Trust and Reputation in Multi-Modal Sensor Networks for Marine Environmental Monitoring

Edel O'Connor B.A.(Mod.) M.Sc.

A dissertation submitted in fulfilment of the requirements for the award of

Doctor of Philosophy (Ph.D.)

to the

Dublin City University
School of Computing

Supervisors: Prof. Alan F. Smeaton and Prof. Noel E. O'Connor

17th January, 2012

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

ID No:

Date:

Abstract

Greater temporal and spatial sampling allows environmental processes and the well-being of our waterways to be monitored and characterised from previously unobtainable perspectives. It allows us to create models, make predictions and better manage our environments. New technologies are emerging in order to enable remote autonomous sensing of our water systems and subsequently meet the demands for high temporal and spatial monitoring. In particular, advances in communication and sensor technology has provided a catalyst for progress in remote monitoring of our water systems. However despite continuous improvements there are limitations with the use of this technology in marine environmental monitoring applications. We summarise these limitations in terms of *scalability* and *reliability*. In order to address these two main issues, our research proposes that environmental monitoring applications would strongly benefit from the use of a multi-modal sensor network utilising visual sensors, modelled outputs and context information alongside the more conventional in-situ wireless sensor networks. However each of these additional data streams are unreliable. Hence we adapt a trust and reputation model for optimising their use to the network.

For our research we use two test sites - the River Lee, Cork and Galway Bay - each with a diverse range of multi-modal data sources. Firstly we investigate the coordination of multiple heterogenous information sources to allow more efficient operation of the more sophisticated in-situ analytical instrument in the network, to render the deployment of such devices more *scalable*. Secondly we address the issue of *reliability*. We investigate the ability of a multi-modal network to compensate for failure of in-situ nodes in the network, where there is no redundant identical node in the network to replace its operation. We adapt a model from the literature for dealing with the unreliability associated with each of the alternative sensor streams in order to monitor their behaviour over time and choose the most reliable output at a particular point in time in the network. We find that each of the alternative data streams demonstrates themselves to be useful tools in the network. The addition of the use of the trust and reputation model reflects their behaviour over time and demonstrates itself as a useful tool in optimising their use in the network.

Acknowledgements

When I first began my PhD in CLARITY (or at the time it was known as the CDVP) I looked at other PhD researchers that were near completion and I never imagined the time would come where I would be sitting in the lab at the stage where I was writing my acknowledgements with all this work behind me! I can safely say it has been a journey to say the least. However it has been a journey I have thoroughly enjoyed with many highs and a whole new set of experiences I would never have had the opportunity to encounter otherwise. There are many people that helped to make this possible and who thoroughly enlightened me along the way.

I would like to greatly thank my supervisors Alan Smeaton and Noel O'Connor. I have learnt a great deal from Alan and Noel over the past few years and I have been very lucky to have them as my research supervisors. They provided constant encouragement and support and a wealth of knowledge and inspiration for which I will always be grateful. They also allowed me to find my own path in terms of my research goals and objectives which is a skill that will stand to me for a long time post-PhD. They allowed this autonomy while also helping me to see the bigger picture and ensuring that I was on track in terms of my research goals. They also often brought new opportunities to my attention which allowed me to achieve the most from my PhD. In the last few months of my writing, Alan provided great advice and support and I have to apologise to him for the initial very long chapters that took over a couple of his weekends!

I would like to thank the Beaufort Marine Research Awards which funded this research and without which this opportunity would not have been possible. It also provided me with the opportunity to work with a number of fantastic researchers at the Marine Institute who provided me with a great amount of knowledge and resources especially at the early stages of this research. Guy Westbrook, Paul Gaughan, Glenn Nolan, Caroline Cusack and Kieran Lyons are some of the MI personnel that I have worked with over the years and they have provided a wealth of knowledge and support for which I am very grateful. Barbara Fogarty, our national co-ordinator for the Advanced Marine Technology Programme also worked tiredlessy to provide an abundance of information on opportunities in the field, and for that I am also very grateful.

In DCU I have been very lucky to work with people from a variety of research centres and disciplines. It has provided me with a breadth of knowledge I could never have imagined possible on beginning a PhD in the School of Computing and being from a computer science background! From biofouling to contaminated shellfish, I've heard it all! This was made possible through working within MESTECH (or when I began - the Beaufort research group). I would like to thank Dermot Diamond and Fiona Regan for the great support and opportunities that have provided me with over the last few years. I am very grateful and the research would not be where it is today without their input and support. I would also like to thank other people I have worked with within MESTECH and the NCSR - Declan, Aoife, Mary, Tim, Edwina, Aine, Cormac and Antóin Lawlor. They have also greatly supported my work along the way and opened up my research knowledge to whole new areas.

I have also been very lucky to work within CLARITY. I have had the opportunity to work with superb researchers and experts in their field who have provided an indispensable amount of knowledge and support throughout the completion of this

research. I have also made some very good friends along the way and who will remain friends for a long time to come. A special thanks to Adam Bermingham who sat beside me for four years and was there for all the highs and lows. Adam helped me to see the bigger picture in days when I was getting stuck in the details and has really been a great support throughout this process and most importantly a great friend. I would also like to thank Graham Healy who was there for most of the late nights in the lab through the final days of this work and was also a great support. Other past and present members of CLARITY have also provided great support and made my experience over the last few years very enjoyable - Paul, Daragh, Colum, Neil, James, Pete, Kevin, Eamonn, Niamh, Sinéad, Cathal, Aiden, Phil, Ciarán and the list goes on!

Last but most certainly not least I would like to thank my family and friends. To the 'home' gang, I can't thank you enough, and even though you tried to derail me a couple of times, in hindsight I can see they were much needed breaks! To my family, I won't even start because if I did I'd never finish! The support you have shown me down through the years makes me one of the luckiest around, and I will forever be grateful for the opportunities you have provided me with. I have to especially thank Miriam for the food runs and allowing me to call her in the middle of the night when I was walking out of DCU in the early hours! I also have to thank my mother Rena who has shown me incredible support over the years, she is always someone I can rely upon no matter what the problem! She never ceases to be there to support me and is constantly looking out for me. I would also like to thank Odhran who has been very understanding and supportive over the last few months. He really helped me to get through the final hurdles and for that I will always be very grateful. Finally my grandparents have been a very inspirational part of my life and it is with great sorrow that my grandad is no longer with us but I know he would be very proud.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The demand from various scientific and management communities for reliable monitoring of our coastal and inland waters has increased. We wish to investigate issues such as climate change, the impact of human activities on ecosystems and water quality and the overall state and well-being of our rivers, lakes and coastal zones. These resources represent vital assets on many levels playing a key role in many environmental processes and supporting a range of commercial and recreational activities. In these environments an array of biological, chemical, geological and physical processes occur over a range of temporal and spatial scales. These are dynamic environments affected by a range of anthropogenic factors as well as naturally occurring processes. Water quality can be affected by a variety of factors such as recreational and commercial activities, runoff from the local watershed, runoff from urban areas or industrial output and in the case of coastal zones, discharge from rivers in the catchment or environmental catastrophes such as oil spills.

The need to continuously protect, regulate and monitor these environments is being recognised with the introduction of a growing body of legislation such as the EU Water Framework Directive[1]. The introduction of legislation such as this is increasing the need for advanced technologies to manage water quality. All EU Member states must now achieve good status in all waters and maintain that status.

[1]http://www.wfdireland.ie

The establishment of this EU directive has led to the need for water managers to adopt a new approach to managing their waters. They are under increasing pressure to monitor inland waters and coastal zones in relation to a variety of water quality variables. In recent years there have been investigations into streamlining marine environmental monitoring with in-situ wireless sensor networks (WSNs) proposed as an effective tool for continuous real time monitoring at greater temporal and spatial scales.

While there have been huge advances, there are a number of limitations with the use of these technologies for achieving the scale of sensing required for monitoring widespread locations. This research proposes that marine environmental monitoring applications would strongly benefit from the use of a multi-modal sensor network utilising visual sensors, modelled outputs and context information alongside the more traditional in-situ wireless sensor networks. Additionally this research proposes the use of a trust and reputation framework to deal with the inherent unreliability associated with these alternative sensing modalities and modelled outputs in order to optimise their benefits to the network. Throughout this thesis the formal 'We' and 'Our' is used to refer to the work completed by the author in carrying out the research required to complete this body of work.

1.1 Motivation

Field measurements for coastal or river environments involves costly, time and labour-intensive on-site sampling and data collection, transportation to laboratories for analysis, and then subsequent evaluation. This type of sampling is too limited on temporal and spatial scales to adequately monitor the quality of water bodies on a long term basis, to model and understand key environmental processes, or to capture dynamic events which may pose a threat to the environment or human health. In the past this type of sampling has also introduced various data quality issues through inadequate quality-control and quality assurance protocols such as

extended holding times before analysis and the use of non-standardised methodologies (Glasgow et al., 2004).

New technologies are helping to streamline the water quality monitoring process and are enabling the collection of more data from more places, and more cost effectively than in the past. In recent years, the use of in-situ WSNs for monitoring our aquatic systems has been investigated to allow continuous real-time remote monitoring of these environments at greater temporal and spatial scales. This provides an opportunity for long-term data collection at scales and resolutions that are difficult or impossible to obtain otherwise. The data collection process is streamlined with a minimisation of human errors and time delays increasing the quantity and quality of data on temporal and spatial scales with a possibility of real-time alert notifications of harmful marine events (Glasgow et al., 2004). Data can also be accessed remotely which negates the need for data collection in sometimes hazardous or hard to reach environments.

1.1.1 The Importance of High Spatial and Temporal Monitoring

Greater temporal and spatial sampling allows environmental processes and the well-being of our waterways to be monitored and characterised from previously unobtainable perspectives. Many important environmental processes demonstrate high frequency spatial variation and are extremely heterogeneous. Observing these processes with high fidelity allows us to create models, make predictions and better manage our environments (Estrin, 2007). Undersampling on a temporal scale can result in masking the variability caused by processes occurring at higher frequencies than the sampling rate (Johnson et al., 2007). Sampling at a limited number of points spatially in the environment can mask the dynamics or trajectory of a phenomenon. It also leads to incomplete understanding of our aquatic environments and the natural or anthropogenic factors influencing processes in that environment.

From an operational perspective, high spatial and temporal monitoring allows the development of rapid detection and response systems to deal with environmental threats such as flooding, harmful algal blooms (HABs), pollution or oil spills (Glasgow et al., 2004).

For example one phenomenon frequently highlighted as causing problems in coastal and river systems is that of eutrophication (Smith et al., 1999; McGarrigle et al., 2010). Eutrophic waters are those having relatively large supplies of nutrients with eutrophication referring to "the process by which water bodies are made more eutrophic through an increase in their nutrient supply" (Smith et al., 1999). This over-enrichment of nutrients can stimulate excessive plant growth which can have serious effects on an aquatic ecosystem. These include issues such as increased biomass of freshwater phytoplankton which in some cases can be toxic, elevated pH and oxygen depletion, and an increased probability of fish kills among others (Smith et al., 1999; Mainstone and Parr, 2002; McGarrigle et al., 2010). In a recent report published by the Environmental Protection Agency of Ireland on water quality in Ireland, eutrophication is outlined as being of major concern in aquatic systems (McGarrigle et al., 2010). The eutrophication status of Irish inland and coastal waters is assessed with recommendations on issues that will need to be addressed for improving water quality e.g. identifying and tackling sources of diffuse pollution, infrastructural investment in new wastewater treatment plants etc.

High spatial and temporal monitoring is necessary to characterise rapid processes occurring in relation to nutrient loading of aquatic environments (e.g. (Donohue et al., 2005), (Moscetta et al., 2009)). Phosphorous and nitrogen are key nutrients that contribute to the eutrophication process (Smith et al., 1999). Although these nutrients occur naturally there are many human activities that lead to excessive amounts being found in our waters (Smith et al., 1999; Mainstone and Parr, 2002). In particular we can distinguish between point and diffuse sources of pollution with point sources typically attributable to outputs from sewage treatment works and diffuse sources mainly attributable to run-off such as that from agriculture (Smith

et al., 1999; Mainstone and Parr, 2002). Diffuse sources of pollution can be hard to identify and thus more difficult to address as shown by McGarrigle et al. (2010).

Measurements of nutrients in natural waters are often carried out through manual collection of samples, followed by laboratory analysis. Not only is this labour intensive, but it provides limited information about spatial and temporal distribution of these nutrients throughout a water system. Increased monitoring may help to identify sources of pollution or offending parties.

1.1.2 Technologies for Marine Environmental Monitoring

New technologies are emerging in order to enable remote autonomous sensing of our water systems and subsequently meet the demands for high temporal and spatial monitoring. In particular, advances in communication and sensor technology has provided a catalyst for progress in remote monitoring of our water systems (Glasgow et al., 2004). In recent years the concept of wireless sensor networks (WSNs) has been the focus of research. The concept is relatively new and involves a diverse range of technologies and disciplines while impacting a wide variety of application sectors (Diamond et al., 2008a). The demand for continuous assessment of nutrient concentrations in coastal and inland waters has also lead to the development of novel analytical instruments using newly emerging technologies (Moscetta et al., 2009; Diamond et al., 2008a). Despite continuous improvements there are limitations with the use of this technology in marine environmental monitoring applications.

These monitoring applications essentially require large-scale low-cost sensor networks that can operate reliably and autonomously over extended periods of time. However there is a significant gap between the current state of the art in both in-situ wireless sensor networks and analytical instruments and what is needed to realise this vision. Aquatic environments can be harsh environments for sustaining in-situ instrumentation. Also in times of extreme events such as flooding, such instrumentation is prone to failure. Sophisticated analytical instrumentation such as chemo-bio sensors are currently not suitable for scaled-up deployments over many months or

years in terms of sustainability, reliability or cost (Diamond et al., 2008a).

A more in depth analysis of this topic is provided in Chapter 2, however the two main issues are *scalability* and *reliability*.

- **Scalability:** The current state of the art in in-situ wireless sensor networks or analytical instruments is not suitable for scaled up deployments suitable to meet the demands of certain marine environmental monitoring applications.
- **Reliability:** Marine environments can be quite aggressive and sensor nodes are subject to failure or damage, especially when not maintained regularly. Failure of in-situ sensor networks may result in faulty data or gaps in coverage.

1.2 A Multi-Modal Sensor Network

In order to address the two main issues with using WSNs in an aquatic environment, our research proposes that environmental monitoring applications would strongly benefit from the use of a multi-modal sensor network utilising visual sensors, modelled outputs and context information alongside the more conventional in-situ wireless sensor networks. Multiple heterogeneous information sources provides increased information, more effective decision making and the more efficient use of sensing technologies in the network.

For the purposes of our work we consider the use of two forms of visual sensors. These are satellite sensors and off-the-shelf webcam type CCTV devices. Webcam-type CCTV devices can provide continuous daylight data for periods extending to decades at a very low cost. They can effectively quantify coastal and river parameters with high resolution in space and time. Unlike point sensors, they can have a view over a wide spatial area. Multiple cameras may be deployed and left for years. Therefore an abundance of data on a wide variety of parameters can be collected over long periods of time without human intervention at a scale which wouldn't be possible with in-situ sensor networks. They can also provide surrogate measurements

for parameters otherwise obtainable with sophisticated in-situ instrumentation e.g. a change in depth may indicate run-off which may indicate nutrient loading etc. Remote sensing from satellite or airborne sensors has proved to be a tremendous tool for studying our environment at large spatial scales and at a much higher frequency than was previously possible without the use of such technology. It offers unique large scale synoptic data to capture the range and variability of many complex processes, essentially providing high scale resolution data that would be economically and technically unfeasible with point sensors at such a resolution.

Despite the advantages associated with each of these data sources, they have their limitations. For example, the coarse spatial and temporal resolution of satellite remote sensing and issues with radiometric calibration (i.e. linking pixel intensities to a physical parameter), limited field of view and outdoor maintenance of webcams can be limiting factors. We are not proposing that these data sources should replace a conventional in-situ sensor network but rather a system than uses these in a complementary fashion can greatly benefit from each of their strengths for a comprehensive environmental event detection system. By the terms "multi-modal" and "heterogeneous" we also refer to the use of models in our network incorporating data from heterogeneous in-situ nodes in the network for the prediction of values of alternative in-situ nodes in the network. If our in-situ nodes are unreliable, and we do not have redundant in-situ nodes in the network available to us, we need to somehow optimise the technology in our multi-modal network to compensate for these deficiencies, and lead to a more efficient and effective use of technology in the network. This leads us to the two key research objectives of this thesis.

1.3 Research Objectives

Our research proposes that the coordination of multiple heterogeneous information sources can allow more efficient performance of the more sophisticated in-situ analytical instruments in the network, subsequently elongating their lifespan and

rendering them more cost efficient and reducing maintenance requirements. This may subsequently render the deployment of such devices more *scalable*, which is one of the two main issues with WSNs in marine monitoring. Here we propose to develop a model using rainfall radar information and in-situ depth information to provide information that can be used by an in-situ analytical instrument to perform more optimally and efficiently.

The second objective of our research relates to the issue of *reliability*. As nodes in in-situ sensor networks are subject to failure, we investigate the ability of a multi-modal network to compensate for such failure, where there is no redundant identical node in the network to replace its operation. We investigate the use of models to predict the values of alternative in-situ sensor nodes. In other words if an in-situ sensor node fails and we have no redundant node available as backup, we investigate how well we can replicate its behaviour from other data in the network, and estimate values for the missing data. In the literature, trust and reputation systems have been proposed as a tool to monitor the behaviour of nodes in a wireless sensor network where all the nodes are homogenous and in order to address the issue of data reliability and make in-network decisions. We adapt a model from the literature known as RFSN[2] (Ganeriwal et al., 2008) for dealing with the inherent unreliability associated with each of the alternative sensor streams in order to optimise their benefits to the network and choose the most reliable output at a particular point in time. More formally we state these research objectives in terms of the following hypotheses and research questions below.

[2]Reputation-based Framework for Sensor Networks

1.3.1 Hypotheses

1. *The use of multiple sensing modalities including visual sensors, context information and modelled outputs will enhance the use of an in-situ sensor network in marine environments.*

2. *A trust and reputation model adapted for use in a multi-modal sensor network will help to deal with the unreliability associated with the visual sensor streams and modeled outputs in the network and optimise their use by choosing the most reliable output at a particular point in time.*

1.3.2 Research Questions

In order to evaluate our hypotheses the following research questions need to be addressed:

1. *What information can be extracted from a low-cost off the shelf camera for complementing the in-situ sensors located at the deployment site? Can we develop models for classifying the relevant features in these images?*

2. *Can satellite earth observation data be used in the context of an operational multi-modal network where it is required to produce data at high temporal scales to complement and enhance in-situ sensors, so that in times of in-situ node failure it can act as an appropriate substitute?*

3. *Can multi-modal context information in the form of rainfall radar images and in-situ depth data be effectively modelled to produce outputs that can improve the efficiency of a more sophisticated in-situ node in the network?*

4. *Can alternative in-situ nodes in the network be modelled to replicate the behaviour of an alternative in-situ node for use in the situation where that node has failed?*

5. *Can a trust and reputation model developed for use in a network where all nodes in the network are homogeneous and there is multiple redundancy be effectively adapted for use in a multi-modal network where all nodes are heterogeneous and there are no redundant nodes?*

6. *Can this trust and reputation model help to address the unreliability associated with the visual sensors, satellite sensors and the modelled outputs and produce the most reliable output in the network at a particular point in time that best replicates the behaviour of the failed in-situ node?*

1.4 Overall Contributions

While the specific contributions of this thesis are outlined in the final conclusions chapter, it is important to point out that this research represents a significant and novel piece of work in relation to coastal and river environmental monitoring networks. It adopts novel and unique techniques for integrating a wide variety of technologies for more efficient and effective low-cost monitoring. It investigates the potential of a variety of sensing modalities for improving the effectiveness of environmental monitoring networks while also investigating techniques for optimising their use in the network. These investigations result in very interesting and promising results and provide a number of avenues for further progress and development. This work also involved the development of significant pieces of technology and a number of research publications which are outlined in Appendix A and Appendix B.

1.5 Thesis Outline

Chapter 1: In this current chapter we introduce the need for high spatial and temporal monitoring of inland and coastal waters, technologies that are helping to meet these needs and their limitations. This provides the motivation for this research and we subsequently introduce our research objectives, hypotheses and

research questions.

Chapter 2: This chapter provides an an overview of some key concepts from the literature in relation to environmental sensor networks, highlighting the difficulties with the current state of the art in achieving reliable large-scale autonomous sensor networks required for environmental monitoring applications. This chapter also provides an overview of trust and reputation models, along with a detailed outline of the model to be adapted for use in our research.

Chapter 3: In this chapter, we introduce our two test sites - the River Lee and Galway Bay and the multi-modal data sources available for incorporation into our network at these particular sites, along with the issues with these data sources.

Chapter 4: This chapter relates to research question 1. Here we demonstrate the use of a visual sensor (camera) in a river environment and how it can be used as a complimentary sensing modality to an in-situ environmental sensor network. The focus of this particular study is on the estimation of depth from the camera images since this particular parameter can be tied in with the in-situ sensor network. Depth is also quite an influential feature at the site and effects the values of other water quality parameters. We develop and evaluate classifiers for four features that can provide an estimation of depth.

Chapter 5: This chapter relates to research question 2 where we investigate the use of satellite imagery as a tool in a multi-modal sensor network that requires near real time cooperation between the remote and in-situ sensor streams, so that in times of node failure it can act as an appropriate substitute for an in-situ sensor. We provide an overview of the literature in this area, while outlining our reasoning for our choice of data streams. We also describe how we extract information from these data streams for use in the network.

Chapter 6: This chapter relates to research questions 3 and 4. The first part of this chapter investigates the development and evaluation of a model to predict average freshwater levels at the River Lee using rainfall radar images and in-situ depth data. These predictions can subsequently be used to control the operation of

a more sophisticated node in the network and improve its efficiency. We present a methodology for integrating pixel information from rainfall radar images and in-situ depth data into an ANN for predicting average freshwater levels. We investigate a number of issues such as the most effective way to present rainfall radar information extracted from a digital image to the network, the effects of rainfall from different points of catchment on the model, the effect of differing lag times on the model, and the effect effect of rainfall and water level information on the model. The second part of this chapter develops and evaluates models incorporating information from a variety of different combinations of heterogeneous sensor nodes for the prediction of an alternative in-situ node in the network in order to investigate if these models would provide a suitable substitute data stream in time of node failure.

Chapter 7: This chapter relates to research question 5. In this chapter we adapt the trust and reputation framework outlined in (Ganeriwal et al., 2008) to a multi-modal sensor network and examine its performance in three diverse application scenarios. In the first application scenario we examine its use for determining the most reliable model for replicating the actions of the in-situ sensor node in question. In the the second application scenario we examine its use for determining the most reliable satellite remote sensor data stream for reflecting the readings of the in-situ sensor node. In the third application scenario we examine its use for determining the most reliable visual data stream, while also using it to evaluate the use of an algorithm for relating visual features to the in-situ depth data.

Chapter 8: This chapter relates to research question 6. Here we examine the performance of the network from an applications perspective. We examine the type of performance that can be obtained from a multi-modal network if an in-situ sensor node were to fail, aware however that the needs of different applications vary in terms of the precision required. Then we carry out a case study evaluating the performance of the multi-modal network in detecting depth events at the River Lee test site. Subsequently we equip the network with a trust and reputation model and evaluate if this brings about improved performance in optimising the use of the

modelled outputs and visual sensor streams in the network.

Chapter 9 We summarise the outcomes of each of the chapters and present our overall conclusions in relation to our research hypotheses. Following this we reflect on the studies carried out and possible directions for future research.

Chapter 2

Overview

In this chapter we provide an overview of some key concepts from the literature in relation to environmental sensor networks. As outlined in Chapter 1 environmental monitoring applications require reliable data sampled at high spatial and temporal scales. However the following highlights a number of difficulties with the current state of the art in environmental sensing technology in achieving this. This overview provides the foundation for the motivation behind this research. It essentially highlights the issues with regards to sensor reliability and the need for developing innovative solutions for smarter, more efficient and effective sensor networks.

2.1 Wireless Sensor Networks

In the last two decades networking technologies have revolutionised the manner in which individuals and organisations exchange information and communicate. This technology is providing a foundation for a new wave of developments that involves observation and control of the physical world (Cerpa et al., 2001). Wireless sensor networks are a significant technology that has emerged as part of this new wave of developments. The advancements in this area have resulted in the development of low-cost, small, low-power multifunctional sensor nodes that consist of sensing,

data processing and communication components (Akyildiz et al., 2002). Chong and Kumar (2003) highlighted how these "cheap, smart devices with multiple onboard sensors, networked through wireless links and the Internet and deployed in large numbers" provide numerous opportunities in a number of application domains, such as in the home, cities and the environment and are a key technology for the future.

Sensor networks essentially provide a gateway through which the digital world can sense and respond to changes in the real world. The introduction of chemo/bio-sensing extends this to sensing the real world at the molecular level (Diamond et al., 2008a). The WSN concept envisages a world of ubiquitous sensing through large scale deployments of self-sustaining WSNs, linked to digital communications continuously monitoring our environment and instantly detecting and reporting changes in the quality of our environment. Internet enabling of these devices essentially leads to internet scale sensing and actuating systems (Diamond, 2004). Envisaged are hierarchical layers of autonomous sensors working collaboratively to monitor specific target parameters, adapting to the detection of specific events and sending personalised information back to the relevant destination.

2.2 Sensor Networks: Issues

Environmental monitoring applications generally require high spatial and temporal sampling. However analytical systems capable of some degree of scale up in the marine environment need to be small, portable, environmentally compatible, robust, inexpensive to own and operate, and capable of providing reliable analytical information over extended periods of autonomous operation (Bowden et al., 2002). Diamond (2004) layers analytical devices into a hierarchy in terms of sophistication, capabilities, operational costs and degree of autonomy, outlining a significant correlation between these factors and density of distribution. The key challenge is outlined as driving devices towards the more densely distributed layers by lowering cost while maintaining their reliability and data quality. The most densely

distributed layer is dominated by the use of physical transducers, such as pressure and temperature sensors. While transducer-based WSNs are very important, it is the introduction of chemo-/biosensing that will really lead to the vision outlined earlier. However the current state of the art in this technology is not ready for large scale deployments. These devices need to provide reliable analytical measurements while matching less sophisticated sensors such as physical transducers in terms of cost, power and ruggedness (Byrne and Diamond, 2006) which is a huge challenge for the materials and analytical science community.

Physical transducers are generally low cost, rugged, reliable and consume very little power. They are generally encased within a rugged encapsulant and can continue to function from within this environment. On the other hand chemo-/biosensing involves much more complex processes than the type of sensing carried out in transducer-based deployments. It invariably involves liquid handling (especially in the marine environment) and intimate binding events or reactions to occur either in solution (e.g. with reagent-based systems) or at the surfaces of sensing devices. This means they are prone to degradation as they depend on active materials and surfaces that interact with the sample. The operating characteristics of these active materials or surfaces may change over time and thus require frequent re-calibration for generating a reliable signal. Chemo-/biosensors generally consume considerably more energy in carrying out these complex processes and are generally quite expensive (Diamond et al., 2008a).

Diamond et al. (2008a) highlights significant challenges for the chemo-/biosensor research community to deliver sensing platforms that are appropriate for integration into scaled-up deployments in terms of reliability, cost and sustainability. These platforms need to be able to operate autonomously and reliably over extended periods of time without requiring maintenance and at a low cost. Diamond et al. (2008a) suggest some medium-term solutions along with possible avenues for further research, some of which is addressed in Section 2.3.

Even without the challenges of chemo/bio-sensing there are a number of chal-

lenges associated with large scale sensing in the environment. Deployments of other such networks are generally limited in number due to issues related to network, power and data handling and they are almost entirely restricted to transducers for detecting physical parameters (e.g. temperature, pressure, light, vibration etc.) (Byrne and Diamond, 2006). Deployments of these devices are still relatively small and there are difficulties in relation to optimising power consumption and communications capabilities (Diamond et al., 2008b). They are also prone to data faults and node failure. Xu et al. (2011) remarks a number of reasons for WSN failure such as hardware or software failures and malfunctions, radio interference, battery depletion or malicious damage. In particular data faults are a key issue in sensor network research if this data is to be beneficial in arriving at meaningful conclusions in a particular application domain. Deployment experiences have shown this to be a major issue that requires further analysis (Ni et al., 2009). Ni et al. (2009) provides a systematically characterized taxonomy of common sensor data faults outlining the most commonly used features to model both data and faults.

2.2.1 Issues with Sensor Networks in the Marine Environment

Even without the issues outlined above, deployment of a sensor network in the marine environment poses unique challenges. Sensors are subject to harsh conditions and often require greater levels of device protection. It is also quite a vast environment and energy consumption can be quite high since it often necessary to cover large distances and communications signals may be attenuated due to the constant motion of the water. Nodes can also be moved around by the tide or waves causing further difficulties. There may also be logistical problems associated with the deployment and maintenance of sensors and the need for additional devices in the marine environment e.g. flotation and mooring devices. The cost of the instrumentation is also often significantly higher than for a land-based WSN (Albaladejo

et al., 2010).

The reliable and successful operation of sensors in a marine environment brings with it a whole range of issues including obvious problems such as the possible effect of salt on the sensors but there also exists other issues that need to be addressed such as that of biofouling. Biofouling, the growth of nuisance or unwanted biofilms on surfaces, is a significant problem in marine environmental sensing. It limits data quality and deployment periods of instruments (Manov et al., 2004). Much research is focused on anti-fouling strategies for sensors in coastal or inland marine environments with a number of strategies proposed in the literature e.g. (Manov et al., 2004; Whelan and Regan, 2006; Sullivan and Regan, 2011). The significant attention this subject receives demonstrates the severe consequences this phenomenon has on autonomous marine sensor networks.

The marine environment can be quite vast and with the cost of marine environmental sensors, there are significant limitations to the spatial coverage that can be achieved with single point sensors with varying degrees of reliability. Marine monitoring applications often require high spatial coverage, hence techniques for increasing the efficiency of deployed sensors and the adoption of alternative sensing modalities such as visual or satellite sensors is worth exploring.

2.3 Sensor Networks: Opportunities for Progress

We now highlight some avenues for progress in order to achieve higher scale chemo/biosensing and also smarter, more efficient and effective sensor networks. Higher scale chemo-bio sensing would allow high frequency measurements from multiple points of our aquatic environments, as opposed to intermittant measurements from a very limited number of spatial areas. This would greatly improve our understanding of the spatial and temporal dynamics of various environmental processes, and our ability to capture dynamic and potentially harmful events.

2.3.1 Autonomous Chemo-/Biosensing with Potential for Scale-Up

Diamond et al. (2008a) examine developments in wireless sensor platforms that are helping deliver reliable autonomous chemo-/biosensing capable of some degree of scale up. We are still a considerable distance away from low cost, reliable, self-sustaining chemo/bio sensing devices (Diamond et al., 2008a,b; Diamond, 2004; Byrne and Diamond, 2006) but many of the interim solutions would benefit from a smarter more adaptive approach, whereby they operate with an awareness of their environment and changes to that environment. The following provides an overview of some these interim solutions and how they may benefit from a smarter more context-aware approach.

According to (Diamond et al., 2008a), the concept of 'micro-total analysis systems' or μTAS was introduced by (Manz et al., 1990) around 1990 and became known as lab-on-a-chip (LOAC). In principle LOAC devices, offer a route to incorporation of sophisticated chemo-/bio processing in a compact, low-power platform (Diamond, 2004). They offer a compromise between existing lab-based instruments and completely self-sustaining miniaturised sensors capable of massive scale up. As outlined in (Diamond et al., 2008b), the key component of such a device from an analytical perspective is the microfluidic manifold through which samples are accessed, reagents are added, measurements are made, and calibration is performed (Reyes et al., 2002). In its ultimate manifestation, this concept provides a route to the generation of field- deployable micro-dimensioned analytical instruments that could operate autonomously over relatively long periods of time.

However these devices can only store a limited amount of reagent and they generally require a lot of power (Diamond et al., 2008a). Thus they only have a limited number of samples before maintenance is required. If these samples could be used more effectively it would improve the efficiency of the device in terms of reagent consumption, power, maintenance, etc. Such a device has been developed

by the CLARITY research centre at Dublin City University (DCU) for monitoring phosphate levels in water (lakes, river, treatment plant outlets etc.) (Slater et al., 2010). Its portability, small size and potential low cost renders it very promising for some degree of scale up in the marine environment. Our work seeks to improve the efficiency of such a device by controlling the sampling rate based on contextual information from other sensing modalities in the environment. The possibilities associated with adaptive sensing are explored further in Section 2.3.3.

2.3.2 Alternative Sensing Modalities

Since the cost base is still quite high for autonomous analyser platforms, other mechanisms are also being proposed in the literature. However some of these other mechanisms involve completely alternative sensing modalities such as visual sensing. For example Fay et al. (2010) presents an approach using a robotic fish with an integrated low-power wireless video camera to provide information on water quality. The robotic fish is used in conjunction with low cost, dispersed colorimetric sensors. When a sensor has been located, the camera 'interrogates its condition' reporting the result via a wireless link. The authors state how this single low power robotic platform can provide analytical information at multiple locations using very low cost sensors.

Visual sensing sensing has been proposed as an alternative sensing modality in a variety of contexts throughout the literature, and not just in relation to chemical or biological sensing. By the term visual sensing we are referring to sensing from cameras or satellite-based imaging instruments. For example in previous studies, coastal video systems have been identified as effective tools for coastal monitoring. A prime example of this is a major European research project entitled CoastView (Davidson et al., 2007). This focused on the development of video systems in support of coastal zone management utilizing Argus technology. Argus stations consist of optical systems developed for nearshore sampling (Holman and Stanley, 2007). The CoastView project demonstrates the use of fixed video remote sensing systems

to partially ameliorate some of the problems associated with in-situ measurements of waves, currents, and morphological change. Davidson et al. (Davidson et al., 2007) refers to some of the research carried out investigating algorithms for the quantitative extraction of geophysical signals from image data including morphology (Aarninkhof et al., 2005), flows (Chickadel et al., 2003) and wave parameters (Holman and Chickadel, 2004) and refers to the scientific literature that has tested and reviewed the reliability, accuracy and versatility of coastal video systems. However despite the benefits there are also some drawbacks with the use of these types of systems. For example they need to be weather proofed and they also need to have sufficient power and network transmission speeds for continuing to operate and to relay back data. Data quality can also be affected by weather and illumination conditions, and can they can only provide data during hours of daylight.

Satellite remote sensing is also continuously evolving with the launch of new satellites and improved processing algorithms. There is much literature which focuses on the validation of earth observation data through match-ups with in-situ observations e.g (Nightingale et al., 2008; Wimmer et al., 2008; Barker et al., 2008). Its use has been investigated in the context of diverse applications in the marine and land environment e.g (Hansen et al., 2008; Ruddick et al., 2008; Stefouli et al., 2008; Brodsky, 2008; Pedrazzani et al., 2008). Its applicability to climate studies has also been addressed with particular emphasis on its role in the investigation of climate change (Knorr et al., 2008).

While visual and satellite sensing poses a number of opportunities in a wide variety of contexts, there are also a number of drawbacks with using this type of data. These issues are addressed further on in the thesis. However they represent a very valuable sensing modality for complementing the use of in-situ sensor networks, and we subsequently examine their use in the context of this research.

2.3.3 Smarter Networks

Diamond et al. (2008b) note how future sensor networks will be heterogeneous in nature, constructed from layers of sensing devices of similar complexity. These may range from very simple, low cost physical/chemical transducers to much more sophisticated devices like that described in Section 2.3.1. At the lowest level of complexity will be the most densely deployed devices consisting of relatively simple sensors which will feed information to the more sophisticated levels of sensing which are less densely distributed due to their cost and maintenance requirements. The less reliable but more abundantly deployed low cost devices will be used for providing early warning information of events to modify the operating characteristics of the more sophisticated nodes (e.g. increase sampling frequency, or wake up). These more reliable nodes can subsequently confirm or dispute the information coming from the less sophisticated layer (Diamond et al., 2008b). This can help to reduce the duty cycle of the more sophisticated nodes reducing the energy required and their overall efficiency and lifetime in the field, whilst also achieving high resolution sensing. The more sophisticated nodes may also provide a mechanism against which the more densely distributed but dumber devices can be remotely calibrated (Diamond, 2004).

Group behaviour strategies may be used to identify anomalous signals, device malfunction and also may provide information about the source, dynamics, trajectory and area of effect of an event. Diamond (2004) notes how in a densely distributed network that monitors water quality for example, it is unlikely an event will be detected on one device only. This event will normally be detected on a number of devices clustered in a particular region, and the temporal development of the signal pattern can be used to follow the dynamics of the event. An autonomous mobile device could subsequently be deployed to that region for locating and validating the event. Corresponding with this line of thought Estrin (2007) highlights that the power of embedded sensor network technology "derives from embedding measurement devices in the physical world and networking them to achieve intelli-

gent, adaptive, coordinated sensing systems". Estrin outlines that their approach to achieving scalable and robust systems is through exploiting heterogeneity and hierarchy in their designs, providing a number of examples of deployments using multi-layer actuated systems. Ganeriwal et al. (2008) interestingly notes how sensor nodes are implemented using inexpensive hardware components that are highly unreliable, however these barriers will not disappear with advances in technology since the driving force in sensor networks is the quantity of nodes as opposed to the quality of each individual sensor. This comment appears to have been made more in the context of large-scale transducer based networks. While marine sensor networks generally consist of a limited number of sophisticated nodes, as outlined above this may move towards a larger number of unreliable devices which can control the placement and the operational characteristics of the more sophisticated devices.Hence we need to understand the limits of these unreliable technologies and operate within these constraints.

Based on experiences in the field there are a number of notable comments made by Estrin (2007) regarding lessons learned and future research. These comments are mainly based around the theme of heterogeneity, hierarchy, and data quality which are central to this research. She states how it is not just about enabling the largest number of smallest, lowest power devices to create fully autonomous systems but it is about optimising the end to end heterogeneous system. She also outlines how rather than minimising energy use and bits transmitted, that its about minimising uncertainty. Consequently they have focused on the development of architectures and algorithms that optimise across components that include power on demand platforms, sensing modalities that combine physical and image sensing devices, and mobile and static sensor nodes, with mobility outlined as being key to addressing the issue of undersampling with static sensors. Another critical theme outlined by Estrin (2007) is that of data quality and system integrity. In-network processing algorithms and statistical techniques are being used as initial tools for calibration, self-test and validation.

The work carried out in our research is very much focused on optimising the technology available to us in the network to achieve a more efficient and effective sensor network. It is specifically focused on optimising a heterogeneous multi-modal sensor network for prolonging the operation of sensor nodes in the network, providing redundancy in the network in the case of node failure and addressing the issues of data quality in these modelled outputs. Key issues in environmental sensor networks are sensor reliability and scale of sensing. Incorporating redundancy in the network is often not an option in marine environmental monitoring due to cost and maintenance issues. Hence we need to investigate methodologies in how we can use available technology in the network more intelligently whilst also considering the use of alternative sensing modalities in the network such as visual sensing, to help to address these issues. The following provides an overview of some sensor network deployments from the literature that reflect some of the issues previously outlined.

2.4 Sensor Network Deployments

Example deployments of mote-based wireless sensor networks for environmental sensing cover a wide variety of applications contexts. Tolle et al. (2005) presents a case study of a WSN deployed across a 70-metre tall redwood tree at a resolution of 2 metres to monitor its microclimate. Another widely cited deployment is that of the use of wireless sensor networks for habitat monitoring at Great Duck Island (Mainwaring et al., 2002; Szewczyk et al., 2004). The aim was to monitor the occupancy of underground nesting burrows by sea birds and the role of factors relating to microclimate in their habitat selection. Another deployment described in (Cardell-Oliver et al., 2004) monitors soil moisture, adaptively adjusting the frequency of measurements to monitored rainfall. In (Werner-Allen et al., 2006b), the deployment of 16-wireless sensor nodes equipped with seismoacoustic sensors over 3 km on a volcano in northern Ecuador is described. An evaluation on the performance of this sensor network in terms of its effectiveness as a scientific instrument is carried

out in (Werner-Allen et al., 2006a). A number of issues were experienced during many of these deployments. For example some experienced problems with battery performance, problems with data delivery, data quality etc. As previously outlined data quality is an issue that requires huge attention in sensor network research if data from these devices are to be used for scientific analysis.

Estrin (2007) highlights deployments from four applications domains - seismic, terrestrial ecology, aquatic and contaminant transport. Many of these deployments are composed of multi-modal (includes sensors, cameras, actuators etc.) and multi-scale actuated sensing technologies. One particular technology that is employed is a robotic sensing node known as NIMS (Pon et al., 2005). Pon et al. (2005) highlights some of the limitations of conventional sensor network deployments to properly characterise environmental phenomena to a high level of fidelity citing issues such as spatiotemporal sampling, sensing uncertainty and their ability to effectively operate over long periods of time. Insufficient spatiotemporal sampling can lead to uncertainties in characterising what has been sensed. They also note that compact, low power, in-situ sensors are not available for monitoring many phenomena, and that it may be more suitable to have samples collected from a mobile device and analysed away from the environment. They introduce the NIMS method where robotic devices are deployed on cable way systems. These systems have been deployed in a wide range of contexts such as microclimate and solar radiation mapping and for measurement of water quality in the Los Angeles area watershed.

2.4.1 Marine Environmental Sensor Networks

Over the last ten years, there has been increased scientific interest in monitoring of the marine environment with the instruments used ranging from small-scale sensor networks such as WSNs to highly complex observation systems. Albaladejo et al. (2010) provide a systematic review of all high quality research evidence relevant to the use of WSNs in oceanographic monitoring. They distinguish two broad cat-

egories of marine wireless networks based on the data transmission medium that they use – WSNs based on radio frequency (RF) aerial communications (A-WSNs) and Under-Water Acoustic Sensor Networks (UW-ASNs). They note how most of the deployments reviewed cover relatively small marine areas (not more than 20km) with the distance between nodes in a range of 100-250m. Data was sampled in most deployments every 5 to 10 minutes. This is in contrast to coastal oceanographic observatories where the spatial resolution is in kilometres and the time resolution is in hours or days, however they often cover much larger areas.

A number of challenges are indicated that will need to be overcome to enable large-scale implementations of this type of network. These include reducing the costs of sensors which can account for up to 70 per cent of the total initial investment, incorporating efficient power supply systems to cover the duration of the deployment, continued improvements of communication systems, incorporation of components that guarantee appropriate levels of insulation and corrosion proofing, and buoy design which considers maintenance issues and eventual dismantling.

Johnson et al. (2007) provide an overview specifically of chemical sensor deployments. They outline how great progress has been made towards autonomous chemical sensing capabilities in the past decade by the marine and aquatic chemistry communities and that a variety of chemical sensing systems are now continuously deployed in marine environments, even through they are still not at the same level of cost or reliability as physical sensors. They outline the issues associated with undersamping of the marine environment and the need for sustained, large-scale, high-frequency observations which is not realistically possible through field sampling programs. However, there are no examples of chemical sensor networks operating at global scale, with some operating at smaller scales.

There are limited examples of chemical sensors that can operate autonomously for long time periods producing reliable measurements. In their review Johnson et al. (2007) describe chemical sensors that have been deployed in the marine environment and have data records of at least a few months. They then go on to describe examples

of networks that use these sensors or analysers. The majority of these deployments are based on sophisticated platforms, which demonstrates the lack of large-scale low cost chemical sensor deployments.

2.5 Trust and Reputation Systems

Trust and reputation systems are generally built around similar principles to that which are used in traditional social situations. For example we tend to only interact with those we consider as having a good reputation. This represents our opinion of another person and it may be based on our own direct experience with that individual, our observations of others experiences involving that individual or on reports of experiences from other trustworthy individuals.

There are a variety of descriptions of trust and reputation provided in the literature. In their survey of trust and reputation systems for online service provision Jøsang et al. (2007) note how trust is challenging to define due to the fact it 'manifests itself in many forms'. They highlight two common definitions of trust – *reliability trust* and *decision trust* and differentiate between reputation and trust through postulating reputation as a more global measure based on the ratings of members of a community, with trust being a more subjective notion based on various factors or evidence. Roman et al. (2009) states that while there is no consensus on the definition of trust that it is usually defined in terms of a *trustor* and *trustee* – the *trustor* is "the subject that trusts an entity or a service" and the *trustee* is "the entity being trusted". It is noted that there are several definitions for reputation highlighting one definition of reputation as "an expectation about an individual's behaviour based on information about or observations of its past behaviour" (Abdul-Rahman and Hailes, 2000). Similarly Aberer (Aberer and Despotovic, 2001) describes reputation as "a measure that is derived from direct or indirect knowledge on earlier interactions of agents and is used to assess the level of trust an agent puts into another agent". Ganeriwal and Srivastava (2004) defines reputation as "the perception that

a person has of another's intentions" and trust as "the expectation of one person about the actions of others". The notion of trust and reputation that will be used in this research is described further on when we are describing the trust and reputation model to be adapted in this work.

The manner in which a system discovers, records, and utilises reputation to form trust, and subsequently uses this trust to influence behaviour and decisions is referred to as a reputation and trust-based system (Srinivasan et al., 2006a). The use of reputation and trust to facilitate decision making has been studied in the context of a variety of diverse domains such as e-commerce, ad-hoc wireless networks, and peer-to-peer networks through various trust frameworks and reputation systems. The following provides an overview of the use of reputation and trust-based systems in these environments. This will provide a basis for an in-depth overview of the model proposed in Ganeriwal et al. (2008) which is the framework from which this research will be based.

2.5.1 E-Commerce and P2P Systems

The Internet has created new opportunities for the buying and selling of goods and services. However this often involves interacting with strangers without trusted third parties and subsequently a certain amount of risk is involved. The mechanisms from which trust is traditionally created are often not available in such an environment (Resnick et al., 2000; Resnick and Zeckhauser, 2002). The use of a reputation system has emerged as a method for fostering trust among strangers in e-commerce transactions (Resnick et al., 2000; Jøsang and Ismail, 2002; Patton and Josang, 2004). A reputation system collects, distributes, and aggregates feedback about participants' past behaviour (Resnick et al., 2000; Patton and Josang, 2004). It seeks to help people decide whom to trust, to encourage trustworthy behaviour, and also to deter participation from dishonest parties (Resnick et al., 2000). In order to understand how these systems can foster trust in such an environment, Resnick et al. (2000) examines how trust builds naturally in long-term relationships. They refer

to an effect called the "shadow of the future" by political scientist Robert Axelrod (Axelrod, 1984) which translates into an expectation that people will consider each other's past in future interactions. However strangers do not have knowledge of one another's past histories and no prospects of interacting in the future. To some degree, their 'good name' is not at stake since they are not subject to a network of informed individuals that will use their past interactions as a basis for future decisions. Therefore the temptation to act deceptively may outweigh the incentive to cooperate. The use of a reputation system seeks to reinstate the "shadow of the future" to transactions in an e-commerce environment by creating an expectation amongst users that their past behaviour will affect their future interactions (Resnick et al., 2000).

Online auction sites such as eBay.com were the first web-sites to introduce reputation systems (Jøsang and Ismail, 2002; Patton and Josang, 2004). In fact eBay's reputation system is one of the earliest and best known Internet reputation systems (Resnick and Zeckhauser, 2002). After a transaction is completed, the buyer and seller have the opportunity to rate each other and leave comments, therefore creating a public history of transactions from which future buyers or sellers can base their decisions. Yahoo! Auction, Amazon, and other auction sites also feature similar reputation systems to that used by eBay with some variations (Resnick et al., 2000). Reputations systems are also used by internet sites other than auction sites such as Bizrate.com, Expertcentral.com, and epinions.com.

Despite their usefulness and importance in an internet environment, reputation systems encounter significant challenges. Problems arise in eliciting, distributing and aggregating feedback (Resnick et al., 2000). For example some people may not bother to provide feedback at all, it can be difficult to ensure that feedback is honest, and it is relatively easy to change one's identity and start over. There are also issues in relation to aggregating and displaying feedback. Some simple numerical ratings may fail to convey important information from previous online interactions as shown by Resnick et al. (2000). A number of reputation systems,

have been proposed in the literature for online environments and agent systems (Yu and Singh, 2000; Zacharia et al., 2000). Some of these have attempted to address the various issues and challenges associated with these systems. For example Chen and Singh (2001) and Dellarocas (2000) attempted to address the issue of quality of the feedback.

According to Jøsang and Ismail (2002), there are two fundamental aspects for consideration in relation to internet reputation systems. Firstly, an engine which calculates the value of the users' reputation ratings and secondly a propagation mechanism which allows entities to obtain reputation values. Jøsang and Ismail (2002) outline how reputation engines proposed in the literature range from quite simple ones which simply accept numerical values and add them together (e.g. eBay) to those which perform more complex mathematical equations e.g (Yu and Singh, 2000). Two approaches are available for user reputation propagation - centralised or decentralised. In the centralised approach reputation values are stored in a central server (e.g. eBay) whereas in the decentralised approach individuals manage the reputation of others themselves (Zimmerman, 1995; Yu and Singh, 2000). Jøsang and Ismail (2002) propose a new reputation engine known as the beta reputation system. This reputation system has a firm basis in the theory of statistics and is based on the beta probability density function. Here they describe a centralised approach, however it can also be used in a distributed environment. The beta distribution has since been the foundation for other reputation systems in diverse domains such as mobile ad-hoc networks and wireless sensor networks including the reputation system outlined in Ganeriwal et al. (2008) which we adapt for use in our research further on in the thesis.

Jøsang et al. (2007) published a survey in relation to developments of trust and reputation systems for online service provision. They provide an overview of various principles used for computing reputation and trust measures such as summation or average of ratings, Bayesian systems, discrete trust models, belief models, fuzzy models and flow models, and describe the most well-known applications of online

reputation systems. All of these have a centralised architecture (as opposed to distributed) and use a simple summation or average for computing ratings apart from two systems which use a flow model e.g. Google's web page ranking system. They highlight problems with these systems and propose some possible solutions. Reputation based trust management systems have also been studied specifically in the context of P2P systems, with the publishing of many well known models in the literature. A widely cited algorithm for reputation management in P2P environments known as *EigenTrust* is outlined in Kamvar et al. (2003). Other well known models include *PeerTrust* (Xiong and Liu, 2003, 2004) and a model proposed by Aberer and Despotovic (Aberer and Despotovic, 2001).

2.5.2 Mobile Ad-hoc networks (MANETs)

Recently efforts have been made to model MANETs as trust and reputation based systems. Many of the issues that arise when unknown entities seek to transact in an e-commerce domain or a P2P system can translate to this environment. A MANET is a self-configuring system of mobile nodes connected by wireless links (Srinivasan et al., 2006b). The nodes are free to move randomly which means the network's topology can be changed rapidly and unpredictably. All network activities are carried out by nodes themselves and since there is no infrastructure to guarantee correct behaviour, the cooperation of all the nodes is required for the efficient functioning of the network. Even a few misbehaving nodes can have a severe impact on the network throughput. A node may agree to forward packets and then fail to do so. This may be for a variety of reasons. For example, a selfish node may not want to spend battery life, CPU cycles, or available network bandwidth to forward packets which are not of direct interest. However it will expect other nodes to forward packets on its behalf. A malicious node may deliberately drop packets to launch a denial of service attack on the network (Marti et al., 2000). Through modelling a MANET as a reputation and trust based system nodes can make reputation and trust guided decisions. For example a node may avoid

forwarding packets through any routes which contain misbehaving nodes or may refuse to forward packets for any nodes that it has 'blacklisted' as noncooperative.

Marti et al. (2000) proposed two techniques that improve throughput in an ad hoc network, where there are a certain number of nodes that agree to forward packets but fail to do so. Dynamic Source Routing (DSR), a protocol developed for routing in mobile ad-hoc networks, is complemented with a Watchdog component that identifies misbehaving nodes and a pathrater that helps routing protocols to avoid these nodes. This resulted in increased throughput in the network. However, using this approach, malicious nodes that do not cooperate are not punished. They are simply relieved of the burden of forwarding packets for others while their messages are forwarded by the non-misbehaving cooperative nodes. Also, the nodes rely on their own watchdog exclusively and do not exchange reputation information with other nodes in the network. Although this means that nodes are robust against false praise and false accusations from other malicious or colluding nodes in the network, the detection of malicious nodes may be slower.

Other protocols were proposed which aim to make misbehaviour unattractive through detecting and isolating misbehaving nodes and hence providing an incentive for cooperation. Ratings from other nodes in the network are also considered in order to speed up detection time. However this may render the system vulnerable to false accusations or false praise.

The CONFIDANT protocol proposed in (Buchegger and Boudec, 2002) seeks to find misbehaved nodes and isolate them so that misbehaviour will not pay off. Malicious nodes are detected through observing other nodes in the neighbourhood and from shared experience of malicious behaviour reported from other nodes in the network. It updates the reputation accordingly and if the reputation rating is bad, action is taken in terms of their own routing and forwarding and other nodes are informed by sending an ALARM message. CONFIDANT consists of a *monitor*, *reputation system*, *path manager* and *trust manager*. The aim of the *monitor* is to detect events or attacks. The trust attributed to warnings from second-hand

observations is controlled by the *trust manager*. The *trust manager* controls how incoming and outgoing second-hand reports are handled. The goal of the reputation system is to create knowledge based on these first and second hand observations. Unlike the approach described in Marti et al. (2000), this approach learns from reported malicious behaviour from other nodes which decreases the detection time of a misbehaving node when compared to relying exclusively on first hand observations. However, this renders the system vulnerable to false accusations where a trusted node makes wrong accusations or a sufficient number of nodes collude to make a wrong accusation. Also CONFIDANT relies solely on negative information; positive observations do not influence the rating.

Michiardi and Molva (2002) propose a generic mechanism based on reputation to enforce node cooperation. This mechanism called CORE is based on a collaborative monitoring technique where each network entity keeps track of other entities' collaboration using reputation. Reputation is calculated and updated through direct observations and through information provided by other nodes. However this scheme defines three types of reputation which are weighted for a combined reputation value - subjective reputation (direct observations), indirect reputation (positive reports by others) and functional reputation (task-specific behaviour). This combined reputation value is subsequently used to make decisions on whether to cooperate with a node or to gradually isolate it. As noted above the indirect reputation only takes positive values by other network entities to prevent malicious broadcasting of negative ratings for legitimate nodes. The potential from learning from the negative experience made by other network entities goes unused, making a similar trade-off between robustness and detection speed as the system proposed by Marti et al. (2000). This system is still also subject to false praise and may result in a good reputation for false nodes.

Buchegger and Boudec (2003a,b,c, 2004) propose a reputation system that makes systems such as CONFIDANT, robust against false accusations or false praise while retaining the benefit of using both positive and negative second hand information.

This reputation system formulates the problem in the realm of Bayesian analytics. RFSN (Ganeriwal et al., 2008), the model which will form the basis for the system developed in our research is developed using a similar Bayesian formulation. We will now provide an overview of this system.

2.5.3 Reputation-Based Framework for High-Integrity Sensor Networks

RFSN (Ganeriwal et al., 2008; Ganeriwal and Srivastava, 2004) is the first reputation and trust based system designed and developed exclusively for WSNs. Ganeriwal et al. (2008) examine the use of a trust and reputation based system for providing data reliability in a WSN. They propose that mechanisms in sensor networks to ensure the reliable relaying of data between trusted parties should also account for the accuracy and trustworthiness of the data itself. In (Ganeriwal and Srivastava, 2004) they argue that cryptography alone is not sufficient to counteract some of the misbehaviours encountered in WSNs. Subsequently they propose the integration of tools from various domains such as economics, statistics and data analysis with cryptography for developing trustworthy sensor networks.

Through the RFSN framework, network entities (i.e. sensor nodes) maintain reputation for other entities in the network. These reputation metrics are continuously built up over time from monitoring the behaviour of other nodes and rating them as being *cooperative* or *non-cooperative*. An entity uses this reputation to evaluate the trustworthiness of another entity and the data it provides, and to predict future behaviour. Based on this evaluation, decisions can be made and behaviours can be changed. The terms *cooperative* and *uncooperative* may adopt a different meaning depending on the system under consideration. Cooperation entails that a node should behave as expected in the network. In this work however the actions of interest are sensor readings and the level of confidence that can be associated with a data reading for a given sensor node.

