asphalt surface of the airfield. The data were acquired when the trolley-mounted gamma probe was positioned over individual grid points. During data collection at any grid point, the gamma probe was held stationary until approximately sixty measurements were acquired. Acquisition of multiple measurements at each point made it possible for us to run the estimation algorithms multiple times choosing data randomly in a Monte Carlo fashion to evaluate their average performance. The exposure time for each radiation dose measurement (effectively the sampling interval) was fixed and was about 0.8 seconds.

Several data sets were collected during the trials but in this Chapter we focus on only three. The data sets are referred to as Test set 1, 2 and 3, and contained 1, 2 and 3 sources, respectively. The source locations in the local Cartesian coordinate system are listed in Table 6.2.

The areal picture of the Puckapunyal airfield site is shown in Figure 6.3. The three green dots show the locations where the radiation sources (Source 1, 2 and 3) were emplaced for this experiment. The X and Y axis markings on the plot show the distances along these directions in meters. The white cross symbols indicate

Table 6.2: *The locations of radiation sources in the field trial*

Test set	Source 1 Location	Source 2 Location	Source 3 Location
1	(11, 10)m	-	-
2	(11, 10)m	(3, 50)m	-
3	(11, 10)m	(3, 50)m	(41, 5)m

the 97 grid positions where the data were planned to be collected during the trial. Not all grid points were visited during each test due to the time constraints, hence measurement positions were not identical between tests.



Figure 6.3: *Aerial image of the Puckapunyal airfield site where the field trial was conducted. The green stars located at coordinates (11, 10)m, (3, 50)m and (41, 5)m of the local Cartesian coordinate system indicate where the three sources were emplaced for Test 3. The numbered grid points marked with crosses are the measurement points*

While the strengths of radiation sources are typically characterised in terms of their activity in GBq, as shown in Table 6.1, in this work, for simplicity, we characterise the strength of a radiation point source by the expected dose rate at a distance

of 1m from the source. The strengths of our three radiation sources characterised in this manner were $1912 \mu\text{Sv}/\text{h}$, $392.35 \mu\text{Sv}/\text{h}$ and $98.1 \mu\text{Sv}/\text{h}$, respectively.

Our gamma probe displayed measured data in the units of $\mu\text{Sv}/\text{h}$. We converted these data into raw count measurements $z_j \in \mathbb{Z}^+$ by multiplication with the appropriate conversion factor. The radiation source strengths, expressed above in the units of $\mu\text{Sv}/\text{h}$, were also converted using the same factor. Therefore, the radioactive strengths of sources, expressed in counts measured at a distance of 1m, are then 9105, 1868 and 467, for Source 1, 2 and 3, respectively.

6.5.2 Algorithm evaluation

To evaluate and compare the performance of the MLE implemented using *fminsearch*, PSO and the two genetic algorithms, each algorithm was run 50 times in a Monte Carlo fashion for one, two and three source cases. We are not showing results from PSO implementations as it did not fulfill the preliminary accuracy. As mentioned above, because 60 measurements had been acquired keeping the sensor stationary at each test grid point, different measurements could be randomly selected from each position for each one of the Monte Carlo runs. Because the algorithms failed to converge in some of the Monte Carlo runs, the following criterion was used to define such divergent runs and these runs were discarded when computing the root mean squared errors of the algorithms.

The R^2 statistic for the goodness of fit between measured data and data predicted at the same measurement positions using the estimated sources was computed for each Monte Carlo run. Those Monte Carlo runs that resulted in R^2 values less than 0.9 were considered as divergent runs, and their source estimates were discarded when

computing RMS errors.

6.6 Parameter estimation results

Tables 6.3, 6.5 and 6.7 list the RMS errors of parameter estimates computed for the three estimation algorithms. Tables 6.4, 6.6 and 6.8 list the number of divergent runs and the average execution time.

Table 6.3 shows that all three implementations have RMS position errors that are nearly equal to the CRB. The RMS error for source strength is larger than the CRB prediction for all implementations.

We see from Table 6.4 that the GA implementations take longer to run than the *fminsearch* implementation. However, *fminsearch* algorithm failed to converge in 11 of the 50 trials, whereas the GA implementations converged successfully in all runs.

Table 6.3: *Test 1 data set (single source): RMS estimation error ($M = 50$ runs)*

	$\sqrt{\text{CRB}}$	RMSE		
		Fmin	GA-b	GA-c
x_1 [m]	0.76	0.73	0.81	0.84
y_1 [m]	0.52	0.64	0.71	0.62
α_1 [$\mu\text{Sv}/\text{h}$]	164.38	280.08	156.31	168.02

Table 6.4: *Test 1 data set (single source) : Algorithm comparison ($M = 50$ runs)*

	Fmin	GA-b	GA-c
Diverg.(out of 50)	11	0	0
Execution time(s)	2.0	45.9	47.16

Results obtained when these algorithms were applied to data collected in the presence of two point radiation sources are summarised in Table 6.6. The RMS errors of the parameter estimates returned by the algorithms for the first source are nearly two times larger than that predicted by the CRB. The estimates of the second source parameters show larger discrepancies; particularly, the source strength is overestimated and shows much larger RMS error than the CRB prediction. The *fminsearch* algorithm again failed to converge in a majority of the trial runs, whereas the the binary GA algorithm diverged in only eight runs. The continuous GA algorithm did not experience any divergences according to the chosen criterion. The *fminsearch* algorithm was run in our tests with a randomly chosen initial guess for each run. The number of divergent runs of *fminsearch* could probably have been improved by running it with a number of different initial guesses.

Table 6.5: *Test 2 data set (two sources) RMS estimation error (M = 50 runs)*

	$\sqrt{\text{CRB}}$	RMSE		
		Fmin	GA-b	GA-c
x_1 [m]	0.31	0.44	0.49	0.54
y_1 [m]	0.48	1.08	1.02	0.85
α_1 [$\mu\text{Sv}/\text{h}$]	80.74	270.1	179.44	125.9
x_2 [m]	2.66	8.20	8.40	4.16
y_2 [m]	0.91	4.31	2.46	2.12
α_2 [$\mu\text{Sv}/\text{h}$]	123.03	1007.8	783.37	330.5

Table 6.6: *Test 2 data set (single source) : Algorithm comparison (M = 50 runs)*

	Fmin	GA-b	GA-c
Diverg.(out of 50)	42	8	0
Execution time(s)	7.09	57.75	46.18

The summary results of 50 runs of the three algorithms using data collected in the presence of three sources are shown in Table 6.7. Both GA-based estimation algorithms were able to estimate the three sources reasonably well, but *fminsearch*-based estimation fared poorly. As in Test 1 and Test 2, the source strength estimates experienced very large RMS errors because the source strengths were vastly overestimated in many runs. Again the GA-c converged in all 50 runs and the GA-b diverged in two runs.

Table 6.7: *Test 3 data set (single source): RMS estimation error (M = 50 runs)*

	$\sqrt{\text{CRB}}$	RMSE		
		Fmin	GA-b	GA-c
x_1 [m]	0.37	24.49	2.67	2.01
y_1 [m]	0.47	7.19	2.28	0.63
α_1 [$\mu\text{Sv}/\text{h}$]	93.57	1413	471.57	399.69
x_2 [m]	1.43	279.3	30.17	2.12
y_2 [m]	0.70	296.45	18.45	1.91
α_2 [$\mu\text{Sv}/\text{h}$]	69.91	4700.5	1077	134.19
x_3 [m]	0.31	218.22	1.89	2.00
y_3 [m]	3.0	59.81	1.80	1.32
α_3 [$\mu\text{Sv}/\text{h}$]	60.99	2575.1	790.65	593

Table 6.8: *Test 3 data set (three sources): Algorithm comparison (M = 50 runs)*

	Fmin	GA-b	GA-c
Diverg.(out of 50)	34	2	0
Execution time(s)	10.47	70.07	67.84

6.6.1 Effect of data collection geometry on radiological source localisation

In this study, we examined the differences of data collection geometries influencing the topography of objective function surfaces in radiological source localisation. It was shown that data gathered along the circumference of a circle around a point radiation source has an associated mirror image source of a different strength that exists outside the circle (See Fig. 6.4 and Fig. 6.5). When the radius of the circle becomes infinity, then the path becomes a straight line, and it results in a mirror image source of identical strength to the original source. These image sources cause the objective function surfaces to have global optima that are identical to the true source optima making source localisation algorithms to sometimes converge to image solutions and return unacceptable parameter estimates. Cramer-Rao bounds computed for the parameter estimators do not reflect this geometry-dependent uncertainty. It appeared that data acquired along an irregular path generated using a random walk eliminate the image sources making source parameter estimation easier.

6.7 Summary

The Chapter described the use of binary and continuous forms of genetic algorithms (GA) to estimate position and strength of radiological point sources. GA was used to implement maximum likelihood estimation (MLE) because it is not possible to carry out MLE analytically for this problem. Real experimental radiation data collected in a recent field trial were used to test the performance of the GA-based MLE and to compare it to another implementation based on *fminsearch* routine in MATLAB.