RFSN borrows some design features from other works in the literature such as eBay (Resnick and Zeckhauser, 2002), CONFIDANT (Buchegger and Boudec, 2002), CORE (Michiardi and Molva, 2002), and peer-to-peer networks (Xiong and Liu, 2003). However as a complete system it differs from each of these. For example these may be limited in scope – e.g. they may only counteract routing selfish misbehaviour attacks; differ in architecture - e.g centralised versus distributed; assume a deterministic model for representing reputation or portray a very high level picture of the probabilistic framework based on debatable heuristics. As previously outlined reputation systems were proposed for MANETs which formulate the problem in the realm of Bayesian analytics e.g. (Buchegger and Boudec, 2004, 2003b,a,c). RFSN is developed using a similar Bayesian formulation which has a strong foundation in statistics and can counter any arbitrary misbehaviour of nodes.

The main properties of a reputation system are the representation of reputation, how the reputation is built and updated and how the ratings of others are considered and integrated (Buchegger and Boudec, 2004). RFSN employs a Bayesian formulation, specifically a beta reputation system, for the algorithmic steps of reputation representation, updates, and integration and trust evolution. Jøsang and Ismail (2002) provide a detailed analysis about reputation systems based on beta distributions. The two key components of RFSN are *Watchdog* and *Reputation*.

Watchdog

The Watchdog monitors the actions of other nodes and classifies these actions as being cooperative or non-cooperative. As previously explained, in the context of this work cooperation is determined by the quality of the actual sensor readings. The Watchdog both collects these observations and carries out the characterisation. Its objective is to detect the presence of invalid data resulting from compromised and faulty nodes. Instead of simply carrying out a binary classification i.e. cooperative or non-cooperative, a level of confidence or probability (any real number between (0,1)) can also be associated with the sensor reading.

The Watchdog is designed as a library of outlier detection protocols which allows easy integration of customized protocols for specific scenarios. A set of data readings is input to the watchdog and a level of confidence that can be associated with each data reading is output as a rating in the range (0,1). Ganeriwal et al broadly classify outlier detection protocols as either model-based or consensus based. Model-based outlier detection looks for data readings that deviate from a specific model where each data reading is evaluated independently against the model. A priori knowledge of this data model is required. Consensus-based outlier detection protocols generally look for consistency among data readings in a set and assign a level of confidence to a data reading proportional to its deviation from the consensus. These protocols require the readings for their functionality (i.e. they do not require any further data models than the readings), however the rating depends on other readings in the set.

In (Ganeriwal et al., 2008), Ganeriwal et al. use consensus or consistency-based outlier detection. More specifically, RFSN uses the density-based approach to consistency-based outlier detection. The prototype implementation of RFSN uses Local Outlier Factor, or LOF (Breunig et al., 2000) the simplest version of density-based outlier detection. For more details on this approach, see Breunig et al. (2000).

Reputation

The Reputation component maintains the reputation of a node. It manages reputation representation, reputation updates based upon the new observations made by the Watchdog, reputation integration based on other available information, reputation aging, and the creation of an output metric of trust. Reputation is maintained as a probabilistic distribution; therefore nodes are not constrained by discrete levels of reputation as with some e-commerce systems previously described such as eBay or Yahoo! Auctions. It is used to statistically predict the future behaviour of other nodes.

Reputation representation

RFSN uses a beta distribution to represent the reputation of a node. This has a strong foundation in statistics and it is indexed by two parameters (α, β). It can be expressed using the gamma function as:

$$P(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{(1-x)^{\beta-1}} \forall 0 \leq x \leq 1, \alpha \geq 0, \beta \geq 0 \tag{2.1}$$

The variable x is a probability variable. The mean of a beta distribution with parameters (α, β) is $\frac{\alpha}{\alpha+\beta}$ and its variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Reputation updates based on direct observation

Bayes theorem is used to calculate the probability of a belief given an observation whereby normalization is a step whereby the numbers are turned into probabilities by normalizing them so that they sum to one.

$$P(Belief/Observation) = \frac{P(Observation/Belief) * P(Belief)}{Normalization} \tag{2.2}$$

In RFSN, belief can be considered to represent the reputation of a node and the observation represents the direct observations made by a node about the other node. Therefore when node i receives output from its watchdog based on a direct observation of node j (D_{ij}), it updates the reputation of node j, R_{ij}, as follows:

$$R_{ij} = \frac{P(D_{ij}/R_{ij}) * R_{ij}}{\sum P(D_{ij}/R_{ij}) * R_{ij}} \tag{2.3}$$

If a node rates the behaviour of another node on a binary scale (cooperative (1), non-cooperative (0)), a classical beta-binomial framework for estimating reputations is adopted by RFSN. Letting θ denote the reputation of node j held by node i, a prior distribution $p(\theta)$ is assigned to θ which reflects the uncertainty about the behaviour of node j before any transactions take place with node i. The beta distribution with parameters (α, β) is used to represent the reputation of a node. Without any

prior information, θ is uniformly distributed over the measurement space, (0,0). Thus $p(\theta)$ = uni (0,1) = Beta (1,1). If node i subsequently interacts with node j in m+n events: out of which it characterises m+n interactions to be cooperative and non-cooperative, given this information, node i wants to predict the behaviour of node j (cooperative/non-cooperative), θ, for the next event - these interactions are modelled using a binomial distribution. Then the posterior distribution of θ is calculated as:

$$P(\theta) = \frac{Bin(m+n, m) * Beta(1,1)}{Normalization} = Beta(m+1, n+1) \tag{2.4}$$

The posterior distribution is also beta. The reputation of node j maintained at node i is subsequently represented by:

$$R_{ij} = Beta(\alpha_j + 1, \beta_j + 1) \tag{2.5}$$

α_j+1 and β_j+1 represents the cooperative and non-cooperative interactions between node i and j respectively. Without any prior observations α_j and β_j both equal zero, hence $R_{ij} = Beta(1,1)$. A reputation table RT_i at node i contains a tuple of the form (α_j, β_j) for every node j that i maintains a reputation metric for. If node i has built up some reputation for node j (R_{ij}) and it again interacts with node j for r+s more events (r cooperative and s non-cooperative), the new reputation of node j can be updated as:

$$P(\theta) = \frac{Bin(r+s, r) * Beta(\alpha_j + 1, \beta_j + 1)}{Normalization} ....or.... R_{ij} = Beta(\alpha_j + r + 1, \beta_j + s + 1) \tag{2.6}$$

Therefore the reputation update is equivalent to just updating the value of the two parameters α_j and β_j as follows:

$$\alpha_j^{new} = \alpha_j + r; \beta_j^{new} = \beta_j + s; \tag{2.7}$$

If interval ratings are used as opposed to binary ratings, and so are assigned some value in $[0, 1]$, Ganeriwal at al describe a more elaborate framework involving Dirichlet processes (Ferguson, 1973). However a detailed discussion in Ganeriwal et al. (2008) demonstrates that the two parameters α, and β can be maintained and that we replace the Bayesian update step with an identical bookkeeping step. Therefore after a single transaction, if the assigned probability of cooperativeness were p $\in [0, 1]$, the beta parameter updates would be:

$$\alpha^{new} = \alpha + p; \beta^{new} = \beta + p - 1 \tag{2.8}$$

Trust

Given a reputation metric R_{ij}, the trust metric T_{ij} is defined as node i's prediction of the expected future behaviour of node j. It is the subjective expectation a node has about another node's future behaviour. It is obtained by taking a statistical expectation of this prediction. Unlike reputation which is maintained as a probabilistic distribution, the trust metric is simply a number.

$$T_{ij} = E[R_{ij}] = E[Beta(\alpha_j, \beta_j)] = \frac{\alpha_j}{\alpha_j + \beta_j} \tag{2.9}$$

This metric can be used by a node in several ways for example in data fusion, node revocation and for decision making e.g. if there is a particular service from some other node in the network, T_{ij} can be used as a decision making criterion for choosing the best node or for choosing whether to cooperate with another node or not.

Reputation Integration and Ageing

As well as building up reputation from direct observations about other nodes in the neighbourhood, in RFSN nodes can also learn about other nodes through the experiences of other nodes in the neighbourhood. Ganeriwal et al classifies these indirect observations as *evidence* and this step as *reputation integration*. For reputation in-

tegration, they use an approach proposed in Jøsang and Ismail (2002), based on the concept of belief discounting (Jøsang, 2001). In their evaluation they do not evaluate this part of the framework and it is also not relevant in the context of this research. Hence no further description is provided here and readers are referred to (Ganeriwal et al., 2008) for further details. Similarly an aging mechanism is also incorporated into RFSN so that a node's trustworthiness is re-evaluated continuously. However this is also not relevant in the context of our research so is not described any further here.

2.5.3.1 Why RFSN ?

As previously outlined RFSN is described as being the first reputation and trust based system designed and developed exclusively for WSNs. While work has been carried out in relation to trust and reputation systems in areas such as P2P and Ad-Hoc Networks, the work in this area in the domain of WSNs is still in a relatively early stage. However Lopez et al. (2010) notes how this research area is becoming quite active, highlighting surveys that have been produced in recent years i.e. ((Aivaloglou et al., 2008; Fernandez-Gago et al., 2007; Roman et al., 2009)). Lopez et al. (2010) in fact attempts to derive trust management best practices in the context of WSNs and analyses the compliance of the current state of the art with these.

There are a number of reasons why the RFSN model was chosen for our research. Firstly, the model is based on the beta reputation system. Jøsang and Ismail (2002) outline how the beta reputation system has a firm basis in the theory of statistics, unlike many other reputation systems which are intuitive and ad hoc. In (Jøsang et al., 2007) the advantage of Bayesian systems is further re-enforced with the authors stating how they provide a theoretically sound basis for computing reputation scores. Secondly, in (Ganeriwal et al., 2008) it is noted that these Bayesian systems can counter any arbitrary misbehaviour of nodes and RFSN is described as a generalised and unified approach for countering all types of misbehaviour (Ganer-

iwal and Srivastava, 2004). Also in (Ganeriwal et al., 2008) RFSN is specifically investigated with regards to providing a generalised and unified approach for providing information about the data accuracy in sensor networks. Data accuracy is the focus of our research, hence choosing a model which focuses on this aspect of the network as opposed to another aspect of the network such as data routing is more appropriate. Thirdly, while RFSN is proposed as a distributed reputation system for sensor networks, it is flexible and can also be run in a centralised manner, which also suits the needs of the application context of our research. Finally, unlike many other models presented in the literature, the authors outline a detailed framework of application and RFSN has been developed as a middleware service on motes and simulated using real data sets. The evaluation carried out in (Ganeriwal et al., 2008) demonstrates its applicability in identifying misbehaving nodes for a variety of fault scenarios identified from previous deployments.

Trust and reputation models have since been implemented in sensor networks in a variety of application scenarios. In the literature they are often proposed from a security perspective due to the fact that nodes are susceptible to malicious attacks from adversaries due to the nature of their distributed deployment. Shaikh et al. (2006) note how trust is a prerequisite of any security implementation and that they are both highly interdependent. There are many reasons why these models are not suitable in the context of our research. Some are proposed in the context of a very different application scenario e.g. (Hu et al., 2011; Srinivasan et al., 2006b; Krasniewski et al., 2005), others are only specifically suitable really in the context of a WSN as opposed to the multi-modal sensor network proposed in the context of our research e.g. (Shaikh et al., 2006; Zia, 2008; Boukerch et al., 2007; Crosby et al., 2006), or suitable especially in the context of dense networks e.g. (Srinivasan et al., 2006b; Zhang et al., 2006). Others have a simplified or unsuitable method of recording and updating reputation e.g (Srinivasan et al., 2006b; Shaikh et al., 2006; Zia, 2008). Other proposed models address issues with RFSN that are not relevant in the context of our research, for example the fact that the *Watchdog* might be

compromised, or an adversary might be able to use the outlier detection scheme to their advantage, or a Bayesian system may be unsuitable if there is latency in the network e.g. (Chen et al., 2008; Zhang et al., 2006; Probst and Sneha, 2007). Other models address issues related to memory overhead e.g. (Probst and Sneha, 2007; Boukerch et al., 2007), while others look at context specific trust e.g. (Probst and Sneha, 2007) or are focused on a very specific problem like data aggregation e.g. (Zhang et al., 2006). Also in most of the proposals there is no real implementation that has been tested on a real network with real data.

2.6 Summary

As outlined in (Estrin, 2007), the basic hardware and software buildings blocks are now in place for wireless sensing systems and their associated applications. While there will continue to be research and innovation in relation to the basic building blocks of WSNs, progress to date has opened up a whole new set of research questions. For marine environmental monitoring applications the key issues that need to be addressed are reliability and scale of sensing. While there is much work being done in the sensor network and materials science communities, we need to investigate innovative ways to optimise technologies that are available to us. Many authors suggest that we need to investigate how we can use a dense number of low cost unreliable devices in a hierarchical network that contains sophisticated nodes for validation purposes. Diamond et al. (2008b) notes that the realisation of a "densely deployed heterogeneous sensor network requires significant effort to understand the level of information content that can be acquired from relatively dumb sensors used in a collaborative manner". We also need to consider how alternative sensing modalities such as visual sensing can contribute to our network, investigating innovative techniques to optimise a heterogeneous hierarchical system.

In line with this theme, our work is specifically focused on optimising a heterogeneous multi-modal sensor network for prolonging the operation of sensor nodes

in the network, providing redundancy in the network in the case of node failure and addressing the issues of data quality in these modelled outputs. This chapter has presented a review of work related to marine environmental sensing. We first covered wireless sensor networks and the issues with using them in a marine environment. We then focus on autonomous chemo-/biosensing and the potential for scale-up of such sensing in a marine environment. One of the central issues in this area is building up a quantified treatment of trust and reputation of sensing in an environment like the marine and to address this we introduce and describe the RFSN model for wireless sensor networks, and show how it can form the basis for a trust and reputation model for sensing in a marine environment. With this background and related work described, we now move on to the core work in the thesis. In the following chapter we introduce our test sites and data sources.

Chapter 3

Test Sites and Data Sources

3.1 Introduction

In order to investigate multimodal sensor networks we need to choose at least one site which has available to it as diverse a range of environmental sensor modalities as possible. Working with the Marine Institute, we have identified two such sites which are very different. One is a river at the point that it flows into the sea, through a city centre location, the River Lee at Cork city. The other is a large seawater bay on the west coast, Galway Bay.

The diversity of these sites and the range of sensors available to us at each of them are described in the following sections in this chapter. We begin with the River Lee and present a site overview in Section 3.2, followed by a description of sensor data sources and issues with these in Section 3.3. In Section 3.4 we present an overview of the Galway Bay site and in Section 3.5 we describe the sensor data and the issues with that data that are available to us from that site.

3.2 River Lee – Site Overview

The River Lee represents one of the largest rivers in the southwest of Ireland. It and its main tributaries – Rivers Sullane, Laney, Dripsey, Bride and Shournagh –

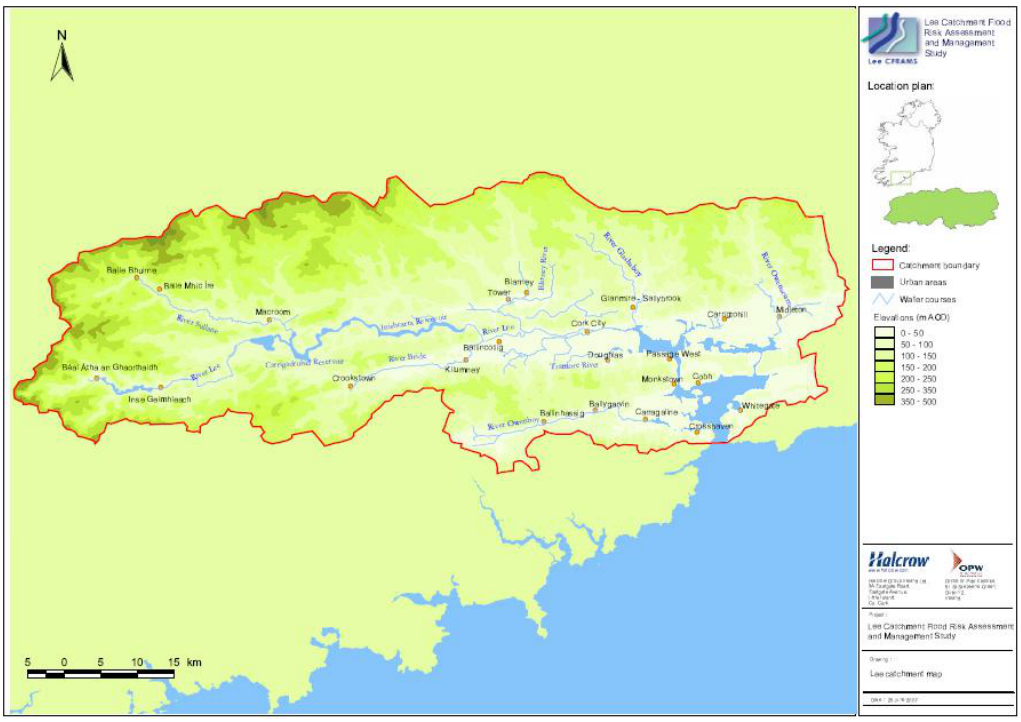

Figure 3.1: Lee catchment Source: Office of Public Works (2008)

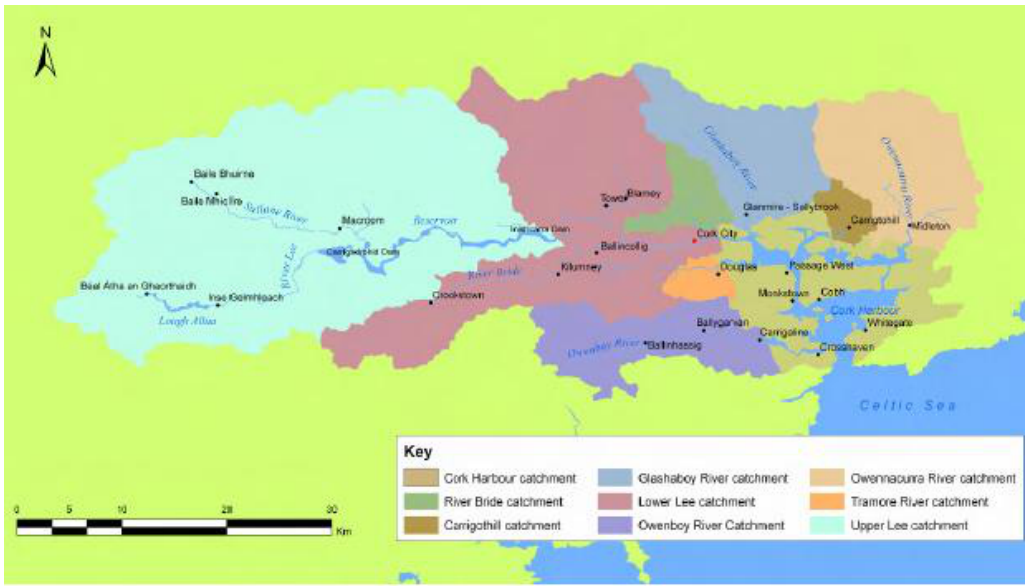

Figure 3.2: Lee catchment. Source: Office of Public Works (2008)

drain a catchment area of approximately 1,100km^2 upstream of Cork City (Office of Public Works, 2008). This catchment known as the 'Lee Catchment' drains into Cork Harbour and is shown in Figure 3.1.

The River Lee and its extended catchment area, including the Glashaboy, Owennacurra, and Owenboy rivers, represents an area with a high level of flood risk with a

number of significant incidents having occurred in the past. One of the more recent of these events took place in November 2009, where significant flooding occurred from the River Lee around Cork city and its hinterland. Information on past flood events for the catchment can be found on the Irish Office of Public Works (OPW) National Flood Hazard Mapping web site[1]. Fluvial, tidal and pluvial flooding where surface water cannot escape due to high river or tide levels all pose threats in the catchment. One of the worst fluvial floods occurred in August 1986, however most of the flood events have occurred in the winter season, especially during the month of November.

In August 2006, the Irish Office of Public Works commissioned a Lee Catchment Flood Risk Assessment and Management Study (Lee CFRAMS). This is outlined as a pilot flood risk assessment and management study in Ireland that will set a framework for future similar studies in other catchments across the country (Office of Public Works, 2008). The main objectives of the study include assessing and mapping the spatial extent and degree of flood hazard and risk in the catchment, examining future issues that may lead to an increased risk of flooding e.g. land use and climate change, building the information repository necessary for making decisions in relation to managing flood risk, and developing an economically, socially and environmentally appropriate long-term strategy for managing flood risk. Our description of the River Lee and its catchment is obtained from the hydrology report commissioned as part of this study (Office of Public Works, 2008)

For hydrological assessment and hydraulic modeling, in (Office of Public Works, 2008), the catchment was broken down into nine sub-catchment areas shown in Figure 3.2. These were further broken down themselves to provide detailed hydrological analysis and detailed hydrological inputs into the hydraulic model for each sub-catchment area. A significant amount of data was collected in undertaking the hydrological assessment including topographical, hydrometric, meteorological, tidal, mapping and historical data. However a lot of this data is not publicly available.

[1](http://www.floodmaps.ie

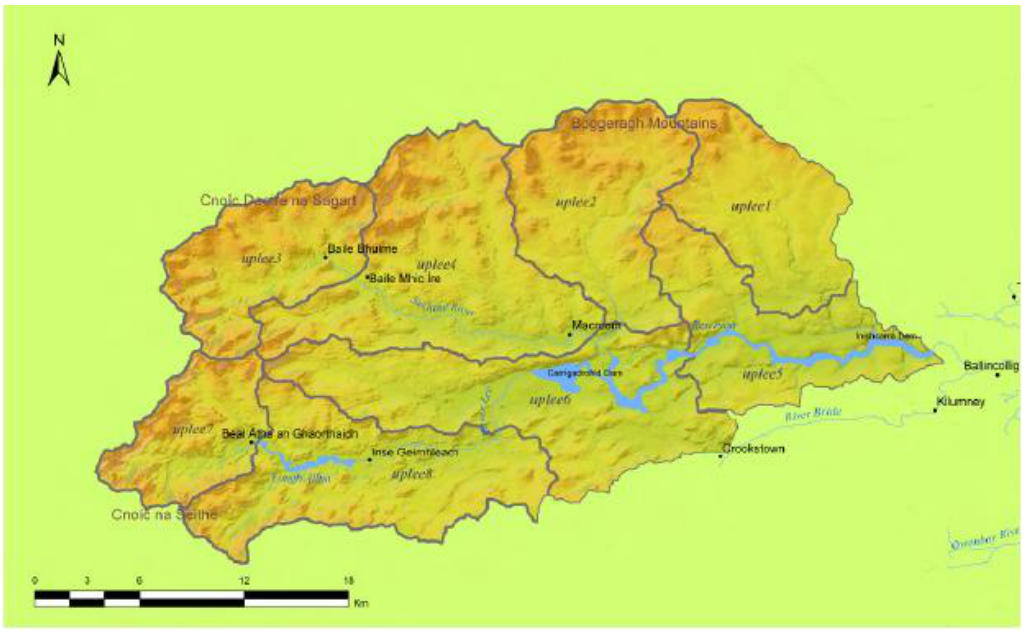

Figure 3.3: Upper Lee catchment broken into sub-catchments for the Lee CFRAMS study. Source: Office of Public Works (2008)

Our study is mainly concerned with the River Lee as opposed to the Lee catchment. In (Office of Public Works, 2008) the Upper Lee catchment is said to encompass an area of 790km^2 extending from the Inishcarra Dam westwards to the Shehy mountains (See Figure 3.3). The main rivers in the Upper Lee catchment include the Lee, Sullane, Foherish, Laney and Dripsey. The catchment uplands consist primarily of exposed rock and sandstone till subsoils, with the majority of the catchment consisting of deep, well-drained mineral soils with areas of peaty topsoil and blanket bogs which can be found in the uplands. Agricultural activities consist mainly of forestry and hill grazing. In the hydrology report it also states how the peat uplands and steep topography give a slightly elevated runoff potential. Based on their analysis of meteorological data, the annual average rainfall (AAR) for this part of the catchment is 1450mm. As discussed further on, there are two dams on the river Lee controlling the flow from the upper Lee catchment. These are the Carrigadrohid dam and the Inishcarra dam, with the Glengariff river and Dripsey river the two main rivers along this reach discharging to the dam reservoir. Approximately 0.3% of the upper Lee catchment is made up of urbanised areas with the

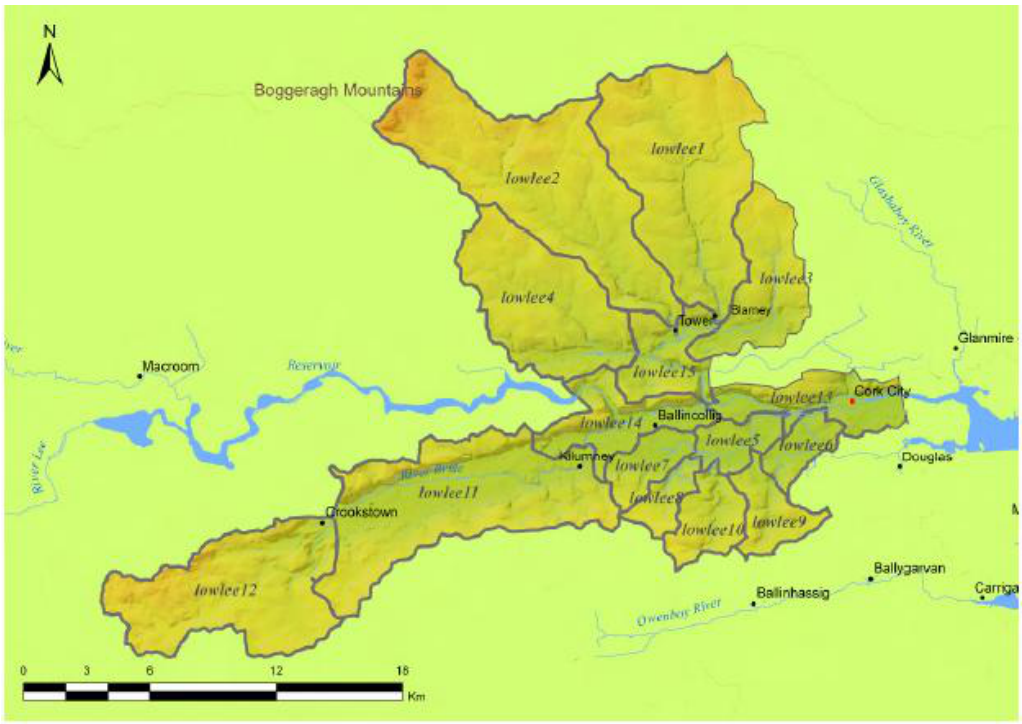

Figure 3.4: Lower Lee catchment broken down into sub-catchments for the Lee CFRAMS study. Source: Office of Public Works (2008)

majority of these being located along the primary watercourses.

In (Office of Public Works, 2008), the sub-catchment area outlined as the Lower Lee extends downstream of the Inishcarra dam to Cork Harbour over an area of approximately 420km^2 (See Figure 3.4). This is drained by a number of watercourses, however the main river draining this sub-catchment is the River Lee. It is stated how the River Lee flows primarily in a west east direction from downstream of Inishcarra dam through Cork City where it then discharges into Cork Harbour. The tidal cycle in Cork Harbour also greatly influences water levels of the River Lee in Cork City. The catchment mainly consists of sandstone till overlain by well drained acid brown earths. The report outlines how the geology and topography of the catchment results in a a lower runoff potential than the upper Lee catchment. The area has an Annual Average Rainfall (AAR) value of 1100 mm. This sub-catchment area is said to consist of good agricultural land mainly used for pasture grazing. Approximately 6% of the lower Lee catchment consists of urban areas, which can lead to increased runoff of rain.

3.2.1 Hydroelectric Dams

As previously outlined, the Carrigadrohid and Iniscarra hydroelectric dams owned by the ESB are located on the River Lee. The Iniscarra dam is located approximately 13 km west of Cork City (with Carrigadrohid Dam a further 14 km upstream) and it is used to generate electricity when there is demand. There are no specific times for releasing water from the dams, they are run to maximise electricity generation which is dependent on the availability of water, flow rate, daily electricity demand, etc. Control of water levels is also dependent on the season in question. During a flood event, the dam is operated according to the Regulations and Guidelines for the Control of the River Lee (Office of Public Works, 2008). However the operation of these dams have proved problematic in the past, with the controversial release of the Iniscarra dam in November 2009 contributing to major flooding in Cork city. One of the outcomes of the study carried out in (Office of Public Works, 2008) is a review of the operation of the hydroelectric dams before or during flood events. Real-time data streams and future predictions of water levels could greatly improve the operation of these dams during high-risk periods.

3.3 River Lee Data Sources

In this section some of the sources of river data we use in our experiments are outlined. These include in-situ sensor data from the DEPLOY project, visual data from a camera and context data from rainfall radar.

3.3.1 The DEPLOY Project

DEPLOY[2] (O'Flynn et al., 2010) is a technology demonstration project showing an implementation of state of the art technology for continuous, real-time monitoring of a river catchment. This demonstration project began collecting data from five sites on the River Lee at 10-15 minute intervals from April 2009 until May 2010.

[2]http://www.deploy.ie

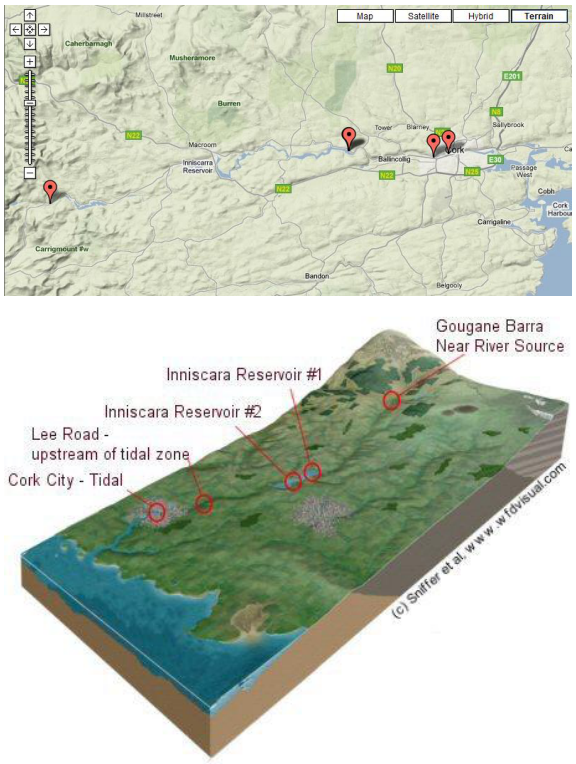

Figure 3.5: DEPLOY sites on the river Lee. Source: DEPLOY - Google Earth and www.wfdvisual.com

The project was co-funded by the Irish Marine Institute and the Environmental Protection Agency (EPA) and was seen to be a step towards the realisation of a wide area network of autonomous sensors for monitoring the temporal and spatial distribution of various water quality and environmental parameters.

The monitoring sites for DEPLOY are located in four zones representative of varying conditions along the river and shown in Figure 3.5. One station is near the source of the river at Gougane Barra, two stations are in the Inniscarra Reservoir, one station is in the main channel of the river (Lee Road) and the final station is in Cork City where the river has entered the estuary (Lee Maltings). This location on the river is tidal and partially saline. This was the site chosen for our study and for the rest of this thesis it is referred to as the Lee Maltings. These zones are considered typical of significant river systems, with stations situated at the source, reservoir, main channel, and an estuary (Lawlor, 2010). Each of the stations were equipped with a number of sensors which are described further in the following section.

3.3.1.1 Lee Maltings

The Lee Maltings site is located on the north channel of the river Lee at the Tyndall National Institute near the upper end of the estuary on a left hand bend of approximately 70° (DEPLOY, 2010). As previously outlined, the site is tidal with a tidal range of approximately 4m. During the Summer months large sections of the river bed tend to dry, however from October this rarely happens. At the site, a pump is situated on the quay side above the high water mark and water is pumped to a sampling tank at fixed intervals. Instruments deployed at this site include off-the-shelf commercial sensors for monitoring conductivity[3], chlorophyll-a-fluorescence, dissolved oxygen, temperature and water depth. We requested data from the project and received data for each of the above parameters from the end of April 24, 2009 to June 04, 2010.

There were a number of reasons for choosing the Lee Maltings site as the location for our study. Firstly, this site is located at the Tyndall National Institute which forms part of the CLARITY research collaboration. Tyndall provide facilities which allows us to instrument the site with other sensing modalities such as a camera sensor network. Secondly, the Lee Maltings site represented the only site with a reliable depth sensor for most of the study. Finally, the Lee Maltings represents an interesting and very challenging site to monitor due to the dynamics at the site caused by the tide and the dam further upstream, and its position on the river (at a bend where the river enters the sea).

The site is also historically important as it was used by the mathematician George Boole in his book "An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities", published in 1854, as a worked example to illustrate how to combine probabilities of independent events (Boole, 1854). In a footnote on page 256 Boole says "Opposite the window of the room in which I write is a field, liable to be overflowed from two causes, distinct,

[3]i.e. the ability of water to conduct electric current which can provide information regarding substances present e.g. dissolved salts can significantly increase the conductivity of water

but capable of being combined, viz., floods from the upper sources of the River Lee, and tides from the ocean." Thus the first attempt to predict water levels at this site go back over 160 years.

3.3.1.2 Other DEPLOY sites

The following provides a brief overview of the four other DEPLOY sites as a comparison to the Lee Maltings site. The Lee road site is situated approximately 500m upstream of where the river becomes estuarine and tidal. The station is powered by a small solar panel and includes sensors for monitoring conductivity, temperature, pH, chlorophyll-a and water turbidity. The Iniscarra Pumphouse station is located on the northern shore of Inniscara Reservoir which is situated approximately 10 km from where the River Lee enters the estuary. The system is physically located in the intake tower of Cork County Council's water treatment plant which is approximately 2km upstream of Inniscarra Dam. The sensors deployed here monitor conductivity, pH, chlorophyll-a, dissolved oxygen and water temperature. The Iniscarra Reservoir station consists of a buoy which is moored in approximately 20m of water upstream of the pumphouse station. The Gougane Barra monitoring station is located approximately 3km downstream of Gougane Barra lake at the source of the river Lee. Different dynamics are experienced at each site with the Lee Road experiencing almost immediate effects from the release of the dam (Lawlor, 2010), eutrophication has been a concern at the Iniscarra Reservoir and the rapid rise of water levels is a dynamic experienced at Gougane Barra (DEPLOY, 2010).

3.3.2 SmartCoast

Prior to the DEPLOY project, an in-situ wireless sensor monitoring system was deployed in the River Lee at the Lee Maltings site as part of the SmartCoast project[4] (O'Flynn et al., 2007). This project was launched in June 2005 as a building block towards a coastal monitoring programme based on sensor technologies capable of

[4]http://www.tyndall.ie/projects/smartcoast/

monitoring the spatial and temporal distribution of important environment parameters prioritised by the European Water Framework Directive (WFD). The project ceased uploading data to the SmartCoast web site in January 2009 however there were issues with the sensing devices long before this date. SmartCoast was an applied R&D Project, co-funded by the Irish Marine Institute and the EPA. Among its aims were the development of a WSN featuring plug and play sensor platform, novel sensors and low power consumption communications for enabling in-situ chemical and physical measurements and subsequently allowing this data to be view in real-time on the Internet.

During the SmartCoast project lifetime key water, environment, and system parameters were measured approximately every 15 minutes, relayed back to a central server and accessed from a portal page. These parameters included turbidity, water temperature, pH, conductivity, water depth, air temperature, relative humidity and daylight level, received signal strength (RSS) and battery voltage.

3.3.3 Issues with In-Situ Data Sources – River Lee

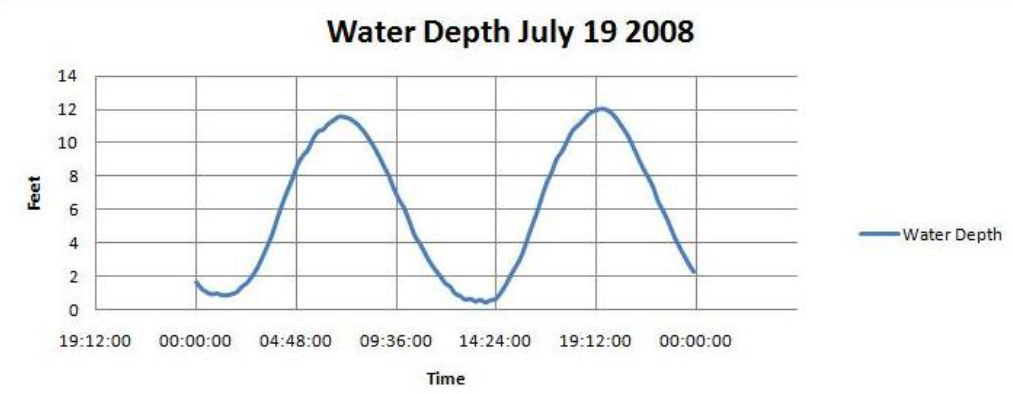

Figure 3.6: Water depth from the SmartCoast depth sensor on July 19, 2008

Despite the numerous benefits provided by the in-situ deployments as part of SmartCoast and DEPLOY, there have been issues which have been highlighted throughout the lifetime of both projects. Firstly there appears to be gaps in the sensing, other problems include irregularities with the actual sensor data itself. This

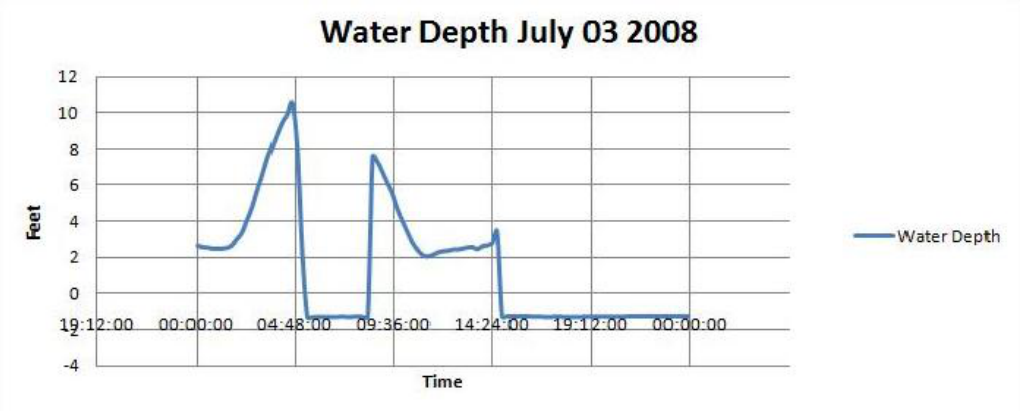

Figure 3.7: Water depth from the SmartCoast depth sensor on July 03, 2008

may be due to the fact that the sensors needed to be maintained more regularly or may have suffered from the harsh conditions that are often associated with deploying sensors in a marine environment. Biofouling is often a major problem especially during the warmer Summer months.

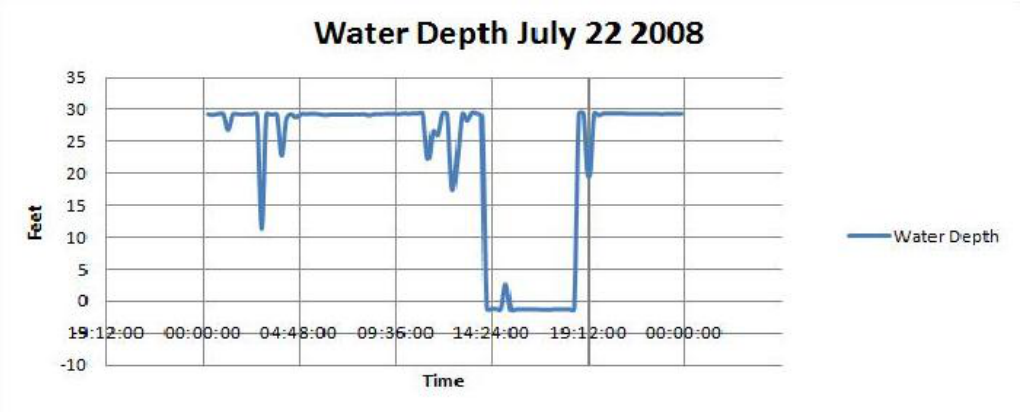

Figure 3.8: Water depth from the SmartCoast depth sensor on July 22, 2008

To demonstrate some of the issues with a marine environmental sensor network, an analysis was carried out on the data collected from the SmartCoast water depth sensor for the month of July 2008. The Lee Maltings site is tidal meaning that the depth data should essentially be periodical, following the sort of pattern similar to that displayed in Figure 3.6 which shows sampled data from the water depth sensor for July 19th, 2008. Sometimes the sensor would intermittently produce erroneous data which indicates a problem with the sensor. Figure 3.7 shows data

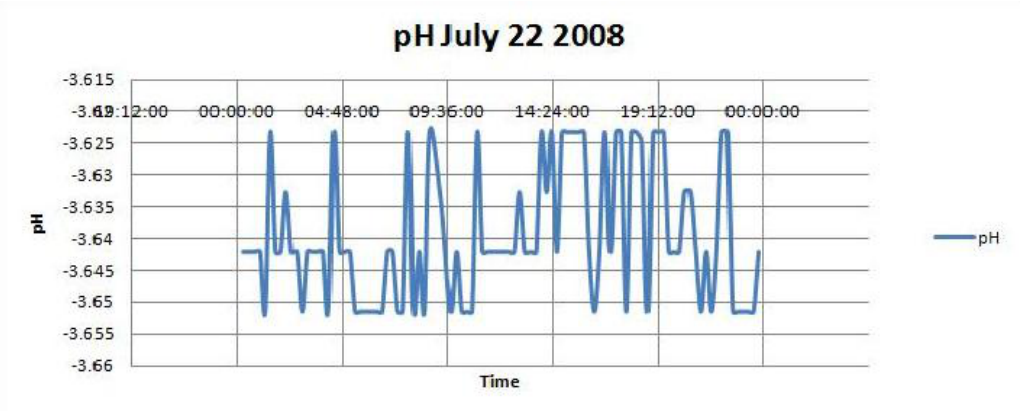

Figure 3.9: Water depth from the SmartCoast depth sensor on July 22, 2008

collected from the water depth sensor on July 3 2008. In this graph it can be seen that the sensor is working correctly until approximately 4:30 - 5:00 am after which it appears to start giving incorrect readings. It then seems to start performing correctly again approximately after 8:30 am. However the sensor starts to produce erroneous negative values again from approximately 3pm onwards for the rest of the day. Figure 3.8 shows data from the SmartCoast depth sensor for July 22nd, 2008. The sensor is reporting depth values of approximately 30 feet and sudden drops in water level. Similarly other sensors also produced erroneous values for periods. For example, Figure 3.9 shows data from the SmartCoast pH sensor for July 22nd that are completely off-scale.

While it is not ideal, this behavior is typical of occasional sensor data errors which are outside our control but with which we have to deal with. Indeed this provides part of our motivation for investigating multimodal sensing in this thesis. Intermittently erroneous data is a fact of life for real word sensing. Some of these sensors had been deployed since 2005 and were perhaps coming to the end of their life or perhaps maintenance was not as rigourous as needed. As previously outlined the project ceased uploading data to the SmartCoast web site in January 2009 and the sensors were taken out of the water. The DEPLOY project then followed on from SmartCoast with deployments at five points along the river, with one of these being the Lee Maltings site.

Even though this was a new project with new state of the art technologies, some of the difficulties with deploying and maintaining an in-situ sensor network in a marine environment remained, as will usually be the case in such situations. In (Lawlor, 2010), it is outlined how there were periods where data was not transmitted to the web server. This was as a result of sensors being taken offline for maintenance but also due to issues such as power failure or as a result of technical issues with the sensors. Lawlor (2010) outlines how the frequency of maintenance was generally governed by the fouling rate of the sensors which depends on location, sensor type, the hydrologic environment and the season. However, maintenance visits were reduced towards the end of the project and this became apparent in the data.

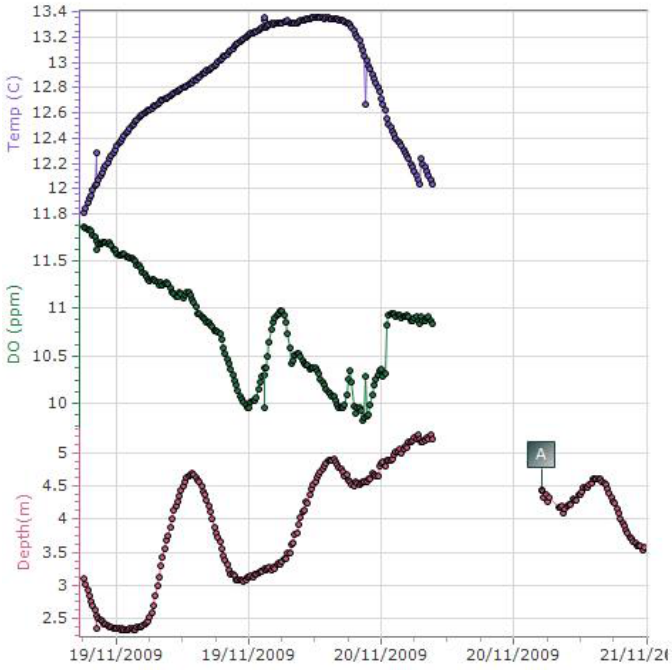

Figure 3.10: Data from the Lee Maltings site of DEPLOY during a period of flooding Source: DEPLOY

Another issue is problems with the sensor technologies themselves. For example on the Lee Road site, a pressure transducer was installed for monitoring depth which was later removed from the site (DEPLOY, 2010). Also in November 2009, there was a flooding event at the Lee Maltings site where the river Lee burst its banks. This resulted in the depth sensor along with other in-situ sensors going offline demonstrating that in-situ sensor nodes often cannot withstand extreme events.

Figure 3.10 shows data from the temperature, conductivity and depth sensors during this time period (DEPLOY, 2010). The depth can be seen to continually rise during this period of increasing depth. It subsequently goes offline. By the time the sensor is back online again water levels have subsided.

All of these issues outlined lead to the conclusion that multiple sensing modalities are desirable in such a marine environment. For example during periods of extreme events such as flooding, a camera overlooking the water will continually collect data whereas as we saw with the DEPLOY project, in-situ sensors may break down. Thus it provides a back-up sensing modality in the network. It can also be used to dispute or validate in-situ readings from a sensor network whereby it is unclear whether events are 'real' or due to errors with the in-situ sensor node. It is also apparent that for continuous datasets redundancy in the network is required as it is not only during extreme events that sensors go offline or produce erroneous data. Frequently this may also occur due to technical issues, power failure or the fact that the site is due a maintenance visit and sensors need to be cleaned etc. However it is often not within the budget of projects to deploy a number of sophisticated sensors each monitoring the same parameter. Hence we need to investigate other manners in which redundancy can be provided in the network through sophisticated modelling techniques and the use of multiple sensing modalities for more efficient and effective sensing.

3.3.4 Visual Data Source - Camera Network

An AXIS PTZ Network camera was deployed overlooking the banks of the river Lee at the Tyndall Research Institute, Cork, Ireland, as described in (O'Connor et al., 2008). It is controlled remotely from a desktop PC at Dublin City University (DCU). A software application was developed within the CLARITY research centre to interface with the camera, since these camera models had been used for previous research projects. Every minute this application moves the camera to four different positions and saving images at pan-right-zoom ($ca{-}wall$), pan-left-zoom ($ca{-}trees$),

Figure 3.11: The angle of the images captured by the camera - labelled as follows - $ca - trees$, $ca - wall$, $ca - sky$, $ca - centre$

pan-up-zoom ($ca - sky$) and finally a full zoom-in ($ca - wall$) on the water. These are shown in Figure 4.3. The images from the camera are stored on a server at DCU for further analysis. Initially the application interfaced with the camera at ten minute intervals but it soon became apparent that a greater sampling frequency was required. This was due to the fact that there was often great changes between the images and missed events with the lower sampling rate. For example, because the site is tidal there could be great changes in depth in a ten minute interval.

The camera was fully deployed and linked up to the network at the Tyndall Research Institute from May 14 2008. Due to initial problems with camera positioning and camera stability, usable data is available for analysis from July 2008. Figure 3.12 presents examples of some of the challenging data taken from the same camera position that we are using demonstrating disparate appearance due to varying river and lighting conditions. Nonetheless, as we explore later, camera data can be a good complement to other in-situ sensors.

Figure 3.12: Examples of the challenging image data we are using, demonstrating disparate appearance due to varying river conditions.

3.3.5 Contextual Data - Rainfall Radar

The Irish meteorological service provides an image sampled every 15 minutes displaying rainfall radar data on their web site[5]. These images display rainfall intensity on a scale from 0 (no rain) to 5 (heavy rain), each pixel area being colour-coded and overlaid on a map of Ireland. Within our research group we continually retrieve and store these images for analysis. Processing these images determines clouds of rainfall along with their location, intensity, speed and direction. This rainfall radar data provides information for us on the rainfall within the catchment areas of our rivers and coastal regions by combining rainfall radar with hydrological maps of the areas. This provides contextual information about likely rises in water levels within the rivers. Rainfall radar data has been gathered by us continuously dating back to

[5] $http://www.met.ie/latest/rainfall_radar.asp$ accessed on May 24 2011

May 2007.

Despite the benefits and potential applications of this data source, there are also some issues with its use. On the web page displaying the rainfall radar image, Met Éireann notes how at longer ranges (generally over 100km), that the accuracy reduces due to the curvature of the earth. The radar may not detect rain at long ranges or report rain at high altitudes that never reaches the ground. Sometimes echoes returned from the ground can also appear on images. The images produced by Met Éireann are a composite of data from two radars - one at the east of the country at Dublin airport and the other in the south-west of the country at Shannon airport. Using a JPEG image as the data source also means that we are not using specific rainfall radar values but rather a selection of 5 classes of rainfall types. The resolution of each pixel is 1.5 km^2, which can also mean a low spatial resolution when trying to estimate rainfall for a particular site or catchment. Further details on this data source and its use in the River Lee catchment are provided in Chapter 6.

3.4 Galway Bay – Site Overview

Ireland's ocean territory constitutes approximately ten times its land mass with approximately $220,000,000$ acres of underwater territory granted by the United Nations Convention on the Law of the Sea. It is generally considered to be one of Ireland's greatest natural resources with resources such as oil and gas, aquaculture, marine transport and shipping, coastal tourism, wind and wave renewable energy etc. all being present (Smeaton, 2010). In 1999 the Irish National Seabed Survey (INSS) was launched which represented one of the largest marine mapping programmes ever undertaken anywhere in the world (Geological Survey of Ireland, 2011b). The successor to the INSS programme was the INtegrated Mapping FOr the Sustainable Development of Ireland's Marine Resource (INFOMAR) also funded by the Irish

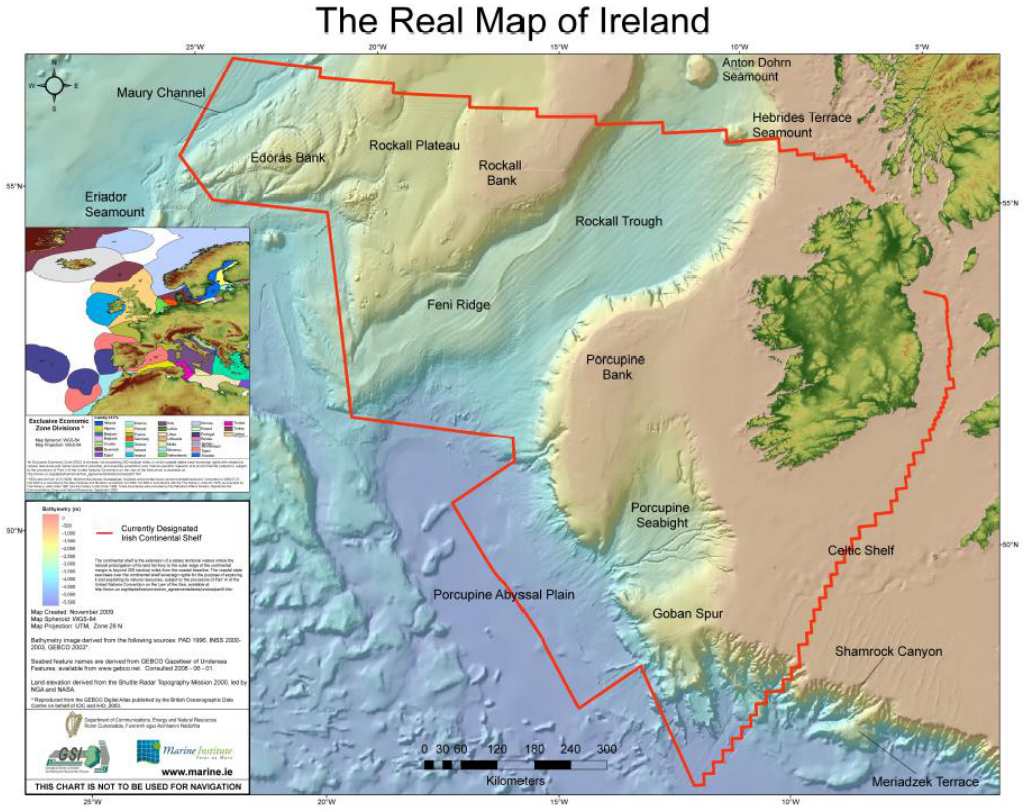

Figure 3.13: The Real Map of Ireland Source: The Marine Institute

Government through the National Development Plan, 2007-2013[6] . This mapping has resulted in what's often commonly referred to as 'The Real Map of Ireland' (shown in Figure 3.13). The main focus of the INSS survey was deep water mapping, whereas INFOMAR concentrates on nearshore surveys (Geological Survey of Ireland, 2011a). During an extensive stakeholder exercise conducted between 2002 and 2005, 26 priority bays and three priority areas were delineated for mapping in the first 10 years of INFOMAR (INFOMAR, 2011b), with one of these bays being Galway Bay.

Galway Bay is located on the west coast of Ireland. The south of the bay is bordered by Co. Clare and it is bordered by Co. Galway to the north. According to (INFOMAR, 2011a) Galway Bay is 62 km long from the Brannock Islands (situated just north west of the Aran islands) in the west to Oranmore in the east. The mouth of the bay is 22km wide from Doolin to Lettermullan, however it narrows at Black Head to 10km. An image of Galway Bay in the context of the British

[6] $http://www.infomar.ie$ accessed on May 24 2011

Figure 3.14: Galway Bay in the context of the British Isles. Source: Bing Maps

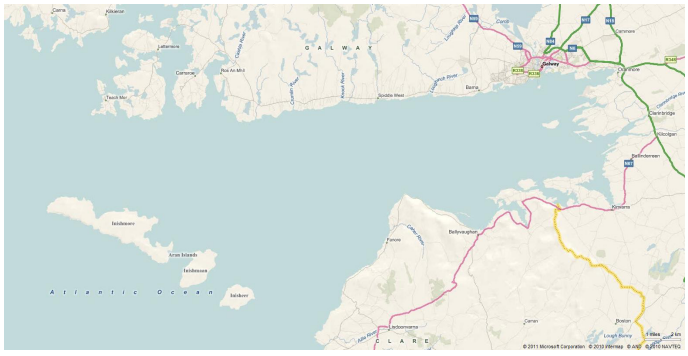

Figure 3.15: Galway Bay, close up. Source: Bing Maps

Isles can be seen in Figure 3.14 and a more close-up view can be seen in Figure 3.15. According to (INFOMAR, 2011a), the main rivers entering the bay are the River Corrib at Galway and the Owenbolisly River at an Spidéal. Along the south coast of Galway Bay there are many minor freshwater inputs, as well as submarine freshwater sources due to drainage patterns associated with the Burren, a karst-landscape region in northwest Co. Clare.

Galway Bay is quite an important resource and supports a full range of mar-

itime activities including environmental monitoring, shipping, inshore and offshore fishing, marine leisure and tourism, aquaculture, and research and development in relation to aquaculture and also research and development in relation to ocean energy. Many research institutes and organisations use Galway Bay as the basis for research programmes and projects, most notably the Irish Marine Institute[7] located in Oranmore, Co. Galway. This is the national agency responsible for Marine Research, Technology Development and Innovation (RTDI). The Marine Institute is the lead implementation agency in Sea Change - Ireland's marine research and innovation strategy (Marine Institute, 2006a,b). This was adopted by the Irish Government in February 2007 and aims to drive the development of marine resources in Ireland in a way that contributes to the knowledge economy. It was developed as an integral part of the Irish government's *Strategy for Science, Technology and Innovation*.