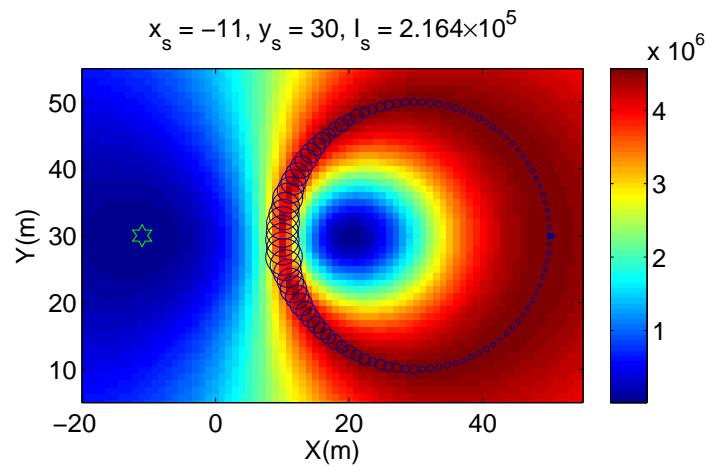


Figure 6.4: Mirror source localised outside

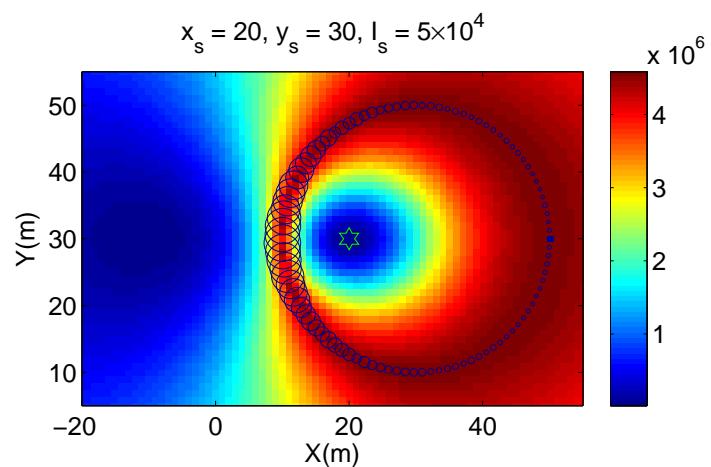


Figure 6.5: Mirror source localised inside.

While the latter algorithm resulted in smaller run times compared to the GA implementations, it failed to converge in many of the trial runs. While the binary GA implementation also failed to meet our convergence criterion in some of the trial runs, the continuous GA implementation did not diverge in any of the runs. In the case of a single point source, the RMS estimation errors produced by the algorithms were in good agreement with the limits predicted by the Cramer-Rao bound (CRB). However, the RMS estimation error in two and three source cases were found to be larger than the CRB. The source strengths were overestimated and resulted in much larger RMS errors than that predicted by the theoretical bound. Because the CRBs were computed based on idealised assumptions such as perfect knowledge of measurement positions, zero background and air attenuation, they are too optimistic.

Chapter 7

Conclusions

Let us begin by reviewing the motivation for the research undertaken in this thesis, and the research questions that we have addressed. As attackers have become more sophisticated in terms of the methods they use to disrupt services in the day to day life, sensor network managers are faced with the problem of how to detect and defend against a new generation of highly coordinated attacks. Given that the evidence of these highly distributed attacks can be spread across different geographical locations, there is a growing need for collaborative hazard detection techniques that can correlate evidence from different geographical locations. Consequently, the focus of this thesis has been on new methods for collaborative hazard detection that are suited to detecting large-scale CBRN attacks with efficient information dissemination from the networks and conserving energy.

In Chapter 2, we surveyed the common types of hazardous attacks that have been reported in the literature review, as well as the relevant approaches to bio-inspired hazard detection that have been reported throughout the literature. Based on this survey, we have focused on three open research problems for detecting hazards, namely : (1) how to support increasing levels of information dissemination in hazard

detection in a CHDS without sacrificing computational efficiency, (2) how to identify characteristics of chemical tracer distribution and wireless sensor network topologies and distribute the computation for different types of sensor network topologies of CHDS, given the need for scalability as we increase the number of sensor node participants in a CHDS, and (3) how to maximize the energy conservation in a CHDS while minimizing the communication and computational overhead to the system.

In Chapter 4 we considered the epidemiological algorithm. The energy conservation in chemical sensor networks is crucial as chemical sensors with air sampling tend to consume significant energy for sensing activity compared to that used for communication, unlike other types of sensors, such as optical or acoustic. When considering the threat environment, the chemical tracers dispersed by turbulent motion in the environment display rather complex and even chaotic properties. Hazardous chemical releases are rare events. If all sensors in a wireless chemical sensor network (WCSN) are left in the active state continuously, it will result in significant power consumption. Therefore, dynamic sensor activation is essential for the durability of WCSNs. Dynamic sensor activation for chemical sensor networks using an epidemiology-based sensor activation protocol has been proposed in the literature. In this research study, we investigate the performance of a variant of epidemiological algorithm, a gossip inspired sensor activation protocol of a WCSN in a chemical tracer field. The simulation framework with the gossip-based protocol is validated against an analytical model. We then perform simulation experiments to evaluate the performance of the gossip-based sensor activation protocol on selected performance metrics: the number of active sensors and the reliability of detection. We show by simulations that, by varying the communication radii of sensors, we can achieve better energy conservation

while maintaining the better performance of a WCSN with a gossip-based activation protocol.

We also investigated the behaviour of a WCSN with a bio-inspired gossip-based sensor activation protocol in a turbulent chemical tracer field, which is highly complex in nature, as we showed our studies in Chapter 3. Using a bio-inspired analytical model for a WCSN, initially introduced by Skvortsov *et al.* [94], we verified the simulation framework for WCSNs (Fig. 5.3). Based on simulation results, we evaluated some important performance metrics: the number of active sensors, as shown in Fig. 5.6, and the reliability of detection as shown in Figure 5.7. Based on our results, we can consider extending the communication range of sensors rather than activating more neighbouring sensors, which is an energy-conserving strategy [38]. The gossip protocol equips sensors with intelligence when selecting neighbour sensors for activation. This system can also be applied to a real system where sensor communications links get dysfunctional with time due to aging. In this, we used centralised aggregation of detected concentration values; we hope to incorporate gossip-based averaging [81] and bio-inspired optimisation technique to optimise the WCSN configuration parameters [53] in our future studies.

7.1 Importance of temporal correlation

Currently we consider only spatial correlation, but it is apparent from our experimentation that temporal correlation is an important factor for the WCSN to have an IE. It is not realistic to have a static slice, therefore we use slices which have some degree of variation with time, but their variation is not controlled.

The distribution of tracers of contamination slices that we generated shrinks with

increasing correlation radius, therefore lowering the values of the contamination detection threshold down to a minimum value of tracers will have no effect. The placement of sensors plays an important role, so a more distributed placement is desired. In our experiments we chose initial random sensors randomly in different iterations.

7.2 Issues in selecting chemical contaminant scenarios

When selecting the R_{corr} contaminant scenarios for the experiments, two Rcorr cases were overlapped due to variance in the R_{corr} value over the time slices; therefore we ignored the $a = 0.01$ case. Some tracers of contaminant slices with values greater than one had negative values; therefore we rejected them as R_{corr} values for the cases that were almost similar and closer to non-correlated case.

For an IE to occur, the detection threshold $C*$ value of a sensor should be above the minimum value of chemical contaminants. $C*$ can be selected so that $P > S/(NG\pi r*)$ for an IE to occur. where S is the area of the rectangular sensor field, P is the probability of detection and G is a constant calibration factor for correlated case, being of order unity. P is a function of R_{corr} and $C*$. The $r*$ should be above the distance between two neighbouring sensors. From (3) we can get that it should be greater than the square root of GPN, where G is a constant for a selected field dimensions and chemical contaminant identified. We affirm the hypothesis that IE depends on ϵ or R_{corr} value.

There are language limitation as well. Some errors encountered can be attributed to the handling of complex numbers and functions converging to negative infinity. Also, parameter G can be adjusted to suit the scenario.

7.3 Study effect of correlation of chemical tracers on performance of WCSN

Several authors [18, 89, 102, 105] have studied the correlated information fusion, but have not particularly considered chemical environments or chemical sensor networks. In research studies related to chemical sensor networks, the performance evaluation of networks with epidemiology-based protocols have been explored in the papers[94, 93]. In the aspect of chemical tracers, Borgas *et al.* in [11] and [9] present models for chemical tracer fume propagation and behaviour. Although many performance attributes of such chemical sensor networks, such as energy utilization, detection delay and false alarm characteristics have been studied in the literature, the effect of characteristics of the environment on the WCSN performance has not been studied specifically.

We currently use a chemical tracer environment with identified spatial correlations to study the effect of the environment on WCSN performance. We will be incorporating temporal-correlation into the environment and moving to a set of localised sources for the chemical tracer emission environment. To check for success, we expect that results from simulations should match analytical results. With the design of robust models for correlated chemical tracers, it is expected to fine tune the WCSN to have shorter detection times and a sufficient number of active sensors to detect chemical tracers reliably.

7.4 Threat characterisation

At this stage, when we consider detecting and activating the WCSN, it will be a worthwhile task to identify further characteristics of the threat. Hall *et al.* state

location, characterization, and identification of threat as important issues in data fusion [46]. Srinivas *et al.* in [55] present a protocol based on a gossip algorithm to extract the aggregates from sensor measurements, and Rao *et al.* [81] propose an extension to the protocol to extract other factors, such as maximum and minimum tracers values of measurement detected by a network. Guan and Duckham propose a method to identify fundamental topological relationships between spatial regions[41]. We considered the problem of estimating the locations and strengths of radiological sources using random measurements using a genetic algorithm.

In this study, we consider using chemical tracer emissions with identifiable source locations and we apply identified techniques used in [81, 41] to localise chemical tracer source centroids and further explore extracting meaningful information, such as chemical threat propagation direction, affecting regions and the speed of propagation (Fig. 7.1). In these studies, we hope to use particle swarm optimisation (PSO)[64], a gossip algorithm[55] and quorum sensing[63].

7.5 Ideal design factors for WCSN

In [19], authors say that there is a lack of fundamental system-level research leading to the development of sensor networks that both maximize protection and minimize the system cost for indoor air protection. In [2], Akyildiz *et al.* claim scalability, deployment costs, sensor network topology and power consumption are important design factors. By identifying ideal design factors, a cost effective, efficient and reliable WCSN can be created.

At the current stage, we were able to make a reasonable estimation of a few optimal parameters for the WCSN. It is as an initial step to design an economical, reliable

and efficient mechanism to detect chemical tracers. Deriving the optimal scalability factor is very important in determining the optimal number of sensors required to detect chemical tracers reliably with lower deployment costs, which can be attempted in further studies.

For maximising information fusion and minimising energy, we have tested the sensor activation protocol similar to the epidemic SIS (susceptible-infected-susceptible)[73]. This protocol needs to be further developed to incorporate transmission based on sensor compromise and communication failures. Then other activation protocols such as gossip algorithm can be considered in application to chemical threats and their performance can be compared with those current.

We currently use the ‘grid’ topology to ensure repeatability, and this will be extended to other topologies: line, circle and hybrids of these geometrical forms. We currently use binary chemical sensors for WCSN, with which we are unable to extract some of the features mentioned. Later we will incorporate the multi-bar sensor model [86] instead of the binary sensor model.

7.6 Refinement of simulation framework

The framework developed has already shown demonstrable progress as a platform for testing epidemiological-based data fusion algorithms, as well as supporting other types of protocols. It is a simple construction in software engineering, being written in Matlab®. The use of this framework, is to analyse realistic synthetic contaminant chemical tracer environments of spatially and temporally correlated and fluctuating concentrations within dispersing plumes, along with models of various detectors in our studies. The EM was constructed as a modular structure to facilitate individual

adaptation and development as individual modules, which are integrated to form the framework. The three modules (i.e., environment, sensor network and individual sensor), will be improved to facilitate the research study by applying the required refinements to modules independent of each other.

7.7 Visualisation and interaction

It will be ideal to produce a tool that allows users not familiar with computers and others to visualise and select the ideal chemical environment, network and sensor parameters, and view the outcome. We currently use animations to show evolving chemical tracer slices with time. Erten *et al.* present a novel way to display graphs evolving with time[32] stacked on each other. It is an ideal visualisation tool to identify WCSN behaviour with chemical tracer distribution change, which can be considered at a later stage of our study. Since our intention is not to develop visualisation tools but to derive an ideal protocol for chemical tracer detection, initially we will use existing Matlab tools to develop a simple graphical user interface (GUI), and attempt complex visual improvement techniques later.

7.8 Reliability of Source localisation

In Chapters 3 to 5, we investigated the activation of sensors and did not consider the location of the hazard source, however in Chapter 6, we explored the possibility of localising a radiological source using bio-inspired genetic algorithms and problems encountered in data acquisition. We considered localisation of point sources of gamma radiation using dose rate measurements. As bio-inspired genetic algorithms, we used binary and continuous genetic algorithms (GA). It is reasonable that the RMS errors

of the parameter estimates computed based on noisy experimental data are much larger. The problem that was considered in this Chapter was that of localising a known number of point radiation sources located in flat terrain without obstacles, which is a highly simplified version of the more complex and challenging problem of estimating arbitrary distributions of radioactive materials in complex terrain in the presence of obstacles and variable background. Results presented in this Chapter show that GA-based source localisation has some merit in benign conditions such as those employed in the field trial. In future work, it will be possible to explore more robust estimation techniques that will enable us to remove some of these simplifying assumptions and tackle more challenging and realistic source localisation problems.

7.9 Other analysis and applications

There are other studies ongoing and running in parallel with the work which may have not been noted but should be considered important. We will be improving the threat model to be as realistic as possible, as presented in [11].

In widening our evaluation and domain scope of the research, we may consider using our framework as a consistent framework for simulation of the so-called ‘national wireless sensor network detection against radiological, chemical and biological threats’. This network is based on public cell phones enhanced with a single, low cost sensing device for the simultaneous detection of biological, chemical and radiological threats and has recently attracted significant attention [22].

In this study, we assume sensor nodes to be homogeneous, which is not a very good representation of an actual scenario. In reality, networks divide into ‘clusters’ or ‘communities’. The epidemiological algorithm which we used in our studies, which

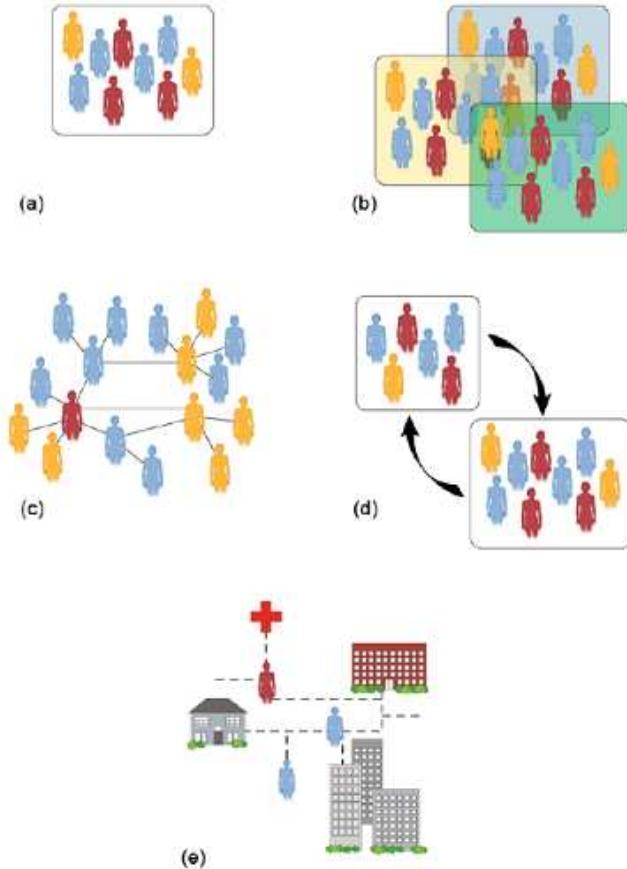


Figure 7.1: Structured ways to model the epidemics: (a) Homogeneous mixing (b) Social structure (c) Contact network models (d) Multi-scale models (e) Agent-based models

governs the activation of the WSN can be further developed to incorporate real-world scenarios with compartments, clusters and adaptability features. As shown in Figure 7.1, Shaw *et al.*[90] give an good overview of currently identified epidemiological structures.

The time and space complexity evaluations of our experiments are important, even though we have not stated them. When introducing a new method or a model, we will be assessing the space and time complexity to reap the maximum outcome from

experiments. Accordingly, we will consider increasing the experimental WCSN field and chemical tracer field dimensions by improving memory management.

7.10 Significance of our research

The research study will highly benefit relevant authorities in identifying and diffusing chemical contaminants, which otherwise would cause a large number of casualties in an attack. It will be a highly useful tool to ascertain the extent of a chemical attack and make use of a collaborative sensor approach, rather than individual chemical sensors or detectors.

Though an analytical ‘equation based’ approach is much simpler and less computationally intensive than the simulation based approach, it deals only with the dynamics of the mean state of the system, rather than with the state of each individual sensor in the network. The simulation will enable addressing the WCSN performance at the sensor level. Also in order to test suitable techniques for optimising WCSN, it would have been necessary to get data from real experimentation which will have health risks. But with the development of simulation tools, there will not be a necessity for manual experimentation, needing risky experimental environments. The optimal parameters derived from simulation will enable the design of ideal cost-effective, reliable and efficient WNCSSs.

We have studied several potential areas of our work, including data fusion in chemical and radiological hazards. This is an ideal framework, which can be applied mainly in defence systems (e.g., military operations, scenario planning, situational awareness), national security(e.g., counter terrorism, border control), hazard management (e.g., bushfires, technological catastrophes), ecological monitoring (e.g.,

air pollution control, wildlife protection, water quality), however by adopting different environmental modules, this can be also used in social network analysis, assess marketing strategies and some industrial applications (e.g., control of technological processes).

Appendix A

Software frameworks

In our studies we considered bio-inspired algorithms. The simulation frameworks with relevant protocols are validated against an analytical models. We then performed simulation experiments to evaluate the performance of sensor activation protocols on selected performance metrics. Here we present the software frameworks created and used in our studies.

A.1 The CHEMSENS software framework

In our information-driven dynamic sensor collaboration (DSC) model, we employed the modelling framework initially proposed in [1][2] for chemical tracers. This framework includes a module of the environment (the temporal realisation of the concentration field of the contaminant caused by turbulent mixing), a module of an individual sensor and a module of the whole network. This modular object oriented design of the modelling framework, as shown in Figure 1, allows for incremental improvements to be made in any of these areas.

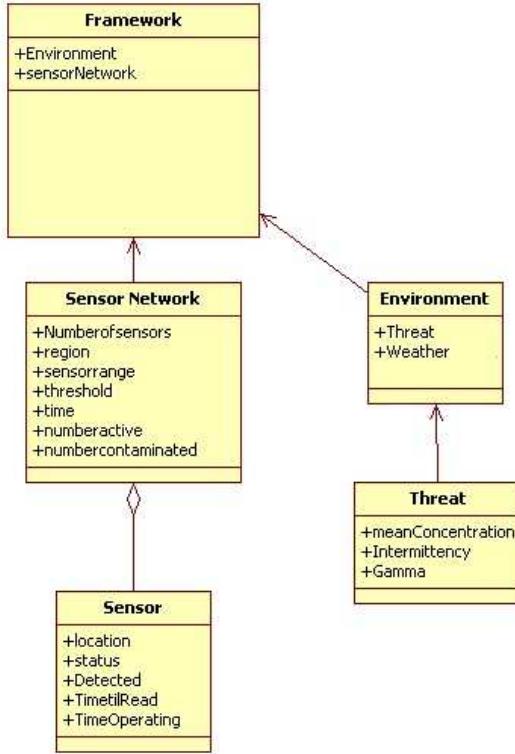


Figure A.1: *WCSN simulation framework*

A.2 The source localisation framework

In application of bio-inspired algorithms for hazard localisation as discussed in Chapter 6, we employed several algorithms. However we used the best performing techniques in the thesis. The framework we developed for source localisation is given here. We have the option of selecting the sensor placement pattern, relevant algorithm and then we extract the results.