The operational programmes of the Marine Institute have included integrated marine exploration, deep sea research, marine climate change, oceanography, phytoplankton monitoring, shellfish microbiology, ocean energy and many more. Galway Bay provides a testbed for many programmes most notably for the Wave Energy Test Facilities and SmartBay which is a national test and demonstration platform facilitating the development of innovative approaches to distributed sensing, communication and data management. SmartBay is a flagship research project for Galway Bay arising from *Sea Change* and it is also a specific deliverable on the National Strategy for Science, and Innovation (Marine Institute, 2010). All this background combined deem Galway Bay a suitable test site for carrying out our own research since it provides a variety of options with regards to in-situ data sources, it represents an interesting site from a research perspective, and it has been delineated as a test bed for a number of innovative new technologies. It is data from the SmartBay buoys that is used for the experiments conducted in our work. Hence the following provides an overview of SmartBay and the available sensor data.

[7]http://www.marine.ie

3.5 Galway Bay Data Sources

3.5.1 SmartBay

Figure 3.16: SmartBay, Galway. Source: Marine Institute

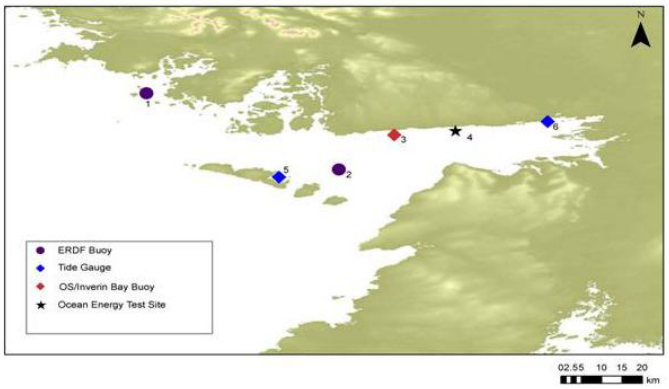

Figure 3.17: SmartBay Pilot Project buoy locations. Source: Marine Institute

The Marine Technology Programme is part of *Sea Change* which as previously outlined is Ireland's marine research and innovation strategy (Marine Institute, 2006a,b), and SmartBay is a significant requirement in support of this programme. SmartBay is a test and demonstration platform for development of innovative products and services for the global maritime sector. On completion, the SmartBay system will consist of a network of seafloor cables, buoys, and infrastructure supporting a range of sensors, information systems, telemetry and other communication technologies, providing the basis for in-situ, real-time monitoring (Marine Institute, 2008), as shown in Figure 3.16.

In 2007, the SmartBay pilot infrastructure was established by the Marine institute (MI) and the Environmental Protection Agency (EPA). This pilot infrastructure consists of a suite of commercially available technology platforms including sensor hardware, communication systems and a network of buoys against which new technologies can be validated. It can be accessed by industrial and academic research groups for test and demonstration purposes.

In Summer 2008, The Marine Institute deployed two SmartBay buoys at Mid Bay, east of the Aran islands, and at an outer bay location – Mace Head, near Carna in Connemara. These locations can be seen in the map shown in Figure 3.17. Each buoy supports an array of advanced sensors for monitoring various oceanic parameters and transmitting real-time information on current conditions.

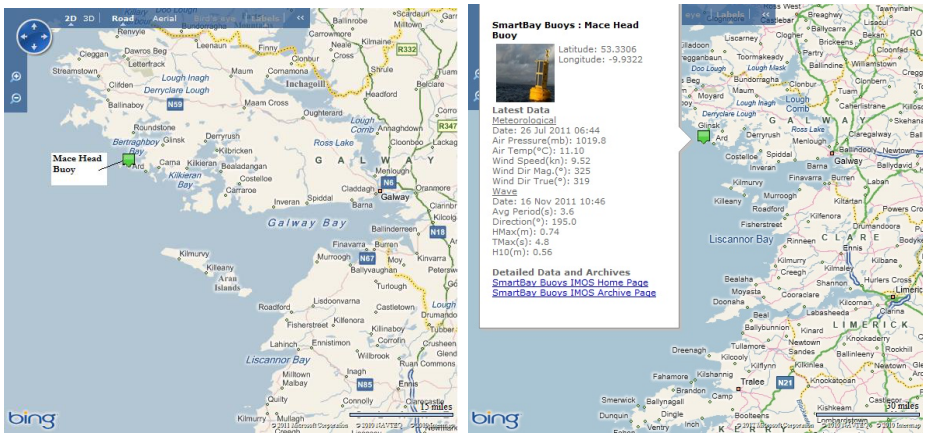

Figure 3.18: Mace Head Buoy at Galway Bay and sample data. Source: Marine Institute

These SmartBay buoys and existing sources of information provide an integrated network of information within Galway Bay. The buoys were equipped with an array of sensors including water quality monitors, wave monitors, weather stations and devices for measuring current. A request made to the Marine Institute in September 2010 for water quality data from the SmartBay buoys returned data for the following water quality parameters for both the Mace Head and the Mid-Bay buoys: chlorophyll, dissolved oxygen, turbidity, pressure, salinity, and temperature. From the SmartBay portal page[8] it is only the Mace Head buoy that appears to be

[8]http://www.marine.ie/home/publicationsdata/data/IMOS/erdfbuoys.htm, accessed Jan 8,

deployed as shown in Figure 3.18. However the last available data appears to date back to July 26 2011.

In addition to the SmartBay sensors, other marine monitoring platforms have also been deployed in Galway Bay providing real-time information on various weather and sea conditions. The Integrated Marine Observations page provides real time data from the National Tide Gauge network and the Irish Marine Weather Buoy Network along with data from other sensors and instruments managed by the Marine Institute [9]. Ireland has one of the richest wave energy climates in the world and the Irish Marine Institute is also promoting research into wave and tide energy. In association with the Sustainable Energy Authority Ireland (SEAI), the Marine Institute established the Galway Bay Wave Energy Test site [10] for testing scaled prototypes of wave energy devices in Galway Bay.

3.5.2 Issues with In-Situ Data Sources – Galway Bay

Similar to the issues outlined in Section 3.3.3 on in-situ sensors on the River Lee, there are gaps in the data from some of the in-situ data sources in Galway Bay which are persistent but irregular in their occurrence as all sensing resources may not be continuously in operation. There also may in an issue with data quality especially in the water quality sensors of the SmartBay pilot project. This was noted by Marine Institute personnel, who noted issues such as significant biofouling between maintenance visits. Often budgetary requirements reduces the number of maintenance operations that can be carried out on the buoys which results in a reduction in data quality. This is because unlike sensors in the River Lee, the Galway Bay buoys are remote and inaccessible and require professional seafarers even to travel to their locations. Again this re-enforces the need for innovative solutions for more efficient and effective sensing and a smarter environmental sensor network, along the lines of the multi-modal approach we are taking.

2012

[9] http://www.marine.ie/home/publicationsdata/data/IMOS/, accessed May 23, 2011

[10] http://www.marine.ie/home/aboutus/organisationstaff/researchfacilities/Ocean+Energy+Test+Site.htm

3.5.3 Galway Bay Visual Data

For the purposes of our research into multi-modal sensor networks for environmental monitoring and to include as diverse a range of sensor sources as possible, the visual data network in Galway Bay consists of satellite imagery. Remote sensing from satellite and airborne sensors has proved to be a tremendous tool for studying our environment at large temporal and spatial scales. It offers unique large scale synoptic data to capture the range and variability of many complex processes. Sophisticated satellite sensors are very effective for monitoring many parameters such as sea surface temperature (SST), sea surface height, ocean currents, turbidity, and chlorophyll pigment concentration (which can subsequently be used to determine the amount of algal growth in the water) along with other water quality parameters such as mineral suspended sediments and yellow substance. A number of such sensors are orbiting the earth on various satellite platforms with differing spatial and temporal resolutions.

Unfortunately the resolution of this data is not particularly suitable for monitoring a relatively small point on the river Lee such as Lee Maltings, hence this type of information was not incorporated into our network for the river Lee study. However, for a larger, open water body such as Galway Bay, satellite imagery offers a very suitable sensing modality for our study. In Chapter 5 we discuss the issues in relation to satellite remote sensing, the products we have chosen for use in our research and the reasoning behind the choice of these products with the specific application context in mind i.e. reliable daily data to enhance the use of an in-situ sensor network. We see that the difficulties with remote sensing data mean that reliable daily satellite information for Galway Bay is difficult to achieve but it is however still desirable for the purposes of our research which concentrates on multi-modal sensor networks, of which satellite images are just one modality.

3.6 Conclusion

In this chapter we have presented site overviews of two diverse locations each with their own challenges. The first is the river Lee, a freshwater river at its tidal point where it enters the sea. This site presents a real challenge in environmental monitoring because of the effects of the tide, a nearby hydro-electric power dam and the fact it is a location susceptible to regular flooding. We also outlined the catchment area of the River Lee, some parts of which has more runoff potential than others. This point is quite important when it comes to analysing the outcome of our experiments in Chapter 6, where we provide a novel analysis of a catchment's response to rainfall using rainfall radar images and in-situ depth data.

We described in-situ data sources available on the River Lee through the DEPLOY project. We focus specifically on the Lee Maltings site from an environmental and practical perspective. It represents a very dynamic and difficult site to monitor, and it also provides a suitable infrastructure for the deployment of a camera. Also it is the only site of the five DEPLOY sites along the river that has a depth sensor as part of its range of in-situ sensor nodes.

The second site we use is Galway Bay on the west coast of Ireland. This bay faces the Atlantic Ocean along a rugged coastline and the Bay hosts a diverse range of commercial and leisure interests, from fishing to sailing. Galway Bay represents an interesting site from a research perspective and it has been delineated as a test bed for a number of innovative new technologies. It is data from the SmartBay buoys that is used conducting the experiments in this thesis.

For each of the sites we have a range of sensors, quite diverse in nature and sometimes not all operational. In this chapter we have highlighted some of the issues which cause periodic faults with sensor data and having to work with such data provides us with a very real testbed on which we can explore multi-modal sensing. These sites and issues lead to very exciting prospects and novel applications for multi-modal sensor networks and how they can be optimised. In the next chapter

we begin examining the use of these multi-modal data sources by investigating the use of a visual sensor in a river environment. This is followed in the next chapter by an examination of the use of remote sensing data in the context of a multi-modal sensor network in a coastal location. Subsequent chapters then present novel studies with very promising results for integrating these diverse data sources and optimising their use to the network for more efficient and effective monitoring of such diverse and challenging environments.

Chapter 4

Extraction of Information from Camera Images

This chapter demonstrates the use of a visual sensor (camera) in a river environment and how it can be used as a complimentary sensing modality to an in-situ environmental sensor network. The focus of this particular study is on the estimation of depth from the camera images at the Lee Maltings since this particular parameter can be tied in with the in-situ sensor network. Depth is also quite an influential feature at the site and affects the values of other water quality parameters. The focus of this chapter is on the development of models to detect various features in the image which through their detection can provide an estimation of depth. In Chapter 7 we move on to investigating the use of these models in the context of a multi-modal network whereby we report methods of relating these features back to the in-situ depth data. Subsequently in Chapter 8, we evaluate the performance of this sensing modality in the detection of high and low depth events. The analysis presented here relates back to research question 1 of the research questions presented in Chapter 1.

Image sensing or what has been termed as 'near-surface' remote sensing by some authors e.g. (Richardson et al., 2009), provides unique perspectives when used in environmental monitoring applications. It can provide high frequency data at a low

cost and use different view angles for increased spatial resolution. They have been used in a variety of contexts in environmental monitoring applications and in the following section we provide an overview of some of the main application scenarios.

4.1 The Use of Cameras in Environmental Monitoring applications

Coastal video systems have been identified as effective tools for coastal monitoring. A prime example of this is a major European research project entitled CoastView (Davidson et al., 2007). This focused on the development of video systems in support of coastal zone management utilizing Argus technology. Argus stations consist of optical systems developed for nearshore sampling (Holman and Stanley, 2007). The CoastView project demonstrates the use of fixed video remote sensing systems to partially ameliorate some of the problems associated with in-situ measurements of waves, currents, and morphological change. Davidson et al. (2007) refers to some of the research carried out investigating algorithms for the quantitative extraction of geophysical signals from image data including morphology (Aarninkhof et al., 2005), flows (Chickadel et al., 2003) and wave parameters (Holman and Chickadel, 2004) and refers to the various scientific literature that has tested and reviewed the reliability, accuracy and versatility of coastal video systems.

Goddijn-Murphy et al. (2009) explore the possibilities of employing a conventional digital camera, as an alternative low-cost technique to satellite imagers or multi-spectral radiometers, to estimate water composition from optical properties of the water surface. This paper presents the method that was used to acquire digital images, derive RGB values and relate measurements to water quality parameters. Measurements were recorded in Galway Bay and in the North Atlantic. Both yellow substance and chlorophyll concentrations were successfully assessed using this method. Iwahashi et al. (2006) investigates detecting water level using a land water boundary. However their work is using a a video signal as opposed to still images.

Also they aim to classify pixels as land or water, where they are assuming that the land region contains solid objects with fine texture full of high frequency components. This is not always the case with the land region in the images in this study, which under certain conditions can partially appear visually similar to water.

Other studies have investigated the use of cameras not only in the context of monitoring a marine environment but also in other forms of environmental monitoring applications. In (Ahmadian et al., 2007), a scalable end-to-end system for vision-based monitoring of avian behaviour during a nesting cycle is presented. The manual collection of phenological data can prove to be labour intensive and thus requires the use of innovative new methodologies such as the use of digital cameras. Graham et al. (2009) investigate the use of visible light digital cameras in determining the dynamics of expanding leaf area for Rhododendron occidentale, a deciduous understory shrub. Richardson et al. (2007) explored whether digital webcam images could be used to monitor spring green-up in a deciduous northern hardwood forest. They concluded that webcams offer an inexpensive means by which phenological changes can be quantified. In (Richardson et al., 2009), they build on previous work to demonstrate the use of networked digital cameras to document spatial and temporal variation in the phenology of forest canopies. Bradley. et al. (2010) outlines the benefits of a multi-modal approach and outlines their development of a web application which incorporates digital camera data and satellite data. They present two cases studies demonstrating its use for looking at cloud duration and height and plant phenology time series.

The work described in the following sections follows on from previous work we carried out which is described in (O'Connor et al., 2009), which also focuses on estimating river water level using visual sensing. The approach described in this chapter again uses the land water boundary in order to determine water level. However it adopts a more sophisticated classification approach and a different camera angle that contains more distinct features. It also takes into account difficulties that were encountered in the work reported in (O'Connor et al., 2009) where it was difficult

to model pixels as water or land due to the varying lighting conditions. It attempts to overcome these difficulties and to take a more adaptable approach that can easily be applied to a variety of images from different camera angles.

4.2 Visual Sensing System

As previously outlined an AXIS PTZ Network camera was deployed at the Tyndall Institute overlooking the banks of the River Lee at the Lee Maltings. Data is available from July 2008. Each minute the camera automatically pans to four different positions in order to save images from the camera at four different angles. A visual based sensing system was subsequently developed with the objective of complementing and enhancing the use of the in-situ sensor network at the site.

4.2.1 In-Situ and Visual Sensing Parameters

The in-situ parameters measured at the Lee Maltings site include conductivity, chlorophyll, temperature, dissolved oxygen and depth. The images from the camera were then analysed in order to highlight image features that could be used to complement the information being retrieved from the in-situ sensor network. Analysing outdoor data is challenging due to the wide-range of environmental conditions and their rapid changes. Varying river and weather conditions, extreme changes on lighting and reflections on the water are representative of some of the challenges presented. Examples of the challenging image data were previously shown on page 59.

Previous studies where cameras were used to estimate water colour were carried out using much more specialised circumstances and not in the case where a camera is just placed in a building overlooking the water e.g. (Goddijn-Murphy et al., 2009). However the purpose of this work is to examine how we can use a relatively low cost off-the-shelf webcam type device for complementing in-situ sensor networks.

Depth can give us an indication of a variety of conditions at the Lee Maltings

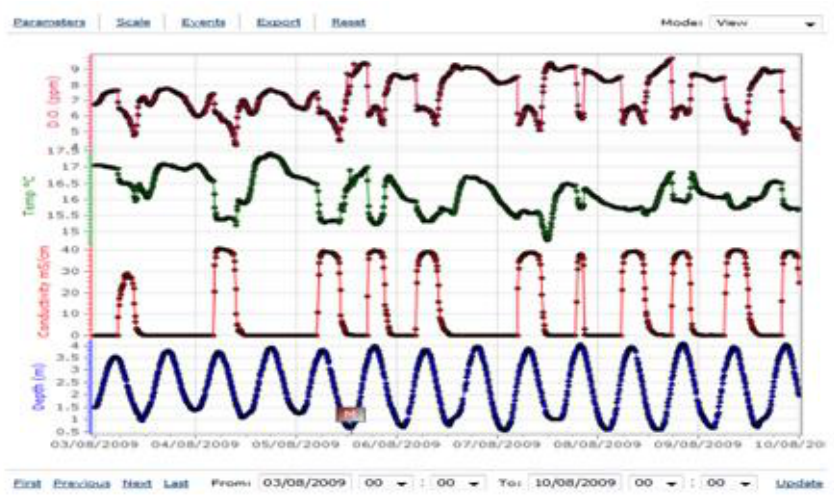

Figure 4.1: The relationship between depth and various in-situ parameters at the Lee Maltings site. The conductivity data also illustrates dilution in the River Lee due to dam releases from the Iniscarra reservoir .

site such as temperature, dissolved oxygen and conductivity readings. Figure 4.1 demonstrates the influence of depth on a variety of the in-situ sensor readings. While it is difficult to quantify this relationship due to the influence of site characteristics, in Chapter 6, we attempt to model the relationship between these parameters in order to predict the values of missing parameters if one of the in-situ nodes fails. The conductivity data also illustrates dilution in the River Lee due to dam releases from the Iniscarra reservoir. Therefore the estimation of depth from the camera images is a really important indicator of conditions at the site.

Our visual sensing system also undertook the detection of other image features such as objects floating on the water, boats, water turbulence etc. However it is really only the extraction of depth that can be linked up with the in-situ sensor readings which is the prime focus of this research. The detection of objects such as boats or floating objects is also an example of how a visual sensor can complement and enhance the use of an in-situ sensor network, as these are objects that cannot be detected or immediately detected by an in-situ sensor network. Figure 4.2 demonstrates images where a boat and some material can be seen floating on the top of the water.

Figure 4.2: A boat on the river and scum floating on the top of the water. Our visual sensing system can detect objects such as these floating on the top of the water

However the focus of this chapter is on estimating water depth using a visual sensor at the Lee Maltings site. As well as providing an indication of present conditions at the site, continuous monitoring of water level is important for flood warnings and also for navigational and recreational safety.

4.3 Estimation of Water Depth using a Visual Sensor

Changing depth is a feature of almost any inland waterway. If the water-land boundary is visible, visual imaging is a practical means for determining the water-level. As previously outlined the camera deployed overlooking the river at the Lee Maltings site, pans to four different angles every minute - pan-right-zoom ($ca - wall$),

Figure 4.3: The angle of the images captured by the camera - labelled as follows - trees, wall, sky, centre

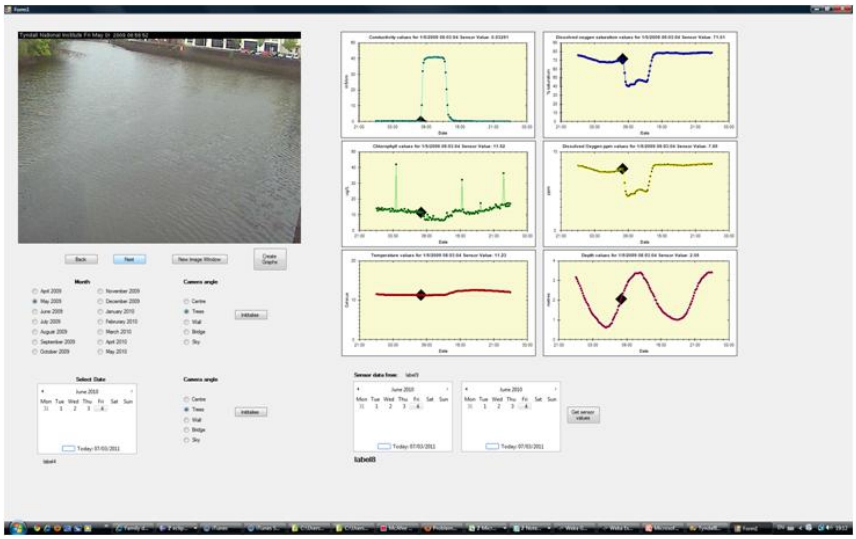

Figure 4.4: Visual sensor analysis tool - enables the analysis of visual data alongside in-situ sensor readings in order to examine features and relationships between features and in-situ sensor data.

pan-left-zoom ($ca-trees$), pan-up-zoom ($ca-sky$) and finally a full zoom-in on the water ($ca-centre$). These four camera angles can be seen in Figure 4.3

In order to analyse the relationship between the sensor readings and features in the images, we developed a software tool to specifically enable the visualisation of the images and the nearest in-situ sensor reading that corresponds to the time that

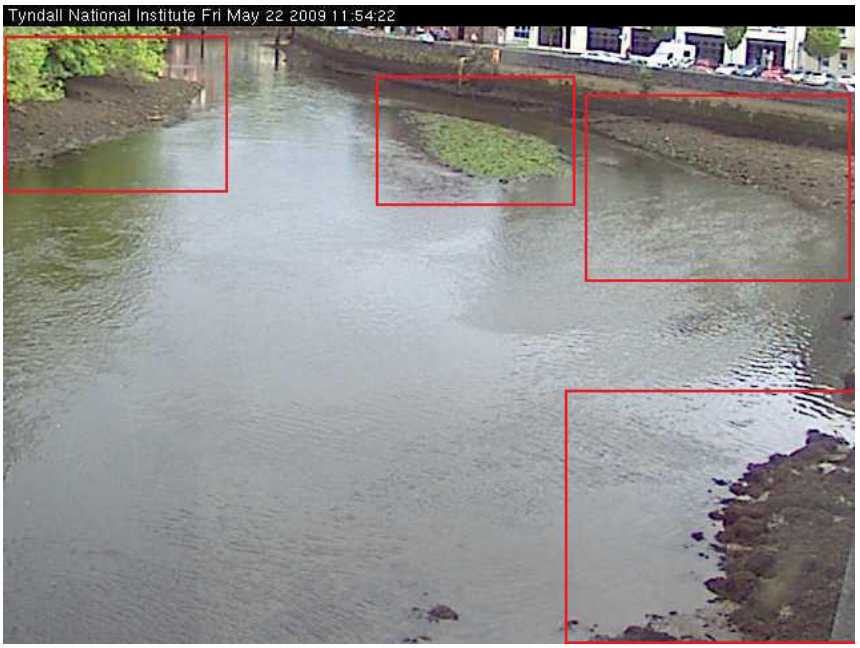

Figure 4.5: The features highlighted in the image become visible in order with changing depth.

image was captured (See Figure 4.4). From analysis of the images from $ca - wall$ along with the in-situ depth readings, it is apparent that certain features in the images become visible in a certain order as water levels are decreasing and can thus provide an estimation of water level at the site. This can be clearly seen in Figure 4.5.

As the depth of the water decreases, the first feature to becomes visible is the appearance of rocks beneath the trees in the far left of the image (*feature 1 - rocks at trees*). The second feature to become visible is the appearance of rocks in the far right of the image (*feature 2 - rocks at far wall*). The third feature to become visible is the appearance of rocks in the near right of the image (*feature 3 - rocks at near wall*) and finally the final feature to become apparent that indicates depth is the appearance of a small island in the middle of the water (*feature 4 - island*). Thus if each of these features can be accurately detected, then this can provide a very good indication of water levels at the site. Each of these features are used to delineate a certain type of water level e.g. the appearance of *feature 1* denotes water level 1, the subsequent appearance of *feature 2*, denotes water level 2, and so forth.

As previously outlined in Chapter 7 we present methods for relating these features back to the in-situ depth data. Then in Chapter 8 we evaluate the performance of these models in detecting a series of depth events.

Previous work demonstrated the difficulty with accurately detecting the boundary of such a feature due to colour changes in the image (O'Connor et al., 2009). Thus an approach which allows the detection of such a feature or a series of features appears to be more robust to these conditions. The following section outlines the approach and results for the detection of each of these features.

4.4 Methodology - Detection of Depth Features

The detection of these features in the image is far from trivial due to the huge lighting changes of the water. Sometimes the water can appear almost black mid-day due to reflections on the water and the appearance of rocks at each of the individual features is unclear even to the human eye. This therefore renders it difficult for an image processing algorithm to accurately detect the appearance of each of the features.

4.4.1 Data

As previously outlined an image is captured approximately every minute by the visual sensor network. This leads to over $25,000$ per week being captured for one camera angle alone if operations are running correctly. Thus for 4 camera angles, there are over $100,000$ images captured per week. Future work may involve methods to deal with this data volume. For example following our determination of the relevant features required to capture events in the image, it may be the case that only this information needs to be stored. Images are currently stored as JPEGs. For the development of the models, images from May 15-27 2009 were chosen due to the clear display of a number of events associated with changes in depth. This data was manually annotated four times over in order to have a set of ground-truth

images for each of the features outlined above.

For each feature, the images were annotated as follows: 1 - feature present, 2 - no feature present, 3 - intermediate. The third annotation is used where the annotation of the image is slightly ambiguous. The feature is at an intermediate stage of appearance whereby it is beginning to appear but not as apparently as when annotated as 'present'. This is due to intermediate changes in the water level. These images were given their own class so that it could be decided when carrying out classification whether it is better to have three-way classification for these features or two-way classification whereby images where the feature is annotated as being intermediate are just classified simply as 'present' or 'not present'.

4.4.2 Classes

Two types of classification are examined in this study - 2-way classification (positive, negative) and 3-way classification (positive, negative, intermediate). For two-way classification, we build a model using only images annotated as positive or negative. These are the extremes we are mainly interested in and it is important that the classifier learns these extremes. It is not really of due concern whether images where the feature is is at an intermediate stage of appearance are annotated as positive or negative. For the 3-way classifier the model is built using sample images across the positive, negative and intermediate ground-truth datasets.

4.4.3 Image Analysis - Feature Sets

The Matlab image processing toolbox (Version R2009A) was used for processing images and extracting relevant image features. For each of the depth features to be detected (i.e. depth features 1-4) a set of image features were extracted from the images at the relevant region of interest using the four sets of ground truth data. A variety of features were extracted including colour features such as *average hue, hue histogram, average saturation, saturation histogram, average value, value histogram*,

texture features such as *average entropy, entropy histogram*, edge features such as *edge histogram, number of pixels marked as an edge in feature area, percentage of pixels marks as an edge in mask area, number of pixels marked as an edge in mask area after a fill operation, percentage of pixels marked as an edge in mask area after a fill operation* and other features such as *average brightness, brightness histogram, average luminance, luminance histogram*. Various feature sets were examined for the detection of each depth feature.

For 2-way classification, feature sets were extracted for 2000 (1000 positive instances, 1000 negative instances) images across the ground truth datasets available for each of the depth features from May 17-27 2009, apart from *feature 4 - island*. Due to the limited number of instances occurring in the ground truth dataset, feature sets were then extracted for 1300 (650 positive instances, 650 negative instances) images across the ground truth dataset. Thus when all the depth features are considered a total of 7300 images are analysed in this part of the study.

For 3-way classification, feature sets were then extracted for 2340 (780 positive instances, 780 negative instances, 780 intermediate instances) images sampled across the ground truth datasets available for each of the depth features from May 17-27 2009, apart from *feature 4 - island*. Again, due to the limited number of instances occurring in the ground truth dataset, feature sets were then extracted for 750 (250 positive instances, 250 negative instances, 250 intermediate instances) images sampled across the ground truth dataset. This leads to a total of 7770 images used for this part of the analysis.

These feature sets were then input into a Support Vector Machine (SVM) classifier which is detailed more below. After initial testing the feature sets outlined in Table 4.1 were the most successful for detecting each of the depth features outlined above. The features in the first row of the table are the features used in all the features sets, with the features in the following rows outlining those that are specific to a particular feature set. Following initial evaluation, there were no great differences between feature sets 1-5. However feature set 5 seemed to perform marginally better

All Feature Sets	average hue, hue histogram, average saturation, saturation histogram, average value, value histogram, average entropy, entropy histogram
Feature Set 1	+ edge histogram
Feature Set 2	+ edge histogram, percentage of pixels marked as edge in feature area after a fill operation
Feature Set 3	+ percentage of pixels marked as edge in feature area after a fill operation
Feature Set 4	+ edge histogram, average brightness, brightness histogram
Feature Set 5	+ edge histogram, luminance histogram

Table 4.1: Various feature sets examined in the study. All feature sets contain the features outlined in row 1 in addition to the features outlined in their specific rows.

in many cases, hence this feature set is used in the remainder of this study.

4.4.4 Support Vector Machine (SVM) - Classification

Support Vector Machines (SVMs) work well with large feature sets and after training they are very quick to classify new observations. With the correct parameters, they are known to work as well or better than most classification methods (Segaran, 2007). Initially a simple thresholding approach was attempted for classifying the presence of the various depth features. However this involved manually testing and setting thresholds for image features for each depth feature under investigation. With an SVM, features can be extracted for the location of interest in the image, formatted and input into an SVM for training or classification. This is a much more efficient and successful approach to classification of each of the depth features.

4.4.5 SVM Parameters

LibSVM[1], an integrated software for Support Vector Classification, is used in the Weka[2] data analysis environment for classification of the presence of the various depth features. The official implementation is described in (Chang and Lin., 2011). Normalisation of features is carried out in the Weka environment and classification is carried out using two different kernel parameters - an RBF and linear. Kernel methods enable a linear classifier to be turned into a nonlinear classifier as long as it uses dot-products for comparisons. Since SVMs use dot-products they can be used with kernels to perform nonlinear classification. The 'kernel trick' involves replacing the dot-product function with a different function that transforms the data to a higher dimensional space and returns what the dot-product would have produced if the data had been transformed before it was applied. There are a number of transformations with the radial-basis function often recommended (Segaran, 2007).

The RBF kernel can handle a nonlinear relationship between class labels and attributes, however it also may be the case that if the number of features is large, performance may not be improved from mapping data to a higher dimensional space. Default parameters in the Weka environment were used for both kernel parameters. However future work may involve optimising these. However satisfactory results were found without optimisation, thus this process was not carried out in the context of this study. Ideally optimisation of the (C, γ) space would be carried out for SVMs using an RBF kernel (*SVM-rbf*) and of the C space for SVMs with a linear kernel parameter (*SVM-linear*). The gamma parameter (γ) used by *SVM-rbf* can be adjusted in order to improve the linear separation for a dataset . The cost parameter (C) is the penalty parameter of the error term. In this study a C value of 1 and a γ value of 0 was used for *SVM-rbf* and a C value of 1 is used for *SVM-linear*. 10-fold cross validation is used for evaluation of the model. This is a standard technique used in the machine learning literature and involves dividing the dataset

[1]http://www.csie.ntu.edu.tw/ cjlin/libsvm/
[2]http://www.cs.waikato.ac.nz/ml/weka/

into 10 parts where each part is held out and the learning scheme is trained on the remaining 9 parts and evaluated on the holdout set. The resulting 10 error estimates are averaged to produce an overall error value (Witten et al., 2011). It is found that *SVM-linear* performs better and it is these results that are reported here. However optimisation of parameters for *SVM-rbf* could result in an increase in performance.

4.5 Results - - Evaluation

As previously outlined, certain features in the images become visible in a certain order as water levels are decreasing and can thus provide an estimation of water level at the site. Thus if each of these features can be accurately detected, then this can provide a very good indication of conditions at the site. The following outlines our results in relation to detection of each of these depth features in the images. For each of the depth features, two types of models are developed. We developed a *3-way classification model* using positive, negative and intermediate images and a *2-way classification model* consisting of positive and negative images. The *3-way classification model* was developed in order to examine the ability of a classifier to distinguish between the three classes if needs be.

Figure 4.6 shows an image where none of the features are present. It also demonstrates the challenges in terms of processing these images. Reflections of the buildings on the water can be seen in this image and in general lighting conditions and reflections on the water vary greatly between images depending on weather conditions. In the following sections, for each classification approach the feature set that produced the most accurate results, the accuracy, the F-measure and the area under an ROC[3] Curve are reported. The F-measure represents the harmonic mean

[3]This acronym stands for emphreceiver operating characteristic, it is a term used to characterise the trade off between hit rate and false alarm over a noisy channel in signal detection Witten et al. (2011)

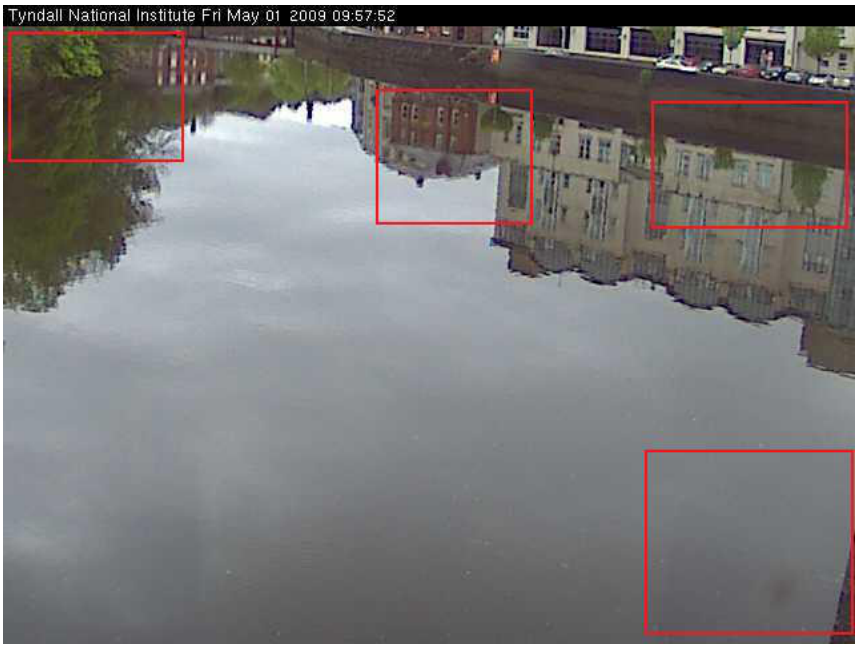

Figure 4.6: Image where none of the depth features are present. It also demonstrates huge image processing challenges with reflections on water and changes in lighting.

of precision and recall and is calculated using the following formula:

$$F = \frac{2 * precision * recall}{precision + recall} \tag{4.1}$$

The area under an ROC curve represents the discrimination ability of the classifier, with a perfect result being 1. The F-measures and ROC areas reported here are the weighted average of the values for each of the individual classes in question for each of the classification scenarios.

4.5.1 Depth Feature 1 - rocks at trees

Figure 4.7 (a) shows a sample image where rocks are appearing under the trees and Figure 4.7 (b) shows an image where rocks are only beginning to appear under the trees (i.e. annotation 3 - intermediate) . From the results shown in Table 4.2, it can be seen the the 2-way classification produces a very high classification accuracy of 93.1%. From the confusion matrix (Witten et al., 2011) outlined in Table 4.3 it can be seen that the model incorrectly classifies 70 positive instances as being negative

Figure 4.7: a) Rocks at trees b) Intermediate rocks at trees

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
2*class*	93.1	0.931	0.069	0.931	0.931
3*class*	77.01	0.77	0.115	0.77	0.828

Table 4.2: Results for each of the classification scenarios for the detection of rocks at the trees

*2*class	rocks	no rocks
rocks	930	70
no rocks	68	932

*3*class	rocks	no rocks	inter. rocks
rocks	646	24	110
no rocks	32	613	135
inter. rocks	90	147	543

Table 4.3: Confusion matrices for the results of the *2class* and *3class* classification for depth feature 1 - rocks at trees

and 68 negative instances as being positive. It produces a true positive and false positive rate of of 0.93 and 0.069 and a large ROC area of 0.931 demonstrating a strong ability to discriminate between the classes. The 3-way classification produces a classification accuracy of 77.01% and an ROC area of 0.828 which is much lower than that achieved by the 2-way classification model. From the confusion matrix outlined in Table 4.3, it is clear that many of the images denoted as having no rocks are being classified as an intermediate detection of rocks and vice versa. This is not surprising since it can be difficult even for the human eye to distinguish between these classes, for various reasons such as lighting.

4.5.2 Depth Feature 2 - rocks at far wall

Figure 4.8: a) Rocks at far wall b) Intermediate rocks at far wall

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
2*class*	95.15	0.952	0.049	0.951	0.952
3*class*	82.56	0.826	0.087	0.828	0.869

Table 4.4: Results for each of the classification scenarios for the detection of rocks at the far wall

*2*class	rocks	no rocks
rocks	957	43
no rocks	54	946

*3*class	rocks	no rocks	inter. rocks
rocks	617	2	161
no rocks	18	661	101
inter. rocks	73	53	654

Table 4.5: Confusion matrices for the results of the *2class* and *3class* classification for depth feature 1 - rocks at the far wall

Figure 4.8 (a) shows a sample image where rocks are appearing at the far wall and Figure 4.8 (b) shows an image where rocks are only beginning to appear at that point in the image. From the results shown in Table 4.4, it is clear that there is improved performance in both classification scenarios compared to the detection of *depth feature 1 - rocks at trees*. From visual analysis of the images, it is clear that the visual distinction of the classes is more apparent than for *depth feature 1*. Therefore these results are consistent with this observation.

The 2-way classification produces a very high classification accuracy of 95.15%, with a true positive rate of 0.952, a false positive rate of 0.049 and a very high ROC

area of 0.952. From the confusion matrix (Table 4.5) it can be seen that the model incorrectly classifies 43 positive instances as being negative and 54 negative instances as being positive. The ROC area demonstrates a strong ability to discriminate between the two classes. The 3-way classification produces a classification accuracy of 82.56% and an ROC area of 0.869 which is quite promising. From the confusion matrix (Table 4.5) it is apparent that many of the incorrectly classified positive and negative instances are being classified as intermediate.

4.5.3 Depth Feature 3 - rocks at near wall

Figure 4.9: a) Rocks ar near wall b) Intermediate rocks ar near wall

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
2*class*	97.75	0.978	0.023	0.977	0.978
3*class*	84.36	0.844	0.078	0.844	0.883

Table 4.6: Results for each of the classification scenarios for the detection of rocks at the near wall

*2*class	rocks	no rocks
rocks	983	17
no rocks	28	972

*3*class	rocks	no rocks	inter. rocks
rocks	723	7	50
no rocks	17	587	176
inter. rocks	39	77	664

Table 4.7: Confusion matrices for the results of the *2class* and *3class* for depth feature 3 - rocks at near wall

Figure 4.9 (a) shows a sample image where rocks are appearing at the far wall and Figure 4.9 (b) shows an image where rocks are only beginning to appear at that

point in the image. From analysis of the images, this feature is more visually more distinguishable than the other two depth features previously considered and this is apparent from the results presented in Table 4.6.

The 2-way classification produces a very high classification accuracy of 97.75%. From the confusion matrix (Table 4.7) it can be seen that the model incorrectly classifies 17 positive instances as being negative and 28 negative instances as being positive. It produces a true positive and false positive rate of of 0.978 and 0.023 with an ROC area of 0.978, which is extremely satisfactory. The 3-way classification produces an accuracy of 84.36% and an ROC area of 0.883. From the confusion matrix (Table 4.7), it is apparent that many of the negative instances are being incorrectly classified as intermediate.

4.5.4 Depth Feature 4 - island

Figure 4.10: a) island feature present b) island feature only intermediately present

Figure 4.10 (a) shows a sample image where an island like feature can be seen in the middle of the water and Figure 4.10 (b) shows an image where this feature id only beginning to appear. The results for the 2-way and 3-way classification can be seen in Table 4.8. The 2-way classification again produces a high accuracy of 96.15% for this depth feature. From the confusion matrix in Table 4.9 it can be seen that 22 positive instances are incorrectly classified as negative and 28 negative instances are incorrectly classified as positive. It produces a large ROC area of

Class	Accuracy	TP Rate	TN Rate	F-measure	ROC Area
2*class*	96.15	0.962	0.038	0.962	0.962
3*class*	82.8	0.828	0.086	0.83	0.871

Table 4.8: Results for each of the classification scenarios for the detection of the island feature

*2*class	rocks	no rocks
rocks	628	22
no rocks	28	622

	island	no island	inter. island
island	213	4	33
no island	5	203	42
inter. island	14	31	205

Table 4.9: Confusion matrices for the results of the *2class* and *3class* for depth feature 4 - island

0.962. Table 4.8 shows that the accuracy of the 3-way classification is quite good at approximately 82.8% and an ROC area of 0.871. From the confusion matrix (Table 4.9) it is apparent that the majority of incorrectly classified positive and negative instances are classified as intermediate.

4.5.5 Relationship between visual and in-situ parameters

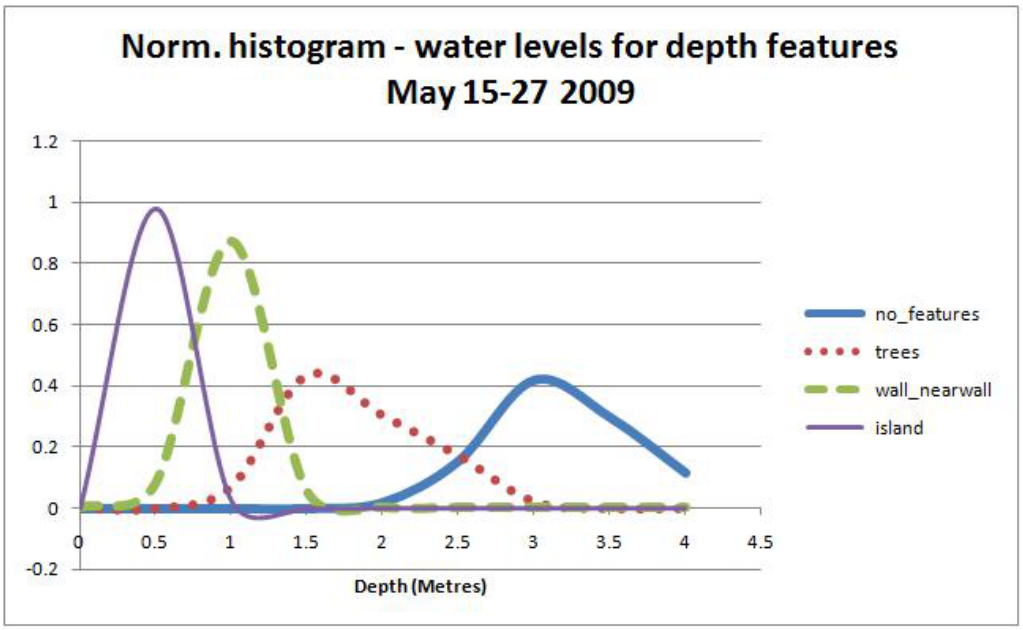

Figure 4.11: Histogram showing the distribution of water levels for the various depth features

It also needs to be examined how information from our visual sensors can be linked to the in-situ depth data. Figure 4.11 shows a histogram demonstrating the relationship between the appearance of the various depth features and water level readings from the DEPLOY water depth sensor. The curve representing *no-features* shows the normalised distribution of water depth values when there are no depth features present in the images from the training set. The curve entitled *trees* shows the normalised distribution of values when *depth feature 1 - rocks at trees* is present or intermediately present **and** none of the other three depth features are present (i.e *depth feature 2 - rocks at far wall*, *rocks at near wall* and *island*). The curve *wall-nearwall* shows the normalised distribution of values when *depth feature 2 - rocks at far wall* **or** *depth feature 3 - rocks at near wall* is present **and** *depth feature 4 - island* is not present. Finally the curve *island* shows the normalised distribution of depth values when *depth feature 4 - island* is present. It is clear from this histogram that there is a clear distinction between the distribution of water depth values output by the depth sensor for the varying appearance of the four depth features.

4.6 Evaluation – Testing

The models developed in Section 4.5 based on data sampled across the period of May 15-27 2009 were tested on data from both similar and alternative times of the year. Firstly the models were evaluated on data from May 1-7 2009. Secondly data was combined from November 8-20 2009, January 8-12 2010, and February 3-7 2010 and tested on the models. The data from November, January and February demonstrated very different lighting conditions and very little samples of the various lower depth features, hence the reason for combining data from a variety of time periods. For the sake of brevity and clarity, this dataset is referred to as *novjanfeb* in subsequent parts of this thesis.

For the 2-way classification, the model for each feature is evaluated using test data containing an equal number of positive and negative instances. In the case of

'intermediate' instances, it is not preferred that the model classify these instances as positive or negative since they are quite ambiguous images at an intermediate stage of an event, however it is preferred that the model be consistent in its classification of these images, allowing the model ultimately to decipher which class these should take on. Thus we evaluate each model using a test set consisting only of intermediate instances under two different scenarios - firstly where we consider all intermediate instances to be positive and secondly where we consider all intermediate instances to be negative and ultimately see how the model performs. For the 3-way classification, the model for each feature is evaluated using test data containing an equal number of positive, negative and intermediate instances.

For the testing period of May 1-7 2009, 400 instances were sampled across the ground truth dataset for each class for each feature. However for *depth feature 4 - island* for testing on the 3-way classification model only 100 instances of each class were used due to a limited number of instances in the 'intermediate' ground truth dataset.

For the testing period using combined data from November, January and February, there were a very limited number of instances of many of the lower depth features, hence only a limited number of instances could be used for testing. For *depth-feature 3 - rocks at near wall* and *depth-feature island* 120 instances of each class were sampled across the ground truth datasets. However there were no instances available of the class intermediate for *depth feature 4 island*. For *depth feature 1 - rocks at trees* and *depth feature 2 - rocks at far wall* 250 instances of each class were sampled from the ground truth dataset.

4.6.1 Depth Feature 1 - rocks at trees

For the 2-class model there is an accuracy of 89.25% with an ROC area of 0.893 for test data from May. For the test data from *novjanfeb*, the accuracy is reduced to 57.6% with an ROC area of 0.576 (Table 4.10). From the confusion matrix for the May test data (Table 4.11) it appears that many of the incorrect classifications are

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
May testing					
2*class*	89.25%	0.893	0.108	0.892	0.893
3*class*	70.58%	0.706	0.147	0.7	0.779
novjanfeb testing					
2*class*	57.6%	0.576	0.424	0.533	0.576
3*class*	39.6%	0.396	0.302	0.34	0.547

Table 4.10: *Depth feature 1 - trees* - Results for the *2-class* and *3-class* models on test data from May and *novjanfeb*

Class	No. output as Pos.	No. output as Neg.	Accuracy
May testing			
positive	385	15	96.25%
negative	71	329	90%
intermediate pos	233	167	58.25%
intermediate neg	233	167	41.75%
novjanfeb testing			
positive	220	30	88%
negative	182	68	27.2%
intermediate pos	206	44	82.4%
intermediate neg	206	44	17.6%

Table 4.11: *Depth feature 1 - trees* - Confusion matrix for the 2-class model on test data from May and *novjanfeb*

for negative instances. It classifies 96.25% of the positive instances correctly and only 90% of the negative instances correctly. However this is apparent from looking at the images where the reflection of the trees and the lighting at that point in the image can often make it visually difficult to differentiate a negative instance from a positive one. For the *novjanfeb* test data, there is an overall accuracy of 57.6%. It classifies 88% of the positive instances correctly and performs quite poorly on the negative instances only classifying 27.2% correctly. In terms of the intermediate instances, in both test sets the classifier performs better when these are considered as positive. However the margin is much greater for the *novjanfeb* test set. For the May test set 58.25% are classified as positive, where as for the *novjanfeb* test set 82.4% are classified as positive (Table 4.11).

Class	No. output as Pos.	No. output as Neg.	No. output as Intermediate	Accuracy
May testing				
positive	373	9	18	93.25%
negative	22	224	154	56%
intermediate	113	37	250	62.5%
novjanfeb testing				
positive	208	30	12	83.2%
negative	124	64	62	25.6%
intermediate	187	38	25	10%

Table 4.12: *Depth feature 1 - trees* - Confusion matrix for the 3-class model on test data from May and *novjanfeb*

When we tested the 3-class model, there is a reduced overall accuracy for both the May test data and the *novjanfeb* test data. The accuracy and the ROC area for the *novjanfeb* test data is quite poor at 39.6% and 0.547 (Table 4.10). Again however this test data contains a reduced number of instances. From the confusion matrices, it can be seen that it is mainly the negative and intermediate instances that are being incorrectly classified (Table 4.12). The model achieves a 93.25% accuracy on positive instances for the May data and a 83.2% accuracy on positive instances from the *novjanfeb* data. However it only classifies 56% and 62.5% of the negative and intermediate images correctly from the May data and 25.6% and 10% of the negative and intermediate images correctly from the *novjanfeb* data. In the case of the May test data, negative images are mostly incorrectly classified as intermediate, in the case of the *novjanfeb* test set, most negative images are incorrectly classified as positive. The incorrectly classified intermediate images are mainly being classified as positive for both test sets.

4.6.2 Depth Feature 2 - rocks at far wall

For the 2-class model there is an accuracy of 85.38% and an ROC area of 0.854 for test data from May. For the test data from *novjanfeb* there is a poorer accuracy of 79.6% with an ROC area of 0.796. (Table 4.13). However it is an improvement on

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
May testing					
2*class*	85.38%	0.854	0.146	0.854	0.854
3*class*	74.42%	0.744	0.128	0.75	0.808
novjanfeb testing					
2*class*	79.6%	0.796	0.204	0.796	0.796
3*class*	64.67%	0.647	0.177	0.637	0.735

Table 4.13: *Depth feature 2 - rocks at far wall* - Results for the *2-class* and *3-class* models on test data from May and *novjanfeb*

Class	No. output as Pos.	No. output as Neg.	Accuracy
May testing			
positive	358	42	89.5%
negative	49	351	87.75%
intermediate pos	272	128	68%
intermediate neg	272	128	32%
novjanfeb testing			
positive	198	52	79.2%
negative	50	200	80%
intermediate pos	125	125	50%
intermediate neg	125	125	50%

Table 4.14: *Depth feature 2 - rocks at far wall* - Confusion matrix for the 2-class model on test data from May and *novjanfeb*

results achieved for the first depth feature. From the confusion matrix for the 2-class model (Table 4.14) it appears to be classifying positive and negative instances almost equally as well for both sets of test data. In relation to the intermediate instances, for the May test data 68% accuracy is achieved when these images are regarded as being positive. For the *novjanfeb* test data, an equal amount of intermediate instances are classified as being positive and negative.

For the 3-class model, similarly to *depth feature 1 - rocks at trees*, there is a reduced overall accuracy for both test sets. A 74.42% accuracy and an ROC area of 0.808 is reached on the May test data. This is reduced to a 64.67% accuracy and an ROC area of 0.735 on the *novjanfeb* test data (Table 4.13). From the confusion

Class	No. output as Pos.	No. output as Neg.	No. output as Intermediate	Accuracy
May testing				
positive	272	3	125	68%
negative	21	315	64	78.75%
intermediate	87	7	306	76.5%
novjanfeb testing				
positive	201	43	6	80.4%
negative	48	179	23	71.6%
intermediate	96	49	105	42%

Table 4.15: *Depth feature 2 - rocks at far wall* - Confusion matrix for the 3-class model on test data from May and *novjanfeb*

matrix (Table 4.15), it appears that it is mainly the positive instances that are being incorrectly classified in the case of the May test data, with many being classified as intermediate. In the case of the *novjanfeb* test data it is mainly the intermediate images that are being incorrectly classified mostly as positive instances.

4.6.3 Depth Feature 3 - rocks at near wall

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
May testing					
2*class*	98.625%	0.986	0.014	0.986	0.986
3*class*	85.17%	0.852	0.074	0.849	0.889
novjanfeb testing					
2*class*	94.5833%	0.946	0.054	0.946	0.946
3*class*	60.28%	0.603	0.199	0.584	0.702

Table 4.16: *Depth feature 3 - rocks at near wall* - Results for the *2-class* and *3-class* models on test data from May and *novjanfeb*

For the 2-class model there is an excellent accuracy of 98.625% with an ROC area of 0.986 for the test data from May. For the test data from *NovJanFeb* an almost equally impressive accuracy of 94.58% is achieved with an ROC area of 0.946 (Table 4.16). These results are much improved on results for the previous two depth features. From the confusion matrices (Table 4.17), and the classification results,

Class	No. output as Pos.	No. output as Neg.	Accuracy
May testing			
positive	400	0	100%
negative	11	389	97.25
intermediate pos	173	227	43.25%
intermediate neg	173	227	56.75%
novjanfeb testing			
positive	119	1	99.17%
negative	12	108	90%
intermediate pos	84	36	70%
intermediate neg	84	36	30%

Table 4.17: *Depth feature 3 - rocks at near wall* - Confusion matrix for the 2-class model on test data from May and *novjanfeb*

it is clear that the model classifies both positive and negative instances quite well. In relation to the intermediate instances, for the May test data 56.75% accuracy is achieved when these images are regarded as being negative. For the *novjanfeb* test data the majority are classified as being positive (70%).

Class	No. output as Pos.	No. output as Neg.	No. output as Intermediate	Accuracy
May testing				
positive	398	0	2	99.5%
negative	12	280	108	70%
intermediate	39	17	344	86%
novjanfeb testing				
positive	117	1	2	97.5%
negative	23	54	43	45%
intermediate	73	1	46	38.33%

Table 4.18: *Depth feature 3 - rocks at near wall* - Confusion matrix for the 3-class model on test data from May and *novjanfeb*

For the 3-class model, an accuracy of 85.17% with an ROC area of 0.889 is achieved on the May test data. A lower accuracy of 60.28% with an ROC area of 0.702 is achieved on the *novjanfeb* test data (Table 4.16). In the case of the May test data it only incorrectly classifies two instances of the positive test instances

(Table 4.18). It achieves the poorest accuracy on the negative instances (70%), mostly incorrectly classifying these instances as being intermediate. In the case of the *novjanfeb* test data, it only incorrectly classifies 3 of the positive instances, however it achieves quite a poor accuracy on the negative (45%) and intermediate (38.33%) images mainly incorrectly classifying the negative as intermediate and mainly incorrectly classifying the intermediate images as positive.

4.6.4 Depth Feature 4 - island

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
May testing					
2*class*	79.25%	0.793	0.208	0.792	0.793
3*class*	65.33%	0.653	0.173	0.654	0.74
novjanfeb testing					
2*class*	53.75%	0.538	0.463	0.464	0.538
3*class*	39.17%	0.392	0.388	0.331	0.502

Table 4.19: *Depth feature 4 - island* - Results for the *2-class* and *3-class* models on test data from May and *novjanfeb*

Class	No. output as Pos.	No. output as Neg.	Accuracy
May testing			
positive	309	91	77.25%
negative	17	383	95.75
intermediate pos	5	95	5%
intermediate neg	5	95	95%
novjanfeb testing			
positive	20	100	16.67%
negative	11	109	90.83%

Table 4.20: *Depth feature 4 - island* - Confusion matrix for the 2-class model on test data from May and *novjanfeb*

For the 2-class model, an accuracy of 79.25% with an ROC area of 0.793 is produced on the test data from May. For the test data from *novjanfeb*, there is a lower accuracy of 53.75% with an ROC area of 0.538 (Table 4.19). From the

confusion matrix for the May test data (Table 4.20) it appears that it is mainly the positive instances that are being incorrectly classified. The model performs extremely poorly on the positive instances from the *novjanfeb* test data achieving an accuracy of only 16.67%. For the May test data, it classifies 95% of intermediate images as negative, meaning that the classifier clearly performs better when these are considered to be negative instances. There are no test images for the class intermediate for the test set from *novjanfeb*.

Class	No. output as Pos.	No. output as Neg.	No. output as Intermediate	Accuracy
May testing				
positive	198	18	184	49.5%
negative	18	336	46	84%
intermediate	0	35	65	65%
novjanfeb testing				
positive	2	79	39	1.67%
negative	14	92	14	76.67%

Table 4.21: *Depth feature 4 - island* - Confusion matrix for the 3-class model on test data from May and *novjanfeb*

For the 3-class model, it achieves an overall accuracy of 65.33% with an ROC area of 0.74 on test data from May. It achieves a very poor accuracy of 39.17% on the *novjanfeb* test data with an ROC area of 0.502 (Table 4.19). For the May test data, the model mainly incorrectly classifies positive (accuracy - 49.5%) and intermediate images (accuracy - 65%), mainly incorrectly classifying the positive images as intermediate and the intermediate images as negative. For the *novjanfeb* test data it only classifies 1.67 percent of the positive images correctly, mainly classifying them as negative (Table 4.21).

4.6.5 Discussion

Since it is really only of interest to detect positive and negative instances in the context of this research, the 2-class model is the best option as it produces the highest accuracies in relation to detecting each of the depth features. From the

evaluation of the 2-class models on the test data is is clear that *depth feature 1 - rocks at trees*, *depth feature 2 - rocks at far wall* and *depth feature 3 - rocks at near wall* are detected to the highest accuracy for the May test data, with *depth features 2* and *3* being detected to the highest accuracy for the *novjanfeb* test data.

The 3-class model performs quite well in certain cases considering the difficulty in distinguishing between these classes even visually due to extreme changes in lighting, weather conditions and water colour (due to reflections on the water or changes in the characteristics of the water e.g. increased turbidity due to runoff or release of the dam further upstream etc). It mainly incorrectly classifies intermediate images as positive for *depth features 1,2* and *3* for both test sets. Negative instances are mainly incorrectly classified as intermediate for all depth features except for the *novjanfeb* data where the majority of incorrectly classified negative instances are classified as positive for *depth features 1 and 2*. Positive instances are mainly incorrectly classified as intermediate for the May test data and mainly incorrectly classified as negative for the *novjanfeb* test data.

In the majority of cases the 2-class and 3-class models perform the best on positive instances except for *depth feature 4 - island*. Also the 3-class model performs better on negative instances for May test data for *depth feature 2 - rocks near wall*. This better performance on the positive instances may be due to the fact that they display a much more distinguishable visual signature than the negative instances, whose visual signature tends to vary much more across different images in the dataset.

As previously outlined it is not preferred that the 2-class model classify the intermediate instances as positive or negative, but ultimately that it be consistent in its decisions. We investigated how each model classifies these types of images in order to ultimately let it decide what they should be considered as. Many of the classification results did not produce as clear of a distinction as we had hoped, and some models classified these intermediate classes mainly as positive and others classified them mainly as negative. Therefore when carrying out the classifications

for experiments further on in the thesis, we decided to use a 2-class model where all intermediate instances are considered to be negative. Since there is no clear distinction, we decided to only detect each feature when it is fully visible in the image, since this provides an indication of a clear event at the site. Despite the fact that these intermediate instances are regarded more as positive instances by the classifiers for some depth features and test sets, we decided to keep a consistent approach in our consideration of these types of instances.

Figure 4.12: February Data - a) Rocks at trees b) Intermediate rocks at trees

Figure 4.13: February Data - a) No Island b) Island

The models generally perform poorer on the test data from *novjanfeb*. This is understandable considering the data for these months demonstrates a different visual signature. This is especially apparent for *depth feature 1 - rocks and trees* and *depth feature 2 - rocks at island*. Figure 4.12 shows an image where rocks are beginning to appear at the trees (i.e. intermediate) and an image where rocks are

present at the trees. Both of these images are from February 2010. Figure 4.13 shows an image where no island is present and an image where an island is present in November. There are no sample images from the annotated test set where an island is intermediately present. The island seen in this image is visually quite different to the sample from May which is green due to the time of year. There are only a very limited number of 'island' sample images in the test set and even at that there are none which are representative of what we see in May. It is this feature which is the most visually different from the May training data and this is apparent in the classification results for test data from *novjanfeb* for *depth feature 4 - island*.

Class	Accuracy	TP Rate	FP Rate	F-measure	ROC Area
rocks at trees					
2class	89.6	0.896	0.104	0.896	0.896
3class	66.27	0.663	0.169	0.665	0.747
rocks at far wall					
2class	97.2	0.972	0.028	0.972	0.972
3class	87.2	0.872	0.064	0.871	0.904
rocks at near wall					
2class	96.25	0.963	0.038	0.962	0.963
3class	91.67	0.917	0.042	0.917	0.938
island					
2class	91.25	0.913	0.088	0.912	0.913

Table 4.22: Models developed for each of the depth features using the *novjanfeb* data and evaluated using ten-fold cross validation

Some of the classification results suggest it may be more appropriate to develop a model for different seasons. Table 4.22 shows the results of a model built for each of the depth features using the test data from *novjanfeb* and evaluated using 10-fold cross validation. These results demonstrate that much higher accuracies are possibly achievable building separate models for use during different time periods. Thus for experiments further on in the thesis, we develop a separate model using data from November, January and February for classifying data from winter/early spring. In fact future work way involve the development of a number of seasonal models.

However it must be noted here also that the evaluation of the models on the *novjanfeb* test data included a lower number of instances and thus cannot be directly compared. In fact the overall test sets used here are much smaller to the datasets initially used in the evaluation in Section 4.5. The evaluation carried out here is to provide a snapshot of performance on unseen test data. Overall this was quite satisfactory for the 2-class model, except for perhaps *depth feature 1 - trees* and *depth feature 4 - island* on the *novjanfeb* test data.

While overall very satisfactory performance was achieved in relation to the detection of each of the depth features, these classifications may be further improved by using data from another camera on the network and viewing features from another angle. In fact when the camera pans to other angles, this also provides an alternative view of certain features. In certain cases this may alleviate some of the problems caused by reflections or limited vision from a particular angle. When there are uncertainties regarding a particular classification, the algorithm could refer to data from an alternative camera angle. Examining data from an alternative angle also enables us to capture events at the site from multiple different perspectives. The aim of this work is to examine what can be achieved from deploying a low-cost camera overlooking a site without having to make any modifications to the site, adding another camera to the network would still render the deployment very cost effective while increasing knowledge about events at the site.

4.7 Alternative Features

In this study we have looked at four depth features as a proof of concept on the how we can use features in the image that can correspond to changes in water depth for providing an estimation of depth. However there are other features that could also be investigated. For example it may be worthwhile investigating into the use of another feature to give an indication of when higher waters levels are changing. A sample feature might be to use different patterns delineated on the wall depending on

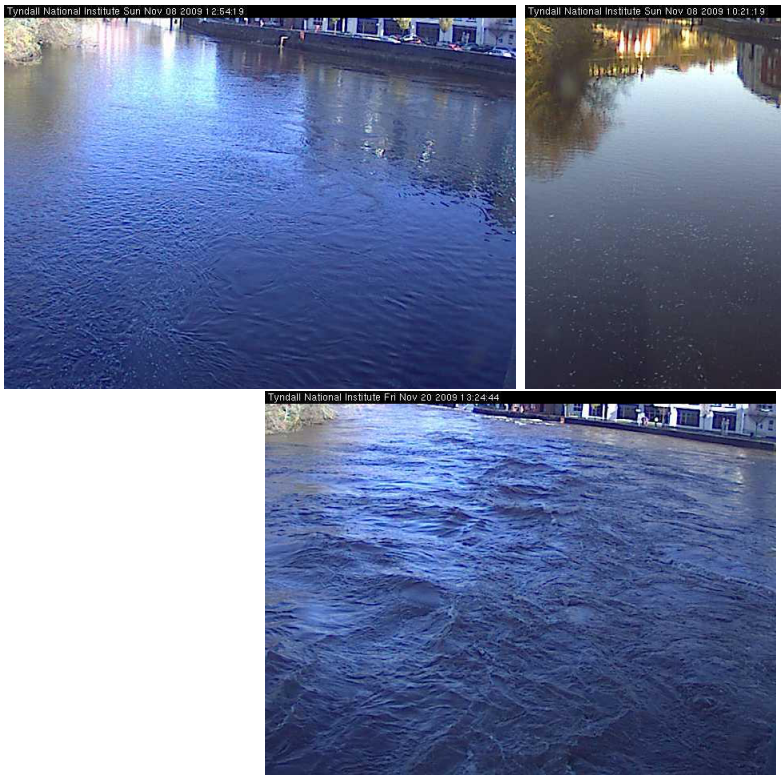

Figure 4.14: Features on the wall showing changes in higher water levels

how much of the wall was showing. Three different levels can generally be depicted by these colours. Figure 4.14 shows three images - the first image demonstrates when three lines of colours are visible due to a lower water level, the second image demonstrates when two lines of colours are visible due to a higher water level and the third image demonstrates a really high water level when only the top colour of the wall is visible.

	low water	middle water	high water
low water	628	165	7
middle water	149	589	62
high water	1	41	758

Table 4.23: Models developed for each of the depth features using the November January and February data and evaluated using ten-fold cross validation

We developed a model using 800 samples of each class from the *novjanfeb* dataset since it contained images with a larger number of sample instances of the higher

water levels. We then used ten-fold cross validation in order to evaluate if the model has the ability to different between these different appearances of the wall depending on the water level. The model achieved an accuracy of 82.29% and an ROC area of 0.867. This is very satisfactory and as can be seen from the confusion matrix in Table 4.23, the model was especially accurate in classifying the higher water levels which is the most important considering this could indicate a flood alert.

4.8 Conclusion

Our objective here was to demonstrate the use of a visual sensor for providing an estimation of depth from the camera images. Here we described four depth features that correspond well to readings from the in-situ depth sensor and to different water levels at the Lee Maltings. We subsequently developed and evaluated four models for classifying these features with very promising results.

We evaluated the fours models on two unseen datasets - one containing data from May and the other combining data from November, January and February (*novjan-feb*). For the 2-class model, accuracies of 89.2%, 85.38%, 98.63% and 79.25% were achieved on the May test data. The results were slightly poorer for the combined test data - *NovJanFeb*, with accuracies of 67.08%, 79.6%, 94.58% 53.75%. From visual inspection of the data, the time period of the second test set displayed very different visual signatures to that of the May test set. An evaluation carried out on data from this time period suggests that better accuracies may be achieved by training models with data specific for this time period. This is subsequently the approach adopted when using these models at a later stage in our study. Also for further experiments, we decided based on our analysis to use a 2-class model where intermediate appearances of a feature would be considered as a negative classification.

Future work may involve improving our analysis of events at the site from looking

at the information from alternative camera angles or even from another camera deployed at the site. Also in this study we have looked at the classification of four depth features as a proof of concept on how image features can be used to provide an estimation of parameters at the site, and linked in with data from an-situ sensor network. However further studies may also investigate the classification of other features at the site. For example, as previously outlined, it may be worthwhile investigating into the use of another feature to give an indication of when higher waters levels are changing.

As previously outlined, the overall objective of this work is to investigate how alternative sensor modalities such as visual sensors can complement the use of an in-situ sensor network, for example as a backup sensing modality in times of node failure. Now that we have developed classifiers for the four depth features, in Chapter 7 we investigate ways of relating these features back to the in-situ depth data. We examine a variety of methods resulting in a number of visual sensor streams. Following this we examine a unique approach for helping us to decide which of these is the most reliable for replicating the work of the in-situ sensor in times of node failure. More specifically we adapt a trust and reputation framework for determining the most reliable data stream at a particular point in time. In Chapter 8, we move on to evaluate the actual effectiveness of this sensing modality for detecting high and low depth events in order to examine its effectiveness as a back-up sensor in the network. However in the following chapter we investigate the role of visual sensing in a coastal region in the form of satellite remote sensing data.

Chapter 5

Satellite Remote Sensing Data

5.1 Introduction

Sophisticated satellite sensors are very effective for monitoring many ocean and coastal parameters such as sea surface temperature, sea surface height, ocean currents, and chlorophyll pigment concentration (which can subsequently be used to determine the amount of algal growth in the water) along with other water quality parameters such as mineral suspended sediments and dissolved organic matter (Miller et al., 2007). A number of these sensors are orbiting the earth on various satellite platforms with different spatial and temporal resolutions.

In the preface to their text on Remote Sensing of Coastal Aquatic Systems, Miller et al. (2007) outline the importance of coastal systems, their complexity and the need for more sophisticated sampling techniques such as that provided by satellite remote sensing. They are dynamic environments affected by a range of factors, both naturally occurring and of anthropogenic influence. They can often be subject to large-scale pollution through discharge from the local watershed, runoff from local areas or industrial output.

However in our work the focus is specifically on satellite remote sensing data that can enhance the use of the in-situ sensor nodes in an operational multi-modal sensor network and provide a form of redundancy in the network. The key objective here is

to investigate the availability of satellite data at an appropriate spatial resolution to provide measurements that can most accurately coincide with data from the in-situ sensor node. This data also needs to be available at a sufficient temporal resolution so that if the in-situ sensor node were to fail, we can still avail of high frequency data. The objectives of this work thus require near real-time cooperation between the two sensor streams. The site and the in-situ data sources of interest here are Galway Bay and data from the SmartBay environmental monitoring buoys.

The satellite remote sensing parameters that are complementary to data from the SmartBay buoys can be divided into two main categories – ocean colour measurements and sea surface temperature (SST) measurements. Hence the discussion proceeds from here according to these two categories of remote sensing data, outlining their issues and applications. Based on this, we outline the satellite data sources chosen for use in this research and the extraction of information from these data sources for use in the network. The analysis provided here forms part of our analysis in relation to research question 2 of our research questions presented in Chapter 1. Due to the large number of acronyms used in this chapter, a glossary is provided at the end of the thesis in order to provide the full terms for satellite sensors and organisation names that have been abbreviated.

5.2 Ocean Colour

In this section we provide an outline of the parameters that can be obtained from ocean colour measurements, a range of satellite sensors used for carrying out ocean colour measurements, issues in relation to choosing an ocean colour sensor and limitations in relation to satellite remote sensing of ocean colour in coastal zones. This highlights the difficulties in relation to satellite ocean colour sensing and in obtaining a suitable data stream in the context of the objectives of our research.

5.2.1 Ocean Colour Parameters

In the Ocean Colour literature the reflectance or optical properties of water are mainly attributed to suspended sediments, phytoplankton, and dissolved organic matter otherwise known as 'yellow substance' or gelbstoff as described in (Dareckia et al., 2003; Hellweger et al., 2004; Darecki and Stramski, 2004; Zimba and Gitelson, 2006; Gitelson et al., 2008). Satellite sensors can correlate the amount of solar radiation reflected by surface water at various wavelengths to estimate water quality parameters such as these.

Chlorophyll is outlined as being one of the most important parameters observable from satellite imagery. Chlorophyll can provide an indication of the phytoplankton content of the water and thus can help detect the occurrence of algal blooms (Rast et al., 1999). Yellow substance or 'gelbstoff' was introduced earlier and consists of decayed organic material (from plants and animals) that has been dissolved in marine waters. It is said to consist of mostly organic molecular polymers and it is usually transported to the sea by rivers (Rast et al., 1999). Finally suspended matter is described in (Rast et al., 1999) as "all suspended particles in water without particular algal pigments and of a size larger than 0.47 μm". It is said to consist mainly of sediments and is an important indicator for marine turbidity studies. Even in low concentrations, it is said to dominate the water colour for wavelengths up to 670nm in coastal waters. In (European Space Agency, 2006) suspended matter is described as being a combination of inorganic particles and detritus (particulate material that enters into a marine or aquatic system), present due to re-sedimentation and advection processes, atmospheric inputs and dead material such as plankton.

5.2.2 Ocean Colour Sensors

In (European Space Agency, 2006), an overview is provided of the history of ocean colour monitoring through satellite sensors where it is outlined that the first observations of ocean colour through satellite remote sensors were carried out by Coastal

Zone Colour Scanner (CZCS) aboard NASA's Nimbus-7 satellite from 1978 to 1986. From 1986 to 1996, there was a subsequent absence of an operational satellite sensor for the production of ocean colour data. In March 1996, India launched the German sensor MOS which provided useful data despite not providing global coverage. In August 1996, the Japanese sensor OCTS and the French sensor POLDER was launched by Japan on the ADEOS mission. This operated until June 1997 when failure of the satellite's solar panel ended the mission. Then in August 1997, the SeaWiFS was launched by the USA. This was to operate as a follow-on sensor to the CZCS which had finished operating in 1986 and it only finished collecting data in December 2010.

Following the launch of SeaWiFS, a number of other ocean colour sensors also began operations and started providing ocean colour data to the Earth Observation community. A selection of sensors in operation that are used for monitoring ocean colour are outlined in Table 5.1. They they are described under various headings such as spatial, temporal and spectral resolution, etc . The range of ocean colour sensors available provides an indication of the range of activity and importance that is being attributed to remote sensing of ocean colour.

Sensor	**Satellite**	**Agency**	**Dates**	**Swath (km)**	**Spatial Resolution (m)**	**Bands**	**Spectral Coverage (nm)**	**Revisit Time**
MERIS	ENVISAT	ESA (Europe)	March 2002 -	1150	300/1200	15	390-1040	1-3
MODIS	Terra/Aqua	NASA (USA)	Dec 1999 -/May 2002 -	2330	250/500/1000	36	400-1440	1-2
OCM-2	Oceansat-2	ISRO (India)	Sept 2009 -	1420	360/4000	8	402-885	2
COCTS/CZI	HY-1B (China)	CNSA	April 2007 -	1400/500	1100/250	10/4	402-12500/433-695	1/7
GOCI	COMS	KARI/KORDI (South Korea)	June 2010	2500	500	8	400-900	1 hour (geostationary)
POLDER-3	Parasol	CNES France	Dec 2004	2100	6000	9	443-1020	
VIIRS	NPP	NOAA/NASA (USA)	October 2011	3000	375/750	22	400-12000	1-2
ETM+ (MS/Pan)	Landsat 7	NASA	April 1999	183	30 (60-thermal, 15-pan)	7/1	450-2350/520-900	16
HRG(MS/Pan)	SPOT 5	CNES	May 2002	60	10/5,2.5	4/2	500-1750/ 480-700	2-3
MS/Pan	Ikonos	GeoEye	September 1999	11/5	4/1	4/1	445-929	3

Table 5.1: Sensors used for monitoring ocean colour. Abbreviations: MS: multispectral, Pan: panchromatic, SWIR - short-wave infrared

While some of these sensors were developed specifically for monitoring ocean colour, land-observing satellite sensors such as IKONOS, Landsat ETM+, and SPOT are also listed in the table as they have also been used for estimating ocean colour parameters, particularly due to their high spatial resolution. However these types of sensors sometimes do not have as high a radiometric sensitivity which is often required for the accurate retrieval of various water quality parameters in optically complex waters (Hellweger et al., 2004; Kutser et al., 2005). There are also issues in that some of these sensors may have a low temporal resolution. These sensors were not developed specifically for measuring ocean colour parameters and thus may not directly produce estimates of geophysical quantities such as chlorophyll-a, TSM etc. However algorithms may be applied to reflectance data in order to estimate these parameters (e.g. (Hellweger et al., 2004; Kutser et al., 2005)). Other satellites such as the AVHRR which are normally used for sea surface temperature mapping have also been used in a variety of studies for observing ocean colour.

Hellweger et al. (2004) provide an overview of the spatial, spectral and temporal resolution of satellite sensors commonly used for water quality studies. They also provide a review of the literature and point out the various sensors that have been used for measuring water quality in inland, estuarine and near-shore waters. They note a predominant use of high spatial resolution sensors (e.g Landsat ETM) in inland waters and low spatial resolution sensors (e.g. AVHRR) in near shore ocean waters. However Landsat ETM would have a lower spectral and temporal resolution and AVHRR would have a lower spectral resolution than ocean colour sensors such as MODIS or MERIS. However it would have quite a high temporal resolution. It should be noted that much of the literature reviewed here is from the 1980's, 1990's and early 2000's. Sensors such as MERIS and MODIS-Aqua were not launched until 2002.

As outlined in Darecki and Stramski (2004) building upon the heritage of the CZCS, significant efforts were made to develop ocean colour sensors with improved spectral and radiometric performance and improved spectral and spatial coverage.

SeaWiFS (Hooker and McClain, 2000) and MODIS (Esaias et al., 1998) were the result of these efforts in the USA. MERIS (Rast et al., 1999; European Space Agency, 2006) was the result of European efforts for developing a sensor more suitable for monitoring ocean colour. The use of data from each of these sensors is investigated by many studies throughout the literature. Data from these three sensors are also used in the satellite data products used in this work which are described in Section 5.4.1.

5.2.3 Choosing an ocean colour sensor

As can be seen from Section 5.2.2 there are a variety of ocean colour sensors and land-observing satellite sensors that can provide data which could be of use to us in our research. However the choice of sensor is very application-specific as sensors differ in factors such as spatial, temporal and spectral resolutions.

Hellweger et al. (2004) outlines how there is a range of satellite sensors suitable for estimating water quality parameters, however there is a tradeoff in spectral, spatial and temporal resolution and the best combination depends on the intended use. They state that no single sensor can have a high spectral, spatial and temporal resolution. Thus the selection of sensors is application-dependent. As outlined in Section 5.2.2 they compiled a list of sensors based on a literature review of satellite measurements of water quality in inland, estuarine and near-shore ocean waters. They outline the spectral, spatial and temporal resolution of each of these sensors and point to a series of studies from the literature highlighting the sensors chosen for these applications.

Differing applications have different requirements. For example a high spectral resolution is generally required to differentiate between substances (e.g. suspended sediments vs. chlorophyll-a). Sensors designed for this like MODIS generally have a lower spatial resolution. Many land applications however require a high spatial resolution. Sensors designed for such purposes such as Landsat-ETM or IKONOS have a high spatial resolution but a lower spectral resolution. MERIS is highlighted

as an example of a sensor that provides a combination of medium spectral, spatial and temporal resolution with MODIS being highlighted as an example of a sensor that provides a range of spatial resolutions.

In the study reported by Hellweger et al. (2004), they are seeking to choose a suitable sensor for observing water quality in New York Harbour. Hence they require a sensor satellite sensor with a high spectral resolution because they are concerned with characterizing water quality. Because the study has taken place in coastal waters this is even more desirable considering these waters are often very optically complex. They require a high spatial resolution and need to deal with issues such as interference from non-water features. A high temporal resolution is also desirable as the New York harbour area which is the target area, is dominated by the tide and the water quality can change in hours. However none of the sensors listed can achieve the temporal resolution required for monitoring such a dynamic so they chose two sensors spanning a range of spectral and spatial resolutions for their study – the Landsat ETM sensor (low spectral/high spatial resolution sensor) and the MODIS sensor (high spectral/low spatial resolution with the 1km spatial resolution) with mixed results.

Other studies also demonstrate very site-specific requirements in relation to their choice of satellite imager. For example Kutser et al. (2005) investigate the use of satellite remote sensing for investigating Coloured Dissolved Organic Material (CDOM) in lakes over large geographic areas. Their study is based in Southern Finland and Southern Sweden. They outline the high CDOM absorption in many boreal lakes requires high radiometric sensitivity due to low reflectance and that the lake size is often small compared to the $1km$ pixel resolution associated with most aquatic satellite sensors. However the spatial resolution of ocean colour sensors such as SeaWiFS, MERIS, and MODIS is described as being inadequate for most lake measurements despite their high radiometric sensitivities and global coverage. They outline the spatial resolution of land-observing satellites such as IKONOS, Landsat ETM, and SPOT as being desirable for boreal lake measurement, however their

radiometric sensitivity is deemed to be problematic for the scenario in question. Instead they investigated the use of a land observation sensor known as Advanced Land Imager(ALI) with improved spectral, spatial and radiometric resolution. The results of their study indicated that CDOM content in lakes over a wide range of concentrations could be mapped using ALI data.

Finally, in another study carried out by Simis et al. (2005), they investigate the use of remote sensing for the detection of the pigment phycocyanin (PC) which can indicate blooms of cyanobacteria. In this study they choose the MERIS sensor due to the availability of a band deemed suitable for the detection of PC, and its 300m resolution being suitable for the moderately sized lakes which were under consideration in the study. Matthews et al. (2010) also choose MERIS for a observing a lake - Zeekoevlei, a small hypertrophic freshwater lake in Cape Town, South Africa which is dominated by cyanobacteria. They outline the sensor specifications, frequent data acquisition, 300m pixel resolution, high signal-to-noise ratio and number and position of spectral bands make MERIS suited to regular/real-time monitoring applications. However in the introduction to their work the authors note that MERIS is not very well suited to observing lakes smaller than $1km^2$ which represents a significant number of the world's lakes.

From this outline it is clear that there are a number of issues to consider when choosing a satellite remote sensor and it is very specific to the application in question. The application context of our research is monitoring the coastal zone where we require high temporal measurements that also produce reliable measurements. In the study outlined by Hellweger et al. (2004) none of the sensors considered could achieve the temporal resolution required. This may also may be an issue for the needs of our research. The following section highlights some of the limitations with regards to the use of satellite remote sensing in a coastal zone.

5.2.4 Limitations of Satellite Remote Sensing for Ocean Colour Monitoring in Coastal Zones

Rast et al. (1999) states that when detecting biological constituents that dominate the optical properties on the surface layer of marine waters, two classes of water are distinguished. These are commonly referred to in the literature as Case 1 and Case 2 waters. They describe Case 1 water as comprising blue, oligotrophic waters (with pigment concentration below 0.1 mg m^{-3}), biologically active waters (with pigment concentration around 1 mg m^{-3}) and green, eutrophic waters (with pigment concentration as high as 10 mg m^{-3}). Case 2 waters are described as consisting of the constituents found in case 1 waters along with particles from runoff, suspended sediments, dissolved organic matter and other dissolved substances originated from anthropogenic influx into the water system. This water type is outlined as occurring mostly in coastal zones with high fluvial runoff, estuaries and shallow offshore zones.

Dareckia et al. (2003) outline how it was Morel and Prieur (1977) who initially classified marine waters in terms of variability of optical properties and coined the terms Case 1 and Case 2 waters. Darecki and Stramski (2004) describe Case 1 waters as waters where substances other than phytoplankton are either optically insignificant or correlated with phytoplankton (also deemed to be somewhat an over-simplification of things), whereas Case 2 waters typically include coastal and inland water bodies where materials such as suspended inorganic matter and/or dissolved organic matter make a significant contribution to the optical properties and vary independently of phytoplankton and each other.

Based on these descriptions, Galway Bay would be classified as Case 2 waters. The reason for providing this description here is due to the fact that the optical complexity of Case 2 waters often renders remote sensing algorithms for estimating ocean colour parameters unreliable in coastal zones and many studies outline the need for improved algorithms for differing water types e.g. (Zimba and Gitelson, 2006; Darecki and Stramski, 2004; Stumpf et al., 2003). For example, Stumpf et al.

(2003) use ocean colour imagery from SeaWiFS as part of their study for monitoring harmful algal blooms (HABs). They process imagery data using coastal-specific algorithms stating how the NASA global chlorophyll algorithm over-estimates in the region.

The potential inability of the satellite sensor to distinguish between the various water quality parameters has been pointed out as a limitation with satellite remote sensing in a variety of studies e.g (Hellweger et al., 2004) and there are numerous studies in the literature investigating and validating models for the estimation of these parameters, especially chlorophyll-a, in Case 2 waters e.g. (Gitelson et al., 2008; Yang et al., 2010). In (Dareckia et al., 2003), they determine the application of a single remote sensing algorithm for Case 2 waters to be infeasible. Other studies have found more promising results. Gons et al. (2002) adapt a chlorophyll retrieval algorithm for use with MERIS for inland and coastal waters. They claim their study demonstrates that this algorithm can be used to retrieve moderate to high concentrations of chlorophyll-a in different water bodies without region-specific calibration.

There also are other limitations with satellite remote sensing for monitoring coastal zones. Ocean colour sensors often have a spatial and temporal resolution often not sufficient for monitoring small-scale features or highly dynamic environments. For example the resolution of satellite ocean colour sensors such as SeaWiFS, MODIS and MERIS is outlined in Kutser et al. (2005) as often inadequate for measurements of many lakes. Hellweger et al. (2004)) find the spatial resolution of the 1km MODIS data water pixels to present a high risk of contamination from non-water features in the site under investigation (New York Harbour). Spatial accuracy of the land-mask and bottom reflection in near-shore shallow waters are also highlighted as issues. They also state how none of the arrays of sensors they had listed have the temporal resolution to resolve the dynamics of a tidally dominated system where water quality can change within hours.

Ocean colour measurements are also hampered by cloud cover and sun glint,

therefore even if a sensor has a re-visit time of three days, this does not always mean there will be data available at this temporal scale. Gregg and Woodward (1998) outline how due to the fact that ocean colour sensors detect upwelling radiances in the visible and near-infrared portions of the spectrum, that they can be limited by atmospheric processes that obscure ocean viewing, such as clouds and sun glint. Investigations have often found it difficult to get coincident satellite and ground observations for investigating the use of satellite remote imagery at a particular site or the development and evaluation of the accuracy and precision of ocean colour chlorophyll algorithms e.g. (Hellweger et al., 2004; O'Reilly and Maritorena, 1998).

To summarise, despite the potential benefits of satellite remote sensing data, there are also a number of limitations. These limitation relate mainly to difficulties for accurately estimating ocean colour parameters in the coastal zone, limited spatial and temporal resolution for monitoring small scale features and highly dynamic environments, and a further hampering of the temporal resolution due to issues such as sun glint and cloud cover. Subsequently with Galway Bay being a coastal zone and a location that is frequently covered by cloud, a satellite remote sensing stream that meets the needs of our application context and can produce data at very high temporal scales is difficult to achieve. However like many application contexts in the literature it may be very beneficial from a contextual point view in providing data that can help to design ground sampling programs. It can provide an overview of dynamics in the bay over a large spatial area that are difficult to obtain otherwise, and subsequently lead to better management decisions for the bay and in the use of the in-situ instrumentation, and a better understanding of environmental processes. The following provides a brief overview of some of the application scenarios for ocean colour data in the literature.

5.2.5 Applications of Sensing Ocean Colour

There are numerous applications for ocean colour data highlighted in the literature. One of its primary uses is monitoring the spatial extent and duration of harmful

algal blooms (HABs) and the environmental conditions surrounding them. Negative effects of these HABs events include discoloration of the water, surface scum, the covering of beaches with biomass or foam, and the depletion of oxygen levels through excessive respiration and decomposition which can affect marine life and lead to fish kills (Sellner et al., 2003). Some phytoplankton blooms can also contain toxin-producing species. For example Karenia Brevis, a toxic dinoflagellate has been known to cause neurotoxic shellfish poisoning, fish and marine mammal deaths, and human respiratory irritation (Stumpf et al., 2003). Research has been carried out into satellite image analysis techniques to specifically detect toxic species or to distinguish between toxic and other species (e.g. Amin et al. (2009), Stumpf et al. (2003), Astoreca et al. (2009), Kutsera et al. (2006), Simis et al. (2005), Kahru and Mitchell (1998), Millie et al. (1997), Staehr and Cullen (2003), Cannizzaroa et al. (2008) and Matthews et al. (2010)). Some studies are also investigating the use of satellite imagery and in-situ field measurements to predict the occurrence of algal blooms in order to mitigate their effects in a timely fashion, for example NOAA's Harmful Algal Bloom Operational Forecast System (HAB-OFS)[1].

Ocean colour data has also been used to monitor river plumes e.g (Baban, 1995; Froidefond et al., 2004; Mertes and Warrick, 2001). This is an interesting phenomenon that can be observed when a river discharges into the sea. Knowledge of river plumes is critical in regional oceanography. Rivers may carry excessive sediment to the coastal ocean which increases the turbidity of the water and hence affects water quality. Excessive sediment load can have an impact on coastal ecosystems and can lead to the transport of pollution pathogens and toxic phytoplankton (Muller-Karger et al., 2007). Satellite remote sensing data often allows the spatial characterisation and dynamics of a river plume to be monitored in a manner which would be difficult or impossible to achieve with a field sampling program.

These examples highlight some of the applications of ocean colour data and provide us with a flavour of the possibilities. However there are many more applications

[1]http://tidesandcurrents.noaa.gov/hab/

with important environmental impacts such as monitoring pollution or tracking the progress of oil spills. The dynamics of such events can be quickly detected and monitored aiding the management and deployment of rapid mitigation strategies. Use of satellite data in this manner is the focus of many operational marine monitoring programs. For example there are many European Union led initiatives that are investigating a capacity for the use of earth observation data for operational monitoring needs such as these e.g. the Global Monitoring of Environment and Security (GMES) programme [2].

5.3 Sea Surface Temperature (SST)

A similar analysis to that provided for Ocean Colour is provided here in relation to SST satellite remote sensing data.

5.3.1 Satellite Sensors for SST Observation and Constraints

Sea surface temperature (SST) has been routinely observed using thermal infrared data from space-borne sensors for many years. A number of these thermal infrared sensors are outlined in Table 5.2 and described under various headings. Satellite instruments such as AVHRR, MODIS and ATSR have the ability to derive satellite SST measurements with accuracies of a few tenths of a degree (Noyes et al., 2006). High accuracy SST measurements are essential for climate research and other studies and the ATSR radiometers were designed with these requirements in mind. The first ATSR, an experimental instrument, was launched on board ERS-1 in 1991. In 1995 ATSR-2 was launched onboard ERS-2. These were followed by the Advanced ATSR (AATSR) launched onboard Envisat in March 2002.

Donlon et al. (2007) provide an overview of the different types of sensors and platforms for measuring SST, outlining the sampling characteristics and absolute accuracy levels. These include infrared and microwave radiometers with varying

[2] $http://www.gmes.info$ accessed October 21 2011

characteristics. A 3-10km spatial resolution is available from infrared radiometers on satellites with a geostationary orbit. These have a revisit time of approximately 30 minutes. However they have a limited field of view and thus only provide regional coverage. A dual-view infrared radiometer on a polar orbiting satellite provides global coverage but only every three days. It has a spatial resolution of 1-2km and a high absolute accuracy which can be used to standardize the bias errors of other sensors. An infrared wide swath radiometer on a polar orbiting satellite has a spatial resolution of 1-4km, global coverage every 12 hours but a lower absolute accuracy than an infrared dual view radiometer. All infrared radiometers fail over cloud and the infrared wide swath radiometer also fails in the presence of atmospheric aerosol. A microwave radiometer on a polar orbiting satellite has a spatial resolution of 25-50km, global coverage every 1-2 days, a lower accuracy than some infrared radiometers but it can still obtain measurements through cloud cover. It is however affected by rain.

The constraints of IR sensors highlighted by authors include the fact that their observations are affected by clouds and aerosols e.g. (Wentz et al., 2000; Chelton and Wentz, 2005). Microwave radiometry offers a solution to these two main constraints. Microwaves at frequencies below approximately 12 GHz can provide a clear view of the sea under all weather conditions except rain and at these frequencies atmospheric aerosols have no effect. Wentz et al. (2000) provide a brief history of satellite microwave sensing stating that the first microwave radiometers operating at these low frequencies were launched in 1978 on SeaSat and Nimbus-7. The Tropical Rainfall Measuring Mission (TRMM) was launched in November 1997 with a microwave imager TMI. Wentz et al. (2000) state that this sensor is the first satellite sensor capable of accurately measuring SST through clouds.

Sensor	**Satellite**	**Dates of Operation**	**Swath (km)**	**Res.(m))**	**ATSR Spectral Bands**	**IR Spectral Coverage (μm^{-1})**	**Orbit**
AVHRR/1-3	NOAA	1978	2800	1100	4/5/6	AVHRR/3 - 0.58-0.68, 0.725-1.0, 1.58-1.64,3.55-3.93, 10.3-11.3, 11.5-12.5	polar
AVHRR/3	Metop	Oct 2006	2800	1100	6	0.58-0.68, 0.725-1.0, 1.58-1.64,3.55-3.93, 10.3-11.3, 11.5-12.5	polar
ATSR/1-2	ERS	July 1991/ April 1995	500	1000	4/7	first 3 bands only ATSR-2: 0.545-0.565, 0.649-0.669, 0.855-0.875, 1.580-1.640, 3.550-3.930, 10.40-11.30, 11.50-12.50	polar
AATSR	Envisat	2002	500	1000	7	0.545-0.565, 0.649-0.669, 0.855-0.875, 1.580-1.640, 3.550-3.930, 10.40-11.30, 11.50-12.50	polar
MODIS	Terra	1999-present	2330	1000	5	(SST) 3.66-3.84, 3.929-3.989, 4.020-4.080, 10.780-11.280, 11.77-12.27	polar
MODIS	Aqua	2002-present	2330	1000	5	(SST) 3.66-3.84, 3.929-3.989, 4.020-4.080, 10.780-11.280, 11.77-12.27	polar
SEVIRI	MSG	2004	+/- 60 deg	3000	3	0.4-1.6 (4 visible/NIR channels), 3.9-13.4 (8 IR channels)	geostationary
VIIRS	NPP	October 2011	3000	375/750	22	400-12000	polar

Table 5.2: Thermal Infrared Satellite Sensors often used to measure SST

They outline the major limitation of microwave SST retrievals to be the relatively coarse resolution (on TMI, the spatial resolution is approximately 50km), compared to infrared SST retrievals which generally have a higher resolution. Also, as previously outlined, SST retrievals may not be accurate when rain is present and so are discarded. Other limitations of the TMI noted by Chelton and Wentz (2005) include an accuracy of approximately $0.5°C$, an inability to measure SST near land, a restriction of measurements to the latitude band of 40°S-40°N, and a degraded accuracy of retrievals of SST below approximately $10°C$. AMSR-E which orbits the earth on NASA's Aqua satellite addresses some of these limitations This sensor began sampling the global ocean in June 2002 with 89% coverage each day and 98% coverage every two days. Again the resolution of this instrument would not be as high as some of the IR sensors, however it leads to improved global sampling through greater coverage.

The launch of WindSat on the Coriolis satellite[3] in January 2003 brought about a new generation of satellite remote sensing instruments (Brown et al., 2007). Measurements over the ocean are used as input to NWP models of the U.S. Navy, NOAA, and the UK Met Office[4]. Despite all of this, microwaves are several orders of magnitude larger in wavelength than the visible portions of the spectrum. Thus the spatial resolution of typical microwave radiometer observations is coarse. They have limited capabilities for monitoring coastal regions, and precipitation in the atmosphere can severely limit the retrieval of certain geophysical parameters (Brown et al., 2007).

Scientists are investigating the best combination of tools from satellite observations, in-situ measurements, and numerical models for global SST analysis as discussed in (Donlon et al., 2001). This seeks to build on the complementary aspects of various satellite instruments by merging data to provide information of increased quality and resolution. Since we require a satellite data stream with high temporal and spatial resolution for use in the context of this work, an SST analysis created

[3]Coriolis mission is jointly sponsored by the DoD Space Test Program and the U.S. Navy (SPAWAR PMW-180)

[4]http://www.nrl.navy.mil/WindSat, accessed June 1 2011

from incorporating a number of satellite data sources (and perhaps in-situ data sources) is used in this study. In Section 5.2.4 constraints associated with the use of satellite ocean colour data in coastal zones was highlighted. Similar constraints in relation to spatial and temporal resolution exist in relation to using individual SST satellite data streams. Hence a combined analysis is more appropriate for the requirements of our research. The following provides a brief overview of some of the applications of satellite SST measurements.

5.3.2 Applications of Sea Surface Termperature Measurement

SST measurements are fundamentally important to a number of studies and applications including forecasting systems, tourism and fisheries research. Glenn et al. (2000) point out the potential applications for real-time observation and forecasting systems in the coastal environment including safe and efficient navigation and marine operations, efficient trajectory prediction for oil and hazardous material spills, monitoring, predicting and mitigating coastal hazards, aiding search and rescue missions, prediction of HABs or other water quality phenomena and for scientific research. Sea surface temperature data is very important to most forecasting models.

Many studies have highlighted the necessity of satellite SST data for creating data sets with the required coverage. For example, Donlon et al. (2007) highlight the need for earth-observing satellite instruments stating that in situ measurements from buoys, ships of opportunity, and voluntary observing ships are inadequate for providing frequently sampled SST maps with the high spatial resolution and global coverage required for input into ocean forecasting models.

Many studies from the literature also highlight the importance of SST observations. Reynolds et al. (2007) outline sea surface temperature to be an important variable for better understanding interactions between the ocean and the atmosphere. They outline purposes for SST analyses including climate monitoring and

prediction and feature tracking. Donlon et al. (2001) outline how SST products are required by operational ocean analysis and prediction systems to "properly constrain the upper-ocean circulation and thermal structure" and that SST is a key parameter in other oceanographic fields such as coastal oceanography and monitoring of biological resources. It is required as an input to numerical models which form the basis for operational ocean forecasting systems. These systems are being used by governments, industries, fisheries etc. and thus serve a wide variety of needs. It is also extremely important in operational coastal monitoring applications as it can have a large effect of ecosystem function e.g. changes in water temperature can result in increased algal growth in the water (Sellner et al., 2003; Elliott and May, 2008) and can changes the solubility of oxygen e.g.(Vega et al., 1998). It may also be indicative of other events such as increased freshwater inflow which may indicate nutrient loading etc.

Many studies investigate the production of SST analysis products which combines data from infrared and microwave sensors along with also possibly in-situ measurements in order to produce high resolution SST estimations e.g. (Beggs et al., 2011). The Global Ocean Data Assimilation Experiment (GODAE) established the GODAE High Resolution SST Pilot Project (GHRSST-PP) which is described in more detail in Section 5.4.2. However one of the key outcomes of this project is the creation of satellite data products in a common format which can be easily ingested into various analysis systems.

For example the UK Met Office developed a new global, operational, high resolution, combined sea surface temperature (SST) and sea ice analysis system (OSTIA). It outputs a daily global coverage combined SST and sea ice concentration product (approximately 6km resolution), which is generated in near-real time. This system takes advantage of the data provided by GHRSST-PP and aims to meet the needs of applications requiring high resolution space-time scales including global numerical weather prediction (NWP) and operational ocean models. It incorporates both in-situ data and remote sensing data from a variety of instruments. This system

is described in Stark et al. (2007). Due to the high temporal coverage that can be achieved by such analysis products, these types of products are examined for use in our research.

5.4 Satellite Data Products for Use in the Galway Bay Network

While there are numerous benefits to satellite remote sensing data products in many application contexts, in our work the focus is specifically on high temporal satellite remote sensing data that can enhance the use of the in-situ sensor nodes in an operational multi-modal sensor network and provide a form of redundancy in the network. Following the previous analysis it is clear that relying on a singular satellite sensor for this type of data is not suitable in the context of this work.

Therefore based on this, the satellite data products chosen for incorporation into this network are mainly products that produce an analysis based on the combination of many satellite and in some cases in-situ data streams. These products can often achieve a higher accuracy with increased spatial and temporal resolution through exploiting the various characteristics of a number of satellite sensors.

In Chapter 3, a description of the in-situ data sources in Galway Bay was provided. Due to the fact that the SmartBay Buoys contain a number of water quality sensors, we chose to align satellite data streams to this in-situ sensor network. Data is available from the SmartBay Buoy at Mace Head from July 16 2008 until March 03 2010. Data from the SmartBay Buoy at MidBay is available from July 16 2008 until December 07 2010. These are the start and end dates for the data sets provided by the Marine Institute upon request. However this does not mean that data is available continuously from each sensor from these start times, there are temporal gaps.

Satellite data products for ocean colour and sea surface temperature (SST) were downloaded for similar time periods to the in-situ measurements. Data was ex-

tracted from these sources using Beam[5] – an open-source toolbox and development platform for viewing, analysing and processing of remote sensing raster data. It was first developed by Brockmann Consult[6] in 2002 under a contract with the European Space Agency and has undergone many release cycles since and has become an accepted platform for the analysis and processing of Earth Observation Data. The Mace Head SmartBay Buoy is at latitude-longitude location of (53.3327, -9.9324). The MidBay SmartBay Buoy is at a latitude-longitude location of (53.1136, - 9.51083). When extracting information from the satellite data products, the Java Beam API was used to select the pixels that best represented the locations of the SmartBay buoys. Due to the proximity of the buoys to the bay, the spatial resolution of the data products and the fact they were mainly analysis products as opposed to individual satellite data products, this appeared to be the best approach for selecting a corresponding value from the satellite data product for use in each of the networks (i.e. the Mace Head and MidBay networks). The following describes the satellite data analysis products used in our Galway Bay multi-modal sensor networks. However it should be noted that the purpose of this research is not to provide a thorough analysis on the use of all available analysis products, but rather to highlight some of the benefits and issues with using a selection of these in such an application context.

5.4.1 Ocean Colour Data

Data from satellite sensors normally occurs in different formats which generally vary in terms of of processing that has been carried out. For example the MERIS Product Handbook European Space Agency (2006) describes three processing levels for MERIS data described as follows:

- **level 1B** - images resampled on a path-oriented grid, with pixel values having been calibrated to match the Top Of Atmosphere (TOA) radiance.

[5]http://www.brockmann-consult.de/cms/web/beam/
[6]http://www.brockmann-consult.de

- **level 2** - images deriving form the level1B products, with pixel values having been processed to get geophysical mesurements.

- **level 3** - a synthesis of more than one MERIS product (and possibly external data) to display geophysical measurements for a time period.

The level 1B products consist of radiance values, the level 2 products consist of geophysical products such as chlorophyll-a, TSM etc. produced by processing the radiance values from the level 1B data using a standard algorithm, and level 3 products incorporate a number of MERIS products sometimes along with external data to produce perhaps weekly or monthly data. The different MODIS data products are described similarly[7]. A level 4 data product is also outlined which is described as being model output or results from analyses of lower level data (such as variables derived from multiple measurements). They outline for example "Ocean primary Productivity"[8] as an example of a level 4 product.

Initially we had considered using level 2 data from MERIS or MODIS. These satellite sensors produce data with a moderate to high spatial resolution. MERIS is thought to be particularly suited to coastal zones and MODIS has quite a high temporal resolution with a re-visit time of approximately 1 day. As part of this research, we developed a system for the efficient browsing, searching and analysis of MODIS chlorophyll data described in (O'Connor et al., 2010). This system was based on searching and analysing the data using the JPEG summary image of the source data. The pilot system was based on images ranging from May 31 2008 to October 28 2009. It was clear from browsing these images and analysing the pixel colours of the summary images that there is a very limited amount of data available from these single sensor products for Galway Bay, mainly due to cloud cover. Two screen shots from this system can be seen in Figure 5.1. Therefore data sources produced through a combined analysis of ocean colour data using information from a combination of sources were considered to be more appropriate in the context of

[7]http://oceancolor.gsfc.nasa.gov/PRODUCTS/product_level_desc.html
[8]http://www.science.oregonstate.edu/ocean.productivity/

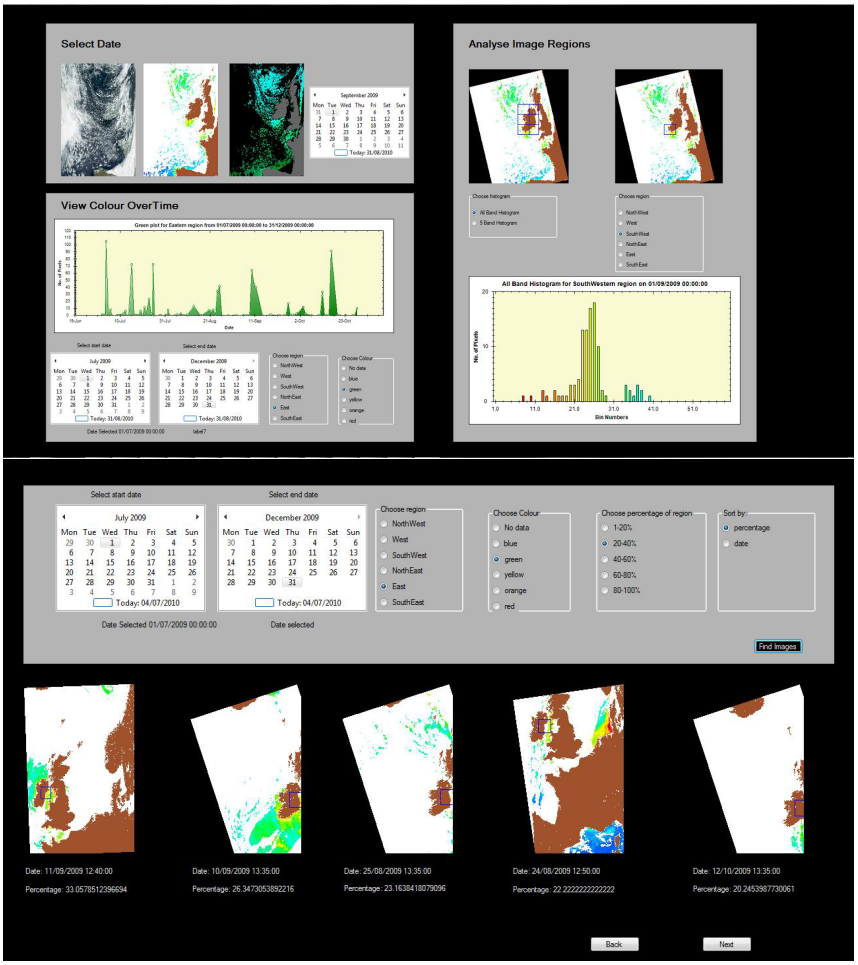

Figure 5.1: Satellite image analysis system developed for this work

our application. We selected two different sources - GlobColour products and an Ifremer product - that produce a combined analysis. These are described in the following sections.

5.4.1.1 GlobColour

GlobColour is a project initiated and funded by the European Space Agency (ESA) to develop a satellite-based ocean colour data service to support operational oceanography and global carbon-cycle research (ACRI-ST, 2008). It was initially a three year project that began in November 2005 under the leadership of ACRI-ST (France). Globcolour has been involved in merging data from SeaWiFS, MODIS, and MERIS.

For our work, data was downloaded for the GlobColour data portal[9]. Figure

[9]http://hermes.acri.fr/ accessed on May 23, 2011

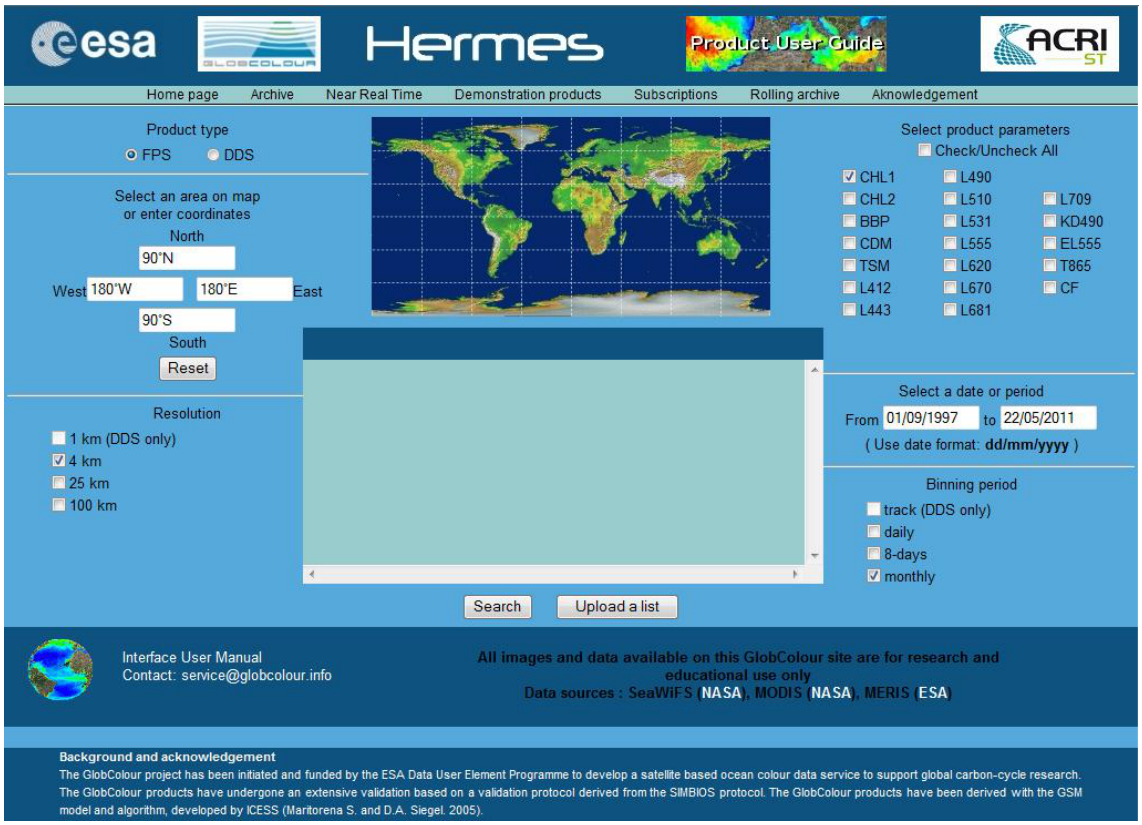

Figure 5.2: Options available on GlobColour data portal

5.2 shows the various options available. Products are provided at varying temporal resolutions - daily, weekly, monthly. Products are also produced at varying spatial resolutions - 1km (DDS only), 4km, 25km, 100km. There are a variety of product parameters available to the user. When searching for available data, the criteria used was - daily 4km Chlorophyll-a data - (CHL1 - chlorophyll-a Case 1 waters and CHL2 - chlorophyll-a Case 2 waters (only available from MERIS)).

These search criteria resulted in a variety of products. The GlobColour merged products are generated using using a variety of merging techniques (simple averaging, weighted averaging, and the GSM (Garver, Siegel, Maritorena) model). The production of the CHL1 derived using the different merging techniques are described in ACRI-ST (2011). CHL2 is only available from the MERIS instrument. A number of products are produced using a varying combination of the three instruments or all three, and differing merging methods. The details can be delineated from the product names.

Based on available files, three data products were selected for use in our analysis, all of which were daily 4km resolution products - the merged MERIS-MODIS-SeaWiFS CHL1 data product using two different merging techniques - weighted average and GSM, and the averaged MERIS CHL2 product.

5.4.1.2 Ifremer Data

CERSAT (French ERS Processing and Archiving Facility) which is part of IFREMER (French Research Institute for Exploitation of the Sea) produces a gridded product for the Atlantic region named *EUR-L4-CHL-ATL-v01*. It produces daily analyses and is a product that is used by the Irish Marine Institute. Again it is a merged MODIS, MERIS, SeaWiFS product. We downloaded this data from the CERSAT ftp site, and then extracted information from the pixels of interest.

5.4.2 SST Data

The SST data chosen for the analysis in this study is from the Global Ocean Data Assimilation Experiment High-resolution Sea-Surface Temperature Pilot Project (GHRSST-PP). An overview of this pilot project can be found in Donlon et al. (2007) or from theGHRSST-PP project page[10]. These data products were chosen since a high resolution diagnostic dataset site (HRDDS) was set up as part of our research.

5.4.2.1 GHRSST-PP Data Products

The GHRSST-L2P product provides all SST data from various agencies and different sensors in a common format, with the addition of ancillary information to assist interpretation. This additional information allows users more flexibility in deciding if a given SST observation is suitable for a particular purpose. It allows the easier evaluation of satellite SST observations against in-situ measurements and

[10]http://www.ghrsst-pp.org

other satellite data, and the easily assimilation of alternative data products into operational forecasting systems, with minimal code change.

The GHRSST-L4 products produce an analysis based on several complementary inputs to provide merged, gridded, and gap-free SST datasets. These products exploit the synergy from using SST from in-situ, satellite microwave and satellite infrared sensors and use all data available in the 24 hour time-period before the analysis. The objective of these products is to provide the best available estimate of SST from a combined analysis of all available L2P and other available sources of SST data. Donlon et al. (2007) note how several L4 production systems are currently operational. These use different analysis methods e.g. Reynolds and Smith (1994), Guan and Kawamura (2004), Murray et al. (1998), and Lorenc (1981). They note that while SST analysis methods predate the initiation of the GHRSST Pilot Project, the availability of L2P data has greatly facilitated their operation and allowed them to exploit more sources of SST.

The High Resolution Diagnostic Data Set (HRDDS)[11] is a service within GHRSST-PP to allow users to interactively compare, analyse and view SST data products, ocean models datasets and auxiliary datasets from the various streams within GHRSST. The HR-DDS system consists of regularly gridded subsets of all available GHRSST SST (Donlon et al., 2007). These are resampled if necessary to a common grid within pre-defined sites. The HR-DDS system examines each GHRSST file (L2P and L4) and an HR-DDS file is produced. These files are subsequently made available via FTP. Approximately 200 of these sites are distributed globally. Two of these sites were established in the context of our work – one at Galway Bay and the other in Dublin Bay. Based on available files, seven of these HRDDS data products from the Galway Bay site were selected for use in our analysis.