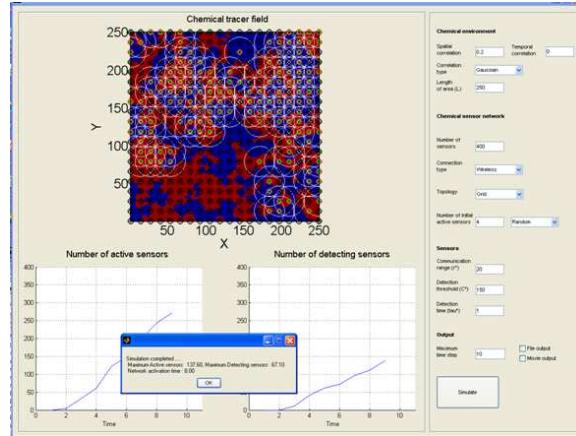


Figure A.2: *GUI for WCSN simulation framework*

A.3 The chemical dispersion simulation framework

To study chemical fields under turbulent conditions, we also tried to use GPU technologies to visualise the scenarios as shown in Figure A.4.

A.4 The random fields

Random fields are which generally mimics environments under turbulent conditions. These help us to understand underpinning issues in application of WSN in an turbulent environment. Figure A.5 shows typical random fields generated by software.

A.5 The physical models

We created a practical implementation of the sensor activation using Lego NXT mindstorms® as shown in Figure A.6, however this is for demonstrations and outcomes were not used to compare with numerical simulations.

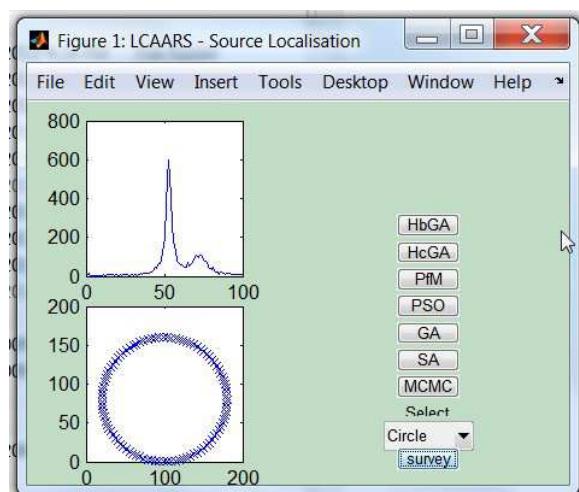


Figure A.3: *Radiological source localisation*

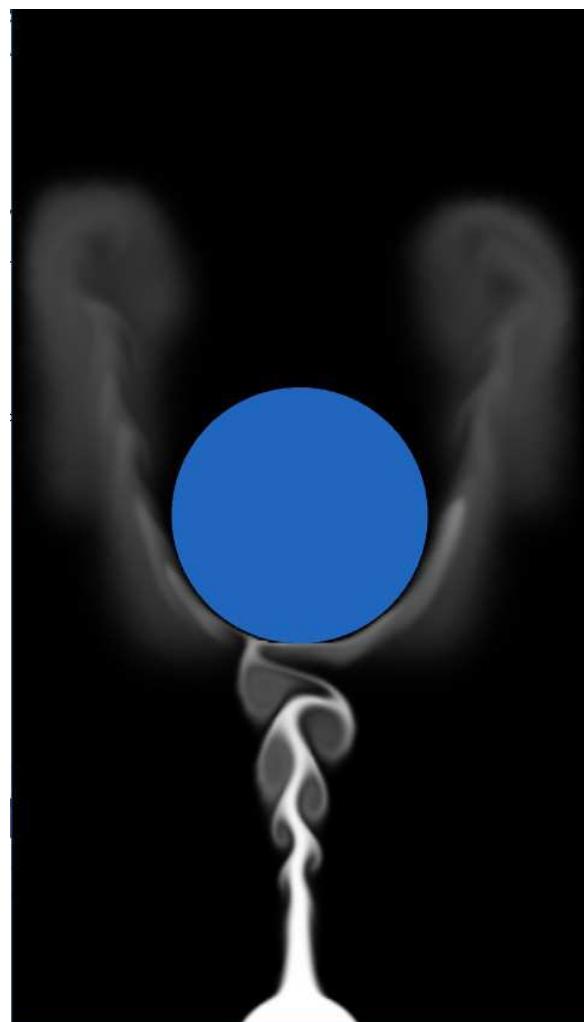


Figure A.4: *Chemical dispersion simulation*

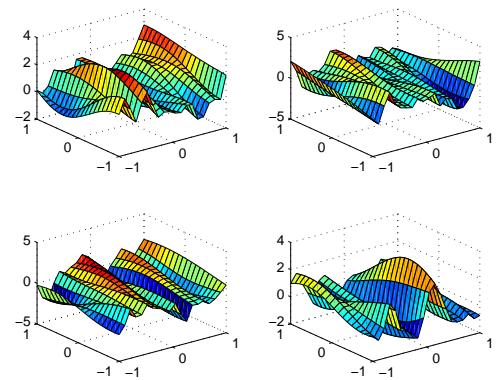


Figure A.5: *Random fields*



Figure A.6: *Practical implementation of sensor activation protocol*

Appendix B

Chemical tracer slices

In reality, the surface-layer contaminant mixing produces a ‘singular’ or rough spatial correlation of concentrations. In our research study, for the purpose of comparison, we considered spatially non-correlated and spatially correlated cases generated by randomness. Figures B.1(a) and B.1(b), respectively, show examples of non-correlated and correlated chemical tracer fields.

Non-correlated chemical pollutants

In study of non-correlated chemical tracer fields, we used a non-correlated chemical contaminant threat model extracted as a random time series which imitates the turbulent fluctuations of the chemical contaminant concentration at each sensor of the network. In [94], the fluctuations in concentration C are described by its probability density function $\rho(C)$, with the mean concentration of the contaminant in the area, C_0 , as a parameter[9], [11].

Here the value $\gamma = 26/3$ has been chosen to make $\rho(C)$ compliant with the theory of tracer dispersion in Kolmogorov turbulence (see [9]), but it may vary with weather conditions. The parameter ω , which models the tracer intermittency in the turbulent

flow, is in the range $[0, 1]$, with $\omega = 1$ corresponding to the non-intermittent case. In general, it also depends on a sensor position within a chemical plume; thus $\omega = 1$ is in the range 0.95 - 0.98 near the plume centroid and may drop to 0.6 - 0.7 near the plume edge. For $\omega \neq 0$, the PDF has a very small impulse in zero, meaning that the measured concentration in the presence of intermittency can be zero on some occasions. It can easily be shown that the PDF integrates to unity, so it is appropriately normalised. In this study, we have used the non-intermittent case for simplicity.

Correlated chemical pollutants

In study of Correlated chemical pollutants, we did not consider atmospheric dispersion and wind flow models in their three-dimensional aspect, but rather as two-dimensional slices. Typical wind-flow and dispersion models that are used routinely tend to provide a ‘best estimate’ of the hazard. They are mean field estimates of state variables of interest, such as mean concentration. Due to the inherently stochastic nature of atmospheric processes, the mean field models cannot provide the current state of a specific threat, but only the average expected over many repetitions of similar events, or an ensemble average.

A spatially correlated model is created using an exponential smoothing filter on a uniformly distributed random field. According to Eqn.(44) of Borgas *et al.*[11], the size of the exponential window is increased with the distance. There are various spatially correlated models, which are beyond the scope of this study. The slices generated were checked by shuffling the order and investigating any change of results. The edge effect of slices was also removed.

In calculating the correlation radius(R_{corr}), we considered only the first segment of correlation for simplicity. We shifted the input data matrix over another copy of the same matrix of the contaminant slice and computed the products at the intersecting positions. Then the sum of these product terms was divided by correlation in x direction $\langle c(x).c(x + r) \rangle / \langle c(x)^2 \rangle$. Since we have isotropic data, we considered only one direction.

We derived the R_{corr} value by extracting the R_{corr} value at correlation 0.5, and by interpolating using a tangent-based method and getting the value of r at correlation 0. We have observed that the correlation has less fluctuation irrespective of the number of slices used. We trialled several methods in deriving R_{corr} using the zeroth correlation location, the interpolation to the zeroth location, and tangent methods used in this study. The tangent method approximates the R_{corr} with no fluctuations. Other methods are prone to errors when correlation is at a minimum. The zeroth method is also unacceptable, as the correlation fluctuates near the zeroth location. We were able to get well-distributed deterministic R_{corr} values with the tangent method. We generated different correlated chemical tracers fields by changing the spatial fraction a . The characteristics of chemical tracer fields are shown in Figures B.1 to B.6.

The correlation coefficient ϵ as defined in Eq. (4.2.3) is assumed to be ideal as a , c parameters cannot be good interpretations of the correlation.

Figure B.7 shows, the influence of detection threshold of sensors (C_*) on the probability of detection (P) of chemical tracers for a selected field. It can be seen that when $C_*/C_0 < 1$, P decreases with correlation radius, and when $C_*/C_0 > 1$, P increases with correlation radius. When $C_*/C_0 = 1$, P is independent of correlation radius.

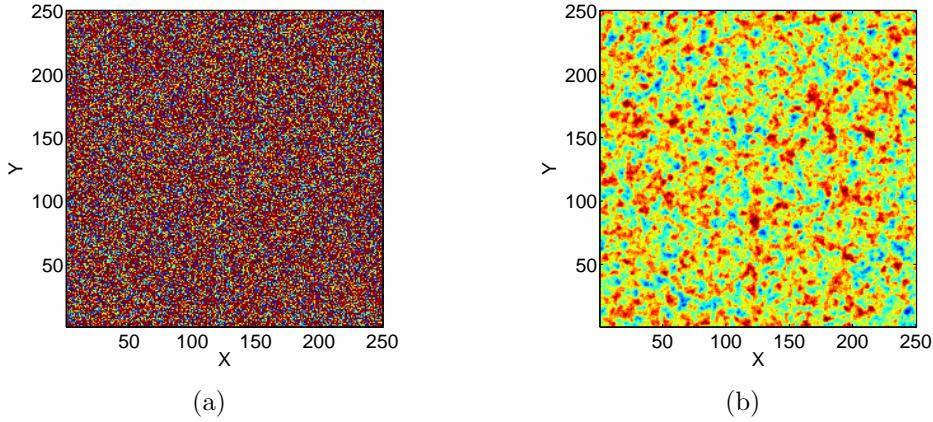


Figure B.1: An example of a random concentration realisation where different concentrations are represented by different colours: (a) non-correlated chemical pollutants (b) correlated chemical pollutants

B.1 Slice generation

We generated 19 chemical slice sets with each containing 150 time slices.

The software interface in generating the chemical slices is shown in Figure B.8. The process of generating chemical tracer environmental slices are illustrated in Figure B.9. The inter-connected of parameters we used are shown in Figure B.10. We used framework shown in Figure B.11 to validate our generated chemical slices.

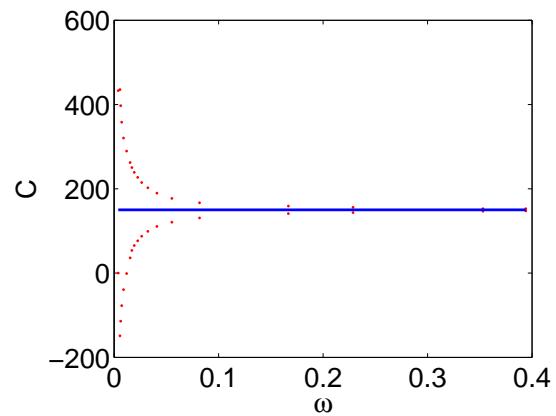


Figure B.2: Characteristics of generated chemical slices

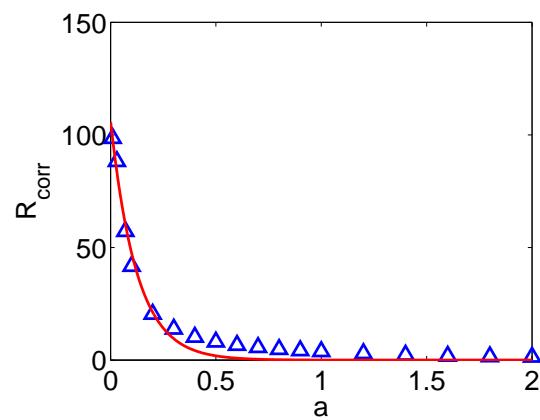


Figure B.3: Variation of R_{corr} with spatial fraction a

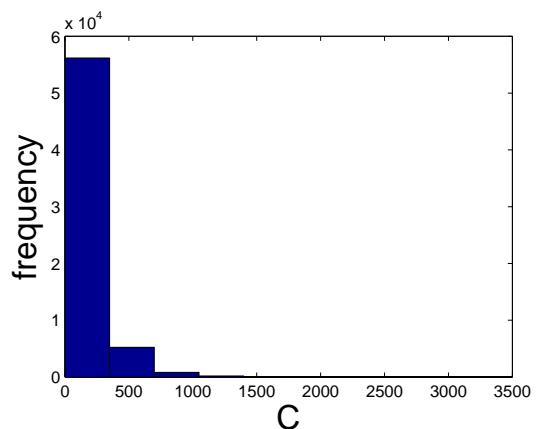


Figure B.4: Extraction of chemical concentration from inverse of CDF

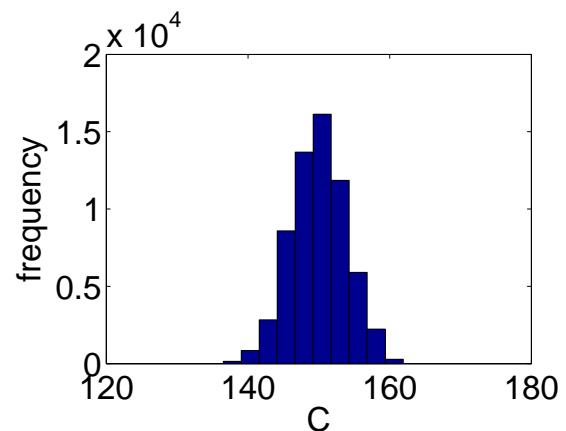


Figure B.5: Preparation of chemical tracer slices using the filter method

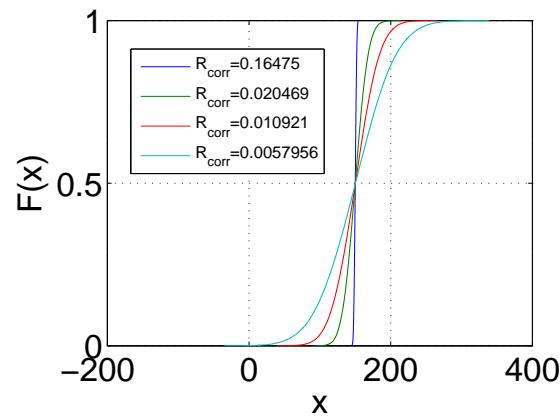


Figure B.6: Compliance of chemical slices to Gaussian distribution

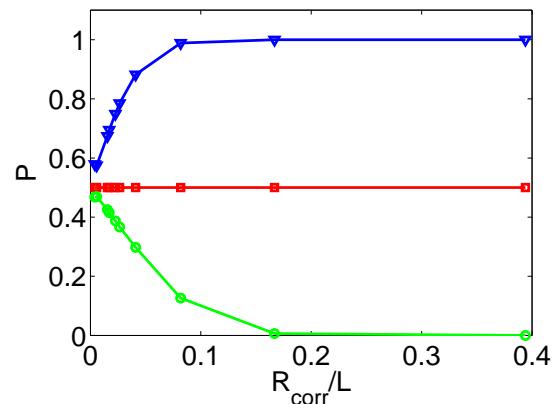


Figure B.7: Probability of detection (P) as a function of correlation radius (R_{corr}). $C_* < C_0$ (blue triangle), $C_* = C_0$ (red square) and $C_* > C_0$ (green circle), L is the size of the field.

Table B.1: *Parameters of generated chemical slices)*

Filter value	R_{corr}	Minimum	Maximum	Mean
0.01	98.51	147.56	152.7	150
0.03	88.25	146.71	152.73	150
0.07	57.18	143.68	155.97	150
0.1	41.7	141.16	158.97	150
0.2	20.44	130.75	166.88	150
0.3	13.83	120.7	177.21	150
0.4	10.24	110.7	189.63	150
0.5	8.09	99.2	202.33	150
0.6	6.61	87.43	214.86	150
0.7	5.65	76.52	227.1	150
0.8	4.82	65.38	238.97	150
0.9	4.26	53.79	250.41	150
1	3.84	36.15	262.45	150
1.2	2.98	-1.01	289.57	150
1.4	2.25	-39.25	320.49	150
1.6	1.81	-77.32	358.29	150
1.8	1.57	-114.13	397.09	150
2	1.42	-148.82	435.55	150
0	1.01	0	433.11	150.02

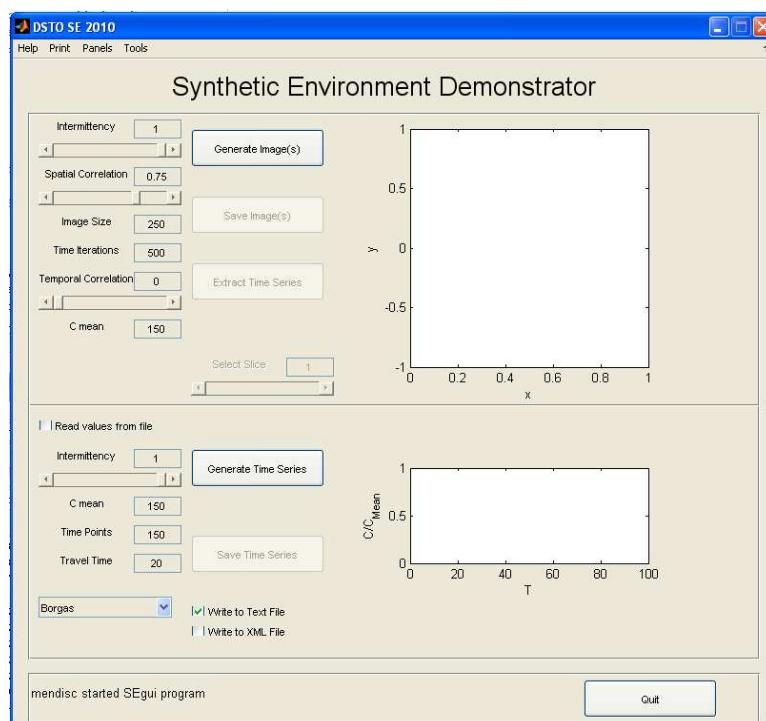


Figure B.8: The graphical user interface used to generate time varying chemical slices [44]