[11] http://www.hrdds.net

5.5 Conclusion

In this chapter we investigated satellite remote sensing data products suitable in the context of our research. In our work the focus is specifically on high temporal satellite remote sensing data that can enhance the use of the in-situ sensor nodes in an operational multi-modal sensor network and provide a form of redundancy in the network. The site and the in-situ data sources of interest here are Galway Bay and data from the SmartBay environmental monitoring buoys. The satellite remote sensing parameters that are complementary to data from the SmartBay buoys are divided into two main categories - ocean colour measurements and SST measurements.

We outlined the parameters that can be obtained from ocean colour measurements and the various ocean colour sensors currently being used to capture these parameters, along with some of their characteristics. We also highlighted a number of considerations when choosing an ocean colour sensor for a specific application scenario and the limitations with the use of this type of data in the coastal zone. However, despite the limitations, there are a number of applications where this type of data is of significant benefit.

Similarly we outlined a range of sensors available for providing SST measurements, differentiating between thermal infrared and microwave radiometers. While microwave radiometers provide data during cloud cover, they often do not achieve the resolution of a thermal infrared radiometer. We also outlined a number of applications of SST measurements.

In terms of our investigation into the availability of high temporal data for use in the context of this research. It appears that a single satellite data product does not meet these requirements. This is mainly due to issues with cloud cover. While airborne sensors may provide a method to overcome this, it still does not provide data at the temporal scale required in the context this work. Hence we have chosen to evaluate satellite products that produce a combined analysis based on the integration

of a variety of data sources. Following this we described the products chosen for use in this research and the extraction of data from these. In Chapter 7 we perform a novel analysis where we investigate the use of a trust and reputation framework for determining the most reliable analysis product to be used at a specific moment in time in the network for complementing the in-situ sensors. Similarly to how an integrated approach can prove very beneficial in the context of satellite remote sensing data, in the following chapter we investigate the integration of data sources at the River Lee site for use in our river environmental monitoring network.

Chapter 6

Integration of Data Sources

In the first part of this chapter we investigate the use of heterogeneous information sources to provide context information to allow the more efficient operation of the more sophisticated in-situ analytical instrument in the network. More specifically we present a methodology for the incorporation of rainfall radar imagery data and water depth data into an Artificial Neural Network (ANN) model for predicting average freshwater levels at the River Lee site for potentially controlling the operation of an in-situ phosphate sensor. The main contributions are a methodology for incorporation of pixel information from rainfall radar images and in-situ depth data into an ANN and the subsequent use of this network to predict average freshwater levels at a tidal point of a river. In carrying out this task we also analyse a number of important issues which could be pertinent to future flood monitoring activities.

In the second part of this chapter we are investigating the ability of the network to optimise the use of its in-situ sensor nodes to compensate for the intermittant failure of a node where there is no redundant identical node in the network to replace its operation. More specifically we investigate the development of models incorporating information from different combinations of in-situ sensor nodes for the prediction of values of an alternative node in the network. In this chapter we address research questions 3 and 4 of the research questions outlined in Chapter 1. Overall in this chapter we are investigating the integration a variety of sensing

modalities for optimising an environmental monitoring network. These studies are quite novel and produce very promising results.

6.1 Combining Heterogeneous Sensing and Context Information for More Intelligent sampling

In Chapter 2 we described how the current state of the art in wireless sensor networks poses many challenges for environmental monitoring applications. In particular we discussed how the more sophisticated sensing devices such as chemical and biological sensors are not suitable for large-scale long-term deployments (Diamond et al., 2008a). An intermediate solution may involve investigating how these devices can be used more intelligently and cost-effectively in an environmental sensor network.

This part of our work seeks to investigate the effects of combining rainfall radar and water depth information as a means for providing contextual information to control the operation of a chemical sensor at the Lee Maltings site in the River Lee, Cork. If a chemical sensor such as the one here can sample more intelligently than just at blind regular intervals, then this leads to improved event detection and a longer lifetime for the sensor.

6.1.1 Study Overview

We investigate the use of an Artificial Neural Network (ANN) for predicting fresh water levels at the site for a given day. The ANN incorporates information from the DEPLOY water depth sensor and rainfall data extracted from rainfall radar images provided by the Irish meteorological service. ANN's have been widely used in the literature for modelling various non-linear hydrological processes. Conventional models require a number of parameters such as catchment topography, river network, soil characteristics and antecedent moisture. ANNs can help to by-pass this due to

their ability to generalize results from unseen data and their suitability to modelling dynamic systems on a real-time basis. We provide an analysis on the output of the ANN for a variety of input models, investigating issues such as:

- the most effective way to present rainfall radar information extracted from a digital image – as opposed to raw data values extracted from rain gauges in a catchment – to the network.
- The effects of rainfall from different points of catchment on the model;
- The effect of rainfall and water level information on the model and the effect of differing lag times on the model;
- The accuracy of the ANN in predicting freshwater levels.

The chemical sensor in consideration is a phosphate nutrient analyser developed as part of CLARITY (Slater et al., 2010). If there is heavy rainfall run-off from further up in the river catchment area leading to a subsequent increase in freshwater level at the Lee Maltings, then this type of sensor should increase its sampling frequency. Rainfall and subsequent run-off may indicate the influx of nutrients into the water especially if the catchment area consists of land mainly used for pasture grazing or cultivation (Anderson et al., 2002; Pote et al., 1999). Thus if a significant change in fresh water level can be predicted then the phosphate sensor should be instructed to increase its sampling frequency in anticipation of a possible pollution event. However during periods with little likelihood of phosphate pollution events, the sensor should remain at a lower sampling frequency. This allows the sensor to operate more intelligently and to prolong its lifetime.

6.1.2 Artificial Neural Networks (ANNs)

An Artificial Neural Network (ANN) is a mathematical model that consists of a network of interconnected elements known as neurons. Signals are presented to the ANN through input units which are then propagated and transformed through the

network towards the output neurons(s). Each neuron has a number of input arcs (coming from other neurons or from outside the network) and a number of output arcs. The output of a neuron is based on the weighted sum of all its inputs, that is then transformed by an activation function. The output of a neuron is then propagated to subsequent neurons, and onwards. Depending on the type of network and training algorithm employed, the activation function may be logistic sigmoid, linear threshold, Gaussian or hyperbolic tangent functions, and can introduce non-linear behaviour to the network. Most studies use the logistic sigmoid or hyperbolic tangent functions. (Dawson and Wilby, 2001; de Vos and Rientjes, 2005; Haykin, 1994)

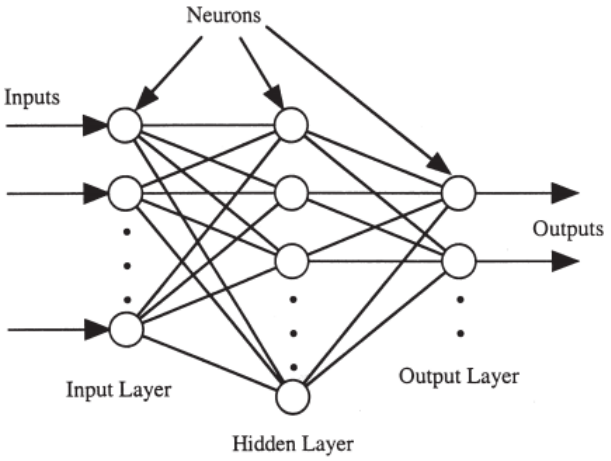

Figure 6.1: Image: (Dawson and Wilby, 2001) The structure of a feed-forward artificial neural network

In feed-forward ANNs, connections flow in one direction between neurons from the input layer, through one or more hidden layers, to an output layer (see Figure 6.1). In the literature two types of feed-forward network are often used in modelling processes similar to rainfall-runoff: the multilayer perceptron (MLP) and the radial basis function network (RBFN) (Dawson and Wilby, 2001). There are many issues that need to be considered and a number of decisions that need to be made in applying ANNs to a problem such as ours. Texts such as Bishop (1995) or Haykin (1999) provide detailed discussions on network types and training algorithms. Comprehensive reviews of the application of ANNs to hydrology have been carried out,

outlining a framework for the development of ANN prediction models in hydrology, along with the options available e.g. (Maier and Dandy, 2000; Dawson and Wilby, 2001). The approach adopted in our work is adapted from the framework outlined by (Dawson and Wilby, 2001) for the application of ANNs to rainfall modelling and flood forecasting.

6.1.3 The use of Artificial Neural Networks in Hydrological Modelling

One of the main research challenges in hydrology is the development of computational models that are able to accurately simulate the response of a catchment to rainfall. These computational models are categorised according to the approach used and de Vos and Rientjes (2005) outline two main categories – knowledge-driven modelling and data-driven modelling. Techniques involved in data-driven modelling are outlined as mainly originating from the field of statistics and artificial intelligence (e.g. time series, empirical regression, fuzzy rule-based systems and ANN modelling), where as knowledge-driven modelling aims to reproduce the real-world hydrological system along with its behaviour in a physically realistic manner. Drawbacks with physically-based models are that they have excessive data requirements, over-parameterisation effects, parameter redundancy effects and large computational demands. Data-driven approaches do not suffer many of the disadvantages associated with knowledge-driven models, however they do have other drawbacks. For example, the range of applications may be limited due to the fact that they are developed from a set of records used for model calibration and thus may not extrapolate well into future situations.

The use of Artificial Neural Networks as a technique has gained significant attention from hydrologists in recent years for modelling water level patterns in a river system. Many authors have highlighted their benefits e.g. (Maier and Dandy, 1996; Thirumalaiah and Deo, 1998; Hsu et al., 1995; Dawson and Wilby, 2001; Dawson

et al., 2002). They do not pre-suppose a detailed understanding of a catchment's physical characteristics or require extensive data preprocessing, and they are also noted to be quite effective for handling incomplete, noisy and ambiguous data. Zealand et al. (1999) highlights their capability for constructing complicated non-linear models for multivariate time-series. They also note issues in relation to the statistical distribution and stationarity of the data. To optimally fit an ARMA-type model to to a time-series, the data must be stationary and follow a normal distribution. On the other hand it is outlined that when developing ANN models, the statistical distribution of the data does not have to be known and that the internal structure of the ANNs implicitly account for non-stationarities in the data, such as trends and seasonal variations. Good generalization capability is also outlined as an advantage of ANNs as unlike ARMA-type models they are relatively insensitive to noisy data and they they have the ability to determine the underlying relationship between model inputs and outputs.

In the literature, ANNs have been demonstrated as a tool capable of modelling various nonlinear hydrological processes. Coulibaly et al. (2000) points to various studies which have demonstrated that they may offer a promising alternative for rainfall-runoff modelling, streamflow prediction, and reservoir in-flow forecasting. They have been demonstrated to outperform traditional statistical models and produce comparable results to conceptual models e.g. (Hsu et al., 1995). Comprehensive reviews on the application of ANNs to hydrology can be found in Govindaraju and Rao (2000) and Maier and Dandy (2000).

6.1.4 Using Artificial Neural Networks for Predicting Changes in Freshwater Levels at the Lee Maltings

In contrast with many other studies in the literature which investigate the use of ANNs in hydrological modelling applications, our objective is not to predict water levels or flow at the site in question, but to predict average freshwater level at a site

which is influenced by the tide and hydroelectric dam further upstream. Predicting average freshwater level for the current day is considered sufficient as it allows the operation of the sensor to be alerted in sufficient time to modify its operating characteristics in order to capture the dynamics of any impending event. The following describes the methodology we used followed by a presentation and discussion of our results which relate to the issues presented in Section 6.1.1.

6.1.5 Methodology

The methodology we applied is adapted from the framework outlined by Dawson and Wilby (2001) for the application of ANNs to rainfall-runoff modelling and flood forecasting. This framework consists of seven stages and the methodology is described under the following headings:

1. Data gathering

Depth data for the Lee Maltings site was gathered from a water depth sensor deployed as part of the DEPLOY project. Data from this sensor is available from April 24, 2009 to June 04, 2010. The sampling rate is approximately once every 10 minutes which results in approximately 144 samples per day. Rainfall data was provided by rainfall radar images downloaded from the Met Éireann web site. These images were gathered from May 15, 2009 until June 30, 2010. They are updated every 15 minutes resulting in approximately 96 images gathered per day. It should be noted that for each of these sensor streams there are some gaps in the data due to issues with the sensor or issues with the data scraping application.

2. Select predictands

As previously outlined, the application of this model requires the prediction of average freshwater level at the site for the current day. Our reasoning behind choosing this as the predictand as opposed to an approximation of freshwater level at a specific moment in time is explained further in the following section.

3. Data pre-processing – stage 1

Dawson and Wilby (2001) outline two steps involved in the data pre-processing stage – data cleansing and the selection of inputs/outputs or predictors/predictands. However certain authors have suggested that extensive data pre-processing does not have to be considered when employing ANNs e.g. (Zealand et al., 1999)) and that it is not considered largely by many studies e.g. (Maier and Dandy, 2000).

Pre-processing of depth data

Initially our approach involved using information from the conductivity sensor deployed at the Lee Maltings site in order to indicate when the tide was out. However this presented a number of difficulties including the issues demonstrated in Figure 6.2. This demonstrates a scenario whereby despite changing water levels with a pattern representative of changes caused by the tide, conductivity values are not changing. This is due to the release of the Iniscarra dam further upstream which is forming a barrier to the entrance of the tidal waters. Figure 6.3 shows trends over a longer time period, demonstrating other issues such as gaps in the data.

We found the most reasonable approach to extracting freshwater levels at the site is to extract the minimum water level from each tidal cycle. Thus for each tidal period the minimum point of that period is extracted as an input to the ANN model. This resulted in approximately two water depths per day since there are generally two tidal cycles occurring within a 24-hour period.

Pre-Processing of Rainfall Radar Images

A system was developed within the CLARITY research centre for the extraction of rainfall data from rainfall radar images for a given catchment. This system also has the ability to carry out short-term rainfall nowcasting or rainfall prediction and is described further in (Wang et al., 2009). The catchment contour for the River Lee is illustrated in Figure 6.4. We divided this catchment area into five strips, each increasingly distant from the point where the river flows into the sea, as shown in

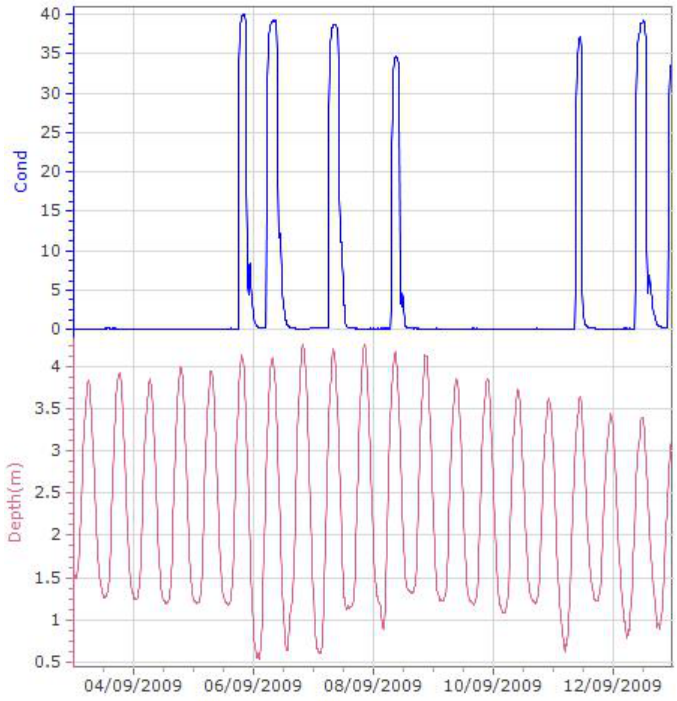

Figure 6.2: A scenario whereby despite changing water levels with a pattern representative of changes caused by the tide, conductivity values are not changing due to the release of water from the Iniscarra dam further upstream

Figure 6.4.

In the rainfall radar images from the Met Éireann website, the map of Ireland is overlaid with coloured pixels indicating five different levels of rainfall intensity namely very light, light, moderate, heavy and very heavy. Each of these types of rainfall/pixel colours are extracted for each of the five strips of the catchment. Processing of the rainfall radar images produces a dataset with five data points for each of the five strips of the catchment for each image. Each data point represents the area of the catchment (in km^2) subject to the type of rainfall in question in that image and each pixel of the catchment represents an area of $1.5km^2$. These data points were also converted into millimetres in order to examine if this provides a better description of rainfall occurring in the catchment for input to the model. Met Éireann provides the rainfall in millimetres per hour for each rainfall type represented in the image and this was used to convert the rainfall value from km^2 to millimetres for each data point.

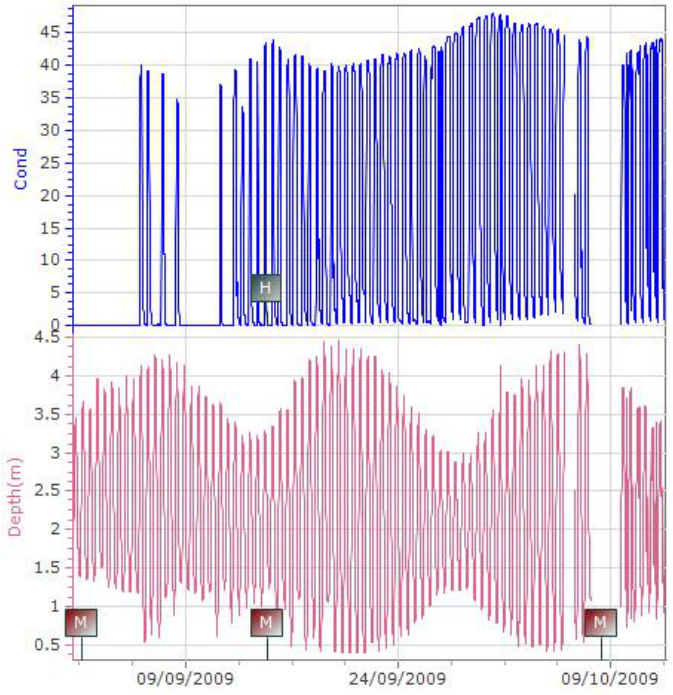

Figure 6.3: Deploy data and conductivity data over a longer time period

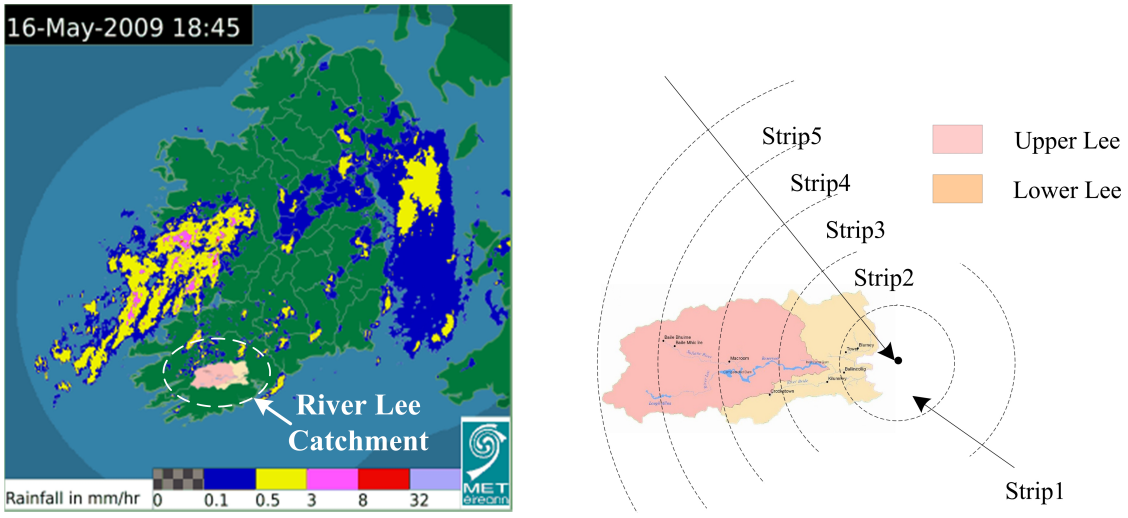

Figure 6.4: Catchment range for River Lee & strips distribution.

This calculation is completed for each of the five rainfall types and aggregated to give a single figure of rainfall in millimetres for that particular strip of the catchment for the image in question.

Aligning the datasets

The rainfall radar dataset is aligned with the water depth dataset for creating a set of instances for input into an ANN. Because of occasional gaps in the rainfall radar data, and the differing sampling rates between rainfall (one every 15 minutes) and freshwater level readings (approximately two per day depending on the tidal cycle), in order to align the two datasets we calculated an average freshwater level and an average rainfall value for each day. Two main categories of data sets were produced:

1. For each day, the average fresh water level + average VLIGHT + average LIGHT + average MODERATE + average HEAVY + average VHEAVY + overall average rainfall in km^2 (average of all rainfall types), per catchment strip;

2. For each day, the average fresh water level + rainfall in mm, per catchment strip.

Selection of input nodes

Maier and Dandy (2000) note how in most ANN applications little attention is given to this task since ANNs belong to the field of data driven approaches which are thought to have the ability to determine which model inputs are critical. However presenting a large number of inputs to an ANN and relying on the network to determine the important model inputs can be problematic (Lachtermacher and Fuller, 1994). Dawson and Wilby (2001) also note the importance of performing some form of data reduction for the input data if it contains many input variables but few points as the model will then have more free parameters to establish than samples to constrain individual parameter values. Suggested techniques include extracting principal components or more simple techniques such as averaging (e.g. averaging data from several rain gauges). In the majority of papers reviewed by Maier and Dandy (2000) input variables were chosen using a priori knowledge. In some cases they were optimised using a variety of techniques such as correlation analysis, prin-

cipal components analysis, a trial and error procedure etc. Appropriate lags were generally using a priori information or a trial-and-error approach or in some papers they were simply chosen arbitrarily. This study is quite novel in that we are investigating the input of pixel information from a colour coded rainfall radar image to the network. Hence a variety of configurations were analysed in order to determine the best input model configuration and to address the other issues outlined in Section 6.1.6.

4. ANN Selection

In this part of the methodology, Dawson and Wilby (2001) outline two tasks – the selection of a network type and a training algorithm. The network type chosen for this work is the multilayer perceptron(MLP), introduced in Section 6.1.2, which comes as part of the WEKA data mining software toolkit[1] (Hall et al., 2009). The MLP provided by Weka can be configured using the associated API or through a user interface and trained and tested using specific feature sets developed by the use for the task in question. The MLP is one of the most popular network types used and it is trained using the error backpropagation algorithm. This training algorithm operates by iteratively changing the network's interconnecting weights so that the overall error is reduced through searching the network's 'weight space' or error function (Dawson and Wilby, 2001). Sigmoidal type functions such as the logistic and hyberbolic tangent functions are the most commonly used transfer functions (Maier and Dandy, 2000). In our work a sigmoid activation function is used.

A number of parameters affect the performance of the training algorithm including step size (Maier and Dandy, 2000). Generally a trial and error approach is used in order to optimise this and it is normally a function of a number of network parameters such as learning rate, momentum, error function, epoch size and the gain of the transfer function (Maier and Dandy, 1998). The epoch size is the

[1]http://www.cs.waikato.ac.nz/ml/weka/

number of training samples presented to the network between weight updates, with the learning rate affecting the size of the steps taken in weight space along with other parameters such as momentum (can be used to help speed convergence to an error minimum as outlined by Dawson and Wilby (2001)), the error function and the type of transfer function used since weight updates are proportional to its derivative (Maier and Dandy, 2000). Dawson and Wilby (2001) mention choosing appropriate values for momentum and learning rate within the range 0.01 and 0.9. Hence we optimised these two parameters which involved evaluating the output of the network for various combinations of these values across a number of hidden node values. Otherwise we used the default parameters in the WEKA toolkit. Following optimisation, we chose a learning rate of 0.1 and a momentum rate of 0.1 as the network appeared to become quite unstable for higher values.

5. Data Preprocessing (stage 2)

The first step in this part of the methodology involves data standardization or normalization. Dawson and Wilby (2001) states that in general, data are rescaled to the intervals [-1,1], [0.1, 0.9] or [0,1]. The Weka toolkit automatically normalises data within the range [-1,1]. The next step is to split the data into training sets and test sets. Training data ranges from May 15, 2009 until January 31, 2010. However due to gaps in datasets, there is a limited number of training instances once the data is aligned and a consecutive number of days of each data source is required for one instance. The training dataset is composed of 129 instances with instances available from the months of May (from May 15 onwards), June, September, December, January and more limited instances then available for July, August, October, November. With limited data availability, a cross-training technique is often adopted (Dawson and Wilby, 2001). 10-fold cross validation is a standard technique used in in the machine learning literature for evaluation of models and this is the technique that is employed in this study. A set of data was also used to test the final model. This data ranged from Feb 1 to June 4 2009. There were less gaps in the data during

this time period and it resulted in 118 test instances. However limited data was a limiting factor in this study.

6. Network Training

This part of the methodology involves specifying the number of hidden layers and the number of nodes in these layers. There exist various approaches for determining an appropriate number of hidden nodes in the network. However certain authors believe the best approach to be via trial and error e.g. (Shamseldin, 1997) and it is one of the most popular approaches adopted (Dawson and Wilby, 2001). We adopt a trial and error approach whereby hidden nodes from 2-50 in steps of 2 are examined using one hidden layer. This range was chosen since initial evaluations demonstrated that over 50 hidden nodes resulted in quite a slow network producing lower correlations. Following the examination of results, we chose to report the results for networks with 2, 12, 22, 32, 42, and 50 hidden node values for each model evaluation. These values can have a large effect on model output.

7. Evaluation

In order to examine the issues outlined correlation values are reported. For evaluation of the best performing models on test data both correlation and mean absolute error (MAE) values are reported. These measurement values are appropriate for the purposes of this study. Other measures that are used in evaluating prediction of flow rates for example are not relevant in this context.

6.1.6 Results and Discussion

Input models consist of rainfall information and water level information. A number of different input models were examined which vary in terms of:

- **rainfall information** – The rainfall information was presented to the network as individual rainfall classes or as an aggregated value across all rainfall classes.

Converting the rainfall for catchments into millimetres was found to have no major impact on network output, hence the results of these models are not presented here.

- **lag times** – Different combinations of lag times were examined with up to 5 days of information presented to the ANN (current day and 4 antecedent days). From the literature this seemed to be an appropriate value and a greater number would have rendered the number of training instances too small as it would have required that number of consecutive days of each parameter to be available throughout the training datasets.

- **catchment strips** – Input models varied in relation to whether they contained information from one catchment strip or a combination of information from all catchment strips. Input models with information from one catchment strip were mainly used to investigate the effect of information from different catchments on the output of the model.

Input Models 1 and 2 – are used where networks are developed separately for each strip of the catchment. This is to examine the effect that rainfall information from each strip of the catchment has on the ANN output. It may become apparent that information from a particular part of the catchment has more of an influence on model output than other strips of the catchment.

- **Input model 1**– Rainfall information from each rainfall type is presented to the network separately for each day considered i.e. VLIGHT, LIGHT, MODERATE, HEAVY, and VHEAVY.

- **Input model 2** – uses averaged information from all rainfall types so one rainfall value is presented to the network for each day.

Input Models 3, 4, 5 and 6 – combine information from all strips of the catchment.

- **Input model 3** – is similar to input model 1 but with information from each strip of the catchment presented to the network for each day considered.

- **Input model 4** – is similar to input model 2 but with this value presented to the network from each strip of the catchment for each day considered.

- **Input model 5** – averages the information for each individual rainfall class across all strips for each day considered.

- **Input model 6** – averages the input values from input model 5 so that one rainfall value is presented to the network for each day considered.

When presenting the results, the number of rainfall days and water level presented to the network is denoted using the format **(rainfall days, water level days)**. For example (3,2) means that 3 days rainfall information and 2 water level information are presented to the ANN. The following presents our analysis in relation to the issues outlined in Section 6.1.1.

6.1.6.1 Effects of Rainfall from Different Parts of the Catchment on Model Output

Rainfall Days	**1**	**2**	**3**	**4**	**5**
Input Model 1	4	4	4	4	4,1,2
Input Model 2	3,2	4,5	4	3,5,4	5,4,3

Table 6.1: Strips of the catchment generally (not always) producing the highest correlation coefficients in predicting freshwater levels at the Lee Maltings site where no water level information is input to the ANN model.

In order to examine if rainfall from specific parts of the catchment have more of an effect than others, for this part of the analysis we use input models 1 and 2 comparing the correlations output by the ANN with a varying number of days of rainfall and water level values for each strip of the catchment. We examined the output for input models with between one and five days of rainfall information

Rainfall Days	**1**	**2**	**3**	**4**	**5**
Input Model 1	4,1	4,1	1,4	1,4	3
Input Model 2	4,5	5,4	5,4	5,4	5,4

Table 6.2: Strips of the catchment generally (not always) producing the highest correlation coefficients in predicting freshwater levels at the Lee Maltings site where 2 days water level information is input to the ANN model

(1-5) and zero or two preceeding days of water level values (0,2). This allows us to examine the effect of rainfall from each strip with and without the additional input of water level information to the ANN. The graphs for this analysis can be found in Appendix C. Figure C.1 shows the results for input models 1 and 2 consisting of 1 day of rainfall information and two different values of water level information (zero or one), Figures C.2, C.3, C.4, C.5 shows the same for input models 1 and 2 with two, three, four and five days of rainfall information. Tables 6.1 and 6.2 summarise these results, highlighting the catchment strips that generally produce the highest correlations for each of the input models across different combinations of the number of rainfall and water level values.

It is clear that **strip 4** appears to be a very dominating strip of the catchment featuring heavily in the results of both input models. For *input model 1* the clearly dominating strips appear to be **1** and **4**. For *input model 2* the clearly dominating strips are **4** and **5**, with **strip 5** generally outperforming **strip 4**. For *input model 1* **strip 4** appears to be the overall best performing strip when no water level information is added to the model. When water level information is added, **strips 4** and **1** dominate the results. For *input model 2* when no water level information is added to the model **strip 3** appears to also dominate the results. However when water level information is added it is clearly **strips 4** and **5** that dominate the results. This is a very exciting and interesting outcome considering the description of the Upper Lee Catchment in Section 3.2 on page 47 outlining a slightly elevated runoff potential due to the peat uplands and steep topography. Overall it is *input model 2* that is generally producing the higher correlation values.

It should be noted here also, that on further analysis, there was a slight change in trend when increased water level information was added to the model most notably from *input model 2*. Figure C.6 demonstrates that when 4 days of water level information is added to the model, it is clear that **strip 3** becomes quite prominent for higher rainfall days in *input model 1* and **strips 2** and **3** becomes quite prominent for higher rainfall days in *input model 2*. Thus this is an area that requires further analysis and investigation, and possibly examined with more training data.

6.1.6.2 The Most Effective Way of Presenting Rainfall Radar Information To The Network

Here we investigate the most effective method to present rainfall radar information to the ANN through examining the correlations output by the various input models. As previously outlined these differ in the manner in which this data is presented to the network. We examine the outputs of the network for combinations of 1, 3, and 5 days of rainfall information and 0 and 2 days of water level information

Input Models 1-2 - strip 4 of the catchment Since **strip 4** seemed to be a dominant strip of the catchment for a variety of scenarios examined in Section 6.1.6.1, we used the results from this part of the catchment for carrying out the analysis here. *Input model 1* presents information from different classes of rainfall separately to the network while *input model 2* averages across these rainfall types resulting in one rainfall value per day as opposed to five values per day with *input model 1*. The graphs displaying these results are shown in Figure C.7 in Appendix C.

Firstly we examine the output of the models when no water level information was provided to the network. For 1 day of rainfall information *input model 1* performs best, producing a maximum correlation value of just over 0.30. However for 3 and 5 days of rainfall information, *input model 2* generally performs best, producing correlation values of just over 0.49 for 3 days of antecedent rainfall and 0.58

for 5 days of rainfall information. Hence it is clear the network is performing better when the rainfall information is summarised before presenting it to the network. This is consistent with the recommendations for data reduction outlined in the step regarding the selection of input nodes when describing the methodology in Section 6.1.5. When water level information is added to the model, for 3 and 5 days of rainfall information *input model 2* performs with a higher correlation, with a maximum correlation value achieved with 3 days water level information (0.93). Hence again it appears the network is performing better when the rainfall information is summarised.

Input models 3-6 - combination of all strips *Input models* $3-6$ which incorporated information from all strips of the catchment, were also compared over these different combinations of rainfall and water level information. The graphs displaying these results are shown in Figure C.8 in Appendix C. When no water level information is input to the network, it appears that *input model 3* is generally producing the lowest correlations and *input model 6* is generally producing the highest correlation values, with *input model 4* also producing similar correlations for 5 days rainfall information. When water level information is added to the network, *input model 5* outperforms other models for 1 day of rainfall information, producing a correlation coefficient of just over 0.93. However when 3 or 5 days of rainfall information is added to the network, this model performs poorer than most. *Input model 3* is generally the poorest performing overall. For 3 and 5 days of rainfall information, *input model 6* overall generally performs with the highest correlation values, reaching just under 0.92 for 3 days rainfall information and just under 0.91 for 5 days rainfall information. Similar trends could be seen with 4 days of water level information, and thus are not outlined here.

What we learn from all this it that it is clear that applying many rainfall values to the network, where a value for each individual rainfall type for each strip of the catchment is presented to the network, generally results in poor correlation

coefficents and aggregating these values in some manner appears to improve results. Following this, our subsequent analysis proceeds using input model 2 for analyses relating to individual strips of the catchment and input model 6 for analyses relating to all strips of the catchment.

6.1.6.3 The Effect of Rainfall and Water Level Information on the Model

Here we investigate if rainfall information actually has any impact on the network outputs or if it is only water level information influencing outputs. The rainfall radar information can be quite noisy and thus it is interesting to see if it provides any benefit to the network. We also investigate whether there are any particular combinations of lagtimes between the two data sources that appear to increase performance. In conducting this investigation we look at the outputs of *input model 2* using rainfall data from **strip 5** of the catchment, as this catchment area seemed to generally produce the highest correlation values in the analysis carried out in Section 6.1.6.1, and *input model 6* which incorporates rainfall information from all areas of the catchment. The graphs showing these results are shown in Figures C.9 and C.10 in Appendix C.

Firstly while input models consisting of no water level information do not produce very high correlation coefficients, it is clear the impact that increased rainfall information has on the output of the network as the greater number of days of rainfall input into the model, the higher the correlation coefficient tends to be. *Input model 2* reaches a correlation of just over 0.61 with no water level information and 5 rainfall values, and when water level information is added to the model another trend becomes apparent, namely that 2 and 3 days of rainfall information seem to generally produce the highest correlations (produce the highest correlation of 0.938). Adding an increasing amount of water level information to the model does not seem to effect model output hugely. Similarly for *input model 6* while input models consisting of no water level information do not produce very high correlation coefficients, it is clear the impact that increased rainfall information has on the

output of the network. Again the addition of water level information after one day does not appear to have a huge impact on the output of the network. For this input model 3 days of rainfall information and 3 days of water level information produces the highest correlation value of 0.9337.

6.1.7 Performance of the ANN for Predicting Average Water Level

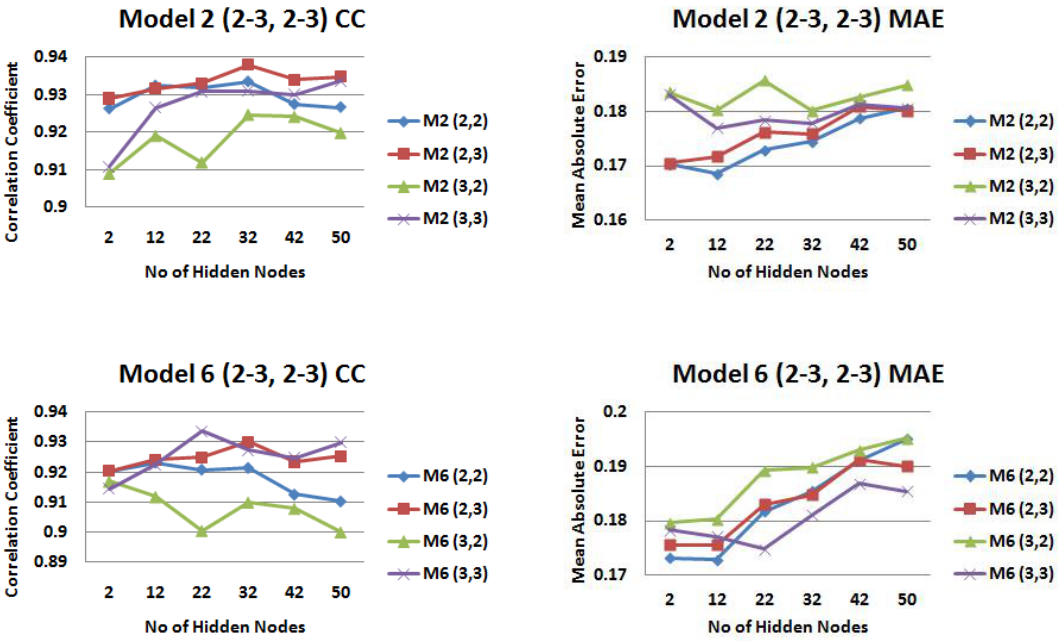

Figure 6.5: Correlation coefficients and mean absolute error values for models 2 and 6 with different combinations of 2 and 3 days rainfall and water level information.

From section 6.1.6.3, we see that quite high correlations can be achieved with the use of an ANN. Based on this analysis, we have chose to concentrate on input models with 2 days rainfall information and 2 and 3 days water level information and 3 days rainfall information and 2 and 3 days water level information for examining the overall performance of the ANN for predicting average freshwater levels since these configurations seemed to be performing well in the previous analysis.

Figure 6.5 shows the correlation (CC) and mean absolute error (MAE) values

of again predicted versus actual outputs, for *input models 2* (strip 5) and *6* with combinations of 2 and 3 days rainfall and water level information. For *input model 2*, an input model consisting of 2 days rainfall information and 2 days water level information (2,2) generally produces the lowest MAE, reaching a lowest value of just under 0.17. For *input model 6*, apart from small hidden node values, the MAE is generally lowest for an input model consisting of 3 days rainfall information and 3 days water level information (3,3), reaching a lowest MAE of 0.1748. For *input model 2*, models with 2 days rainfall information are generally producing the highest correlations with a highest correlation value of 0.938 (2,3). For *input model 6*, models with 3 days water level information are generally producing the highest correlations with model (3,3) producing the highest correlation value of 0.9337. As shown in Table 6.3, apart from during a event in November where the in-situ sensors went offline, the range of daily average freshwater levels was from 0.35 metres to 2.48 metres in the training set, hence the performance being produced is very satisfactory, with error values of just under 0.17 being reached.

	Max	**Min**	**Avg.**	**Stdev**
Training Data	4.1*	0.34765	1.0953	0.5641
All Test Data	1.8937	0.51315	0.8068	0.2439
Feb	1.8937	0.68405	0.9547	0.3018
Mar	1.3635	0.6722	0.8336	0.2132
Apr	1.6945	0.51315	0.2508	1.0754
May	0.9366	0.54325	0.653	0.0818

Table 6.3: Statistical properties of the training and test data for data for average freshwater level, * This value is 2.48 not including the 3 daily averages for the 3 flood days

Thus for testing our models we chose to use an input model with 2 days rainfall and water level information for *input model 2* and an an input model with 3 days rainfall and water level information for *input model 6*. These models were tested using data from February 1 2009 to June 4 2009, resulting in 118 test instances. From Figure 6.6, it is apparent that the input models are performing with a lower CC and MAE on the test data. *Input model 2* produces a CC between 0.7572 and 0.7653 and a MAE between 0.0967 and 0.1144. *Input model 6* produces a CC

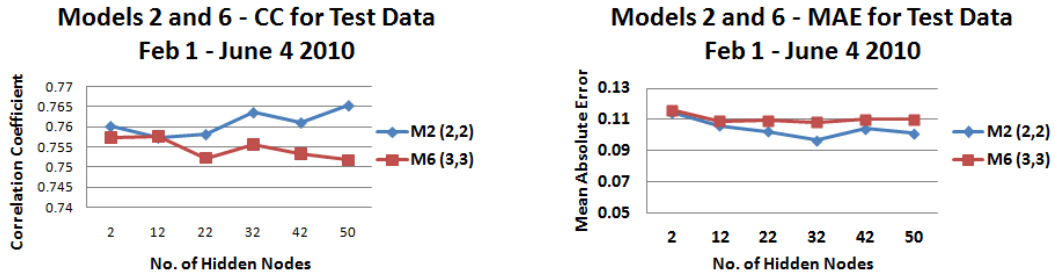

Figure 6.6: Correlation coefficients and mean absolute errors for input models 2 (2,2) and 6 (3,3) when tested on data from Feb 1 to June 4 2009

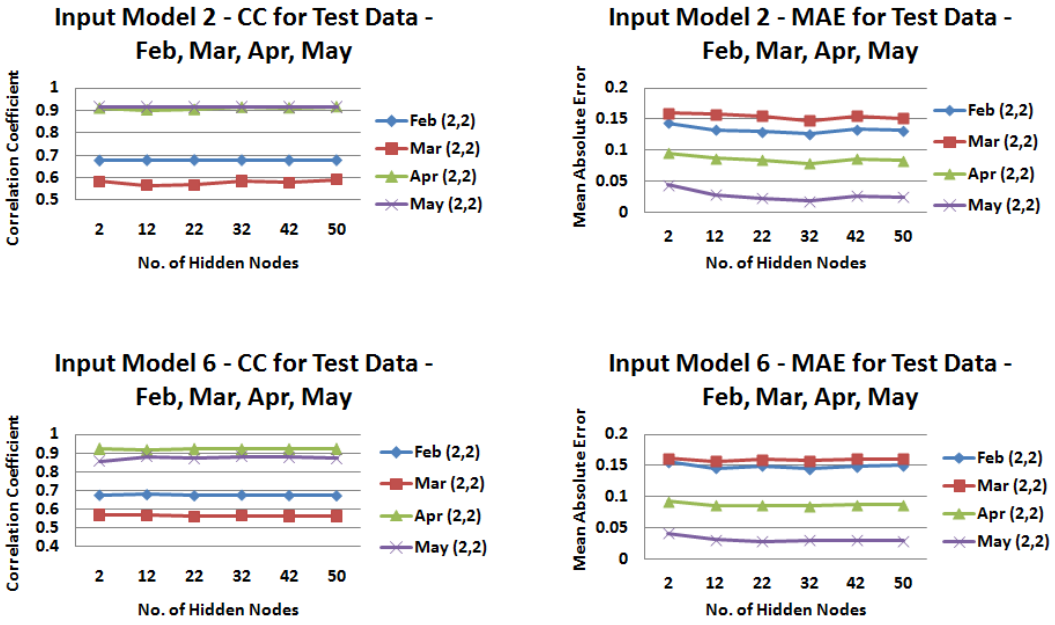

Figure 6.7: Correlation coefficients and mean absolute errors for input models 2 (2,2) and 6 (3,3) when tested on data from individual months - February, March, April, May

between 0.7517 and 0.7576 and a MAE between 0.1084 and 0.1161. As shown in Table 6.3, the range of daily average freshwater levels was from 0.51 metres to 1.89 metres in the test set. The range is smaller than that for the training set, which may explain the reduced correlation but lower error values. However overall the error ranges described above are again very satisfactory for the application context in question.

In order to investigate the reduced performance in terms of correlation between

the actual values and predicted output, the test set was broken down into its component months - February, March, April, May. June was omitted since there were only 4 days data available for this period. Each of these test sets are limited to 18 instances, hence more data would probably be needed in order to carry out a more thorough investigation. These results can be seen in Figure 6.7.

It is unclear what is causing the reduction in performance in February and March. It may be due to operations of the dam further up stream whose operation cannot be predicted or accounted for in the model. It also may be due to lack of training data. There is a limited amount of training data available to the model and none for the period of February and March. However it should be noted that this is also a very limited test set with only 18 instances per month. Despite the reduced performance, the application of this model is to predict a sufficient change in average fresh water level to initiate a change in the sampling rate of a sensor and the MAE's produced by this model are sufficient to meet this criteria. That particular application does not require precise measurements of freshwater level. The statistical properties of the average freshwater level data from which each of the datasets are drawn are outlined in Table 6.3.

6.2 Using Heterogeneous Sensor Nodes to Provide Data Redundancy in the Network

In this part of our work we evaluate whether a set of heterogeneous in-situ sensors can be modelled to predict the value of another in-situ sensor node in a network, thus providing a form of data redundancy. We now describe the data-filling models we developed for the Lee Maltings site.

As previously outlined, sensors deployed as part of the DEPLOY project provide in-situ real-time monitoring of conditions at the Lee Maltings site. Like most in-situ sensor networks in a marine environment, these sensors can produce noisy data and are subject to failure. The deployment of multiple sensors to provide redundancy

in the network is not an option due to budgetary constraints and often there is a limitation to the types of sensors that can be deployed at each site. As outlined in Chapter 3, the range of sensors deployed at the Lee Maltings site included sensors for monitoring conductivity, chlorophyll-a-fluorescence, dissolved oxygen, temperature, and water depth.

Because of the complex processes in operation at the Lee Maltings site, the site is predominantly influenced by the effect of the tide and the release of water from the Iniscarra dam further upstream. Hence there are a number of dynamic processes in operation which are difficult to account for and to model. However, notwithstanding that, our aim is to investigate how far an environmental sensing network can get by using simple input model scenarios with limited data. It is by no means suggesting that such models with such simple assumptions can account for every scenario at the site or could replace the activities of a sensor, It is merely an investigation into how far we can get with limited data resources, into replicating the activity of a sensor while there may be a possible gap in the data or a fault in the network. For predicting water quality parameters such as dissolved oxygen, many studies in the literature use very long datasets e.g. (Shaghaghian, 2010) or complex input parameters (Areerachakul et al., 2011). The aim of our work is to investigate what can be achieved from using current (or a limited number of preceding) values of other simple water quality parameters in the network, without any complex data pre-processing or complex input models.

6.2.1 Methodology

The following describes the various steps of the methodology involved in developing models for the prediction of in-situ parameters.

6.2.1.1 Data

DEPLOY data is available from April 24, 2009 until May 30, 2010. A range of training data was aggregated over this time period for training the models including

Month	Weeks
May 2009	3, 4
June 2009	1, 3, 4
July 2009	2, 3, 4
August 2009	1, 3, 4
September 2009	2, 3, 4
October 2009	1, 3, 4
November 2009	2, 3, 4
December 2009	1, 3, 4
January 2010	2, 3, 4
February 2010	1, 3, 4
March 2010	2, 3, 4
April 2010	1, 3, 4
May 2010	3, 4

Table 6.4: DEPLOY data from which a training set was built for model development

data from the weeks outlined in Table 6.4. The sampling rate of the DEPLOY sensors at the Lee Maltings site is approximately once every 10 minutes or approximately 144 samples per day and 1008 samples per week when operating correctly. Thirty-seven weeks are set aside for training (as outlined in Table 6.4) resulting in over $37,000$ data points in the training dataset for each of the five sensors in operation at the site, though as we see in the next section, we used only four of these five.

6.2.1.2 Feature Sets

For the Lee Maltings site, four parameters were chosen for use in the input models namely conductivity, dissolved oxygen, depth and temperature. Water depth was chosen since it is a parameter that is also being monitoring by our on-site camera and it is an influential parameter because of the tide which affects other water quality parameters like conductivity, which was also included. There are well known relationships between temperature and dissolved oxygen (e.g. (Weiss, 1970)), so these were included also. Chlorophyll-a-fluorescence has a more complex relationship with the other parameters i.e. the growth of algae in the water depends on, for example temperature and dissolved oxygen over a long period of time and our models

	Cond Prediction	**Depth Prediction**	**DO Prediction**	**Temp Prediction**
M1	Depth	Cond	Cond	Cond
M2	DO	DO	Depth	DO
M3	Temp	Temp	Temp	Depth
M4	Depth-DO	Cond-DO	Cond-Depth	Cond-DO
M5	Depth-Temp	Cond-Temp	Cond-Temp	Cond-Depth
M6	DO-Temp	DO-Temp	Depth-Temp	DO-Depth
M7	Depth-Temp-DO	Cond-Temp-DO	Cond-Depth-Temp	Cond-Depth-DO

Table 6.5: Input models evaluation for predicting values of the four parameters. These models are analysed using 1, 5, 10, and 20 values of the input parameters to the model.

for data gap-filling are more short term, so we did not include chlorophyll.

Input models for the prediction of each parameter are created using values from one, two or three of the other parameters. Feature sets are created investigating the use of one alternative value for the prediction of a parameter, a combination of two alternative values for the prediction of a parameter, and a combination of three alternative values for the prediction of a value. These input models are then investigated again by including not only one value from the parameter(s) being used in the model but 5 values, 10 values and 20 values, in order to examine their effect on performance. The input models evaluated for the prediction of each of the four parameters are outlined in Table 6.5. Ten thousand instances of each feature set were created from the training set outlined in Section 6.2.1.1 on page 158. Instances were created using data from across the training set so that the models are representative of changes that may occur with the time of year.

6.2.1.3 Numerical Model

The Weka Data Mining software (Hall et al., 2009) was again used for carrying out numerical prediction tasks. At the initial stages of this study a number of numerical prediction models were evaluated for use. Due to the powerful nature of the ANN and its ability to capture any non-linear relationships in the data, the use of an

MLP was examined. Support Vector Machines (SVM), another form of machine learning, are also known to be a powerful tool for numerical prediction, hence an SVM regression model using both a linear and a radial basis function (RBF) kernel were examined (using the LibSVM[2] toolkit). Simpler numerical prediction models such as linear regression were also examined. However none of these numerical prediction tools performed as high as that of a regression tree for this particular scenario. A selection of decision trees were examined, however it was the REPTree in the WEKA toolkit that produced the best performance. REPTree builds a regression (or decision) tree using information gain/variance reductions. It prunes the tree using reduced-error pruning. It only sorts numeric attributes values once and it is optimised for speed (Witten et al., 2011). The standard settings for parameters in the WEKA toolkit were used. Thus the minimum total weight of the instances in a leaf was set to 2, the maximum tree depth was set to -1 (i.e. no restriction), the minVarianceProp was set to 0.0010 (i.e. the minimum proportion of the variance on all the data that needs to be present at a node in order for splitting to be performed), the number of folds used for pruning was set to 3, and a seed of 1 was used to randomise the data. Decision trees have been used in varying contexts in the literature e.g for determining the limitation factors for phytoplankton (East, 2006), for the prediction of multiple chemical parameters of river water using biological information (Blockeel et al., 1999), for the classification of water quality classes in canals Areerachakul and Sanguansintukul (2010) etc. This research examines their use in an alternative application context.

In order to boost performance we investigated a machine learning approach which combines the output of several different models. We investigated a number of ensemble learning approaches but settled on a meta-classification scheme known as bagging, which appeared to outperform other approaches. As outlined in Hall et al. (2009) the termed bagging for "bootstrap aggregating" was introduced by Breiman who investigated the approach for classification and numeric prediction. For infor-

[2] *http : //www.csie.ntu.edu.tw/ cjlin/libsvm/*

mation on this approach see Breiman (1996). The default parameter settings in the WEKA toolkit were used. The size of each bag as a percentage of the training set size is set to 100, the *calcOutofBag* (i.e. whether the out-of-bag error is calculated) is set to false, the base classifier is set to REPTree with the parameters outlined above. The number of iterations is set to 10 and the random number seed is set to 1.

6.2.2 Results

The following describes the results obtained for the prediction of the four parameters The evaluation was carried out on the datasets of 10,000 instances described in Section 6.2.1.2 using 10-fold cross validation.

6.2.2.1 One Parameter Prediction

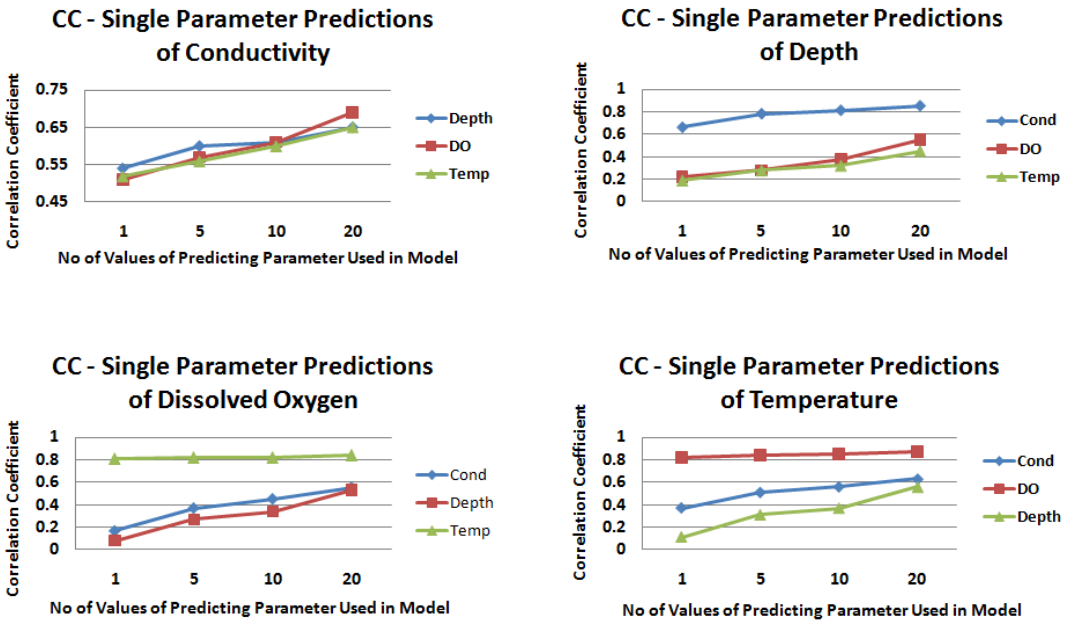

Figure 6.8: Single parameter prediction: Correlation coefficients for models consisting of information from one water quality parameter for the prediction of another with a varying number of input values (1, 5, 10, 20)

Figure 6.8 shows the correlation coefficients (CCs) produced by input models

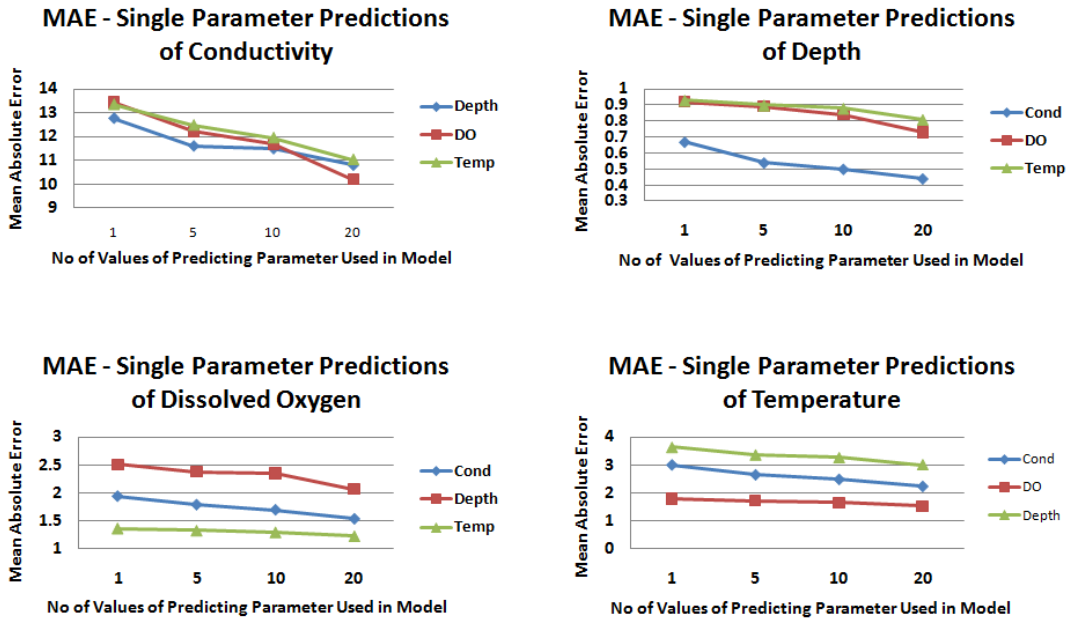

Figure 6.9: Single parameter prediction: Mean absolute errors for models consisting of information from one water quality parameter for the prediction of another with a varying number of input values (1, 5, 10, 20)

consisting of one parameter with a varying number of input values (1,5,10,20). Figure 6.9 shows the mean absolute errors (MAEs).