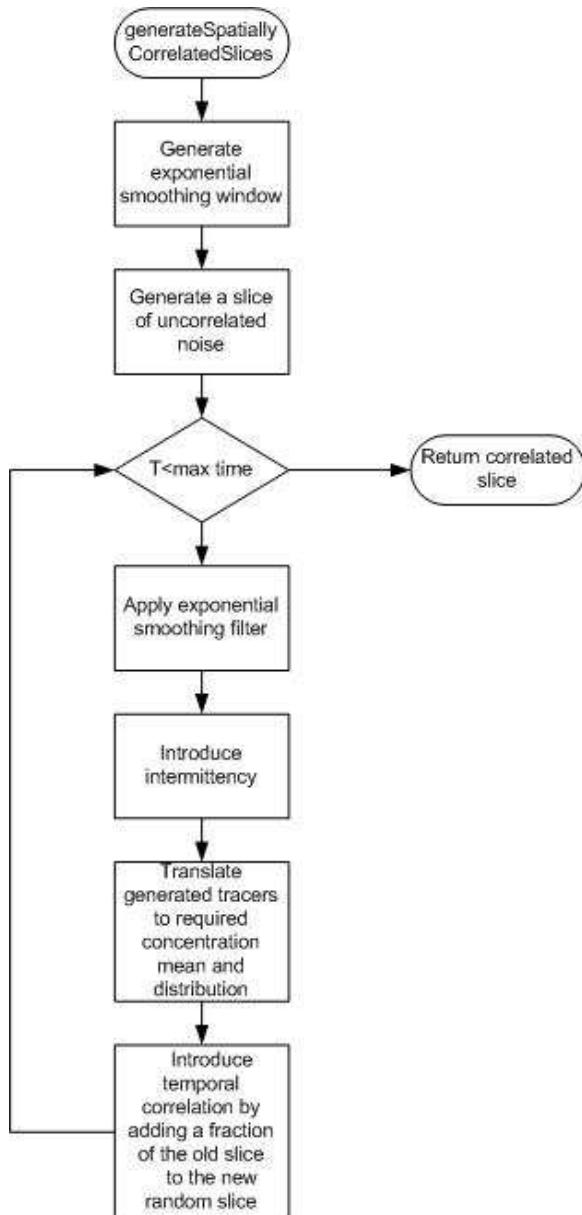


Figure B.9: *The process of generating chemical slices*

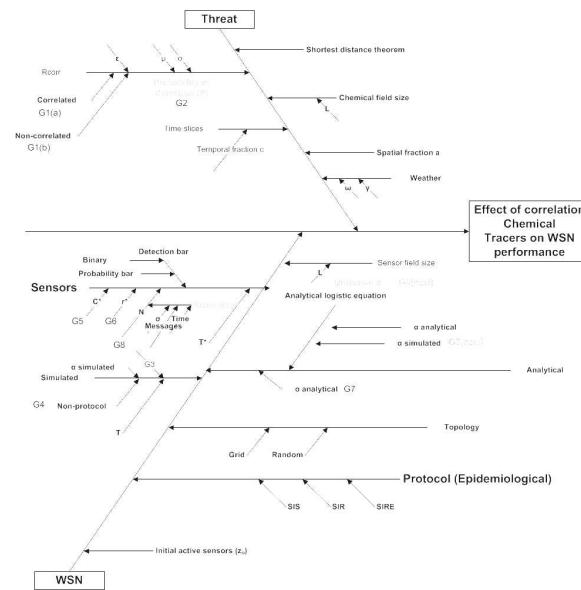


Figure B.10: *Dependency of parameters involved in WCSN performance analysis*

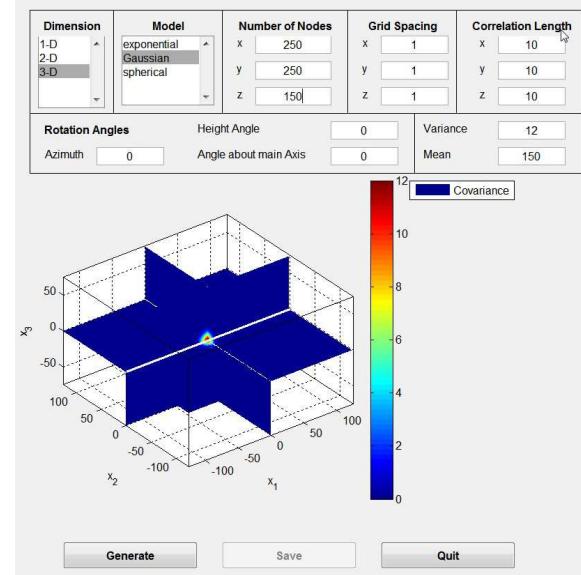


Figure B.11: *chemical slice generation framework [21]*

Appendix C

WCSN performance metrics

C.1 Threshold sensitivity

The simulation parameters were $N = 400$, $r^* = 20$, $R_{corr} = 20.44$, varying C^* and simulation was carried out only once as this is a 'grid' topology and positions of sensors were not changed. The contaminant and sensor field were 250x250 units. The sensors in the four corners of the field were initially activated.

Table C.1: *Concentration, non correlated case and rise time*

C^*	τ	$N+$
40	11	371.36
140	13	367.63
150	15	362.48
151	15	362.17
152	15	361.94

Table C.2: *Concentration, correlated case and rise time*

C^*	τ	N+
40	11	371.36
140	12	372.33
150	28	318.01
151	32	289.86
152	1	0.04

C.2 Communication range sensitivity

The simulation parameters were $N = 400$, $r^* = 20$, $R_{corr} = 20.44$, varying C^* and simulation was carried out only once as this is a 'grid' topology and positions of sensors were not changed. The contaminant and sensor field were 250x250 units. The sensors in the four corners of the field were initially activated.

Table C.3: *Communication range, non-correlated case and rise time*

r^*	τ	N+
13	1	0.04
14	54	261.72
20	15	362.47
40	8	383.65
120	4	391

False alarms

False alarms occur if a detector responds when a chemical agent (CA) is not present, false positive, or it fails to respond to a CA that is present, false negative, the alarm levels for a detector are deliberately set low to ensure a minimal number of false negatives, however this means that false positives are more likely.

Table C.4: *Communication range, correlated case and rise time*

r^*	τ	N+
13	1	0.04
14	60	95.58
20	28	318.01
40	7	383.42
120	4	391

False positive alarms are usually observed when the targeted compound is in the presence of an interference, which may be a chemical molarity similar to a CA, or a substance which may contain elements that are also present in CAs. For example, pesticides containing sulfur or phosphorus would generate a false positive CA alarm when an FPD-based detector is used.

The occurrence of false positives in a civilian setting may have serious implications as it could lead to extreme disruption and possibly panic. More importantly however, repeat false positive alarms could lead to future 'real' alarms being ignored. At present detectors are prone to give false positive alarms as most detect multiple compounds with none being completely selective for a specific CA or class of agents. To overcome this problem another detector, based on a different technology, can be used to confirm any alarm.

False negative alarms are more problematic than false positive alarms because the failure to produce an alarm may lead to dangerous situations. The failure of a detector to alarm to a CA that is present may be due to any number of reasons including operator error, changing environmental conditions, humidity effects, detector malfunction such as software quirks, and the presence of chemical interferences which may mask normal detection capabilities. Ideally, false alarm rates should be zero

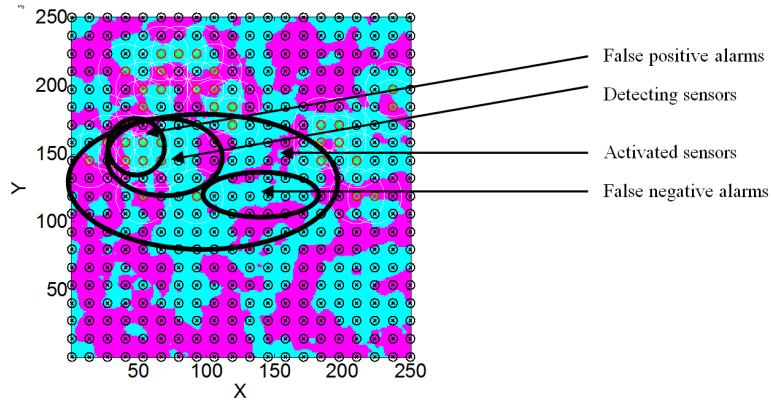


Figure C.1: WCSN behaviour with false alarms in correlated chemical tracer environment.

but in practice this is rarely so. It is therefore imperative that the likely incidence of false responses and the detectors' ability to resist interferences be thoroughly explored prior to its deployment [88]. Figure C.1 shows the correlated tracers regions and sensor activation with false alarms.

Bibliography

- [1] K. Akkaya and M. Younis. A survey on routing protocols for wireless sensor networks. *Ad Hoc Networks*, 3(3):325 – 349, 2005.
- [2] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *Communications Magazine, IEEE*, 40(8):102 – 114, Aug. 2002.
- [3] G. Anastasi, M. Conti, M. D. Francesco, and A. Passarella. Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks*, 7(3):537 – 568, 2009.
- [4] S. Appadwedula, V. Veeravalli, and D. Jones. Energy-efficient detection in sensor networks. *Selected Areas in Communications, IEEE Journal*, 23(4):693 – 702, April 2005.
- [5] M. Bahrepour, Y. Zhang, N. Meratnia, and P. J. Havinga. Use of event detection approaches for outlier detection in wireless sensor networks. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference on*, pages 439–444. IEEE, 2009.
- [6] R. Bakhshi, D. Gavidia, W. Fokkink, and M. Van Steen. A modeling framework for gossip-based information spread. In *Quantitative Evaluation of Systems (QEST), 2011 Eighth International Conference on*, pages 245–254. IEEE, 2011.