Conductivity: For the prediction of conductivity, there is no single parameter that clearly outperforms another. When 1 or 5 preceding values are input to the model, depth performs with the highest correlation and lowest mean absolute error ($1 - CC = 0.54, MAE = 12.76, 5 - CC = 0.6, MAE = 11.6$). However when 10 values are input to the model, dissolved oxygen (DO) produces a similar correlation and error to depth in predicting conductivity. With 20 input values DO produces a higher correlation and lower error than the other input parameters (20- $CC = 0.69, MAE = 10.19$). Adding additional values to the model appears to improve performance in all cases.

Depth: Conductivity clearly outperforms all other parameters for single parameter prediction of depth, with performance improving the more values that are added to the model. Correlation reaches 0.85 when we input 20 preceding data values into

the model. $(1 - CC = 0.66, MAE = 0.67, 20 - CC = 0.85, MAE = 0.44)$.

Dissolved Oxygen: Temperature clearly outperforms all other parameters for single parameter prediction of dissolved oxygen, with performance improving the more values that are added to the model $(1 - CC = 0.81, MAE = 1.37, 20 - CC = 0.84, MAE = 1.23)$. When using only one temperature value to predict DO, a CC of 0.81 and a MAE of 1.37 can be achieved.

Temperature: Dissolved Oxygen clearly outperforms all other parameters for single parameter prediction of temperature, with performance slightly improving the more values that are added to the model $(1 - CC = 0.82, MAE = 1.8, 20 - CC = 0.87, MAE = 1.53)$. Similarly to the case for DO prediction, these results are quite good for using only one parameter to predict temperature. Even when only inputting one value to the model it is producing a CC of 0.82 and a MAE of 1.53. This clearly reflects the well known relationship between these parameters.

6.2.2.2 Two- and Three-Parameter Prediction

Figure 6.10 shows the correlation coefficients (CCs) produced by input models consisting of two or three parameter with a varying number of input values (1, 5, 10, 20). Figure 6.11 shows the mean absolute errors (MAEs).

Conductivity: All three parameters combined (Depth, Temp, DO) produces the highest correlation and lowest error. Additional values to the model improve performance with 1 value producing a correlation of 0.87 and an error of 5.55 and 20 preceding values improving this to a correlation of 0.91 and an error of 4.62. The best two-parameter model generally appears to be the combination of depth and DO, again performance increasing with the number of values added to the model $(1 - CC = 0.75, MAE = 8.47, 20 - CC = 0.85, MAE = 6.17)$.

Depth: Again, all three parameters combined (Cond, Temp, DO) produces the highest correlation and lowest error $(1 - CC = 0.79, MAE = 0.52, 20 - CC = 0.92, MAE = 0.33)$. Both two-parameter models incorporating conductivity seem to perform similarly with conductivity and DO outperforming for higher numbers

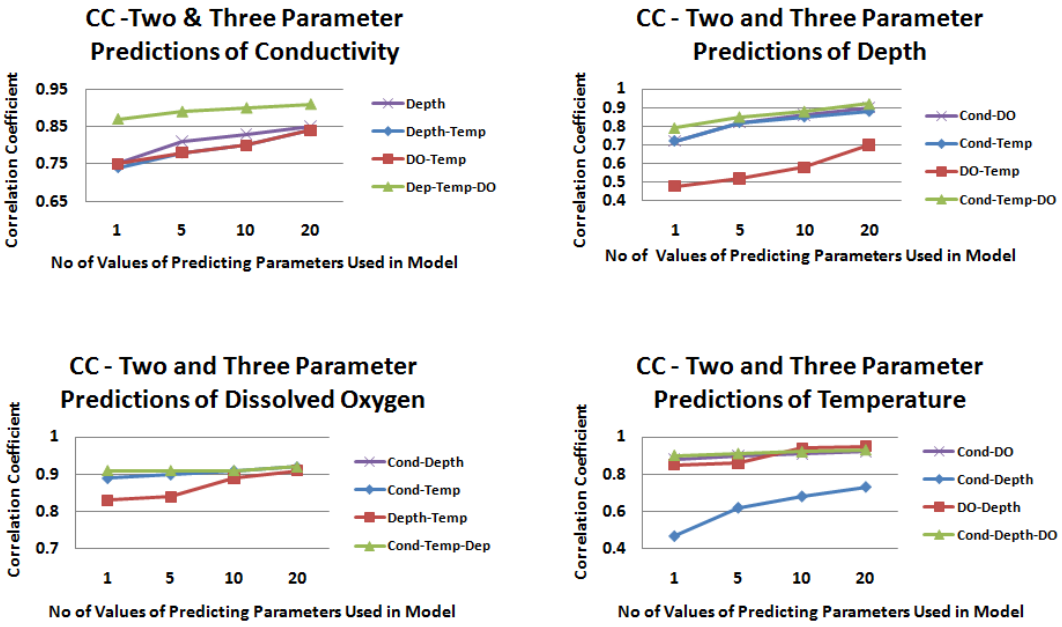

Figure 6.10: Two- and three-parameter prediction: Correlation coefficients for models consisting of information from a combination of two or three water quality parameters for the prediction of another with a varying number of input values (1, 5, 10, 20).

of input values ($1 - CC = 0.72, MAE = 0.6, 20 - CC = 0.9, 20 = 0.35$).

Dissolved Oxygen: For predicting missing values of dissolved oxygen, input models consisting of temperature data way outperform the one model not containing temperature information (Cond, Depth). When 1 and 5 values are input to the model, the three-parameter prediction (Cond, Temp, Depth) performs with the highest correlation between predicted and observed and the lowest error ($1 - CC = 0.91, MAE = 0.75, 5 - CC = 0.92, MAE = 0.68$). A two-parameter prediction model consisting of conductivity and temperature can reach similar performance figures when 10 and 20 values are input to the model ($10 - CC = 0.91, MAE = 0.76, 20 - CC = 0.92, MAE = 0.68$).

Temperature: When a higher number of values is input to the model, a two-parameter combination of DO and depth produces the highest correlation between predicted and observed (CC= $10 - 0.94, 20 - 0.95$) and the lowest error (MAE= $10 - 1.1, 20 - 0.99$) for predicting temperature. Although when 20 values are input

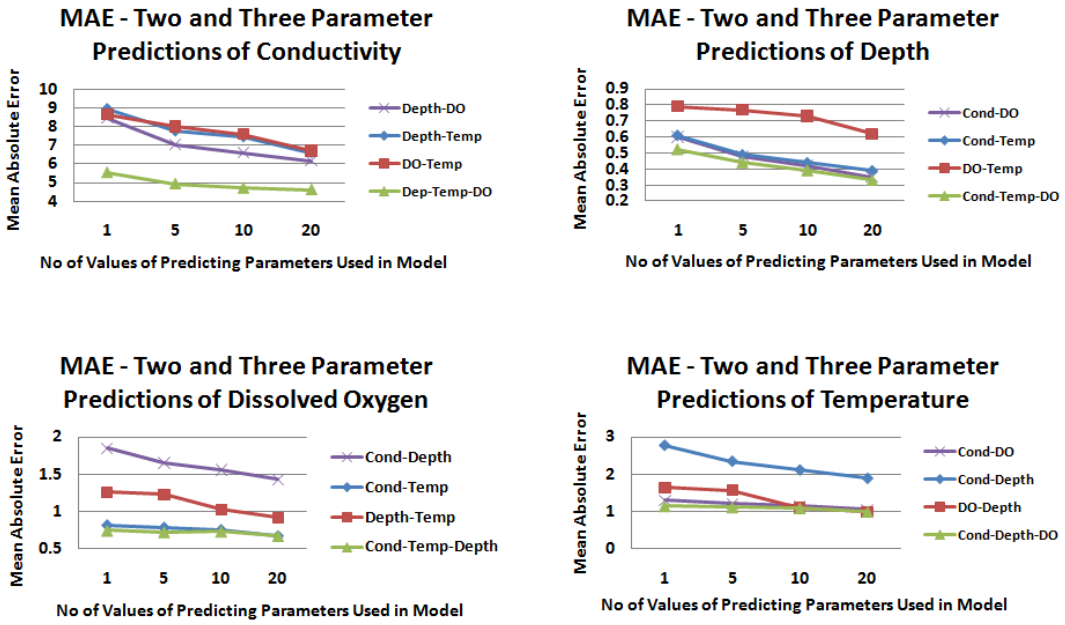

Figure 6.11: Two- and three-parameter prediction: Mean absolute errors for models consisting of information from two or three water quality parameters for the prediction of another with a varying number of input values (1, 5, 10, 20).

to the model, the three-parameter combination (Cond, Depth, DO) matches the error of depth and DO. For input values of 1 and 5, conductivity and DO produces similar correlation and error values to the three parameter combination - *Depth-DO* ($1 - CC = 0.88, MAE = 1.3, 5 - CC = 0.9, MAE = 1.21$) and *Cond-Depth-DO* ($1 - CC = 0.9, MAE = 1.15, 5 - CC = 0.91, MAE = 1.12$).

6.2.3 Test Data

The weeks of data not used in the training data were tested on a selection of the models evaluated above. Since the models incorporating three parameters for prediction generally performed the best in terms of CC and MAE, these models were chosen for evaluation here. 3,000 instances were selected across the test set and evaluated on the model. These instances were The resulting correlations and MAE's can be seen in Figure 6.12 and summarised as follows:

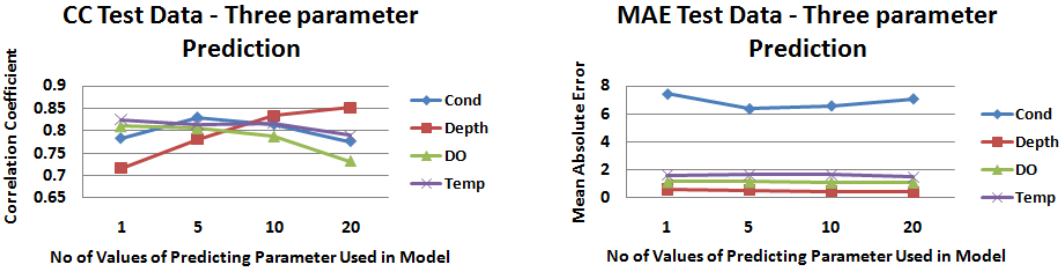

Figure 6.12: Three-parameter prediction: Correlation coefficients and mean absolute errors for models consisting of information from three water quality parameters for the prediction of another with a varying number of input values (1, 5, 10, 20), tested on alternative data to that used in the training set.

Conductivity	**Max**	**Min**	**Avg.**	**Stdev**
Training Data	48.02	-0.0780	19.9488	18.9940
Test Data	45.24	-0.07802	20.1514	18.9793
Depth				
Training Data	5.28	0.3412	2.3209	1.098
Test Data	6.916	0.4036	2.2531	1.0754
D.O.				
Training Data	16.64	1.30	8.08730	3.03992
Test Data	15.98	1.01	7.99327	3.05101
Temp.				
Training Data	21.17	2.86	12.0468	4.1227
Test Data	19.44	3.31	11.6904	3.8775

Table 6.6: Statistical properties of the training and test data for data from the Lee Maltings site

- The prediction of *conductivity* using depth, dissolved oxygen and temperature information produces a CC of between 0.7757 and 0.8294 and a MAE between 6.3597 and 7.4311. Increasing the number of input values to the model does not necessarily result in an increase in performance, with models with 5 and 10 values performing the best.

- The prediction of *depth* using conductivity, dissolved oxygen and temperature

information produces a CC of between 0.7168 and 0.8518 and a MAE between 0.4309 and 0.6258. The output of this model appears to benefit from an increased number of input values.

- The prediction of *dissolved oxygen* using conductivity, depth and temperature information produces a CC of between 0.7323 and 0.8104 and a MAE between 1.1073 and 1.164. An increased number of values does not necessarily improve performance in terms of correlation, however the highest number of input values does marginally result in the lowest MAE.

- The prediction of *temperature* using conductivity, depth and dissolved oxygen produces a CC between 0.7897 and 0.825 and a MAE between 1.5052 and 1.6706. An number of values does not necessarily improve performance in terms of correlation or MAE, however the highest number of input values does marginally result in the lowest MAE.

Thus while there is a reduction in performance compared to correlations and MAE's achieved on the training set, this performance is very promising for estimating data at the site when a sensor goes offline or data is missing. The reasons for a reduction in performance may be due to over-fitting of the model, resulting in a reduction in generalisation ability. Also as we know, the Lee Maltings site is an extremely difficult site to monitor with the dynamics of the tide and the release of the dam further upstream. Furthermore there may be irregularities in the training or test data that cannot be accounted for along with gaps in the data which are all real-world problems with all environmental datasets. The statistical characteristics of the training data and test data are outlined in Table 6.6. If such a model were applied to a river system without these influences, it may indeed be very powerful considering the performance that is being achieved here. However the performance obtained here demonstrates the powerful role these models may play in estimating missing data in a low cost manner with limited data availability and simplified input model scenarios that to not require expensive data sources.

What we can conclude from this analysis is that if our model were used in a prediction scenario, it is sufficient to estimate missing data values at the site without providing precise readings. Considering the limited data and simplified notions of dynamics from the site which we have incorporated into the model, its performance is quite good, and may provide a powerful tool in a environmental monitoring network. Post-processing on model output might further improve performance, for example by examining a window of model outputs. However this model would also need to be evaluated for the time of year in question before its output was adopted in a network. In Chapter 7 we attempt to further optimise the use of these models in a network using a novel and unique approach. We also examine their performance during different times of year in carrying out this task.

6.3 Conclusion

In the first part of this chapter we investigated a methodology for the use of rainfall radar information and in-situ depth data to predict average freshwater levels at the Lee Malting for potentially controlling the operation of an in-situ phosphate sensor. We investigated a variety of issues, some of which are novel in relation to the use of rainfall radar data as an input data stream to an ANN for this purpose. Despite the noisy data, it appears that the rainfall radar information does have an impact on the network and very interestingly it is determining areas of the catchment corresponding to those implied in Chapter 3 as influencing runoff more than others which is a very promising outcome. The evaluation of the model on test data produced a reduced performance for certain time periods but overall the performance of the model on test data is very satisfactory for the application context and can be used for determining a change in the sampling rate of a sensor. With more training data, it may be worthwhile developing seasonal models and evaluating the issues here further. Further data may also help to investigate the particular events for which the models do not seem to be able to capture along with those that they

respond well to. However overall our study demonstrates that with limited training data, a system for controlling the sampling rate of the nutrient sensor can be set up quickly and cost effectively at a deployment and can improve the efficiency of the more sophisticated nodes of the sensor network. It also has the potential of providing the basis for an early warning system for a flood prone region.

We also used data from the Lee Maltings site to explore whether we could train a model to learn missing data values, an inherent characteristic of wireless sensor networks caused by sensor or network problems. We focused on data values from conductivity, depth, dissolved oxygen and temperature sensors, also varying the amount of preceding data used to train the model. Our results found that each of the four sensors we chose can have missing data predicted using readings from the other 3 sensors, with correlations between predicted and observed varying from 0.7168 (water depth missing) to 0.8518 (also water depth missing) and small error rates. These results statistically exploit the dependencies that do exist among these characteristics of a river, where, for example, the levels of dissolved oxygen depend on water temperature and also water depth which in turn is related to conductivity which varies as the tide comes in/out. It it is very promising that we can exploit these relationships to increase the effectiveness of our environmental monitoring network. Considering the difficulties in modelling the dynamics at the Lee Maltings site, the performance obtained here demonstrates the powerful role these models may play in estimating missing data in a low cost manner with limited data availability and simplified input model scenarios that to not require expensive data sources. However from these results it is apparent that some of these models are more effective than others. With many producing very promising results, we need to somehow explore how they can be used most effectively in the network. In the next chapter we investigate a novel approach for optimising the use of these models in the network through adapting a trust and reputation framework to chose the most reliable model at a particular point in time.

Chapter 7

Application of a reputation and trust based framework to a multi-modal network

In the preceding chapters we introduced a number of data streams which can be incorporated into a multi-modal sensor network for marine environmental monitoring. These data streams range from in-situ sensor networks incorporating heterogeneous sensors monitoring a variety of parameters, to visual sensors covering cameras, satellite imagers and rainfall radar images.

In Chapter 4, we modelled the appearance of certain features in the camera images from the Lee Maltings site that can provide us with an estimation of depth. In Chapter 5, we outlined our ideal requirements of a satellite remote sensing data stream with high accuracy at high temporal and spatial scales for near-realtime cooperation with other sensor streams in the network. The difficulties in achieving this are acknowledged and the extraction of information from satellite remote sensing data most suited to the needs of this work are subsequently outlined. In Chapter 6, we developed a model incorporating rainfall radar information and water depth information for controlling the operation of a chemical sensor at the Lee Maltings. We also modelled data streams from heterogeneous in-situ sensor nodes and we used

these models for predicting values of one of the *missing* in-situ nodes in the network for times when it becomes unavailable. The extraction of information from the visual data streams along with these models represents additional sensor streams for use in our multi-modal sensor network.

However it is apparent from our analysis that each of these sensor streams are individually unreliable as is the nature of a physical sensor. Our research proposes that a reputation and trust-based model applied to a multi-modal environmental sensor network can help to deal with the unreliability associated with these additional sensor streams and optimise their benefits to the network. Our research proposes to adapt the model outlined in Ganeriwal et al. (2008) for use in a multi-modal sensor network. This model was described in Chapter 2 and the following describes how we adapt this model for use in the context of our research. The analysis presented in this chapter relates back to research question 5 from Chapter 1.

7.1 Differences in applying RFSN to a multi-modal sensor network

In Chapter 2, we described how Ganeriwal et al. (2008) investigates a generalised and unified approach for providing information about the data accuracy in sensor networks through allowing sensor nodes to establish a community of trust. They developed a Reputation-Based Framework for Sensor Networks (RFSN) where each node develops a reputation for other nodes in the neighbourhood through assessing their actions. They subsequently used this reputation value for each of the nodes to evaluate their trustworthiness when making decisions within the network. The authors outline how RFSN is not fundamentally different from the reputation systems outlined in the domain of ad-hoc networks i.e. (Buchegger and Boudec, 2002, 2003a,b; Michiardi and Molva, 2002). Their two main contributions are the applicability of such a reputation system to sensor networks and the development of a middleware service that can counter the faulty misbehaviour of nodes caused by

errors in data.

In this chapter, we adapt the RFSN system for use in a multi-modal sensor network constituting a selection of the data streams outlined earlier. There are a number of fundamental differences to the use of such a model in the network considered here as opposed to the type of network scenario described in Ganeriwal et al. (2008). These include the following:

- **The notion of nodes:** In our network the notion of a sensor node does not automatically imply a physical node deployed to monitor a particular physical phenomenon like water depth or temperature. It can also refer to visual sensor such as a camera or satellite sensor or a rainfall radar image. It may also imply a 'soft sensor' where the node is not a physical device but a mathematical model developed to provide estimates of a particular parameter using values of other parameters monitored in the network.

- **Heterogeneity:** The network may not consist of an array of homogoneous sensor nodes monitoring the same phenonomen. The 'network nodes' are all heterogeneous consisting of a variety of sensing modalities and modelled outputs. There are no scenarios whereby there are identical nodes monitoring the same phenomenon in the network e.g. there is only one in-situ sensor monitoring depth as opposed to an abundance of in-situ sensor nodes monitoring depth, etc.

- **Limited redundancy:** There are no duplicate or homogoneous nodes in the network, hence there is no redundancy from this perspective. There are heterogeneous nodes either directly monitoring or modelled to provide an estimate of the same phenomenon. However there are a limited number of these with varying reliability.

- **Centralised:** In Ganeriwal et al. (2008), RFSN was developed as a middleware service to be run on resource-constrained sensor nodes. The aim was

to develop a lightweight reputation system for sensor networks that was distributed and localised, with each node in the network running the software framework and storing reputation information for nodes in its neighbourhood. However it is noted that RFSN is flexible and can be run in a centralised manner, which is the manner in which it is used in our work. RFSN is ran on one 'control' or 'abstract node' which carries out all processing and subsequently updates reputation values for each node in the network. This node consists of a program running on a desktop computer, and thus is not subject to the constraints of resource-constrained motes.

- **Application context:** With the network and application scenario being fundamentally different from the previous RFSN model, a variation in the design of certain aspects of the system is required e.g. the outlier detection protocols used in the *Watchdog*. In Ganeriwal et al. (2008), a consensus based outlier detection protocol was used which relies on the redundancy of sensor data information in the system. It was evaluated in relation to faults which were representative of scenarios in existing sensor network deployments. Scenarios relating to the current application context are evaluated here. The application contexts considered in our work are elaborated on further in the following sections.

Thus the main contribution and objective of this chapter is to adapt the trust and reputation framework outlined by Ganeriwal et al. (2008) to a multi-modal sensor network and examine the possible application scenarios. In the following sections three application scenarios are considered. In Section 7.2 the application of such a framework for determining the most reliable model for the prediction of in-situ parameters at the Lee Maltings is considered, in Section 7.3, a similar application scenario is considered whereby such a framework is used to determine the most reliable SST satellite remote sensing data product in Galway Bay. Finally in Section 7.4, its use in determining the reliability of a visual sensing stream is

considered. The notion of 'trust' or 'reliability' is application-dependent.

7.2 Application scenario 1 - Modelling heterogeneous sensor nodes to provide redundancy in the network

In Chapter 6 (page 157), it was noted how it is often infeasible or impractical to deploy an array of homogoneous sensors in a marine environment monitoring the same parameter, thus if a sensor fails, the network is left with a gap in the data. Hence we examined whether heterogeneous in-situ sensor nodes can be modelled to predict the value of other in-situ nodes in the network thus providing a form of redundancy. However as previously noted there are a number of complex processes in operation at the Lee Maltings site (e.g. unscheduled release of a dam further upstream, influence of the tide etc.) which can be difficult to account for and to model. The aim of this work was to investigate what can be achieved using simple input model scenarios with limited data in replicating the activity of a sensor whilst there may be a possible gap in the data or fault in the network.

The output of our investigation resulted in a number of different models with varying degrees of reliability. Some of these models may produce better results than others depending on the time of year or particular trends happening at the site. Data quality may vary between months depending on the frequency of site maintenance visits for example, which may render one stream more appropriate than another for a particular time period for predicting the parameter in question. It is often not clear from a simple visualisation of the model outputs, which model is more appropriate for the purpose in question.

This raises the issue of which model should be chosen to replicate the in-situ sensor in cases where there is a sensor failure of some sort. Well known performance metrics such as correlation and mean absolute error, produced when evaluating the

various models may not accurately predict how they will behave during a particular time period representative of varying trends and scenarios. However evaluating a model's performance over a particular time period or examining its behaviour prior to the need for a substitute data source following the fault of a sensor, may provide a good indication as to which model is the best substitute. Therefore in this application scenario the purpose of a trust and reputation based framework in the network is for the purposes of evaluating which model is the most reliable at that time for replicating the work of the in-situ node in the case of sensor failure. Thus in this context the notion of 'trust' or 'reliability' represents the ability of a model to replicate the work of the in-situ node most effectively.

7.2.1 Framework Design

As previously outlined, the adaptation of the model proposed in Ganeriwal et al. (2008) for use in a multi-modal sensor network is run in a centralised manner with one 'abstract' or 'control' node, collecting data from the in-situ sensor in question and the models developed to predict values of this particular sensor. This control node runs the trust and reputation based framework. Similar to the model proposed in Ganeriwal et al. (2008), this framework consists of a *Watchdog* component and a *Reputation* component.

7.2.1.1 Watchdog Component

The *Watchdog* in Ganeriwal et al. (2008) is designed as a series of outlier detection protocols. In their evaluation they adopt a type of consensus-based outlier detection protocol. Specifically they use a version of density-based outlier detection named Local Outlier Factor or LOF as described by Breunig et al. (2000). Consensus-based outlier detection protocols rely strongly on the redundancy of sensor information in the system. However in this scenario, there is only one in-situ sensor node specifically designed to monitor the parameter in question, rendering this approach not applicable.

Models developed using information from alternative in-situ sensors in the network are trained using data from this in-situ sensor node. Hence in this application context this in-situ sensor node is presumed 'trustworthy' i.e. presumed to be producing accurate data. The goal here is to estimate which model replicates best the work of this sensor in the case of node failure. Thus the *Watchdog* assigns a level of cooperation to each model based on the deviation of its output value to that output by the sensor. This value is within the range of [0,1], with higher values indicating greater cooperation. The calculation of this value is described specifically for each application context in their respective sections of the chapter. As previously outlined in this application context the notion of 'trustworthiness' represents the ability of a model to replicate the work of the in-situ node most effectively.

7.2.1.2 Reputation Component

The *Reputation* component updates the reputation of each of the models based on the output of the Watchdog. Based on the description provided in Chapter 2, this consists of two update steps. After a single output or 'transaction', if the assigned probability of cooperativeness is p $\in [0, 1]$, the beta parameter updates would be:

$$\alpha^{new} = \alpha + p;\ \beta^{new} = \beta + p - 1 \tag{7.1}$$

The calculation of this value p is specific to the particular application scenario. Therefore in our descriptions of the application of the framework to each of the application scenarios, we explicitly state how we obtain this value p. Reputation integration or reputation aging are not relevant in our application context. As previously described in Chapter 2 an output metric of trust can be calculated using the following formula:

$$T_{ij} = E[R_{ij}] = E[Beta(\alpha_j, \beta_j)] = \frac{\alpha_j}{\alpha_j + \beta_j} \tag{7.2}$$

Thus based on a node's prior behaviour, this is the subjective expectation a

node has about another node's future behaviour. In the context of this application scenario, this trust metric can be used by the control node to determine the most suitable model or 'node' to replace the actions of a particular in-situ node in the network in the case of sensor failure. All nodes begin with an initial trust value of 0.5.

7.2.2 Models

	Cond Prediction	**Depth Prediction**	**DO Prediction**	**Temp Prediction**
Model 1 - M1	Depth	Cond	Cond	Cond
Model 2 - M2	DO	DO	Depth	DO
Model 3 - M3	Temp	Temp	Temp	Depth
Model 4 - M4	Depth-DO	Cond-DO	Cond-Depth	Cond-DO
Model 5 - M5	Depth-Temp	Cond-Temp	Cond-Temp	Cond-Depth
Model 6 - M6	DO-Temp	DO-Temp	Depth-Temp	DO-Depth
Model 7 - M7	Depth-Temp-DO	Cond-Temp-DO	Cond-Depth-Temp	Cond-Depth-DO

Table 7.1: Models evaluated using a reputation based framework for the prediction of parameters at the Lee Maltings

In Chapter 6, we investigated the prediction of conductivity, depth, dissolved oxygen and temperature by developing input models for the prediction of each using values from one, two or three of the alternative parameters. This was based on using 5 values, 10 values and 20 previous values. In order to investigate the use of a trust and reputation based framework for determining the most appropriate model in the network we choose to evaluate all models developed using 20 preceding values of each input parameter since this often produced better results. Thus in total this results in 7 models being evaluated (3 single parameter models, 3 double parameter models and 1 triple parameter model). Table 7.1 presents the models and their input

parameters for the prediction of each of the four parameters of conductivity, depth, dissolved oxygen and temperature. So, for example, in Model 1 (M1) we use Depth to predict conductivity whereas in Model 4 (M4) we use conductivity and dissolved oxygen to predict depth (these examples are underlined in Table 7.1)

7.2.2.1 Model Training and Testing

Data from five different months - May, June and July 2009, and January and February 2010 - were selected for use in this investigation. Figure 7.1 shows raw data from the four sensors for each of these time periods. For all months except for June, it is is the first two weeks of the month that are used for testing. Two different training datasets are used for training the models used for the prediction of each of the four parameters.

Generalised Model: Each of the models are first trained using 10,000 instances created from across the year's dataset. This dataset consisted of either the first two weeks or the last two weeks of each month in the year. When selecting these weeks the content of the data was not considered, they were chosen arbitrarily. Thus even if there happened to be gaps in the data during the particular two weeks of the month selected for training, this did not matter.

Monthly Model: Secondly, a more focused training approach was carried out. In this case predictions to be made within a particular month used a model trained using data from a similar time period, as opposed to a generalised model. For each month two weeks of data was set aside for training. This resulted in just over 2,000 instances. Thus each training set was limited to 2,000 instances for consistency (except for January, where there where gaps in the data and the training set was not limited to any number but all available data was used for training). The other two weeks of the month were tested on the models developed using data from that particular month. For each of the five months apart from June, the first two weeks of the month were used for testing and the last two weeks of the month were used for training. Thus for each parameter, seven models were created for making predictions

for each month, resulting in 35 models altogether for the five months.

There are various reasons for choosing to train models in this manner. In certain scenarios and for the prediction of certain parameters, monthly models perform better on the test data. This may be due to the fact that because the training data is specifically focused on a particular time of year, it may have a similar distribution and present similar trends (which should be noted is not always the case). It is also interesting to examine how a model performs with limited training data. However the monthly model may not be appropriate in certain cases where trends vary greatly or there is an issue with data quality or gaps in the data for the training period of the month. Also the generalised model is representative of a larger number and a wider distribution of data samples.

Whether the generalised model versions of M1-M7 or the monthly model versions of M1-M7 are used for the prediction of a particular parameter and the subsequent evaluation of the application of a reputation and trust-based framework is dependent on the performance on the test data for that particular month. In a real-world scenario, the benefit of being able to test the data before it actually occurred in order to decide which model to use, is of course not possible. However the purpose of this study is not to decide how to choose a training approach but to evaluate the use of a trust and reputation framework. Hence the set of models which generally perform best and aid the evaluation of such a framework on the test data are chosen. The statistical properties of each of the training and test sets is shown in Table 7.2.

Figure 7.1: DO, Temperature, Conductivity and Depth for May, June, and July 2009 and January and February 2010. Source: DEPLOY

	May-test	May-train	June-test	June-train	July-test	July-train	Jan-test	Jan-train	Feb-test	Feb-train	All
Cond.											
Max	44.6	44.87	46.73	44.1	45.57	45.76	42.22	41.81	42.970	44.82	48.02
Min	-0.078	0.014	0.033	0.18	0.143	0.125	0.014	0.014	0.014	0.069	0.014
Avg	13.981	17.834	22.763	23.98	18.7	20.323	24.521	20.239	19.595	30.279	21.297
StdDev	17.757	19.29	18.643	18.469	19.108	19.236	16.999	18.063	17.731	15.924	18.872
Depth											
Max	4.066	4.13	4.418	4.18	6.916	4.547	4.657	4.461	4.696	4.631	5.28
Min	0.404	0.429	0.423	0.558	0.533	0.341	0.745	0.720	0.709	0.683	0.341
Avg	2.076	2.122	2.158	2.178	2.301	2.211	2.409	2.478	2.252	2.363	2.332
StdDev	1.088	1.057	1.159	1.089	1.074	1.131	1.071	0.974	1.087	1.142	1.088
DO											
Max	11	9.9	10.69	11.03	9.410	10.610	15.980	13.590	11.630	11.710	14.8
Min	2.42	2.4	1.5	2.58	1.01	2.05	8.970	8.630	8.250	7.970	2.05
Avg	6.222	5.748	5.052	5.239	5.526	6.35	12.726	10.417	10.192	9.774	8.069
StdDev	1.444	1.282	1.641	1.151	1.871	1.484	1.46	0.932	0.562	0.647	2.588
Temp.											
Max	14.57	17.120	19.73	17.91	19.66	18.56	10	10.89	8.250	7.910	18.56
Min	11.19	11.89	14.69	13.18	16.22	15.03	2.86	3.26	4.910	5.01	3.26
Avg	12.243	13.487	16.79	15.285	17.638	16.511	4.577	6.331	6.876	6.943	11.922
StdDev	0.63	1.049	0.815	1.007	0.604	0.810	1.03	1.199	0.738	0.559	4.049

Table 7.2: Statistical properties of the datasets used for training and testing purposes.

7.2.3 Conductivity at the Lee Maltings

The following examines the application of a trust and reputation framework in the scenario where there are a series of models (M1-M7) designed to predict conductivity at the Lee Maltings. However firstly the need for a real-time or retrospective analysis tool is highlighted using the prediction of conductivity from these models as an example. This highlights the benefits of applying a trust and reputation framework in such an application context.

7.2.3.1 The need for a real-time or retrospective analysis tool

We now provide a brief overview on some examples of differences in trends. However it should be noted that this is not a direct comparison of performance between models. The generalised models use 10,000 training instances, the June and February monthly models use 2,000 training instances and the January models use 637 training instances. There are also a different number of test instances in each month - May has 1,995, June has 2,305, July has 1,955, January has 926 and February has 1,498. These differences alone render a direct comparison infeasible. However since our objective is to apply a reputation and trust-based framework, how that particular model is trained is not relevant. The purpose of the following is to highlight some differences in trends which highlight the difficulties in estimating the performance of a model on unknown test data, and hence the need for a tool which allows real-time evaluation of the performance of a model or retrospective analysis of how a model performs in different scenarios. Conductivity prediction was chosen for highlighting some of these difficulties.

For conductivity the generalised model versions of M1-M7 was used for predictions in May and July. The monthly model versions of M1-M7 was used for predictions in June, January and February. Table 7.3 demonstrates the results of a ten-fold cross validation carried out on these models in order to estimate how they perform (i.e. June, January and February models and then the generalised model used for carrying out predictions on test data from May and June). Table 7.4 shows

	June		Jan		Feb		General	
	CC	MAE	CC	MAE	CC	MAE	CC	MAE
M1 (Depth)	0.95	2.96	0.84	6.29	0.91	3.50	0.77	8.02
M2 (DO)	0.87	5.83	0.85	6.34	0.65	8.89	0.80	7.67
M3 (Temp)	0.83	6.88	0.84	6.52	0.95	2.55	0.80	8.02
M4 (Depth-DO)	0.97	2.23	0.90	5.04	0.93	2.88	0.95	3.15
M5 (Depth-Temp)	0.96	2.71	0.93	3.72	0.98	1.93	0.94	3.53
M6 (DO-Temp)	0.94	3.90	0.96	2.94	0.96	2.22	0.92	4.38
M7 (Depth-DO-Temp)	0.97	2.16	0.96	2.93	0.98	1.76	0.97	2.5

Table 7.3: Results of 10-fold cross validation of each of the training datasets for conductivity prediction

the performance of these models on test data. As can be seen, it can be difficult to estimate how a model will perform

For example when looking at the evaluation of the generalised model compared to its performance on the test data from May and July, the output of the generalised model would suggest that M7 may perform best which is clearly not the case when the model is used for predictions in May or July. It would also suggest that dissolved oxygen (DO) or temperature produce a better output at estimating conductivity than depth, which is again not the case when applied to test data from May. When this model is applied to test data from July however DO does perform better than depth at predicting conductivity, with a huge drop in the performance of temperature in predicting conductivity compared to the output of the analysis. We see models that are producing performances along similar lines in the evaluation can perform very differently to each other on the test data. A similar scenario can

May	CC	MAE	RMSE
M1	0.7183	8.411	13.1564
M2	0.625	10.3183	14.5688
M3	0.3127	18.4659	23.6164
M4	0.8902	4.6703	8.5782
M5	0.7217	8.9054	15.291
M6	0.5017	14.8531	20.3492
M7	0.6938	8.5348	15.0045

June	CC	MAE	RMSE
M1	0.9177	4.2657	7.4219
M2	0.2923	14.8692	20.7109
M3	−0.3308	23.4366	28.5563
M4	0.8829	5.1154	8.8103
M5	0.8987	7.7289	8.212
M6	0.0913	17.1897	22.2718
M7	0.8844	5.1089	8.7782

July	CC	MAE	RMSE
M1	0.6817	10.403	13.9772
M2	0.8114	7.7175	11.2423
M3	−0.0339	18.3334	25.4813
M4	0.9037	5.2465	8.353
M5	0.1293	17.1816	24.8577
M6	0.6337	9.2093	15.4598
M7	0.5954	12.3181	18.1068

Jan	CC	MAE	RMSE
M1	0.652	8.9179	13.4423
M2	0.1191	14.4345	18.2036
M3	0.2308	21.7863	26.2124
M4	0.6644	9.7714	13.1314
M5	0.603	15.0295	18.4994
M6	−0.0233	22.3177	26.5591
M7	0.6121	15.4332	18.6943

Feb	CC	MAE	RMSE
M1	0.7058	9.6884	15.7553
M2	0.1711	19.2569	24.2633
M3	0.7523	7.1209	13.6179
M4	0.6706	12.0994	17.4465
M5	0.7906	7.0838	12.8208
M6	0.7482	8.1024	13.6729
M7	0.7892	7.8265	13.0498

Table 7.4: Performance of each of the models when evaluated in the corresponding test data for the prediction of conductivity

be seen when examining the evaluation outputs of the training models from June, January and February (Table 7.3) compared to that of the performance of the test data (Table 7.4). Some models may be performing better than others depending on the month. For example M3 (Temp) performs quite poorly on the test data in June and January, but performs much higher in February and outperforms M2 (DO), which hadn't been the case in June.

However its not always the case that high performance in the evaluation of the training model doesn't reflect high performance on the test data. For example in June M1, M4, M5 and M7 perform exceptionally well in the evaluation of the training model, and also perform very well on the test data from June. In February those that perform higher in the evaluation generally performed higher on the test data (e.g. M2 produced a lower performance than other models in the evaluation

period and on the test data, the others were similar and also had similar enough performances on the test data).

The variation in performance may be due to a range of factors such as a change in the statistical properties of the test data compared to that of the training data due to a change in trend, data quality issues such as gaps in the data etc. Issues with just one data stream or changes in the relationship between one data stream and the other can affect the performance of associated models. Also as previously outlined a different number of instances were used in training the generalised, June, January and February test models. There were also a different number of test instances for each of the five months. In general lower performance overall is achieved in January and February. However from Figure 7.1, irregularities such as gaps in the data are apparent suggesting issues with the sensors, especially in January.

It is apparent from this brief analysis, that it is difficult to estimate how a model will perform on real unknown test data. Some may perform better than others at different times of year, during the appearance of difference trends or simply by using a different training dataset. We cannot always estimate based on the evaluation of the training model what type of performance will be achieved on the test data. Hence it is apparent that there is a need for some real-time update on the performance of a model in order to select which appears to be operating most effectively. It is also useful for a retrospective analysis purposes also, so that if the sensor went completely offline in the future, that the performance of each of the models could be analysed in a variety of scenarios and the best one chosen depending on the time period in question or a variety of other criteria. If we have models on similar values of performance metrics such as CC, MAE, Root Mean Squared Error (RMSE - square-root of the average of the squared errors), etc. during an evaluation, it is useful to evaluate the most appropriate model based on analysing their performance during a variety of scenarios.

7.2.3.2 Application of the Trust and Reputation-Based framework

At each epoch, the *Watchdog* assigns a level of cooperation to a model based on the deviation of its output from the output value of the sensor (i.e. p value as described in Section 7.2.1.2). The calculation of this value is described in Equation 7.3. To calculate this cooperation value for updating the reputation of the model, an absolute error value is calculated through subtracting the output of the model from the corresponding sensor output. This error value is normalised to [0-1] by dividing by the max-min range of the full training set — [48.02-0.014]. This range was chosen even if the model in question is a monthly model trained on a monthly dataset where the range did not vary hugely across months as can be seen in Table 7.2. This value is then subtracted from 1, so that values closer to 1, indicate 'greater cooperation'. The *Reputation* component then uses this output of the *Watchdog* to update the reputation of the model (i.e. Equation 7.1 on page 177). The trust metric of a model is the statistical expectation of the reputation function calculated as described in Section 7.2.1.2. Unlike reputation which is maintained as a probabilistic distribution, the trust metric is simply a number (calculated from Equation 7.2 on page 177). Thus this number is updated after each output of the model. If there is no output value of the sensor to compare the output value of the model to at a particular epoch, the reputation and trust information remains unchanged.

$$p = 1 - \frac{Abs(ModelOutputValue - SensorOutputValue)}{Normalisation} \tag{7.3}$$

7.2.3.3 Results

For demonstration purposes, the output of each of the models for the first two weeks in May along with the actual values output by the conductivity sensor are shown in Figures 7.2 and 7.3 . The associated trust values of each of the models at each epoch is shown in Figure 7.4. At the beginning of May a phenomenon whereby conductivity levels flatten out (while depth values continue to vary periodically) can be seen. This is thought to be caused by the undocumented release of water from the

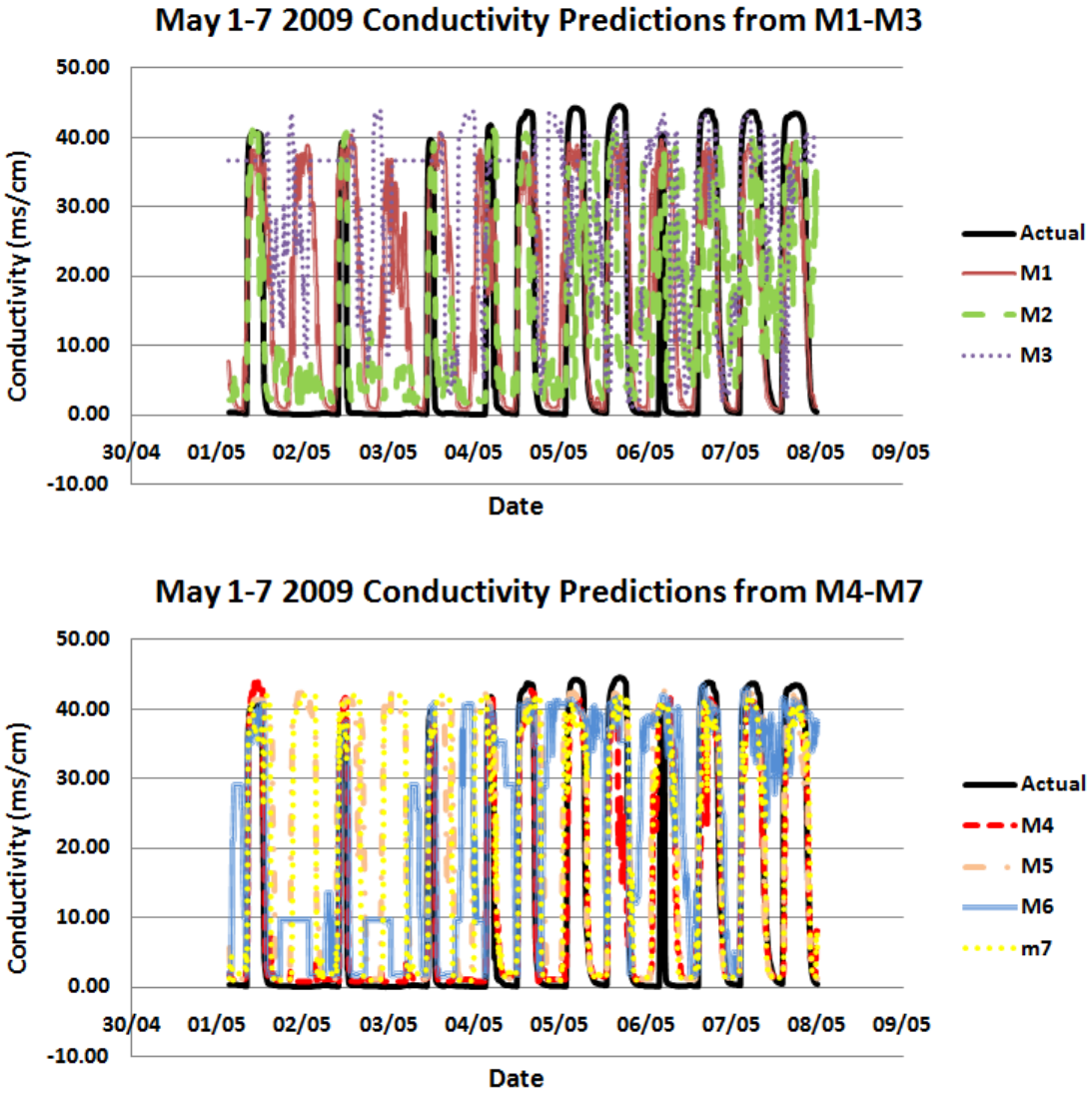

Figure 7.2: Graphs showing the conductivity values output by M1-M3 (Depth, DO, Temp) and M4-M7 (Depth-DO, Depth-Temp, DO-Temp, Depth-Temp-DO) compared to the actual values output by the conductivity sensor sensor for May 1-7 2009

dam further upstream. However Figures 7.2 and 7.3 demonstrates that a selection of the models can somewhat replicate the activity of the conductivity sensor during this time period.

Of the single parameter models (M1-M3), M2 (DO) seems to perform the best in relation to detecting this phenomenon, and begins then to slightly disimprove in performance towards the end of the first week. M1(Depth) continues to fluctuate

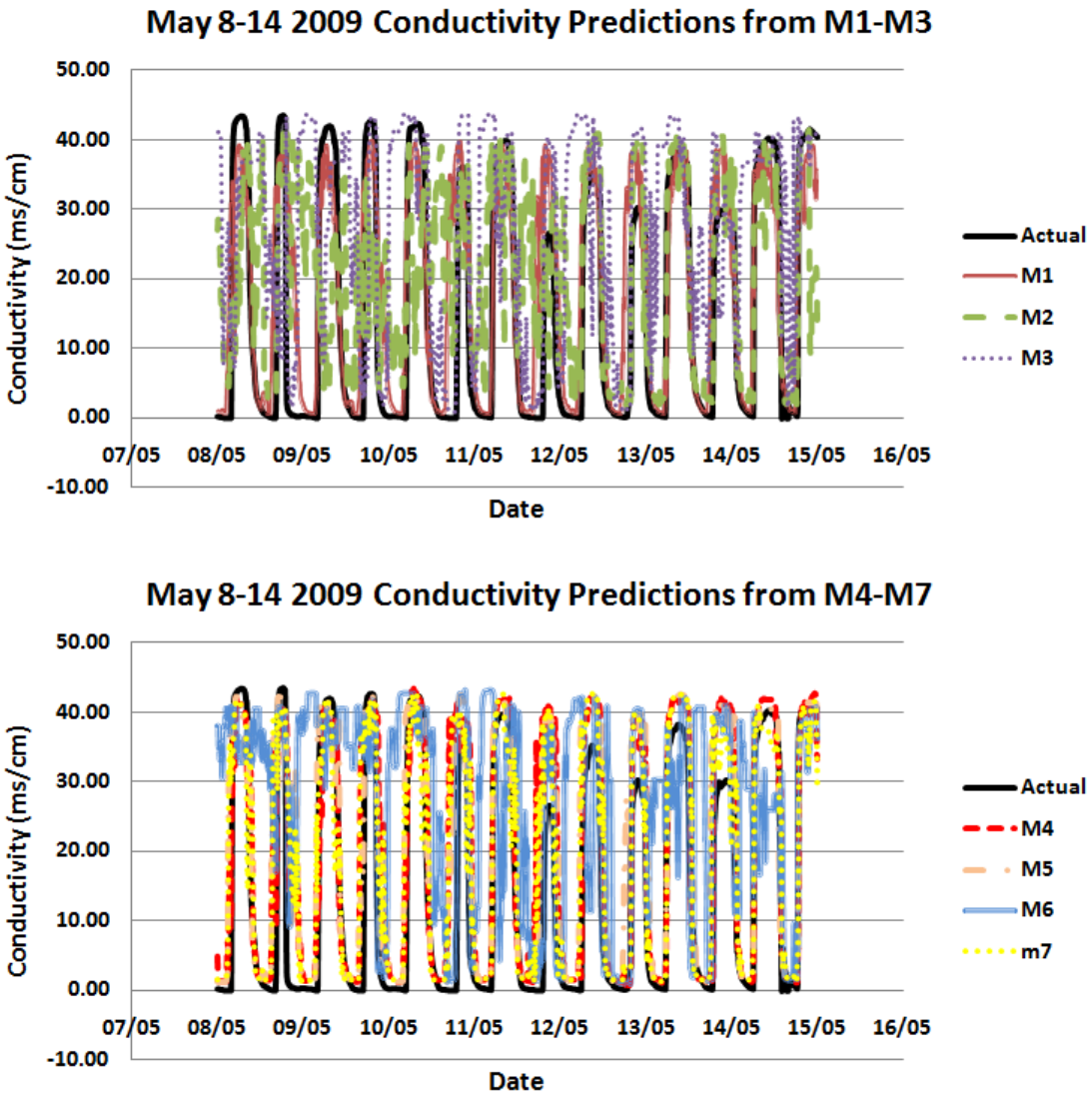

Figure 7.3: Graphs showing the conductivity values output by M1-M3 (Depth, DO, Temp) and M4-M7 (Depth-DO, Depth-Temp, DO-Temp, Depth-Temp-DO) compared to the actual values output by the conductivity sensor sensor for May 8-14 2009

periodically and doesn't seem to pick up on this phenomenon. However when the conductivity sensor operations returns to normal towards the end of the first week, it seems to replicate its values quite well. M3 (Temp) appears to be the poorest performing in terms of replicating the activities of the conductivity sensor and seems to be extremely unreliable. In the second week (May 8-14), the output of M2 appears a little unreliable at the beginning of the week but then improves and appears to output patterns similar to the conductivity sensor. M1 appears to reflect the patterns of of the conductivity sensor quite well this week, with M3 improving but still appearing to be the most unreliable of the three.

If we look at the trust values associated with each of these models (M1-M3) in Figure 7.4, it is clear that they reflect the pattern described above from examining the output of the models in comparison to the conductivity sensor. It is apparent that while M2 starts out at being more trustworthy than M1 in week 1 due to its ability to to accurately reflect the the flattening of the conductivity values, after this its trust slowly disimproves, while that of M1 improves as the days go on and surpasses that of M3. The very poor performance of M3 is also reflected in its trust values which are continuously very low, although after having initially dipped in performance at the beginning of week 1, its performance does appear to continually improve. M1 produces a CC of 0.7183 and a MAE of 8.411, M2 produces a CC of 0.625 and a MAE of 10.3183 and M3 produces a CC of 0.3127 and a MAE of 18.4659. Looking at the performance metrics alone does not provide a clear evaluation of which model is the most appropriate during different scenarios. It is clear that M2 is most appropriate during a scenario such as that occurring at the beginning of the first week in May but during so called 'normal' operation of the conductivity sensor, it appears to be M1 that performs the best at replicating its values.

Of the two and three parameter models (M4-M7) it is clear that M4 (Depth-DO) seems to very accurately detect the phenomenon with the conductivity sensor at the beginning of May and seems also to accurately replicate the normal operation of the conductivity sensor towards the end of the week. M5 (Depth-Temp) and M7

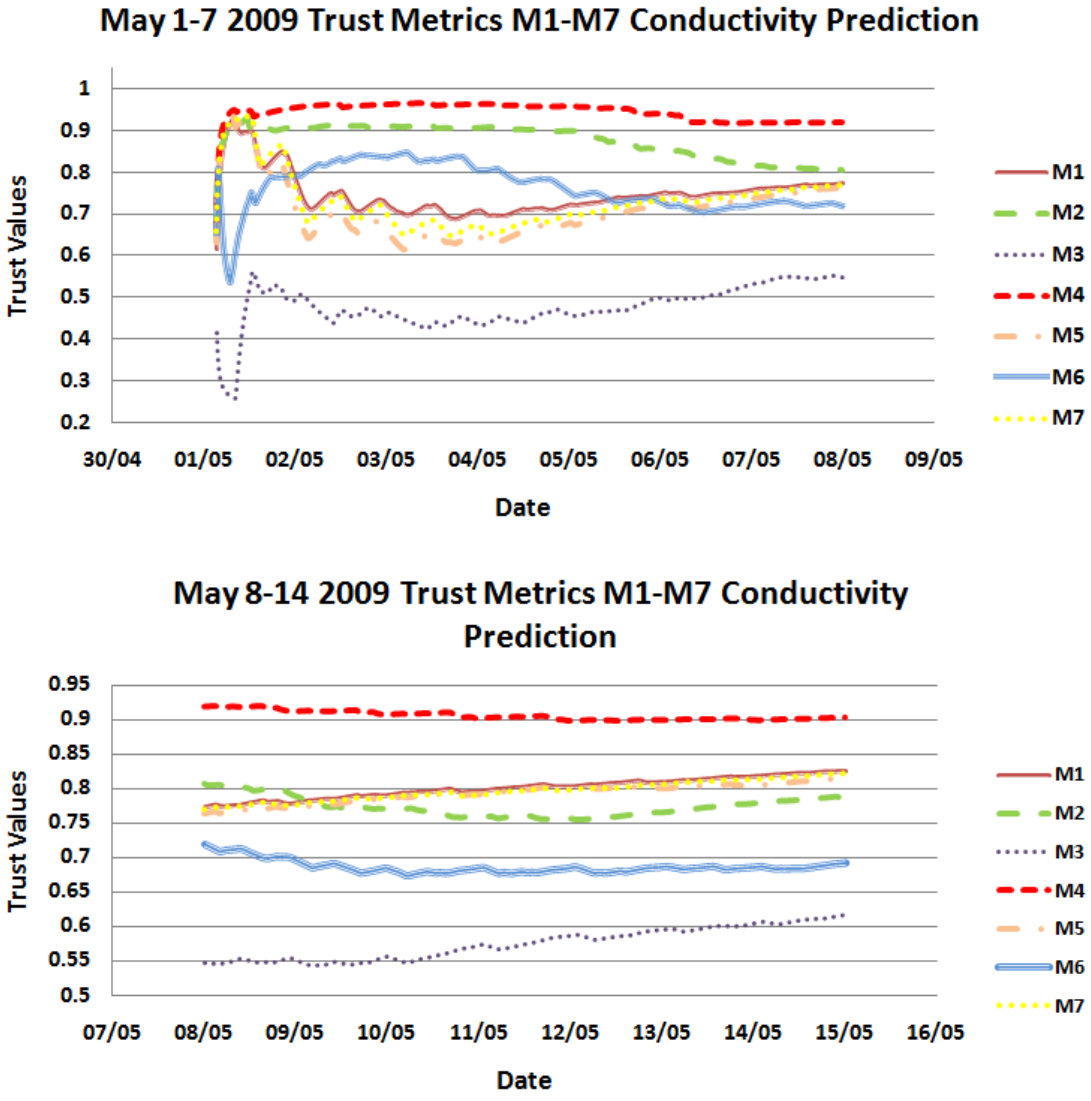

Figure 7.4: Graphs showing the trust values for M1-M3 (Depth, DO, Temp) and M4-M7 (Depth-DO, Depth-Temp, DO-Temp, Cond-Temp-DO) for predicting conductivity for May 1-14 2009

(Depth-DO-Temp) do not appear to have the ability to detect the phenomenon at the beginning of the week but appear to operate quite well after this. M6 (Depth-Temp) somewhat replicates the scenario in place at the beginning of the week, however during the normal operation of the sensor appears to perform quite poorly.

Again this behaviour is reflected in the trust values shown in Figure 7.4. M4 has the highest trust ratings overall, the trust ratings of M5 and M7 are reduced at the start due to their inability to replicate the activities of the conductivity sensor during

the phenomenon at the beginning of the week. However their trust values improve following their improved performance during the normal operation of the sensor. While the trust values of M6 begin quite high due to its somewhat replication of the phenomenon at the beginning of the week, its trust continually decreases right into the second week based on its poor performance after this, finishing just above M3 in the rankings by the end of the second week. From Table 7.4 it is can be seen that M4 produces the best performance metrics which is apparent from the analysis carried out here (CC-0.8902, MAE-0.8902, RMSE - 8.5782). M6 produces the poorest (CC - 0.5017, MAE, 14.8531, RMSE - 15.291), and M5 and M7 produce very similar performance values (M5 - CC - 0.7217, MAE - 8.9054, RMSE - 15.291, M7 CC- 0.6938, MAE - 8.5348, RMSE - 15.0045). The patterns of these performance metrics (i.e. M4 - the highest, M6 - the lowest, M5 and M7 - similar) are reflected in the trust values which subsequently demonstrates itself as a real-time update tool for evaluating the performance of a model.