- [7] S. Basu, E. Foufoula-Georgiou, and F. Porté-Agel. Synthetic turbulence, fractal interpolation, and large-eddy simulation. *Phys. Rev. E*, 70(2):026310, Aug. 2004.
- [8] A. Benavoli, B. Ristic, A. Farina, M. Oxenham, and L. Chisci. An application of evidential networks to threat assessment. *Aerospace and Electronic Systems, IEEE Transactions on*, 45(2):620–639, 2009.
- [9] V. Bisignanesi and M. Borgas. Models for integrated pest management with chemicals in atmospheric surface layers. *Ecological Modelling*, 201(1):2–10, 2007.
- [10] J. Bongard. Biologically inspired computing. *Computer*, 42:95–98, 2009.
- [11] M. Borgas. Multifractal coding for turbulent mixing of plumes. *New Journal of Physics*, 2009.
- [12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *Information Theory, IEEE Transactions on*, 52(6):2508 – 2530, June 2006.
- [13] S. M. Brennan, A. M. Mielke, and D. C. Torney. Radioactive source detection by sensor networks. *IEEE Trans. Nuclear Science*, 52(3):813–819, 2005.
- [14] E. H. Callaway. *Wireless Sensor Networks: Architectures and Protocols*. CRC Press, Inc., Boca Raton, FL, USA, 2003.
- [15] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic. Towards optimal sleep scheduling in sensor networks for rare-event detection. In *Proceedings of the 4th international symposium on Information processing in sensor networks, IPSN '05*, Piscataway, NJ, USA, 2005. IEEE Press.
- [16] R. Cardell-Oliver. Rope: A reactive, opportunistic protocol for environment monitoring sensor networks. In *Embedded Networked Sensors, 2005. EmNetS-II. The Second IEEE Workshop on*, pages 63 – 70, May 2005.

- [17] Centre for Excellence in Emergency Preparedness (CEEP). What is CBRN?, July 2013.
- [18] J. Chan, J. Bailey, and C. Leckie. Discovering correlated spatio-temporal changes in evolving graphs. *Knowl. Inf. Syst.*, 16(1):53–96, 2008.
- [19] Y. L. Chen and J. Wen. Sensor system design for building indoor air protection. *Building and Environment*, 43(7):1278 – 1285, 2008.
- [20] J.-C. Chin, D. K. Yau, N. S. Rao, Y. Yang, C. Y. Ma, and M. Shankar. Accurate localization of low-level radioactive source under noise and measurement errors. In *Sensys '08: Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 183–196, New York, NY, USA, 2008. ACM.
- [21] O. A. Cirpka, M. Rolle, G. Chiogna, F. P. de Barros, and W. Nowak. Stochastic evaluation of mixing-controlled steady-state plume lengths in two-dimensional heterogeneous domains. *Journal of Contaminant Hydrology*, 138139(0):22 – 39, 2012.
- [22] R. Colwell and J. Peeters. Creating a nationwide wireless detection sensor network for chemical, biological and radiological threats. *Gentag White Paper*, 2007.
- [23] D. J. Daley and J. Gani. *Epidemic Modeling*. Cambridge University Press, U.K., 2000.
- [24] P. De, Y. Liu, and S. K. Das. An epidemic theoretic framework for evaluating broadcast protocols in wireless sensor networks. In *Proc. MASS 2007: IEEE International Conference on Mobile Adhoc and Sensor Systems*, pages 1–9, Pisa, Italy, 2007.

- [25] F. De Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac. A graph-based model for disconnected ad hoc networks. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pages 373–381, May 2007.
- [26] S. De Vito, P. Di Palma, C. Ambrosino, E. Massera, G. Burrasca, M. Miglietta, and G. Di Francia. Wireless sensor networks for distributed chemical sensing: Addressing power consumption limits with on-board intelligence. *Sensors Journal, IEEE*, 11(4):947–955, April 2011.
- [27] A. Dekker and A. T. Skvortsov. Topological issues in sensor networks. In *Proc. MODSIM 2009:International Congress on Modelling and Simulation*, Cairns, Australia, 2009.
- [28] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In *Proceedings of the sixth annual ACM Symposium on Principles of distributed computing*, PODC ’87, pages 1–12, New York, NY, USA, 1987. ACM.
- [29] L. Dinga and Z.-H. Guan. Modeling wireless sensor networks using random graph theory. *Physica A*, 12:3008–3016, 2008.
- [30] S. Dorogovtsev and J. Mendes. *Evolution of Networks*. Oxford University Press, UK, 2003.
- [31] P. K. Dutta. *On random event detection with wireless sensor networks*. PhD thesis, The Ohio State University, 2004.
- [32] C. Erten, P. J. Harding, S. G. Kobourov, K. Wampler, and G. Yee. Exploring the computing literature using temporal graph visualization. *Visualization and Data Analysis 2004*, 5295(1):45–56, 2004.

- [33] E. Ertin, J. W. Fisher, and L. C. Potter. Maximum mutual information principle for dynamic sensor query problems. In *Information processing in sensor networks*, pages 405–416. Springer, 2003.
- [34] S. Eubank, V. A. Kumar, and M. Marathe. Epidemiology and wireless communication: Tight analogy or loose metaphor? In *Bio-Inspired Computing and Communication*, pages 91–104. Springer, 2008.
- [35] P. T. Eugster, R. Guerraoui, S. B. Handurukande, P. Kouznetsov, and A.-M. Kermarrec. Lightweight probabilistic broadcast. *ACM Trans. Comput. Syst.*, 21:341–374, Nov. 2003.
- [36] J. P. Fitch, E. Raber, and D. R. Imbro. Technology challenges in responding to biological or chemical attacks in the civilian sector. *Science*, 302(5649):1350–1354, 2003.
- [37] R. Fuentes-Fernández, M. Guijarro, and G. Pajares. A multi-agent system architecture for sensor networks. *Sensors*, 9(12):10244–10269, 2009.
- [38] G. U. Gamm, M. Sippel, M. Kostic, and L. M. Reindl. Low power wake-up receiver for wireless sensor nodes. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010 Sixth International Conference on*, pages 121–126. IEEE, 2010.
- [39] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [40] M. González, H. Herrmann, and A. Araújo. Cluster size distribution of infection in a system of mobile agents. *Physica A: Statistical Mechanics and its Applications*, 356(1):100–106, 2005.

- [41] L.-J. Guan and M. Duckham. Decentralized computing of topological relationships between heterogeneous regions. In *GIScience 2010: Proceedings of the Sixth international conference on Geographic Information Science*, New York, NY, USA, 2010. Springer.
- [42] A. Gunatilaka, B. Ristic, and R. Gailis. On localisation of a radiological point source. In *Information, Decision and Control, 2007. IDC'07*, pages 236–241. IEEE, 2007.
- [43] A. Gunatilaka, B. Ristic, A. Skvortsov, and M. Morelande. Parameter estimation of a continuous chemical plume source. In *Proc. Fusion 2008: 11th International Conference on Information Fusion*, Cologne, Germany, 2008.
- [44] A. Gunatilaka, A. Skvortsov, et al. Chemical source backtracking in turbulent boundary layer (TBL). In *Proc. The 5TH WMO Symposium on Data Assimilation*, Melbourne, Australia, Oct. 2009.
- [45] Z. Haas, J. Halpern, and L. Li. Gossip-based ad hoc routing. *IEEE-ACM Transactions on Networking*, 14(3):479 – 491, 2006.
- [46] D. Hall and J. Llinas. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1):6 –23, Jan. 1997.
- [47] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [48] R. L. Haupt and S. E. Haupt. *Practical genetic algorithms*. John Wiley & Sons, 2004.
- [49] S. E. Haupt. A demonstration of coupled receptor/dispersion modeling with a genetic algorithm. *Atmospheric Environment*, 39(37):7181–7189, 2005.

- [50] S. E. Haupt, C. T. Allen, and G. S. Young. A genetic algorithm method to assimilate sensor data for homeland defense applications. In *Adaptive and Learning Systems, 2006 IEEE Mountain Workshop on*, pages 243–248. IEEE, 2006.
- [51] S. Hicks and B. Welch. Sensor network area protection system, July 2013.
- [52] J. H. Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. University of Michigan Press, 1975.
- [53] S. Karunasekera, A. Skvortsov, and A. Gunatilaka. Minimizing the operational cost of chemical sensor networks. In *Proceedings of the 6th International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2010*, pages 37–42. IEEE Press, Dec. 2010.
- [54] S. Kashyap, S. Deb, K. V. M. Naidu, R. Rastogi, and A. Srinivasan. Efficient gossip-based aggregate computation. In *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS ’06, pages 308–317, New York, NY, USA, 2006. ACM.
- [55] S. Kashyap, S. Deb, K. V. M. Naidu, R. Rastogi, and A. Srinivasan. Efficient gossip-based aggregate computation. In *PODS ’06: Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 308–317, New York, NY, USA, 2006. ACM.
- [56] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection theory*. Prentice Hall, 1998.
- [57] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4, Nov 1995.