Best Output

	CC	MAE	RMSE
May	0.8902	4.6702	8.5781
June	0.9120	4.2535	7.6663
July	0.9039	5.2310	8.3487
Jan	0.6853	8.8685	12.8167
Feb	0.7711	7.0258	13.0556

Table 7.5: Performance metrics when the output of the most trustworthy model is selected at each epoch for conductivity prediction

For each of the test sets - May, June, July, Jan, Feb - we evaluated the scenario whereby at each epoch, the value output by the most trustworthy model was selected. This is referred to as TM (TrustModel). The results of this evaluation can be seen in Table 7.5. In each case TM produces almost identical or slightly improved performance metrics to the best performing model for the month. In February however TM has a slight reduction in performance in terms of CC and RMSE but

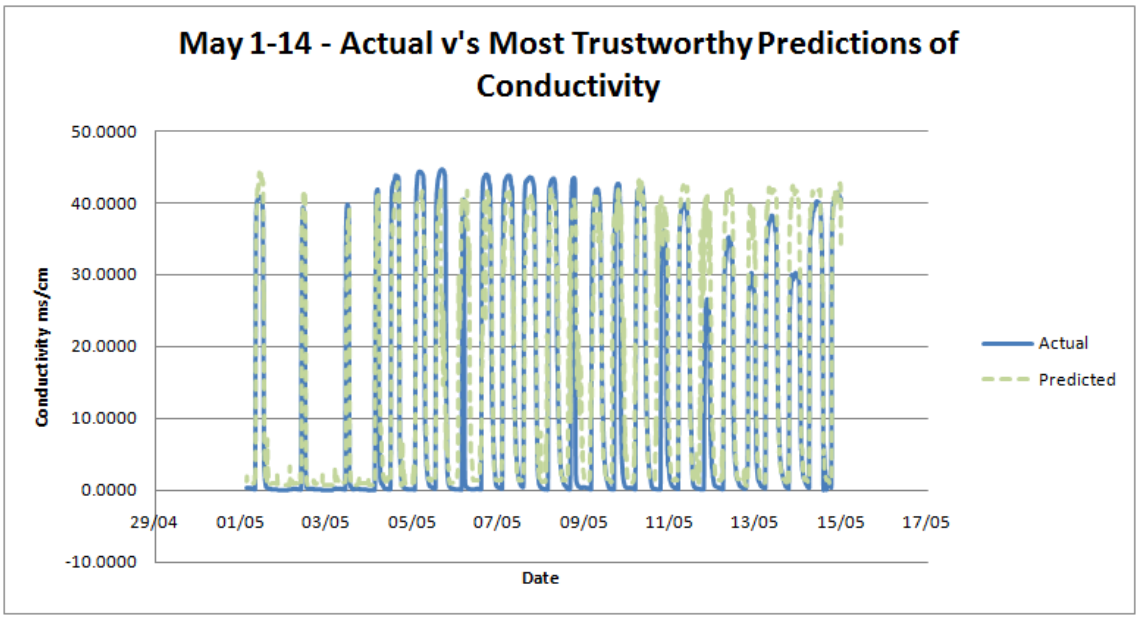

Figure 7.5: Actual v's predicted values for conductivity when the value from the most trustworthy model is chosen at each time epoch for May 1-14 2009

has a slight increase in performance in terms of MAE (May - M4, June M1, July - M4, January - M1, M4, February - M5). Figure 7.5 shows the conductivity values from the in-situ sensor plotted against the predicted values for conductivity when the value from the most trustworthy model is chosen at each time epoch over a period of time.

7.2.4 Depth at the Lee Maltings

For the prediction of depth the generalised model versions of M1-M7 were used for predictions in May and July. The monthly model versions of M1-M7 was used for predictions in June, January and February. The generalised models use 10,000 training instances, the June and February monthly models use 2,000 training instances and the January models range from 488 to 2569 training instances (M1, M4, M5, M7 - 488, M2, M6 - 2550, M3 - 2569). The number of test instances in each month are - May - 2004, June - 2287, July - 1682, January - 2090, February - 2108. The performance of the models on the test data is shown in Table 7.6.

May	CC	MAE	RMSE
M1	0.7803	0.517	0.7352
M2	0.481	0.7837	0.9685
M3	0.1139	1.0564	1.2891
M4	0.7774	0.5261	0.7187
M5	0.8307	0.4948	0.6154
M6	0.3784	0.8561	1.0604
M7	0.79	0.5169	0.6781

June	CC	MAE	RMSE
M1	0.9363	0.3225	0.4118
M2	0.1048	1.1138	1.4498
M3	−0.1537	1.2279	1.4709
M4	0.927	0.3407	0.4387
M5	0.9347	0.3216	0.4183
M6	0.0785	1.1563	1.489
M7	0.9314	0.3271	0.4256

July	CC	MAE	RMSE
M1	0.8405	0.4166	0.6138
M2	0.433	0.776	1.015
M3	0.1455	0.9356	1.1023
M4	0.8591	0.4053	0.5884
M5	0.8422	0.4085	0.5835
M6	0.0401	1.0188	1.2223
M7	0.8594	0.3915	0.5579

Jan	CC	MAE	RMSE
M1	0.3747	0.857	1.0784
M2	0.0518	1.0677	1.3097
M3	−0.1888	1.1905	1.4557
M4	0.3821	0.8853	1.036
M5	0.3886	0.8172	1.0427
M6	−0.0521	1.1268	1.396
M7	0.3745	0.8598	1.0455

Feb	CC	MAE	RMSE
M1	0.6655	0.664	0.8294
M2	0.4612	0.8552	1.0849
M3	0.4263	0.8526	1.102
M4	0.6823	0.6532	0.824
M5	0.6756	0.6574	0.8168
M6	0.5634	0.7126	0.9245
M7	0.6674	0.6622	0.8382

Table 7.6: Performance of each of the models when evaluated in the corresponding test data for the prediction of depth

7.2.4.1 Application of the Trust and Reputation-Based framework

Similar to the system for conductivity prediction, at each epoch the *Watchdog* assigns a level of cooperation to a model based on the deviation of its output from the output value of the sensor (See Equation 7.3 on page 187). The deviation of the error value is normalised to a range of [0-1] by dividing by the max-min range of the full training set - [5.28-0.34], and this value is then subtracted from 1. Similar to the case for conductivity prediction, this range was chosen even if the model in question was a monthly model trained on a monthly dataset considering the range did not vary hugely across months as can be seen in Table 7.2. Again the output of the Watchdog is used to update the Reputation component and there is the subsequent calculation of trust (See Equation 7.1 on page 177 and Equation 7.2 on page

177). Again if there is no output value of the sensor to compare the output value of the model to at a particular epoch, the reputation and trust information remains unchanged.

7.2.4.2 Results

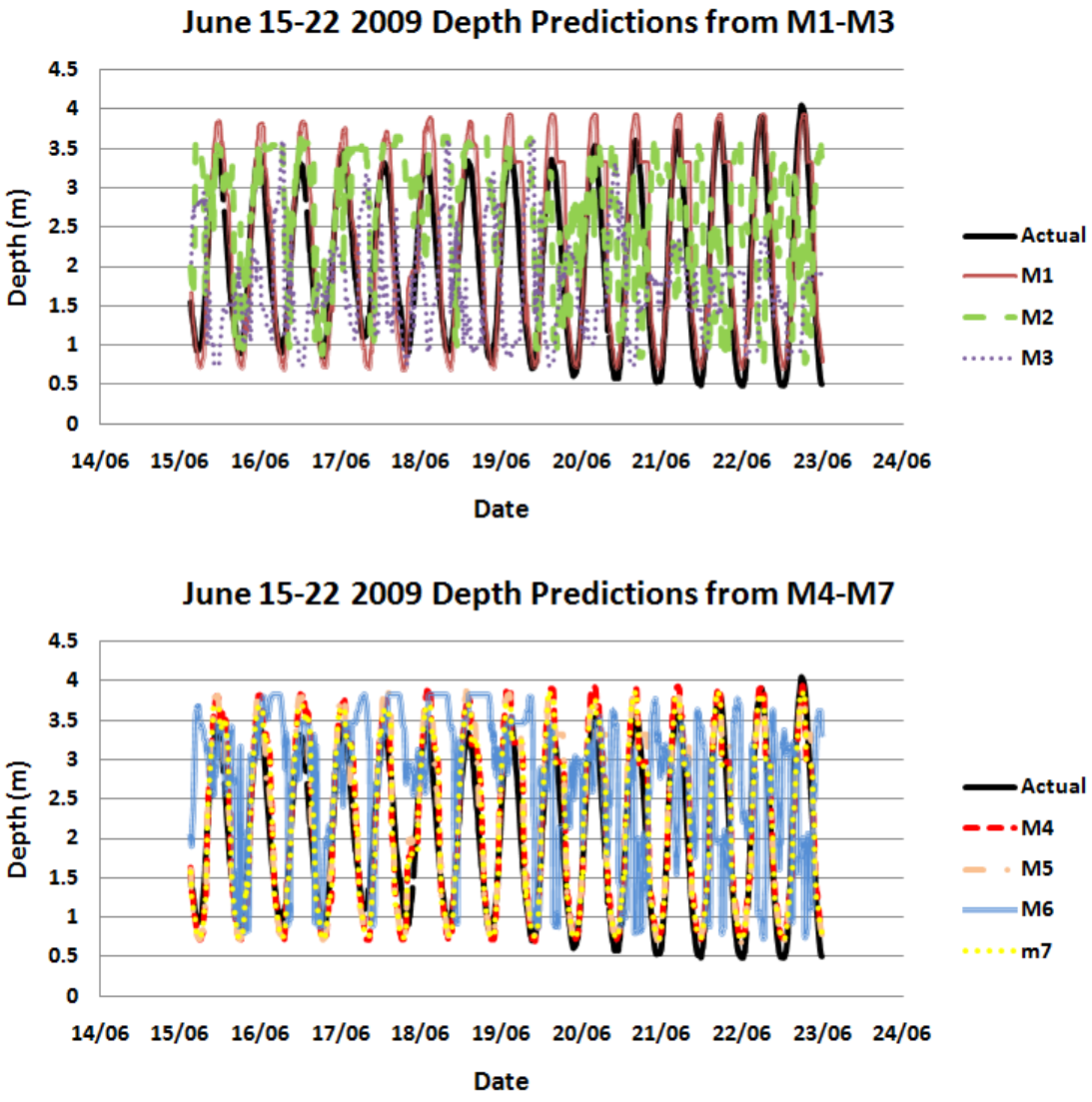

Figure 7.6: Graphs showing the depth values output by M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Temp, DO-Temp, Cond-Temp-DO) compared to the actual values output by the depth sensor sensor for June 15-22 2009

For demonstration purposes, the output of each of the models for the last two

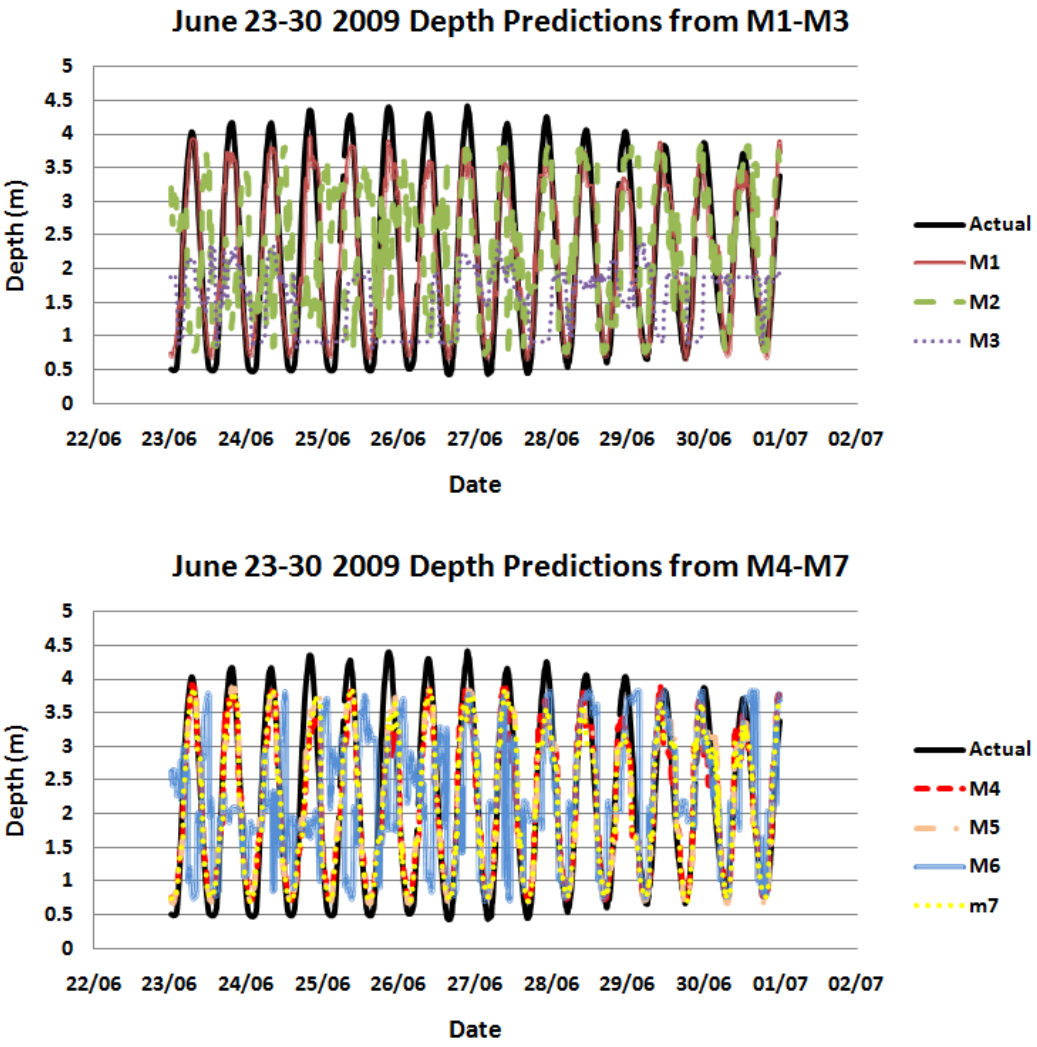

Figure 7.7: Graphs showing the depth values output by M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Temp, DO-Temp, Cond-Temp-DO) compared to the actual values output by the depth sensor sensor for June 23-30 2009

weeks in June along with the actual values output by the depth sensor are shown in Figures 7.6 and 7.7. The associated trust values of each of the models at each epoch is shown in Figure 7.8.

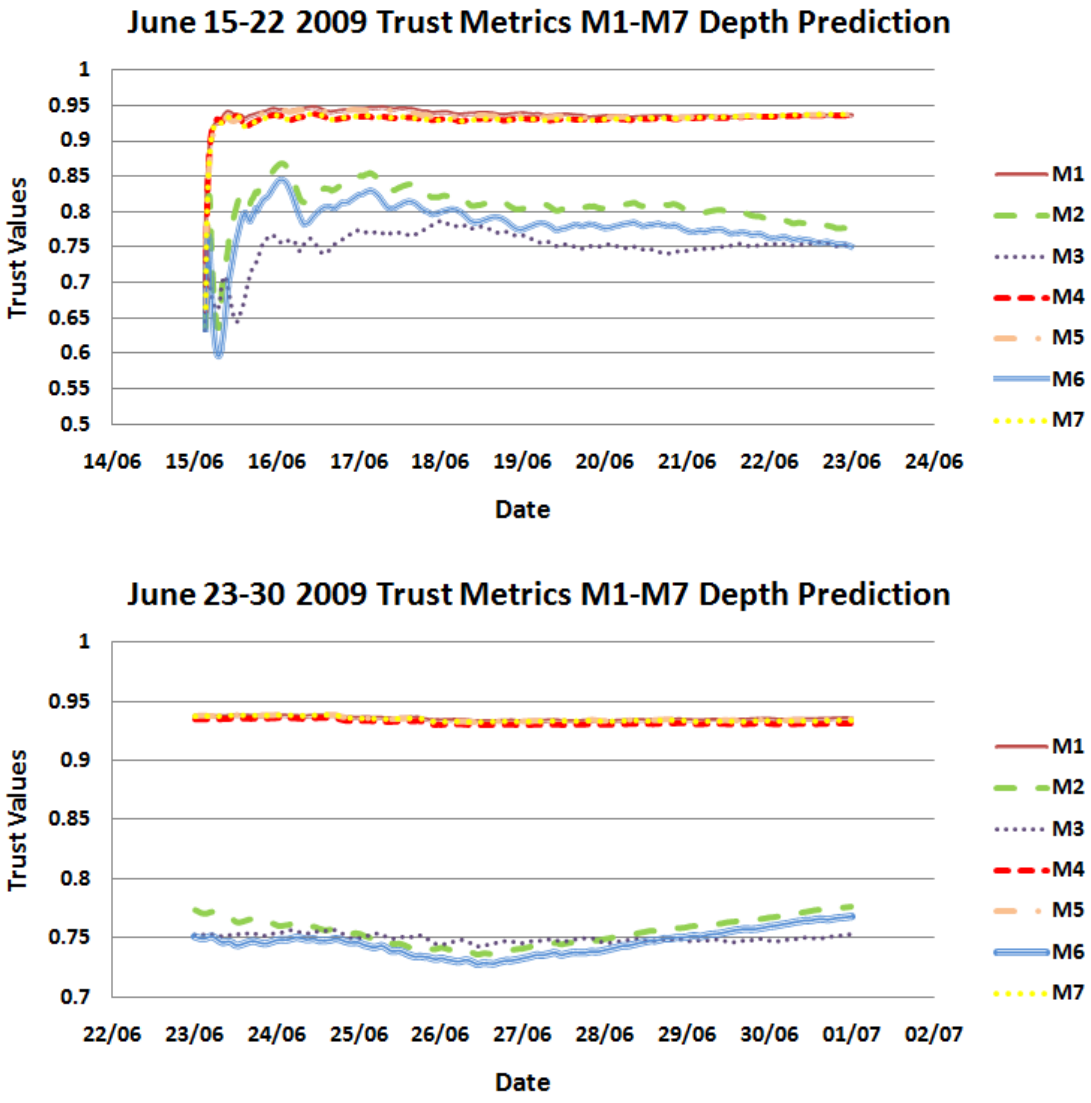

Figure 7.8: Graphs showing the trust values for M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Temp, DO-Temp, Cond-Temp-DO) for predicting depth from June 15-30 2009

Of the single parameter models (M1-M3), it appears that M1 (Conductivity) replicates the activity of the depth sensor quite well. In the first of the two weeks it appears to overestimate the value of depth at the peak of the cycle and sometimes output lower values at the trough of the cycle (See Figure 7.6). At the beginning

of the second week, it appears to do the opposite, then moving on to output the maximum and minimum points of the depth cycle better at the end of the week. M2 (DO) is not very accurate but appears to perform better than M3 (Temp) which towards the end of the first week and for the second week, is very unreliable (See Figure 7.7).

This is reflected in Figure 7.8, where the evolution of trust values demonstrates M1 to be clearly the most reliable, with M2 much lower and M3 lower again in the first week. A slight peak can be seen in the trust values for M3 which then falls towards the end of the first week reflecting the pattern mentioned above. The trust values also can help to pick up patterns which are difficult to see from graphing of the output values. For example from Figure 7.8 it can be seen that by the second week M2 and M3 are almost equally unreliable. Of the two and three parameter models (M4-M7), M4 (Cond-DO), M5 (Cond-Temp), and M7 (Cond-Temp-DO) appear to reflect the behaviour of the depth sensor most accurately, with M6 (DO-Temp) appearing to be the most inaccurate. This is again reflected in the trust values shown in Figure 7.8.

Overall it can be seen that the most trustworthy models are M1, M4, M5, and M7. From Table 7.6, it can be seen that these clearly produce the best performance metrics in comparison to the other models (M1 - CC- 0.9363, MAE- 0.3225, RMSE - 0.4118, M4 - CC- 0.927, MAE- 0.3407, RMSE- 0.4387, M5 - CC- 0.9347, MAE- 0.3216, RMSE- 0.4183, M7 - CC- 0.9314, MAE- 0.3271, RMSE- 0.4256). The clear differentiation between these models and the others in terms of trust values demonstrates the ability of this framework to reflect these performance values and presents a a real-time update tool for evaluating the performance of a model.

Best Output

Again for each of the test sets - May, June, July, Jan, Feb - we evaluated the scenario wherby at each epoch, the value output by the most trustworthy model was selected. Again this is referred to as TM (TrustModel). The results of this evaluation can

	CC	MAE	RMSE
May	0.8315	0.4942	0.6146
June	0.9329	0.3210	0.4223
July	0.8552	0.3890	0.5665
Jan	0.4182	0.8140	1.0258
Feb	0.6870	0.6509	0.8087

Table 7.7: Performance metrics when the output of the most trustworthy model is selected at each epoch for depth prediction

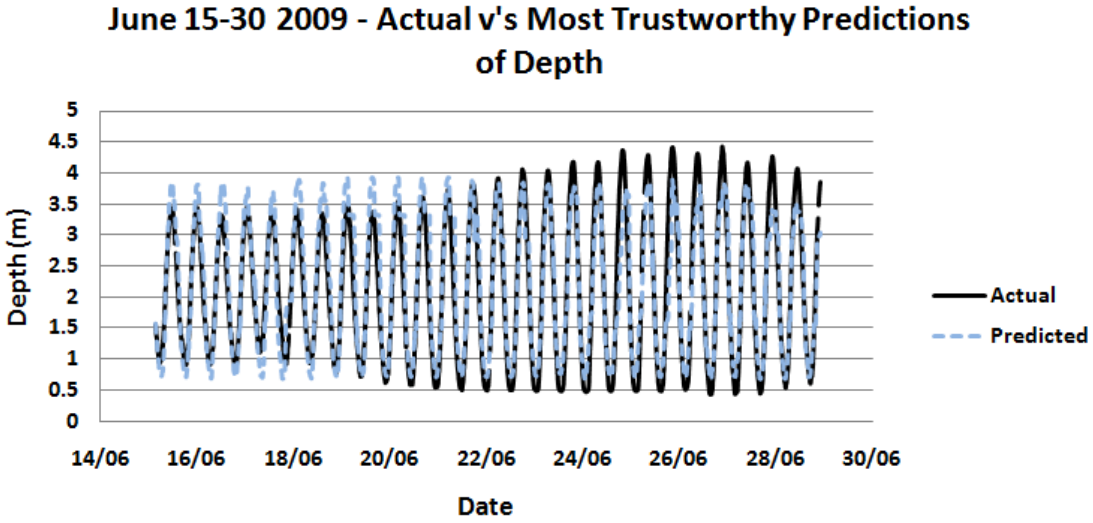

Figure 7.9: Actual v's predicted values for depth when the value from the most trustworthy model is chosen at each time epoch for June 15-30 2009

be seen in Table 7.7. Again TM produces either very similar or slightly improved performance metrics to the best performing models of each of the months (*May*- M5 , *June*- M1 and M5, *July*- M7, *January*- M5, and *February*- M4) with the most noticeable improvement for the January test data. Figure 7.9 shows the values from the in-situ depth sensor plotted against the predicted values for depth when the value from the most trustworthy model is chosen at each time epoch over a period of time.

7.2.5 Dissolved Oxygen and Temperature at the Lee Maltings

We applied the trust and reputation framework to the models for the prediction of dissolved oxygen and temperature. However for the sake of brevity this analysis can be found in Appendix D. However there are some points to note from this analysis. For example due to the fact that the range of the dataset seems to vary more from month to month for dissolved oxygen and temperature than for depth or conductivity, the deviation of the error value is normalised to a range of [0-1] using a different range depending on whether a generalised or a monthly model is used. We demonstrate the difference in the trust values when both types of normalisation are used. It can be seen that while overall patterns don't generally change, the magnitude of difference between models and the height of trust values varies. Using a smaller range to carry out normalisation results in greater magnitudes of differences between the models and overall lower trust values (For example, see Figures D.2 and D.3 which demonstrates the trust values when normalisation using the range of the full dataset is used and normalisation using the range of the February training set is used for models predicting DO).

The application of the framework to the models for DO prediction also reinforces how this framework can inform us of the magnitude of unreliability of one model in replicating the activities of the sensor in comparison to another, allowing us to decipher trends which are difficult from a simple plot of the outputs. For example, in Figure D.1 we see a scenario whereby it is very difficult to determine which model is most reliable from a plot of the outputs. This is the output of the models for DO prediction for the first week in February. This is reflected in the output of the trust model whereby the trust values for each of the models are close but it does show differences in the level of performance of the models which is difficult from simply looking at a plot of the outputs (See Figure D.2). We also see the ability of TM to capture the best performing nodes even in times of noisy output, since the

performance of the models here is generally lower than that achieved by the models for conductivity or depth (See Tables D.1 and D.2).

The output of the framework for temperature prediction demonstrates how the reputation and trust framework is accurately reflecting the type of method that is being implemented to determine cooperativeness or reliability, in that models which correlate highly with the oscillations of the temperature sensor but at a higher range are deemed to be less reliable than models that do not correlate as well with these oscillations but produce closer values to the actual output values of the temperature sensor. Again, it is also apparent here that the trust and reputation model allows us to pick up trends which are difficult to decipher from simply plotting the data and also to see the difference in the magnitude of trust between models (See Figures D.8, D.9 and D.10). The TrustModel matches the best performing models in terms of MAE's as opposed to CC's while sometimes improving CC (See Tables D.3 and D.4).

7.3 Application scenario 2 - Satellite Sensor Information

Another application of RFSN is for determining the most reliable satellite data analysis product for incorporation into data fusion from multiple sources at a particular site. From Chapter 5 it is apparent that there are a number of issues with the use of satellite remote sensing data such as the reliability of algorithms being site-specific, the accuracy of estimation of chlorophyll can be difficult in coastal areas, pixels near land can be contaminated by land borders, temporal or spatial resolution may be low, cloud cover is a problem for certain sensors, etc.

In Chapter 5, we described the selection of appropriate remote sensing data products for estimating chlorophyll and SST, for use in this work. We then described the extraction of data at two sites in Galway Bay, the Mace Head and MidBay test sites. For chlorophyll, Globcolour data products and an Ifremer data product were

selected. For SST, a variety of HRDDS products are available for download at a variety of resolutions. Seven of these were selected for use in this study.

A similar analysis is carried out here to that in Section 7.2 for the Lee Maltings site. In this case, the reliability of the satellite products is estimated in terms of their closeness to the output values of the in-situ sensors. Hence this is similar to application scenario one, albeit with different input data streams but is used to stress the RFSN model. Our analysis could only be carried out on a selection of the HRDDS SST data products which provided daily satellite readings. Unfortunately there was only one chlorophyll satellite data product which provided regular estimates (Ifremer) which rendered this data unsuitable for this particular study.

7.3.1 SST at the MidBay and MaceHead Test Sites in Galway Bay

As previously outlined, data for MidBay and MaceHead was extracted from seven HRDDS data products. In order to align the in-situ and remote sensing datasets, the in-situ SST values from the Mace Head and MidBay SmartBay buoys were averaged. Available in-situ and satellite data from each the data products data for 2009 were subsequently aligned and incorporated into the analysis. From examining the output of each of the satellite products, three of the products appear to produce regular data with four of the products regularly producing a value which is indicative in the product of an erroneous measurement, or no value at all. Hence the three products that produce regular data are used for the analysis.

7.3.1.1 Application of the Trust and Reputation-Based Framework

The objective here is to examine how much the satellite data products coincide with the in-situ sensor. Hence we use a similar method of calculating the level of cooperation for a data product as we did for the in-situ models in Section 7.2. Again the *Watchdog* assigns a level of cooperation to each model based on the deviation

of its output value from the output of the in-situ sensor (i.e. p value as described in Section 7.2.1.2). The calculation of this value is described in Equation 7.3. This value is within the range of [0..1], with higher values indicating greater cooperation. As seen from previous scenarios, normalisation affects the magnitude in difference of reliability between the different data streams and the height of the trust values. From examining the outputs of the data streams for 2009, the difference between the in-situ values and the HRDDS data analysis products may be quite small and is always under 3 Kelvin. Hence in order to capture the greater magnitude of differences in the reliability of each of the satellite data streams, the differences were normalised to a range of [0..1] by dividing by 3. Similar to the previous application scenario, the output of the *Watchdog* is used to update the *Reputation* component and there is the subsequent calculation of trust (See Equation 7.1 on page 177 and Equation 7.2 on page 177) . Again if there is no output value of the sensor compared to the output value of the product to at a particular epoch, the reputation and trust information remains unchanged. Similar to the previous application scenario the notion of 'trust' or 'reliability' is in terms of its correspondence with the in-situ readings. Also similar to previously, all nodes begin with an initial trust value of 0.5.

$$p = 1 - \frac{Abs(DataProductValue - In - situSensorValue)}{Normalisation} \tag{7.4}$$

7.3.1.2 Results

In-situ and satellite SST measurements from the three HRDDS data products for Mid-Bay and MaceHead are shown in Figure 7.10. From both these graphs, HRDDS product 3 appears to be more closely aligned with values from the in-situ sensor than products 1 or 2. This is reflected in the output of the trust and reputation model shown in Figure 7.11 which also demonstrates the magnitude of variation between the three products and the height of trust assigned to each. While products 1 and 2 seem to produce similar outputs, as reflected in the trust values, product 3 seems

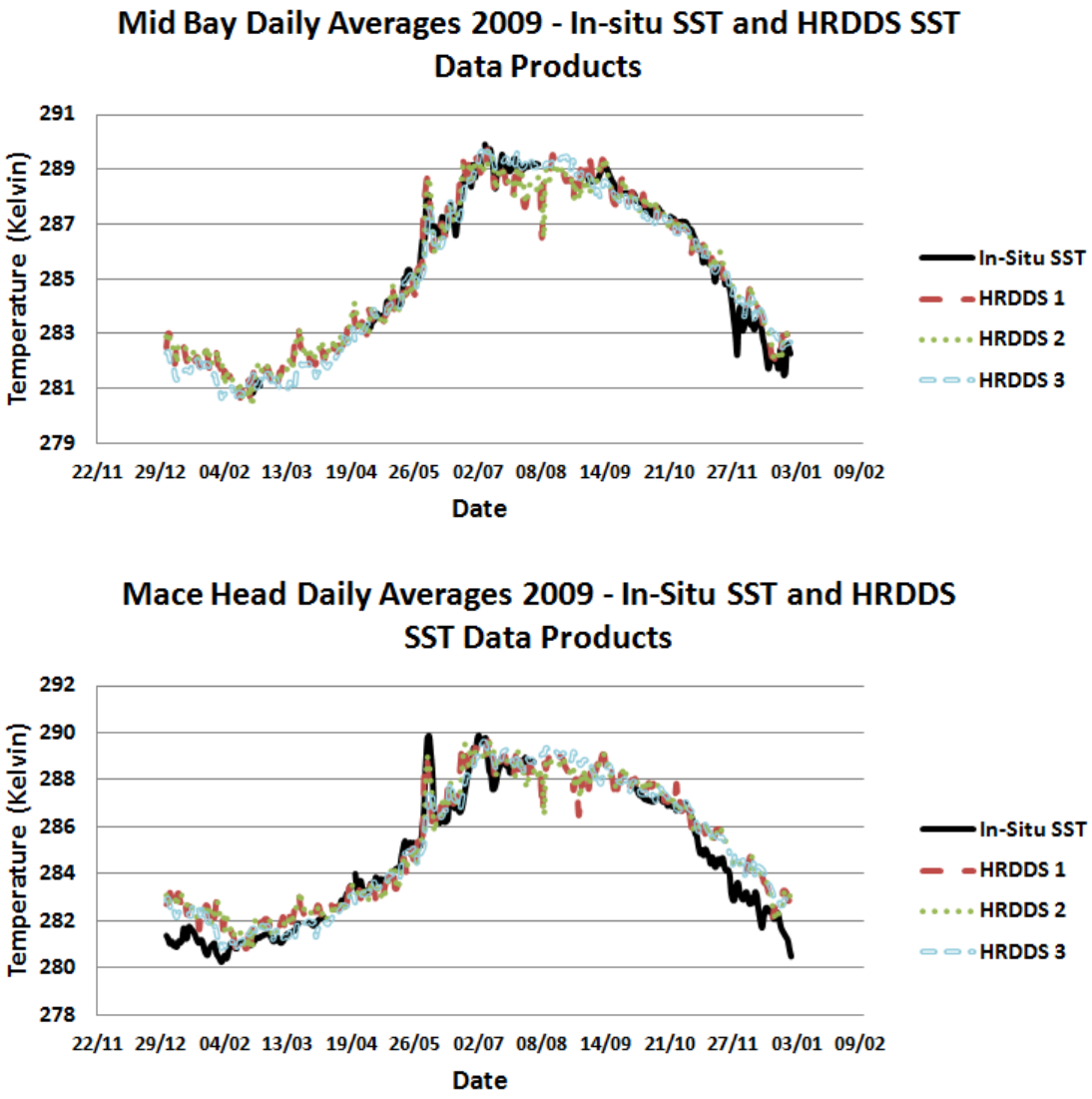

Figure 7.10: Graphs shwoing the average daily SST values for the in-situ sensor and for three HRDDS products at the Mid Bay and Mace Head test sites in 2009

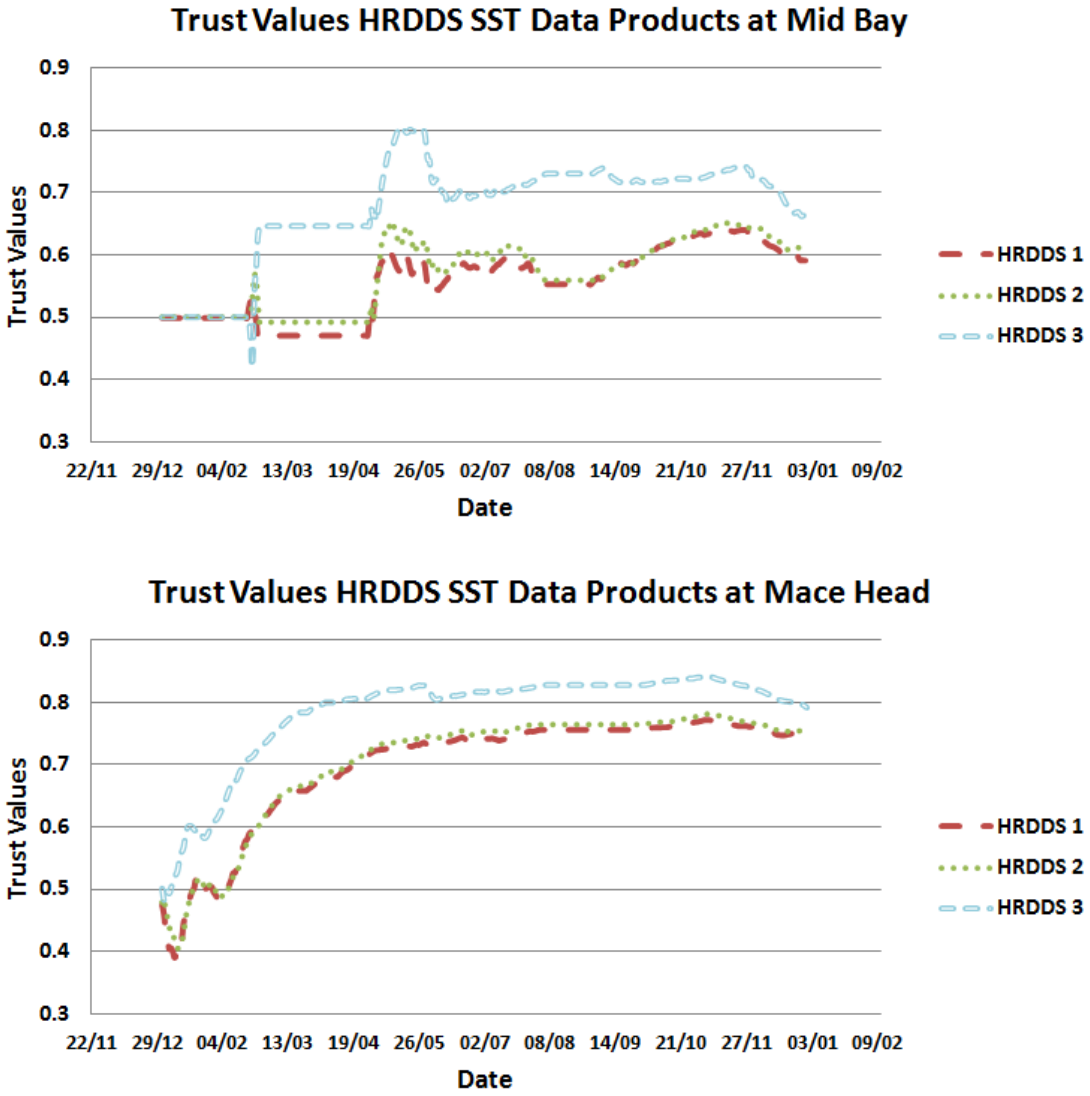

Figure 7.11: Graphs showing the trust values for the three HRDDS products at the Mid Bay and Mace Head test sites for SST outputs in 2009

to be well differentiated from these.

There are many gaps in the in-situ sensor data from the Mid-Bay SmartBay buoy in 2009. This can be seen in Figure 7.10 and it is also reflected in the trust values for the satellite products at Mid-Bay (See Figure 7.11) through a flattening of the curve (trust values of the products are not updated when there is no output value of the sensor to compare the output value of the product).

It is also interesting to note that at both the Mid-Bay and MaceHead sites, there is a dip in SST values from HRDDS products 1 and 2 which is not reflected by HRDDS product 3 around August-September. This dip in SST values appears to be slightly more significant at the Mid-Bay site. There is a gap in the in-situ data at both sites also around this time period. However at the Mid-Bay site, before this gaps occurs, this reduction in SST values is not reflected by the in-situ SST sensor resulting in a reduction in the trust values for products 1 and 2 whereas the trust value for product 3 increases. There is then a subsequent flattening of the trust values due to the gap in the data from the in-situ sensor.

Mace Head	CC	MAE	RMSE
HRDDS Product 1	0.9603	0.7683	0.9632
HRDDS Product 2	0.9618	0.7523	0.9494
HRDDS Product 3	0.9648	0.6165	0.8403

Table 7.8: Correlation (CC), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values for HRDDS products compared to the in-situ SST sensor at Mace Head

MidBay	CC	MAE	RMSE
HRDDS Product 1	0.9781	0.407	0.5238
HRDDS Product 2	0.9808	0.3982	0.5084
HRDDS Product 3	0.9845	0.3353	0.4719

Table 7.9: Correlation (CC), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values for HRDDS products compared to the in-situ SST sensor at Mid Bay

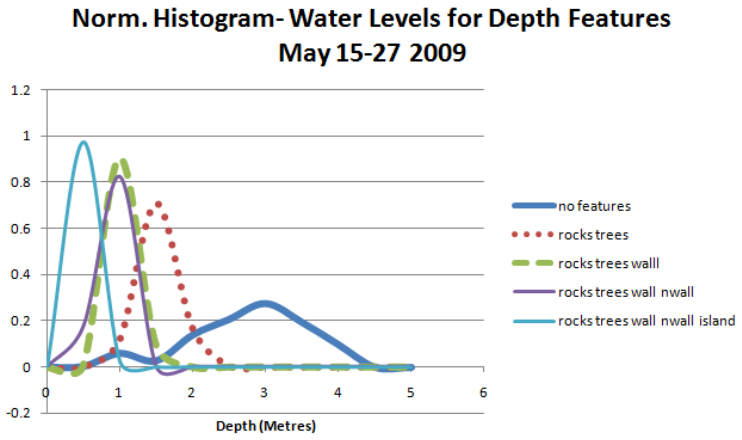

Figure 7.12: Normalised histogram of water levels for depth features for a selection of images from May 15-27 2009

The relationship between each of the products and the in-situ SST sensor values from the the Mace Head and Mid-Bay SmartBay buoys in terms of correlation, MAE and RMSE can be seen in Tables 7.8 and 7.9. These values generally reflect the analysis above. From examining these products through the Beam Toolbox, it is clear that product 3 has a higher resolution to the other products reflecting its increased performance in relation to the in-situ readings.

7.4 Application scenario 3 - Visual Water Depth Estimation

In Chapter 4, we examined the estimation of water depth using a visual sensor through the detection of certain features in the image such as rocks at trees, rocks at the far wall, rocks at the near wall and the appearance of an island like feature in the middle of the water. As previously outlined, generally as the water depth is lowering, these features begin to appear in the order provided above. Hence as these features appear it can provide us with an estimation of water depth. However some manner in which to relate this information back to the the depth values is required.

Figure 7.12 shows the range of water depth values corresponding to the various

	Accuracy	TP Rate	FP Rate	F-Measure	ROC Area
rocks at trees	85.4%	0.854	0.145	0.852	0.854
rocks at far wall	38.93%	0.389	0.417	0.367	0.487
rocks at near wall	91.76%	0.918	0.019	0.924	0.949
island	94.49%	0.945	0.167	0.947	0.889

Table 7.10: Performance metrics for each of the models for classifying the appearance of the four depth features from May 1-7 2009

features where *rocks-trees* indicates rocks at trees only *rocks-trees-wall* indicates rocks at trees and the far wall but no rocks at the other two features etc. From this graph it can be seen that there is a distinction between the range of depth values associated with different depth features, except for the features *rocks-trees-wall* and *rocks-trees-wall-near-wall* which occur in similar ranges. This distinction can be used along with the application of a trust and reputation framework in order to assess the performance of our visual sensing tool or our in-situ sensors. The following describes the methodology, the application of a trust and reputation framework and the outcome.

7.4.1 Classification of Depth Features

The models developed in Chapter 4 using data from May 15-27 2009 were used for classifying images from May 1-7 2009. Classifications were carried out by each of the four models for detecting each of the features on over 5,400 images from this time period (resulting in approximately 22,000 classifications). The overall results of these classifications for each of the models can be seem in Table 7.10. These models performed extremely well except for the model for detecting rocks at the far wall. From analysing the output it appears that it incorrectly classifies many of the negative instances as positive. However during this time period, there were often shadows or a brown reflection on the water which left it very visually indistinct from when rocks were present. This could possibly explain the poor performance of the classifier for discriminating between the two classes. When evaluating the

performance of the classifier in Chapter 4, there is an equal amount of positive and negative instances. In this scenario there a large majority of negative instances. It should be noted also that in each of the instances a 'slight' detection of any of the features as described in Chapter 4 is regarded as a non-detection.

7.4.2 Alignment of the Data Streams

Following the classification of each of the images, these then need to be somehow aligned to data from the in-situ depth sensor. An in-situ depth reading is taken every 10 minutes, where as an image is produced from the camera sensor approximately every minute. Hence for each in-situ depth reading occurring in the May 1-7 time period, the images (i.e. the classifications for each of the depth features) that are within 10 minutes of this reading were aligned with the time of the depth value. Two approaches were then used in order to decipher the appropriate classification for each of the depth features for this particular time. In the first scenario a *max* approach is used whereby if there were any positive instances at all within the classifications, then the classification for this time is considered to be positive. In the second scenario, a *majority* approach is used whereby the classification for this time for a particular feature is the majority classification of all the classifications aligned with this time. It is assumed that the *majority* approach will be the best considering that one poor classification in the other approach may result in an error. The output of this part of the methodology consists of two arrays consisting of the timestamp, the depth sensor value, and the classification of each of the depth features using the *max* approach or the *majority* approach. For each of these arrays a similar array is also output where the actual annotations of each of the images is used as opposed to the classifications when carrying out either of the approaches.

- timestamp, sensor-value, trees-max, rocks-wall-max, rocks-nearwall-max, island-max
- timestamp, sensor-value, trees-major, rocks-wall-major, rocks-nearwall-major,

island-major

7.4.3 Application of Trust and Reputation-Based Framework

Following the alignment of the sensor streams, we are left with a sensor value and classifications for four different depth features, so how do we establish a cooperative relationship between the two outputs for future determination if they are in agreement and how do we examine how well each of these visual sensor streams are cooperating with each other? Ideally after the appearance of one feature we should see the subsequent appearance of the next appropriate feature relevant to whether depth is moving up or down. Also how do we examine which approach to use when deciding on the most suitable classification for a particular time or which depth features are affecting the algorithm?

The application of a trust and reputation framework can help with these issues. This again involves a *Watchdog* component and a *Reputation* component. In this case the *Watchdog* carries out a binary rating with regards to 'cooperation' as opposed to the interval rating using in the previous application scenarios (i.e. in this case the cooperation value or p value (as outlined in Section 7.2.1.2) receives a value of 0 or 1 as opposed to a value in the interval [0..1]). A thresholding approach is used to determine this cooperation value. From examining the training data, it is apparent that if the water is below or between certain levels, certain features should be appearing and others not appearing. Two algorithms are used in the *Watchdog* – one algorithm incorporates all depth features, the other algorithm leaves out the *rocks at far wall* depth feature since the output of its classification is so poor. The algorithm is quite simple: if there is water between a certain level and the appropriate relevant features are detected and there is no detection of the non-relevant features, then a positive cooperation value of 1 is assigned, or else a negative cooperation value of '0' is assigned to this input. Reputation and trust are updated using

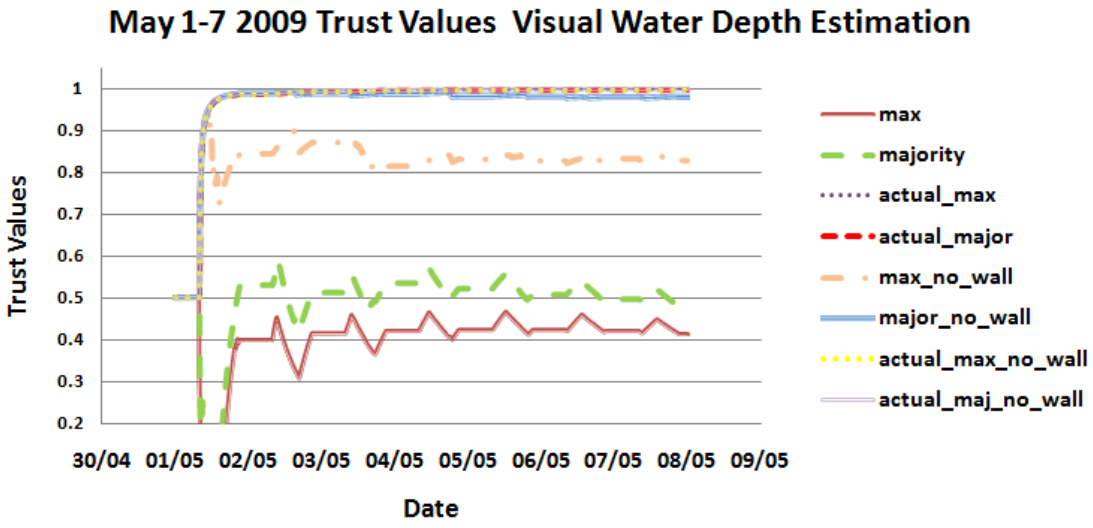

Figure 7.13: Output of the trust and reputation framework for each of the visual sensing models

this output as in other application scenarios outlined earlier in this chapter (See Equation 7.1 on page 177 and Equation 7.2 on page 177). In this context the notion of 'trust' or 'reliability' is the cooperation between sensing modalities in terms of outputs.

7.4.4 Results

Figure 7.13 shows the trust values for eight visual sensing models. These models are as follows:

- *max* this is the output of the algorithm used in the Watchdog when the max approach is used for determining the inputs to the Watchdog component;
- *majority* is when the majority approach is used for determining the inputs to the Watchdog component;
- *actual-max* and *actual-major* are when the max and majority approaches are used on the true annotations for the images as opposed to the classifications (these are used as benchmarks);

- The next four models are all of the above except evaluation of the appearance of the depth feature *rocks at wall* is left out of the *Watchdog* due to its poor classification accuracies.

It is apparent that algorithms using the actual values produce maximum trust values demonstrating that our algorithm in the *Watchdog* is appropriate in determining the relationship between the in-situ and visual sensor streams. It is very promising that our visual model using classifications from our depth feature models is highly effective when using the *majority* approach but leaving out the underperforming depth feature. The incorporation of this feature into the *Watchdog* algorithm has a huge impact on establishing a working cooperation between the outputs of the two sensing modalities. Overall the *majority* approach appears to be the most effective for determining input to the *Watchdog*.

7.5 Conclusion

In this chapter we adapted the RFSN trust and reputation model to deal with the unreliability associated with the additional data streams to be used in a multi-modal environmental monitoring network introduced in Chapters 4-6. This is a novel approach to optimising the use of these additional data streams in the network, enabling us to choose the most reliable output at a particular point in time, or during specific events.

We applied the modified RFSN model to three specific application scenarios to stress test its resilience across three very different applications in environmental monitoring. In the first application scenario we adapted this framework to determine the most reliable model for the prediction of in-situ parameters at the Lee Maltings site. We provided an analysis demonstrating the difficulty in determining the performance of a model on un-seen test data based on the results on an evaluation. This demonstrated the need for a real-time update tool for determining the most appropriate model at a particular point in time or during a specific event.

The trust values output by the modified RFSN model reflected the varying abilities of the the different parameter combinations to replicate the activities of the in-situ sensor in question and captured the most reliable models across time.

In the second application scenario we adapted this framework for determining the most reliable satellite data analysis product at the Mid-Bay and Mace Head test sites in Galway Bay. However some of the satellite data products were not suitable for use within such a framework since it requires regular data readings. Hence only a selection of the HRDDS SST data products were used for this analysis. There was only one chlorophyll satellite data product which provided regular estimates, therefore rendering these data products unsuitable for use within the framework. However we saw very promising results for the use of this modified RFSN model for determining the most reliable satellite data products to use at a particular site, where there are regular estimates produced by the product and the in-situ sensor.

Finally in the third application scenario we applied the RFSN model to the visual sensing streams. It helped us to demonstrate the appropriateness of our algorithm in the *Watchdog* for determining the relationship between the in-situ and visual sensor streams and subsequently to determine the most reliable algorithm for relating the two sensing modalities. Subsequently it may also be used to indicate a problem with the in-situ sensor data or visual sensing tool if there is a lack of correspondence between the output of the two sensing modalities.

Overall each of these application scenarios required a tailoring of the RFSN model but each was faithful to the RFSN model in that each included a *Watchdog* and a *Reputation* component, though the interpretation of what constituted each varied depending on the application. Additionally the application of this model allowed us to decipher trends which are difficult to determine otherwise from a simple plot of the outputs. For example, it can inform us of the magnitude of reliability of one model in comparison to another. In the next chapter we move on to event detection where we bring together the outputs of all the sensor sources, plus the execution of the RFSN model, to detect events in the environment. We examine the usefulness

of the RFSN model in improving the output of the network through carrying out a case study in relation to the detection of depth events.

Chapter 8

Event Detection

In preceding chapters we carried out investigations into the use of a wide range of sensing modalities for complementing and enhancing the use of an in-situ sensor network. We investigated both the singular and integrated use of a variety of these data sources with a view to improving the scalability and reliability of an environmental monitoring network. We also modified the RFSN trust and reputation framework for optimising the use of a variety of these models in the network, with a view to selecting the most reliable model at a particular point in time. If an in-situ sensor fails in the network, this enables the selection of the most reliable model or data stream to replicate its behaviour at a particular point in time or during a specific event.

In this chapter we focus on the performance of this multi-modal network from an applications perspective. Firstly we recognise that the needs of different applications vary. Some require high precision in relation to output values. Others may only require a high-level overview of conditions and are much more tolerant of an error margin. In response to this we carry out an analysis of a variety of the models and sensor streams in relation to various threshold values of the in-situ sensors. We examine the percentage of their outputs that are within various margins of the in-situ sensors in order to capture their possible suitability over a range of application contexts.

Following this we carry out a case study in order to examine the effectiveness of a multi-modal network from a specific applications perspective. In the first part of this case study we examine the ability of these alternative sensing modalities including the models developed in previous chapters to compensate for the failure of a node in the in-situ network. In the second part of the case study, we equip the network with a trust and reputation framework to select the most reliable model or sensor stream at a particular point in time and we investigate if this improves performance.

Our case study relates to the detection of depth events at the Lee Maltings site. Depth is chosen for two main reasons. Firstly it is an important parameter and it can provide an indication of overall conditions at the site. Secondly the visual sensor was modelled to provide an estimation of depth and hence it provides another sensor stream for analysis. We examine the performance of the in-situ data models and the visual sensor data streams in detecting a series of high and low depth events at the site. Following this we carry out an evaluation of the performance of the network in detecting these events when it is equipped with the modified RFSN model described in Chapter 7. The analysis carried out in this chapter relates back to research question 6 in Chapter 1.

8.1 Sensor Stream Performance

We begin by examining the outputs of the models described in Chapter 6 which use values from heterogeneous in-situ sensor nodes for predicting the value of an alternative in-situ node in the network. Then we examine the outputs of seven different HRDDS products providing an analysis of sea surface temperature (SST). Finally we look at the output of four different products providing an analysis of chlorophyll – one Ifremer product and three GlobColour products.

8.1.1 Modelled Outputs at the River Lee

In Chapter 7, we provided various analyses of the performance of these models on test data from five different time periods. For reference purposes, Table 7.2 on page 182 presents the statistical properties of the datasets used for training and testing purposes. Table 7.1 on page 178 describes the components of each of the models for the prediction of the four parameters (Depth, Conductivity, DO, Temperature) and Tables 7.6, 7.4, D.1 and D.3 on pages 194, 185, 279 and 289 provide the correlation, RMSE and MAE values produced by these models on the test data from the five months.

Here we provide an analysis on the outputs of the models predicting depth, conductivity, dissolved oxygen and temperature for two different times of year (June and February). The graphs depicting this analysis can be seen in Appendix E, along with graphs produced for the other months of May, July and January. For models predicting depth, the graphs shows the percentage of values within an error margin of 0-5m in steps of 0.01m, for conductivity the error-margin is from 0 to 48 ms/cm in steps of 0.1 ms/cm, for DO it is from 0 to 10 ppm in steps of 0.02 ppm, and the same range is used for temperature in degree Celsius. From these graphs it is apparent that a curve highly skewed towards the upper left-hand corner of the graph is desirable, indicating the more successful the model is at predicting values within a close range of the in-situ sensor. The faster the curve can converge on the higher percentage values the better. Variations can be seen in the performances of the models in this regard similar to how we saw variations in performance in previous analyses. The statistical properties of the datasets, which provide a reference point on which to evaluate performance, for example by showing the maximum and minimum values of each of the parameters, were previously provided in Table 7.2 on page 182.