- [58] A. Khelil, C. Becker, J. Tian, and K. Rothermel. An epidemic model for information diffusion in manets. In *Proceedings of the 5th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, pages 54–60. ACM, 2002.
- [59] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science, New Series*, 220(4598):671–680, 1983.
- [60] R. J. Klein, S. E. Proctor, M. A. Boudreault, and K. M. Turczyn. Healthy people 2010 criteria for data suppression. *Statistical Notes*, 2002.
- [61] R. V. Kulkarni, A. Frster, and G. K. Venayagamoorthy. Computational intelligence in wireless sensor networks: A survey. *Atmospheric Environment*, 13(1):68–96, 2011.
- [62] Y. Ma, Y. Guo, X. Tian, and M. Ghanem. Distributed clustering-based aggregation algorithm for spatial correlated sensor networks. *Sensors Journal, IEEE*, 11(3):641 –648, March 2011.
- [63] J. Mathieu, G. Hwang, and J. Dunyak. Transferring insights from complex biological systems to the exploitation of netted sensors in command and control enterprises. In *Command and Control Research and Technology Symposium (CCRTS), June 20-22, 2006*. The MITRE Corporation, 2006.
- [64] C. Mendis. *Optimised Sink Trajectories for Sensor Networks*. LAP Publishing, Germany, 1st edition, 2009.
- [65] C. Mendis, S. M. Guru, S. Halgamuge, and S. Fernando. Optimized sink node path using particle swarm optimization. In *Advanced Information Networking and Applications, 2006. AINA 2006. 20th International Conference on*, volume 2, pages 5–pp. IEEE, 2006.

- [66] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.
- [67] K. L. Mills. A brief survey of self-organization in wireless sensor networks. *Wireless Communications and Mobile Computing*, 7:823–834, 2007.
- [68] H. Mohtadi and A. Murshid. *A Global Chronology of Incidents of Chemical, Biological, Radioactive and Nuclear Attacks: 1950-2005*. U.S. Department of Homeland Security, U.S. Department of Homeland Security, Washington DC, USA, 2006.
- [69] H. Mohtadi and A. Murshid. *A Global Chronology of Incidents of Chemical, Biological, Radioactive and Nuclear Attacks: 1950-2005*. The University of Minnesota, 1st edition, 2006.
- [70] T. Monahan and J. T. Mokos. Crowdsourcing urban surveillance: The development of homeland security markets for environmental sensor networks. *Geoforum*, 49(0):279 – 288, 2013.
- [71] M. Morelande, B. Ristic, and A. Gunatilaka. Detection and parameter estimation of multiple radioactive sources. In *Information Fusion, 2007 10th International Conference on*, pages 1–7. IEEE, 2007.
- [72] Y. Moreno, M. Nekovee, and A. Vespignani. Efficiency and reliability of epidemic data dissemination in complex networks. *Phys. Rev. E*, 69(5), May 2004.
- [73] J. D. Murray. *Mathematical Biology*, volume 1. Springer, USA, 2002.
- [74] NATO Civil Emergency Planning Civil Protection Committee. *Guidelines for first response to a CBRN Incident*. Civil Emergency Planning, Operations Division - NATO International Staff, 1st edition, 2013.

- [75] R. W. Nelson. Concept of operations for cbrn wireless sensor networks. Master's thesis, Naval Postgraduate School, USA, 2012.
- [76] R. Nemzek, J. Dreicer, D. Torney, et al. Distributed sensor networks for detection of mobile radioactive sources. *IEEE Trans. Nuclear Science*, 51(4):1693–1700, 2004.
- [77] M. Newman, A.-L. Barabási, and D. J. Watts. *The structure and dynamics of networks*. Princeton University Press, 2006.
- [78] S. Olariu and A. Y. Zomaya. *Handbook Of Bioinspired Algorithms And Applications*. CRC Press, Inc., Boca Raton, FL, USA, 2006.
- [79] W. K. H. Panofsky. Nuclear proliferation risks, new and old. *Issues Sci. Technol. Librariansh.*, 194(SLAC-REPRINT-2003-035):73–75, 2003.
- [80] K. Premkumar and A. Kumar. Optimal sleep-wake scheduling for quickest intrusion detection using wireless sensor networks. In *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pages 1400 –1408, April 2008.
- [81] I. Rao, A. Harwood, and S. Karunasekera. An unstructured peer-to-peer approach to aggregate node selection. In *Proceedings of the 17th international symposium on High Performance Distributed Computing*, HPDC '08, pages 209–212, New York, NY, USA, 2008. ACM.
- [82] N. S. Rao, M. Shankar, J.-C. Chin, D. K. Yau, C. Y. Ma, Y. Yang, J. C. Hou, X. Xu, and S. Sahni. Localization under random measurements with application to radiation sources. In *Information Fusion, 2008 11th International Conference on*, pages 1–8. IEEE, 2008.
- [83] W. Ren, Q. Zhao, and A. Swami. Connectivity of heterogeneous wireless networks. *Information Theory, IEEE Transactions on*, 57(7):4315–4332, 2011.

- [84] M. Richards, M. Ghanem, M. Osmond, Y. Guo, and J. Hassard. Grid-based analysis of air pollution data. *Ecological Modelling*, 194(1):274–286, 2006.
- [85] B. Ristic, A. Skvortsov, and M. Morelande. Predicting the progress and the peak of an epidemics. In *Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 54–60, Taiwan, 2009.
- [86] P. Robins, V. Rapley, and P. Thomas. A probabilistic chemical sensor model for data fusion. In *Information Fusion, 2005 8th International Conference on*, volume 2, pages 7–pp. IEEE, 2005.
- [87] C. Rosales and C. Meneveau. A minimal multiscale Langrangian map approach to synthesize non-gaussian turbulent vector fields. *Physics of Fluids*, 18(7):075104, 2006.
- [88] R. Sferopoulos. *A Review of Chemical Warfare Agent (CWA) Detector Technologies and Commercial-Off-The-Shelf Items*. Defence Science and Technology Organisation, Fishermans Bend, Victoria 3207, Australia, 2009.
- [89] F. K. Shaikh, A. Khelil, B. Ayari, P. Szczytowski, and N. Suri. Generic information transport for wireless sensor networks. *Sensor Networks, Ubiquitous, and Trustworthy Computing, International Conference on*, 0:27–34, 2010.
- [90] L. B. Shaw and I. B. Schwartz. Fluctuating epidemics on adaptive networks. *Phys. Rev. E*, 77(6):066101, June 2008.
- [91] X. Sheng and Y.-H. Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Trans. Signal Processing*, 53(1):44–53, Jan. 2005.
- [92] B. I. Shraiman and E. D. Siggia. Scalar turbulence. *Nature*, 405(6787):639–646, 2000.

- [93] A. Skvortsov and B. Ristic. Modelling and performance analysis of a network of chemical sensors with dynamic collaboration. *International Journal of Distributed Sensor Networks*, 2012(656231), 2012.
- [94] A. Skvortsov, B. Ristic, and M. Morelande. Networks of chemical sensors: a simple mathematical model for optimisation study. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference on*, pages 385–390. IEEE, 2009.
- [95] D. Stauffer and A. Aharony. *Introduction to percolation theory*. Taylor and Francis, 1991.
- [96] D. L. Stephens and A. J. Peurrung. Detection of moving radioactive sources using sensor networks. *IEEE Trans. Nuclear Science*, 51(5):2273–2278, 2004.
- [97] P. D. Stroud, S. J. Sydoriak, J. M. Riese, J. P. Smith, S. M. Mniszewski, and P. R. Romero. Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Mathematical biosciences*, 203:301–318, 2006.
- [98] Y. Sung, H. Poor, and H. Yu. How much information can one get from a wireless ad hoc sensor network over a correlated random field? *IEEE Transactions on Information Theory*, 55(6):2827–2847, 2009.
- [99] S. Suram and K. M. Bryden. Solution of a two-dimensional inverse radiation problem using evolutionary algorithms. In *ASME 2004 Heat Transfer/Fluids Engineering Summer Conference*, pages 819–823. American Society of Mechanical Engineers, 2004.
- [100] P. Thomas, M. Bull, and G. Ravenscroft. Wmd hazard predictionblending three ai techniques to produce a superior defence capability. In *Applications and Innovations in Intelligent Systems XI*, pages 125–136. Springer, 2004.

- [101] H. L. V. Trees. *Detection, Estimation and Modulation Theory (Part I)*. John Wiley & Sons, 1968.
- [102] M. C. Vuran, O. B. Akan, and I. F. Akyildiz. Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45(3):245 – 259, 2004.
- [103] B. Walsh. Markov chain monte carlo and gibbs sampling. *Lecture Notes for EEB 581*, 26(1):1–24, 2004.
- [104] F. Wuhib, M. Dam, R. Stadler, and A. Clem. Robust monitoring of network-wide aggregates through gossiping. *Network and Service Management, IEEE Transactions*, 6(2):95 –109, June 2009.
- [105] L. Zhang, X. Wang, and S. Li. A routing algorithm based on clusters and balanced use of energy for wireless sensor networks. *Journal of Control Theory and Applications*, 8:81–85, 2010. 10.1007/s11768-010-9195-7.
- [106] F. Zhao and L. Guibas. *Wireless Sensor Networks: An Information Processing Approach*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [107] F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration for tracking applications. *IEEE Signal Processing Magazine*, 19(2), 2002.
- [108] C. C. Zou, W. Gong, D. Towsley, and L. Gao. The monitoring and early detection of internet worms. *IEEE/ACM Transactions on Networking (TON)*, 13(5):961–974, 2005.
- [109] A. M. Zungeru, L.-M. Ang, and K. P. Seng. Classical and swarm intelligence based routing protocols for wireless sensor networks: A survey and comparison. *Journal of Network and Computer Applications*, 35(5):1508–1536, 2012.