Tables 8.1 and 8.2 show the threshold value for which 80% and 95% of model outputs lie within. Similar to the scenario described in previous chapters, model performance can vary between months and the overall performance in the prediction of a parameter can also vary between months. For example for depth prediction in

Depth	M1	M2	M3	M4	M5	M6	M7
May	**0.81**	1.3	1.82	0.84	**0.81**	1.39	0.83
June	0.5	2.11	1.96	0.53	**0.49**	2.13	0.5
July	**0.59**	1.3	1.46	0.63	0.63	1.64	0.63
Jan	1.48	1.77	1.97	1.41	1.42	1.73	**1.36**
Feb	1.11	1.44	1.44	**1.07**	1.08	1.19	1.08
Conductivity	M1	M2	M3	M4	M5	M6	M7
May	14.4	18.9	36.6	**6.4**	15.2	31.5	12.7
June	**5.7**	33.6	40.3	6.7	6.4	33.9	6.8
July	18.3	12.5	39.1	**9**	39.6	14.8	27.4
Jan	16.8	30.1	35.5	**16.3**	25.7	35.6	25.2
Feb	18.1	36.3	**8.5**	22.5	8.7	8.7	8.8
Dissolved Oxygen	M1	M2	M3	M4	M5	M6	M7
May	1.72	2.52	2	2.02	**1.52**	2.28	1.68
June	**2.08**	3.52	3.28	2.3	2.24	3	2.34
July	**2.06**	3.54	3.36	1.74	2.1	3.32	2.1
Jan	4.02	3.74	2.32	4	3.56	**2.28**	3.6
Feb	**1.06**	0.98	1.14	**1.06**	1.16	1.08	1.12
Temperature	M1	M2	M3	M4	M5	M6	M7
May	1.68	**1.58**	1.68	1.82	1.64	1.72	1.62
June	2.32	2.68	4.6	2.86	**2.28**	2.82	2.84
July	1.76	2.1	**1.56**	1.96	1.74	1.94	2
Jan	3.44	1.58	2.96	2.92	3.66	**1.56**	2.92
Feb	0.64	1.02	0.76	0.68	**0.62**	0.88	0.7

Table 8.1: Models - Threshold margin within which 80% of predicted values lie

May, 80% of values output by M1 are within 0.81 of the range of the value output by the in-situ depth sensor, whereas in June this value is even better at 0.5 and in July this value is 0.59. However in January and February this threshold is raised to 1.48 and 1.11. A similar scenario is seen for M4, M5 and M7. Threshold values appear to be higher across the board for depth prediction in January and February. In the case of conductivity, temperature and DO, model performance also varies across the five months. Given that applications vary in their requirements, some of these models display a very promising performance that would be suitable for many application contexts. The best performance for each parameter each month is highlighted in bold and underlined in Tables 8.1 and 8.2. Although threshold values generally rise for a 95% cutoff point, some impressive performances are notable especially for

Depth	M1	M2	M3	M4	M5	M6	M7
May	1.36	1.64	2.14	1.27	**1.07**	1.83	1.22
June	**0.67**	2.57	2.41	0.71	0.67	2.65	0.7
July	0.92	1.68	1.69	0.89	0.89	1.96	**0.83**
Jan	1.87	2.26	2.49	**1.66**	1.83	2.45	1.77
Feb	**1.4**	1.86	1.84	1.42	1.37	1.58	1.43
Conductivity	M1	M2	M3	M4	M5	M6	M7
May	25.7	27.8	39.8	**12.7**	33.8	37.6	35.6
June	**9.1**	39.1	42.2	14	13.4	37.9	14.2
July	23.3	19.9	40.6	**14.1**	41.3	33.2	35.8
Jan	**26.1**	31.1	36.6	24.4	30.7	36.7	29.5
Feb	35.3	39.8	32.1	35.7	**28.4**	32.7	28.9
Dissolved Oxygen	M1	M2	M3	M4	M5	M6	M7
May	2.22	3.1	2.78	2.56	**1.94**	2.84	2.04
June	**2.52**	4.38	3.98	2.62	2.82	3.58	2.84
July	2.64	4.92	4.36	**2.48**	2.54	4.44	2.62
Jan	4.6	4.4	**2.62**	4.68	4.12	2.72	4.08
Feb	1.3	**1.2**	1.38	1.26	1.34	1.28	1.3
Temperature	M1	M2	M3	M4	M5	M6	M7
May	2.06	**1.9**	2	2.3	1.94	2.02	1.92
June	**2.64**	3.12	5.4	3.46	2.68	3.46	3.5
July	2.26	2.52	**1.88**	2.42	2.24	2.38	2.42
Jan	3.82	**1.84**	3.32	3.3	3.92	1.92	3.3
Feb	0.96	1.32	0.96	0.94	**0.88**	1.1	0.92

Table 8.2: Models - Threshold margin within which 95% of predicted values lie

depth and also DO and temperature in February.

8.1.2 HRDDS SST Analysis Products

In this section a similar analysis is carried out on the following HRDDS products providing an SST analysis:

- HRDDS 1 - AMSRE-JAXA-dcu001
- HRDDS 2 - AMSRE-REMSS-dcu001
- HRDDS 3 - NAR17-SST-EUR-dcu001
- HRDDS 4 - NAR-METOP-A-EUR-dcu001

- HRDDS 5 - NCDC-L4LR-AVHRR-AMSRE-dcu001
- HRDDS 6 - NCDC-L4LR-AVHRR-dcu001
- HRDDS 7 - UKMO-L4HRfnd-GLOB-dcu001

In Chapter 7, we applied a trust and reputation model to the last three products listed above. The other four products did not produce regular estimations, hence they were not included in our analysis. For reference purposes, the relationship between the estimations of SST provided by these three products and the in-situ SST sensor values from the the Mace Head and Mid-Bay SmartBay buoys in terms of correlation, MAE and RMSE are outlined in Tables 7.9 and 7.8 on pages 206 and 206.

Figure 8.1 shows the percentage of outputs that are within various thresholds of the in-situ SST sensors at Mace Head and Mid-Bay for each of the HRDDS products. The graphs show the percentage of values within an error margin of 0-5 Kelvin in steps of 0.01. Similar to previously it is apparent that a curve highly skewed towards the upper left-hand corner of the graph is desirable. From these graphs we see that at both the Mace Head and Mid-Bay sites the three products producing regular outputs perform better than the other four products. These four products fail to converge on the 100% mark even within an error margin of 5 kelvin. At Mace Head, HRDDS product 2 (AMSRE-REMSS-dcu001) does not produce any valid output at that site for the year, hence a continuous value of zero is seen on the graph. At Mid-Bay, this is the case for both products 1 (AMSRE-JAXA-dcu001) and 2 (AMSRE-REMSS-dcu001). Thus these products are most especially not suitable for providing an analysis at the site in question.

Tables 8.3 and 8.4 show the threshold value within which 80% and 95% of the HRDDS values lie compared to that of the in-situ SST sensors at the Mace Head and Mid-Bay sites. The first four products have no outputs at these cutoff points within an error margin of 5 Kelvin. What is interesting to note is the lower threshold value for which 80% and 95% of analyses provided by HRDDS products 5, 6 and 7 at the

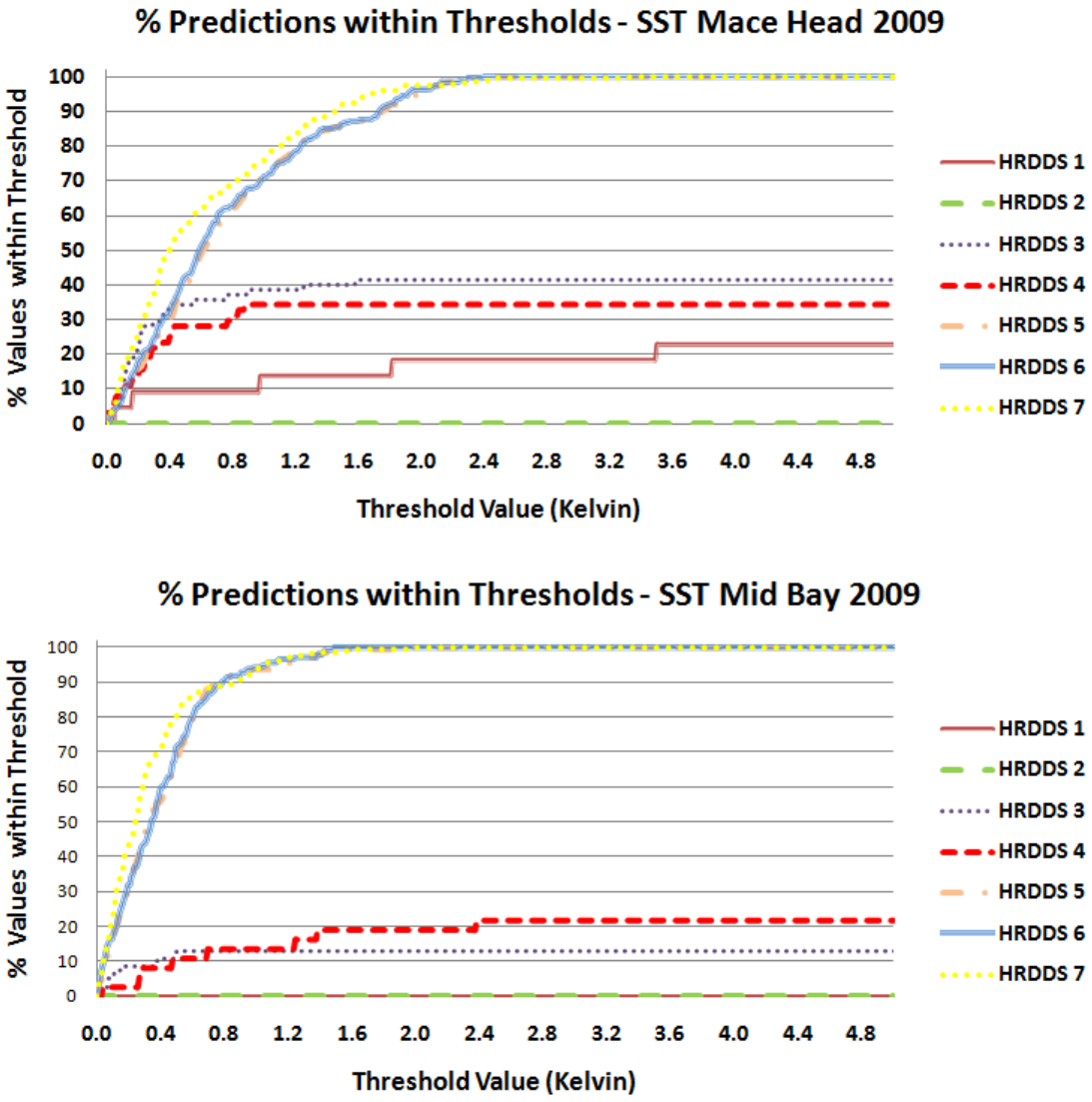

Figure 8.1: HRDDS SST products for 2009 at Mace Head and Mid-Bay - The percentage of values of values within various thresholds of the in-situ sensor reading.

Mid-Bay site compared to Mace Head. 80% of analyses at Mid-Bay are within 0.62, 0.61 and 0.51 Kelvin of the in-situ sensor. At Mace Head these values are raised to 1.23, 1.24, and 1.11. 95% of analyses at Mid-Bay are within 0.75, 0.79 and 0.87 of of the in-situ sensor. At Mace Head these values are raised to 1.76, 1.74, and 1.47. It is also interesting to note that HRDDS product 7 has the lowest threshold at Mace Head and Mid-Bay for an 80 % cut-off point. However at a 95% cut off point, it has the lowest threshold for Mace Head, but not at Mid-Bay, with HRDDS products 5

SST	H1	H2	H3	H4	H5	H6	H7
Mace Head	0	0	0	0	1.23	1.24	1.11
MidBay	0	0	0	0	0.62	0.61	0.51

Table 8.3: SST - Threshold value within which 80% of the HRDDS values lie compared to that of the in-situ SST sensors at the Mace Head and Mid-Bay sites

SST	H1	H2	H3	H4	H5	H6	H7
Mace Head	0	0	0	0	1.76	1.74	1.47
MidBay	0	0	0	0	0.75	0.79	0.87

Table 8.4: SST - Threshold value within which 95% of the HRDDS values lie compared to that of the in-situ SST sensors at the Mace Head and Mid-Bay sites

and 6 having lower thresholds. Thus at Mid-Bay, HRDDS product 7 is slower to converge on the very high percentage values.

As previously described in Chapter 5, temporal resolution can be problematic in relation to satellite data products. Hence it is interesting to look at the percentage of missing values attributed to each of the data products across the year. Table 8.5 presents these values which represent the percentage of days of 2009 where no valid data reading is available for the pixel which best represents the Mace Head and Mid-Bay sites. The percentage of days where in-situ data is unavailable is also provided. From Table 8.5 it is apparent that there are more days with no in-situ sensor readings in Mid-Bay than Mace Head. HRDDS products 6, 7, and 8 have a low percentage of days with no valid data reading. Whilst the other products have a very high percentage, with some even with no valid data reading available for the pixel which best represents the Mace Head and Mid-Bay sites in 2009.

8.1.3 Ifremer and GlobColour Chlorophyll Analysis Products

In this section a similar analysis is carried out on an Ifremer data product providing chlorophyll estimations and three GlobColour data products also providing

	In-situ	H1	H2	H3	H4	H5	H6	H7
Mace Head	22.47	92.33	100	73.42	78.9	4.11	2.74	1.64
MidBay	39.18	100	100	84.11	81.64	4.11	2.74	1.64

Table 8.5: SST - Percentage of days where data is unavailable across 2009

chlorophyll estimations. These products are:

- Ifremer - EUR-L4-CHL-ATL-v01
- GlobColour1 - AV-MER-CHL2
- GlobColour2 - AVW-MERMODSWF-CHL1
- GlobColour3 - GSM-MERMODSWF-CHL1

The first GlobColour product is a chlorophyll 2 data product (i.e. algorithm designed for estimation of chlorophyll in Case 2 waters). The other two are chlorophyll 1 data products but differ in the methodology used to combine data from the three different input satellite data products from MERIS, MODIS, and SeaWiFS.

As described in Chapter 7, the application of a trust and reputation framework was not particularly suitable for the chlorophyll data products since it is only the Ifremer product that produces regular data readings. Also the application of the trust and reputation framework in the context of this research assumes the trustworthiness of the in-situ sensor. Chlorophyll sensors can be notoriously unreliable when not maintained very regularly due to biofouling etc, thus rendering such an assumption difficult in such a scenario.

Figure 8.2 shows the percentage of outputs that are within various thresholds of the in-situ chlorophyll sensors at Mace Head and Mid-Bay for each of the four products. Tables 8.6 and 8.7 show the threshold value within which 80% and 95% of the Ifremer and GlobColour values lie compared to that of the in-situ chlorophyll sensors at the Mace Head and Mid-Bay sites. At a first glance it appears that the

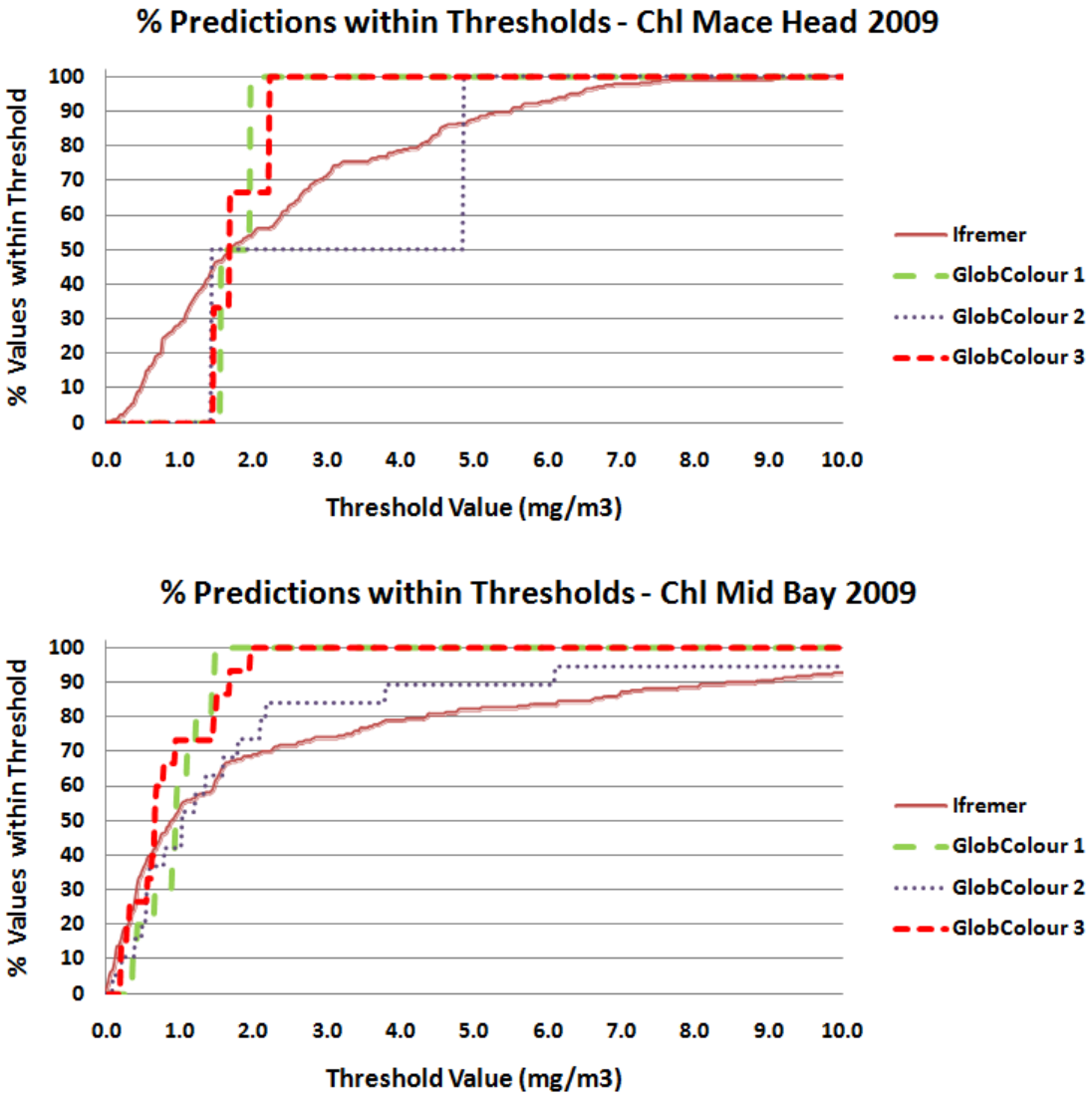

Figure 8.2: Ifremer and GlobColour products for 2009 at Mace Head and Mid-Bay - The percentage of values of values within various thresholds of the in-situ sensor reading.

Ifremer data product is the poorest performing. However unlike the case with SST readings from in-situ sensors which are generally quite stable, we often cannot trust the values from the in-situ chlorophyll sensor so it may be the case that it is the Ifremer data product is producing a more accurate analysis. Also even though the GlobColour data products appear to converge more quickly to the higher percentage values, if we look at the percentage of days of no valid values in Table 8.8, it is clear that the Ifremer has an extremely low rate, whilst the other products have extremely

Chl	Ifremer	Glob1	Glob2	Glob3
Mace Head	4.26	1.96	4.86	2.22
MidBay	4.34	1.22	2.18	1.46

Table 8.6: Chlorophyll - Threshold value within which 80% of the Ifremer and GlobColour values lie compared to that of the in-situ chlorophyll sensors at the Mace Head and Mid-Bay sites

Chl	Ifremer	Glob1	Glob2	Glob3
Mace Head	5.5	1.96	4.86	2.22
MidBay	8.42	1.44	6.1	1.68

Table 8.7: Chlorophyll - Threshold value for which 95% of the Ifremer and GlobColour values lie compared to that of the in-situ chlorophyll sensors at the Mace Head and Mid-Bay sites

high rates of missing values. The GlobColour products are obviously not particularly suited to monitoring at this particular point.

8.2 Case Study: Depth Detection

As outlined in the introduction to this chapter, event detection in relation to depth has been chosen as a case study in relation to our two main research objectives since it is quite an important parameter at the Lee Maltings site and we also have another sensing modality in the form of a visual sensor providing estimations of depth.

Before we move to the analysis, it needs to be determined what is meant by an event or how we classify an event in this context. Research has been carried out in determining events from time series data, or identifying points in a time series at which a behaviour change occurs e.g. (Guralnik and Srivastava, 1999). However in our context we are concerned with a more high-level analysis, and are only concerned with data readings that require management notification. Consider the case of high data readings, which may indicate that water levels are sufficiently high that caution needs to be taken in relation to the release of a dam further upstream. Also with

	In-situ	Ifremer	Glob1	Glob2	Glob3
Mace Head	22.53	0.55	98.63	99.18	98.90
MidBay	39.29	0.55	95.33	91.76	93.96

Table 8.8: Chlorophyll - Percentage of days where data is unavailable across 2009

the case of low data readings, this may indicate a warning for the passing of boats through this part of the river. Hence in the context of this case study, two types of events are denoted. A high water event is considered to be any data reading that is greater than or equal to 3m. A low water event is defined as being any data reading less than or equal to 1m. We had considered setting these thresholds to 3.5 and 0.5, however due to the fact there are times when the depth sensor rarely went over or under these values, the thresholds outlined are used in order to enable a better evaluation.

We first examine how each of these heterogeneous nodes in the multi-modal network would perform in picking up these events had the in-situ depth sensor failed. Secondly we examine if using the trust and reputation model to select the most trustworthy data stream at a particular point can lead to increased performance.

8.2.1 Performance of Heterogeneous Nodes

The performance of the models described in Chapter 6 are evaluated in terms of their ability to detect both types of events in terms of precision and recall.

Precision is a measure of the percentage of events detected by the system that are real-events. It is calculated using the following formula:

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{8.1}$$

Recall is a measure of the percentage of ground-truth events that were identified.

It is calculated using the following formula:

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{8.2}$$

For example if a model predicts 100 events and only 50 of these events are true events, it receives a precision score of 50%. On the other hand, if there exists 100 events to be detected and a model only predicts 50 of these events, then it receives a recall score of 50%. In this scenario, similar to other analysis carried out in this thesis, we use the in-situ sensor to provide the ground-truth since we are investigating how well a particular model can replicate its behaviour if it were to fail.

8.2.1.1 Event 1 - High Water Event

Models

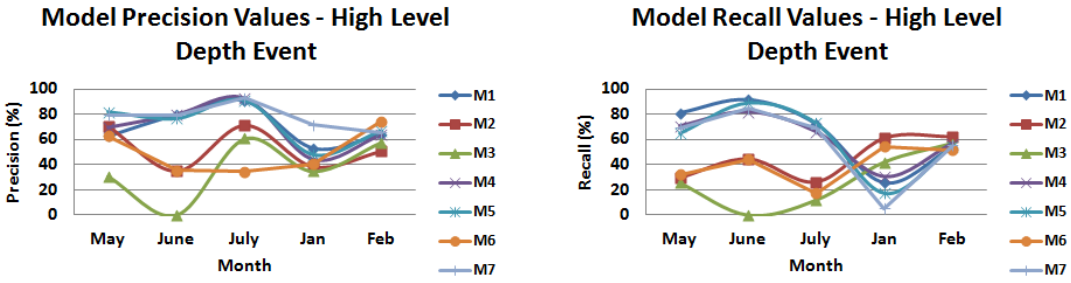

Figure 8.3: Precision and recall values for models M1-M7 in detecting high level water events

Figure 8.3 shows the precision and recall values for models M1-M7 for detecting high level depth events (i.e. depths greater than or equal to 3m) for each of the five test months. It is clear that overall performance in January is quite poor which has been discussed previously. Similar to previous analyses there is a clear difference in the performance of various models especially in May, June and July.

In terms of precision values, in May, it is M5 and M7 that produce the highest values of 81.88% and 79.31%. In June and July M1, M4, M5 and M7 all perform

very well, with values of 79.92% (M4, June) and 92.88% (M4, July) being achieved. In January and February it is M7 and M6 achieving the highest precision scores of 71.43% (M7, Jan) and 77.12% (M6, Feb). As can be seen from Figure 8.3 there is a large discrepancy between the highest and lowest precision values that can be achieved in a month, with M3 in June achieving zero precision. There is also variability in the performance of models across months, even though there are generally four models that stand out in terms of precision values, it is not always the same model that performs the best each month. This reinforces the notion that real-time model selection can be quite important.

In terms of recall, a similar scenario is seen whereby is generally models M1, M4, M5 and M7 producing the highest recall values. However it is interesting to note that in the lower performing months of January and February, it is the alternative three models that appear to be producing the highest recall values. The highest recall values are produced by M1 in May (80.85%), M1 in June (91.37%), M5 in July (72.97%), M2 in January (60.73%), and M2 in February (61.83%).

Visual Sensor Streams

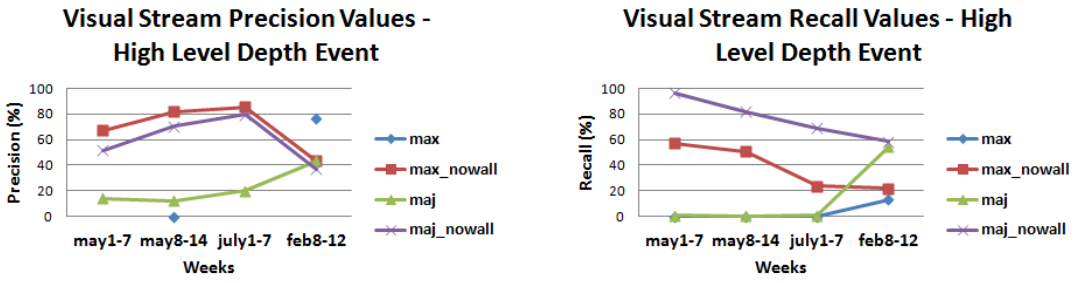

Figure 8.4: Precision and recall values for visual data streams in detecting high level water events

Figure 8.4 shows the precision and recall values for the visual sensor streams. The manner in which it was determined if a high level event had been detected by a visual sensor stream was if none of the visual features described in Chapter 4 were classified as being present. At readings great than or equal to 3m, none of these

features should be visible. In fact this happens somewhere in the over 2m mark. The different visual streams represent the different algorithms described in Chapter 7 — the *max* and *majority* algorithms. The *max-nowall* and *majority-nowall* represent the output of the two different algorithms but where the outputs for the *wall* feature is ignored. As previously outlined, the performance of this feature has been poor and hence it can corrupt classifications.

From Figure 8.4 it can be seen that in the weeks from May and July, it is the max and majority algorithms that do not consider the *wall* feature producing the highest precision and recall values. For these weeks, it is the *max-nowall* algorithm producing the highest precision values of 67%, 81.89% and 85.29% and it is clearly the *majority-nowall* algorithm producing the highest recall values of 96.58%, 81.95%, and 69.11%. In February, the *max* algorithm produces the highest precision value of 76.92%, and the highest recall value is produced by the *majority-nowall* algorithm of 58.44 %, closely followed by the *majority* algorithm with a recall value of 54.55%. In the other months, the *max* and *majority* algorithms do not really feature at all.

For example, in other months the *max* algorithm either produced a precision value of zero, or didn't produce any valid number since it had not detected any positive events at all. The poor precision value of the *max* algorithm makes sense since if any of the images taken within the timeframe of the in-situ sensor value were reported as positive, then this would not be detected as a high level event in this context. During these particular months the precision value of the *majority* algorithm is also quite poor, however never as bad as that for the *max* algorithm. Recall values for both algorithms are also very poor for these time periods with the *max* algorithm producing a recall value of zero for each of the three weeks and the *majority* algorithm producing a maximum recall value of 0.85%. Taking the *wall* feature out of each of these algorithms greatly boosts their performance with *max-nowall* having a greater precision and lower recall than *majority-nowall* due to the fact it is probably detecting more features and therefore less events.

8.2.1.2 Event 2 - Low Water Event

Models

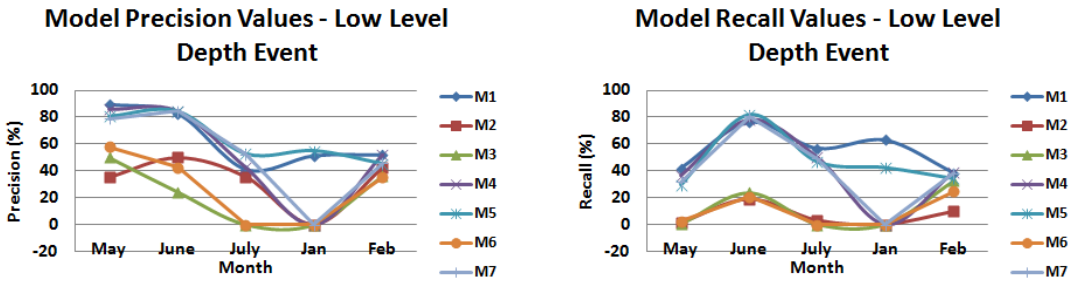

Figure 8.5: Precision and recall values for the models M1-M7 in detecting low level water events

Figure 8.5 shows the precision and recall values for models M1-M7 in detecting a low level depth event. Similar to the high level event, it is M1, M4, M5 and M7 that appear to be performing the best in terms of precision scores overall. The highest precision is 89.59% in May (M1), 84.65% in June (M7), 52.67% in July (M5), 54.96% in January (M5), and 51.9% in February (M4). There are scenarios in July and January where some models produce a precision score of zero. In terms of recall it is the same models generally producing the highest recall scores. The highest recall score achieved is 41.95% in May (M1), 81.73% in June (M5), 56.8% in July (M1), 62.94% (M1) in January, and 38.34% in February (M1 and M7). These recall and precision scores are overall not as good as those achieved for the high level event.

Visual Sensor Streams

Figure 8.6 shows the precision and recall values for the visual streams in detecting a low water depth event. A low water event is detected by the visual stream if either the *island* feature or the *near wall* feature is detected. Since the *wall* feature is not used in this algorithm there is no need for a visual data stream that does not consider this feature, hence we are left with two visual streams represented by the

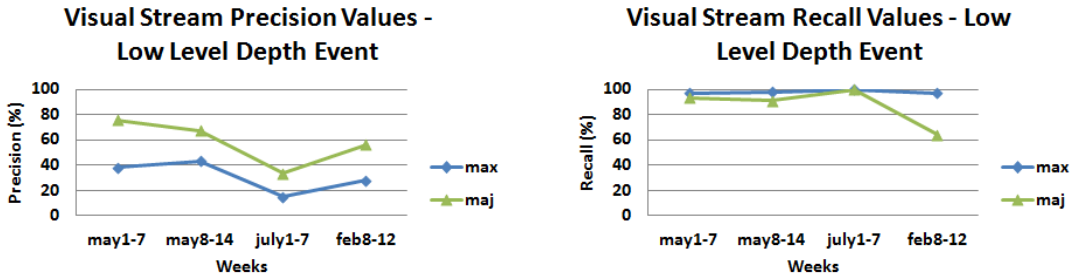

Figure 8.6: Precision and recall values for visual data streams in detecting low level water events

max and *majority* algorithms described in Chapter 4

As can be seen from Figure 8.6, these algorithms perform exceptionally well in terms of recall. However compared to the high level event where at least three features had to be classified correctly, only one feature has to be classified correctly in this scenario. This probably explains the lower precision values. It is the opposite scenario also in that it is the *max* algorithm that is producing higher recall and lower precision values which is reasonable since this event relies on the visual streams positively detecting a feature as opposed to not detecting a feature at all and the max algorithm is more inclined to positively detect a feature. The highest precision values for each of the weeks are 75.78%, 67.40%, 33.33%, and 56.10%. These are all produced by the *majority* algorithm. The highest recall values for each of the weeks are 97.12%, 97.76%, 100%, and 87.22%. These are all produced by the *max* algorithm, although the 100% recall value in July is matched by the *majority* algorithm.

8.2.1.3 Overall Performance of the Models and Visual Sensor Streams

Overall the highest precision and recall values that can be achieved by the models and the visual sensor streams are very satisfactory for the high level event, with improvements perhaps desirable in January and February, even though the highest values achieved may be suitable for many application contexts. The low level event resulted overall in lower precision and recall scores from the models. However, it

resulted in higher recall scores from the visual sensor streams at the expense of precision.

In some cases alternative models or visual streams achieved the best performance in the lower performing months (i.e. January and February for the models, and January for the visual streams) compared to those generally producing the best performance in the higher performing months. In general it is apparent how there is a clear difference between the highest and lowest performing models and that this can change between months. Even though there is generally a set of models that perform best, this can change from month to month along with the model producing the best score. Hence a real-time update tool on the performance of models would be quite beneficial. We now address this through investigating the benefits of incorporating a trust and reputation model for real-time selection of the most appropriate model.

8.2.2 Addition of a Trust and Reputation Framework

There are some issues that need to be taken into consideration before we incorporate a trust and reputation model for selecting the most suitable model or visual sensor stream to replicate the in-situ sensor. Firstly the trust and reputation model that was described in Chapter 7 was not optimised for the events considered here. Rather it was optimised to regard the models producing values closest to the in-situ sensor values to be the most trustworthy. However models largely overestimating and underestimating the depth curve may produce better event detection results in this scenario. Hence it is interesting to see how a trust and reputation model not optimised for the specific task at hand performs in this scenario.

Secondly the time point at which the in-situ sensor fails determines the particular model that is chosen i.e. the most trustworthy model at that particular point in time, reflecting its ability to cope with current trends at the site. Hence the analysis of the performance of the trust and reputation model is sensitive to the particular time epoch that the most trustworthy model is chosen. Once the in-situ sensor fails,

	June		Jan		Feb		General	
	CC	MAE	CC	MAE	CC	MAE	CC	MAE
M1 (Cond)	0.9755	0.1765	0.9045	0.3338	0.9365	0.2606	0.9305	0.3051
M2 (DO)	0.8425	0.411	0.8484	0.3861	0.7698	0.526	0.6921	0.6104
M3 (Temp)	0.8261	0.4425	0.7679	0.4868	0.8117	0.4943	0.6336	0.6799
M4 (Cond-DO)	0.9803	0.1581	0.9427	0.2787	0.9462	0.2422	0.9659	0.2141
M5 (Cond-Temp)	0.9818	0.1477	0.9272	0.2916	0.9393	0.2507	0.9598	0.2293
M6 (DO-Temp)	0.9202	0.2583	0.9158	0.2785	0.9314	0.2817	0.8291	0.4685
M7 (Cond-Temp-DO)	0.983	0.144	0.9555	0.2536	0.946	0.2395	0.9743	0.1845

Table 8.9: Results of 10-fold cross validation of each of the training datasets for depth prediction

the continued trust can no longer be calculated. Therefore our analysis proceeds as follows:

Selection of Models: From the results of the cross-validation evaluation carried out on each of the training datasets for depth prediction, the best performing models are chosen for use in the two scenarios described below (See *Random-model*, *Random-value*). In a real-world scenario, viewing the performance on test data is not possible hence models are chosen based on cross-validation evaluation. In Chapter 7 the generalised model is used for carrying out predictions on test data in May and July and the monthly models are used for carrying out predictions on test data from June, January and February. Hence we continue with this methodology here. From Table 8.9, it is clear that there are five models with correlations over 0.9 for January, February and June – M1, M4, M5, M6 and M7, and four for the generalised model – M1, M4, M5, and M7.

Evaluation Measure: In Section 8.2.1 the evaluation metrics of precision and

recall were used. Here performance is measured in terms of the F-measure which represents the harmonic mean of precision and recall, since it is desirable to maximise both. It is calculated using the following formula:

$$F = \frac{2 * precision * recall}{precision + recall} \tag{8.3}$$

Models to be Evaluated: We use the term epoch in order to refer to a time point for which there exists a data reading. For each test set from the five test months, the evaluation proceeds by selecting a window of 500 classifications at the first epoch in each test set, from the following different models:

- **M1**, **M2**, **M3**, **M4**, **M5**, **M6**, **M7**
- **Random-model**- Randomly selects one of the four or five best performing models outlined above at the first epoch of the evaluation and continues to use the values from this model at each epoch for the remainder of the window.
- **Random-value**- Randomly selects an output from one of the four or five best performing models outlined above for each of the 500 values at each epoch.

Trust Models to be Evaluated: The trust model is used to select the classifications output by the most trustworthy model at the start epoch. Based on analysis of the trust outputs from Chapter 7, it is clear that there are often a number of models that are the most trustworthy or that have trust values that are quite close. Hence a second approach takes the classifications of the top four trustworthy models at the start epoch and fuses their classification outputs in four different ways which are outlined below. Hence the *Trust Models* that are evaluated are as follows:

- **TM** – Selection of the most trustworthy of all 7 models M1-M7 at the beginning of the evaluation window, and subsequent evaluation of that models outputs across the evaluation window.
- **TM-max** – Selection of the top four most trustworthy of all 7 models M1-M7 at the beginning of the evaluation window, and selecting the maximum output

from each of these at each point in the evaluation window for subsequent evaluation (i.e. if any of the top 4 trustworthy models reports a positive output, the output is *positive* at that particular point in the evaluation window)

- **TM-min** – Selection of the top four most trustworthy of all 7 models M1-M7 at the beginning of the evaluation window, and selecting the minimum output from each of these at each point in the evaluation window for subsequent evaluation (i.e. if any of the top 4 trustworthy models reports a negative output, the output is *negative* at that particular point in the evaluation window)

- **TM-maj** – Selection of the top four most trustworthy of all 7 models M1-M7 at the beginning of the evaluation window, and outputting a positive result at each point in the evaluation window only if a majority of these agreed on this classification (i.e. 3 or more)

- **TM-maj-equal** – Selection of the top four most trustworthy of all 7 models M1-M7 at the beginning of the evaluation window, and outputting a positive result at each point in the evaluation window if 2 or more of the models have output positive.

Evaluation Methodology: As previously outlined, we use the term epoch in order to refer to a time point for which there exists a data reading and the evaluation proceeds by selecting a window of 500 classifications at the first epoch in each test set from each of the models outlined above. An F-measure is calculated for each of the models based on their classifications. The window then moves forward 1 epoch in the test set and selects the following 500 classifications and subsequently another F-measure is calculated for each model. This continues until an epoch is reached in the test set where there are not 500 epochs following. The evaluation subsequently terminates and results in a series of F-measures for each of the 14 models for each of the test sets. A window of 500 was chosen as this seems to represent a sufficient amount of time both to contain a large enough number of events and also to determine the performance of the trust models when the in-situ

sensor is offline for a long time period and updated trust values subsequently cannot be calculated. A window of 500 values corresponds to approximately 4 days. The particular time that an in-situ sensor goes offline determines the model chosen, hence the evaluation over a series of 500 value windows.

For the visual sensor a similar evaluation is carried out using all four streams outlined for the high level event in Section 8.2.1.1 and both streams outlined for the low level event in Section 8.2.1.2. However in this scenario a window size of 150 is used as there are less values due to the fact that the visual sensor does not produce estimations at night. However a value of 150 represents approximately two days, considering hours of darkness are not being evaluated.

8.2.2.1 Models

	M1	M2	M3	M4	M5	M6	M7	Rand-mod	Rand-val
May	0.7391	0.3492	0.2889	0.7293	0.7222	0.3757	**0.7435**	0.7342	0.7343
June	*0.8405*	0.3154	0	0.8213	0.8232	0.3192	0.8243	0.7244	0.7276
July	0.8266	0.4178	0.1672	0.8189	0.8178	0.1964	**0.8354**	0.8229	0.8266
Jan	**0.4045**	0.3305	0.2335	0.366	0.308	0.3347	0.1314	0.3062	0.3371
Feb	0.4472	0.4883	0.4827	0.4460	0.4553	0.4945	0.4412	0.4591	0.4604

Table 8.10: Mean F-measure for each of the 7 in-situ data models in detecting a high water event

	TM	TM-Max	TM-min	TM-Maj	TM-Eq-Maj
May	0.7234	**0.7785**	0.3052	0.6242	0.7516
June	0.8195	0.8259	0.1392	*0.8408*	0.8272
July	0.78	**0.8890**	0.2720	0.8097	0.8645
Jan	0.3596	0.4508	**0.4565**	0.3683	0.4002
Feb	0.4054	0.4514	0.3654	0.4414	0.4407

Table 8.11: Mean F-measure for each of the 5 trust models using data from the 7 in-situ data models in detecting a high water event

As previously outlined the evaluation described in the previous section results in a series of F-measures for each of the 14 models. Subsequently we calculated

the mean F-measure for each of the models. The mean F-measure is the mean for 1505 epochs in May, 1788 in June, 1183 in July, 1591 in January and 1609 in February. Since the evaluation results in a series of F-measures corresponding to the performance of each model on each window of 500 values, a paired t-test is used to determine whether the series of F-measures from two models differ from each other in a significant way. We compare the output for the best performing data model with that of the best performing trust model, in order to determine that if there is a difference in the mean F-measure for the best performing trust model, whether this difference is significant. The null hypothesis in this context is that the mean of population 1 is equal to the mean of population 2 (i.e. $\mu_1 = \mu_2$) and the alternative hypothesis is that the mean of population 1 is not equal to that of population 2 (i.e. μ_1 ¡ μ_2). We use a two-tailed paired t-test where p ¡ 0.01.

Table 8.10 shows the mean F-measure for each of the models using no trust and reputation framework in detecting a high water event and Table 8.11 shows the mean F-measure for each of the models incorporating the trust and reputation framework in detecting a high water event. Despite not being optimised for this scenario and a window size of 500 epochs (over 4 days, in other words using the most trustworthy model without regular updates on its performance over 4 days), the trust and reputation model performs very well. Simply choosing the most trustworthy model generally performs better than the *Random-model* or *Random-value* approach (in 3 out of 5 cases) and produces output among the higher scores achieved by M1-M7 but never the best. However when we adopt a late-fusion approach whereby we combine the output of the top four trustworthy models, this results in an improvement in mean F-measure over the nine models not incorporating the framework in all cases except for February (i.e. 4 out of 5 cases). Using a two-tailed paired t-test as described above, this is a significant difference in all cases highlighted in bold (May, July and January). While there is an increase in June, this was not shown to be significant. Adopting a fusion approach increases the performance of the trust models in all cases. The fusion approach that performs the best is not consistent,

however *TM-max* produces the best score in 3 out of 5 cases, amongst the five trust models.

	M1	M2	M3	M4	M5	M6	M7	Rand-mod	Rand-val
May	0.4953	0.0185	0.0034	0.4582	0.3578	0.0275	0.3749	0.4233	0.4243
June	0.7975	0.2096	0.2125	0.8148	**0.8323**	0.2084	0.8267	0.6826	0.7225
July	0.4955	0.0678	0	0.4909	0.5237	0	**0.5417**	0.5101	0.5070
Jan	0.4154	0	0	0	0.3507	0	0	0.1781	0.2236
Feb	0.2366	0.0481	0.1589	0.2034	0.1560	0.0989	0.1711	0.1761	0.1736

Table 8.12: Mean F-measure for each of the 7 in-situ data models in detecting a low water event

	TM	TM-Max	TM-min	TM-Maj	TM-Eq-Maj
May	0.3578	0.4699	0.3577	0.3040	0.4180
June	0.8179	**0.8391**	0.1243	0.8087	0.8271
July	0.4513	0.5322	0.1643	0.4954	**0.5573**
Jan	0.2385	0.3394	0.1826	0	0.2828
Feb	0.1786	0.2186	0.1632	0.1310	0.1746

Table 8.13: Mean F-measure for each of the 5 trust models using data from the 7 in-situ data models in detecting a low water event

Table 8.12 shows the mean F-measure for each of the models using no trust and reputation framework in detecting a low water event and Table 8.13 shows the mean F-measure for each of the models incorporating the trust and reputation framework in detecting a low water event. Similar to the previous analysis in Section 8.2.1, performance is generally poorer for the low water event. Similar to the previous event, in 3 out of 5 cases simply choosing the most trustworthy model generally performs better than the *Random-model* or *Random-value* approach and produces scores mostly among the higher scores of M1-M7, however again never producing the highest score. However a fusion approach again appears to increase performance, with a trust model producing a better score than any of the models not incorporating the framework in 2 out of 5 cases. Again using a two-tailed paired t-test as described above shows that this is a significant difference in both cases. Again adopting a

fusion approach increases the performance of the trust models in all cases, with *TM-max* producing the highest F-measure amoongst the five trust models in 4 out of 5 cases.

Overall the performance of the trust models are very satisfactory since they have not been optimised for this purpose. They result in an improvement over models not incorporating the framework in 6 out of 10 cases, and a significant effect in 5 cases. The fusion approach clearly seems to perform better than simply selecting the individual most trustworthy model, since it increases the performance of the trust models in all cases. The fusion approach that performs the best is not consistent, however *TM-max* produces the highest F-measure amongst the five trust models in 7 out of 10 cases. Overall it appears that the trust models perform best in months where a selection of some of the individual models perform better. In January for example there is an F-measure of 0 for many models.

8.2.2.2 Visual Streams

	Max	Max-nowall	Maj	maj-nowall	Rand-mod	Rand-val	TM
May1-7	0	0.6218	0.0294	**0.6115**	0.3217	0.4334	**0.6115**
May8-14	0	0.5585	0.0116	**0.7494**	0.3174	0.4330	**0.7494**
July1-7	0	0.4009	0.0181	**0.7518**	0.2962	0.3716	**0.7518**
Feb8-12	0.1588	0.2092	**0.4123**	0.3654	0.2894	0.3188	**0.3654**

Table 8.14: Mean F-measure for each of the 7 visual stream models and the trust model in detecting a high water event

Table 8.14 and Table 8.15 show the mean F-measure for each of the 7 scenarios (*max*, *max-nowall*, *majority*, *majority-nowall*, *random-model*, *random-value*, *trust-model*) for the high level event and the 5 scenarios (*max*, *majority*, *random-model*, *random-value*, *trust-model*) for the low level event above for each of the test periods. The mean F-measure is the mean for 417 epochs in the first test week in May, 403

	Max	Maj	Rand-mod	Rand-val	TM
May1-7	0.5306	**0.7860**	0.6521	0.6247	**0.7860**
May8-14	0.5542	**0.7117**	0.6328	0.6199	**0.7117**
July1-7	0.2267	**0.4211**	0.3179	0.2941	**0.4211**
Feb8-12	0.3812	0.3495	0.3605	**0.3841**	**0.3495**

Table 8.15: Mean F-measure for each of the 5 visual stream models and the trust model in detecting a low water event

in the second test week in May, 247 in the July test week, and 126 in the January test week.

In this case for the high-water event the trust model is initialised with *majority-nowall* as the most trustworthy since this is clearly the best performing and there may be an initialisation situation during darkness where for many epochs all streams are on the same trust value of 0.5 before they have produced any values. In other words, if we begin calculating trust values during darkness, the image processing algorithm will be unable to detect image features and hence all visual streams will be initialised with the same trust value. Thus when choosing the most trustworthy model at the beginning of a time window in order to use its values for the rest of that time window, it won't necessarily be the most reliable visual stream if we use a random approach. This could be detrimental to the analysis here, especially if all visual streams are on the same trust value at the beginning of a series of windows, which is possible considering that in this analysis the beginning of the next window moves forward just one epoch each time and the most trustworthy model at this particular epoch is used to provide the measurements for the rest of the window. The low-water event is initialised with the *majority* as most trustworthy.

For both events, in all cases except for one (February) the trust model is consistent with the highest score achieved by the best performing visual sensor stream. It is clear that the trust model is choosing this best performing visual stream as the

most reliable at nearly all epochs. The best performing visual sensor stream here is generally consistent with the best performing in the results from Section 8.2.1.

8.3 Conclusion

In this Chapter we examined the performance of the multi-modal network from an applications perspective. Firstly in recognising that requirements vary in the context of different applications, we analysed the output values of a variety of the models and sensor streams in relation to various threshold values of the in-situ sensors. Some of the in-situ data models demonstrated excellent performance even within a 95% threshold margin of the in-situ sensors at the Lee Maltings. However there were variations in how the models performed across the five months the best performing model for each month varied. Overall each month, some of the models displayed a very promising performance that would be suitable for many application contexts.

In our analysis on the satellite data products, some of the HRDDS data products providing SST values also demonstrated excellent performance with 95% of values within extremely low margins of the in-situ SST sensors at the Mid-Bay and Mace Head test sites. However other products produced a much poorer performance in terms of this analysis. These tended to be products that had a low daily data availability. When analysing the chlorophyll data products, it is clear that the GlobColour products are not particularly suited to monitoring at this particular point. Hence incorporation of other products would need to be investigated. As previously outlined the purpose here is not to carry out a thorough investigation of all available products but rather to highlight the issues with the use of such products. It was difficult here to come to any conclusive results regarding the performance of the Ifremer product. The GlobColour products appeared to converge more quickly to the higher percentage values, however they had extremely low data availability compared to the Ifremer data product which had almost daily data availability. The unreliability of the in-situ chlorophyll sensors made it difficult to determine which

was the more accurate analysis.

Following this we carried out a case study examining the performance of the network in order to examine the effectiveness of a multi-modal network in detecting depth events at the Lee Maltings site. In the first part of this case study we examined the ability of these alternative sensing modalities including the models developed in previous chapters to compensate for the failure of a node in the in-situ network. In the second part of the case study, we equipped the network with a trust and reputation framework to select the most reliable model or sensor stream at a particular point in time so see if this improved performance.

For the detection of the depth events, overall the highest precision and recall values that can be achieved by the models and the visual sensor streams is satisfactory in the context of a variety of application scenarios. For the models there is a poorer performance in general for the low water level event. On the other hand this event resulted in higher recall scores from the visual sensor streams at the expense of precision. In general there was a clear difference between the highest and lowest performing models and changed between months. Even though there was generally a set of models that performed best, this often changed from month to month along with the model producing the best score. This demonstrated the benefits of a real-time update tool on the performance of the models.

Subsequently when we equipped the network with a trust and reputation framework (i.e. the modified RFSN model) we found that it resulted in an improved performance of the in-situ data models in 6 out of 10 cases with this difference being significant for 5 of these cases. Adopting a late-fusion approach improved the performance of the trust models in all cases. *TM-max* produces the highest F-measure amongst the five trust models in 7 out of 10 cases. For the visual sensor streams, for both depth events in all cases except for one (February) the trust model is consistent with the highest score achieved by the best performing visual sensor stream. These analyses and results lead us into the final chapter of the thesis, where we address our overall research conclusions.

Chapter 9

Conclusions

In this thesis we proposed that marine environmental monitoring applications would strongly benefit from the use of a multi-modal sensor network utilising visual sensors, modelled outputs and context information alongside the more traditional in-situ wireless sensor networks. We carried out investigations into the use of a variety of models and sensing modalities to improve the efficiency and effectiveness of an environmental monitoring network with very exciting and promising results. However it was also apparent that each of these alternative data sources can be unreliable for the detection of certain events. Subsequently we modified a trust and reputation model from the literature known as RFSN for optimising the use of these data sources in the network and choosing the most reliable model at a particular point in time. We evaluated the use of this unique approach for three diverse application scenarios. From this it was apparent that the modified RFSN framework could be used to reflect the behaviour of a model or data source over time. We carried out a case study in relation to the detection of depth events where we evaluated the performance of the network both with and without the RFSN framework and it was shown to improve performance. In this chapter we provide a summary of each of the chapters of this thesis which leads us into a section dedicated to the overall conclusions in relation to our research hypotheses. Following this, we reflect on future avenues of research following the outcomes of our experiments and the

analysis and research conducted in meeting our research objectives and evaluating our hypotheses.

9.1 Summary

In our introductory chapter we highlighted the need for high spatial and temporal monitoring of our water systems and technologies that are helping to streamline the water quality monitoring process in the form of in-situ wireless sensor networks and novel analytical instruments. We outlined the limitations with the use of these technologies summarising the main issues in terms of *scalability* and *reliability*. We proposed the use of a multi-modal sensor network utilising visual sensors, modelled outputs and context information to enhance the use of a conventional in-situ sensor network. We also proposed the use of a trust and reputation framework to deal with the unreliability associated with each of the alternative data streams in order to optimise their benefits to the network.

In Chapter 2, we provided an overview of some key concepts from the literature in relation to environmental sensor networks. We highlighted a number of difficulties with the current state of the art in environmental sensing technology in meeting the needs of environmental monitoring applications. We then highlighted some avenues for progress in order to achieve higher scale chemo/bio-sensing and more efficient and effective environmental monitoring networks. This lead us on to an overview of trust and reputation systems. We provided a detailed outline of the model to be adapted for use in our research and our reasoning behind choosing this particular model.

In Chapter 3 we introduced our two test sites - the River Lee and Galway Bay - both representative of different water systems and challenging from an environmental monitoring perspective. For each of the sites we described a range of sensors, quite diverse in nature and sometimes not all operational. We highlighted some of the issues with the in-situ sensors and how this provided us with a very real testbed on

which to explore multi-modal sensing.

Chapter 4 presented our work in relation to research question 1. We demonstrated the use of a visual sensor (camera) in a river environment investigating how it could be used as a complimentary sensing modality to the in-situ environmental sensor network. Since depth could provide us with an indication of a variety of conditions at the River Lee site and could be tied in with readings from the in-situ sensor network we chose to estimate depth from the camera images. We investigated features in the image that correspond to readings from the in-situ depth sensor and highlighted the appearance of four particular features that coincided well with different changes in water level and different ranges of depth values. We subsequently developed and evaluated four software classifiers for the detection of each of these features with very satisfactory results.

Chapter 5 presented our analysis in relation to research question 2. The key objective here was to investigate the availability of satellite data at an appropriate spatial resolution to provide measurements that can most accurately coincide with data from the in-situ sensor node. This data also needs to be available at a sufficient temporal resolution so that if the in-situ sensor node were to fail, we can still avail of high frequency data. With a specific focus on ocean colour and SST data our review on satellite remote sensing data highlighted the difficulties in achieving this. Following this analysis, we outlined satellite remote sensing data products chosen for use in our research.

Chapter 6 presented our work in relation to research question 3 and research question 4. In the first part we presented a methodology for the incorporation of rainfall radar imagery data and water depth data into an ANN model for predicting average freshwater levels at the River Lee site for potentially controlling the operation of an in-situ phosphate sensor. We examined a number of issues including the most effective manner to present rainfall radar information extracted from a digital image to the network, the effects of rainfall from different parts of the catchment on the model, the effect of differing lag times on the model and the benefits or influence

of rainfall radar data or water depth depth to model output. Our analysis produced some very interesting results and overall the study demonstrated that with limited training data, a system for controlling the sampling rate of a sensor can be set up quickly and cost effectively at a deployment. This system may also have a potential use as an early warning system for floods. In the second part of the chapter we developed and evaluated models incorporating information from a variety of different combinations of heterogeneous sensor nodes for the prediction of values of an alternative in-situ node in the network at the River Lee. We investigated what we could achieve with limited data sources into replicating the activity of a sensor whilst there may be a possible gap in the data or fault in the network, and whether we could somewhat estimate the missing data. We developed and evaluated a variety of models for the prediction of depth, conductivity, dissolved oxygen and temperature with very promising results. The performance obtained demonstrates the powerful role these models may play in estimating missing data in a low cost manner with limited data availability and simplified input model scenarios.

In Chapter 7 we presented our work in relation to research question 5. Following our analysis in preceding chapters, it was clear that while some of these models and additional data streams could provide very good performance, some were more effective than others. Hence we needed to explore how their use could be optimised in the network. In this chapter we adapted a trust and reputation framework from the literature for dealing with the unreliability associated with each of the modelled outputs and the visual sensor streams, in order to improve their benefit to the network. We examined its application in three diverse scenarios. First we demonstrated the need for a real-time analysis tool to monitor the behaviour and performance of the various in-situ data models. Subsequently we described our adaptation of the trust and reputation framework for use in this scenario and examined its use. We demonstrated its effectiveness in reflecting the ability of the models to replicate in-situ sensor outputs over time and in providing information that would be difficult to determine otherwise. In the second application scenario we adapted this trust and

reputation model for determining the most reliable satellite data analysis product at the Mid-Bay and Mace Head test sites. We highlighted its unsuitability for use with the chlorophyll data products but we demonstrated its use for the SST data products. Finally in the third application scenario we demonstrated it as a tool for relating visual data stream outputs at the River Lee to the in-situ data and determining the most reliable models.

In Chapter 8 we presented our work in relation to research question 6. This chapter was dedicated to evaluating the performance of a multi-modal network from an applications perspective. Firstly we analysed the output values of the in-situ data models and the satellite data products in relation to various threshold values of the in-situ sensors. Some of the in-situ data models and SST data products demonstrated excellent performance, however the analysis of the chlorophyll data products raised many issues. Following this we carried out a case study evaluating the performance of the multi-modal network in detecting high and low depth events. We wanted to determine how each of these alternative sensor streams would perform in picking up these events had the in-situ depth sensor failed. Overall, good precision and recall scores were achieved. Finally we equipped the network with a trust and reputation model for real-time selection of the most appropriate output in the network. Overall we found this brought about improved performance in optimising the use of in-situ data models and visual sensor streams in the network.

9.2 Conclusions

At this point it is important to refer back to the two main hypotheses of this research presented in Chapter 1. We first summarise our conclusions in relation to hypothesis one and then in relation to hypothesis two.

Our first hypothesis states that the use of multiple sensing modalities including visual sensors, context information and modelled outputs will enhance the use of an in-situ sensor network in the marine environment. In order to evaluate this

hypothesis firstly we investigated the use of a camera as an alternative sensing modality in a river environmental monitoring network. We chose to evaluate the estimation of depth from the camera images since it is an important and influential parameter at the site and it can also be tied in with the in-situ depth sensor deployed at this location. We identified four features that could provide an estimation of depth. We developed classifiers for each of these depth features which performed very well considering the challenging data in question. For the 2-class model, accuracies of 89.2%, 85.38%, 98.63% and 79.25% were achieved on the May test data. The results were slightly poorer for the combined test data - *NovJanFeb*, with accuracies of 67.08%, 79.6%, 94.58% 53.75%, leading us to us training models with data specific for this time period for further experiments. Since our evaluation determined that we could successfully detect each of the four depth features, we related these features back to the in-situ sensor readings and we carried out a case study evaluating the use of a visual sensor as an alternative sensing modality to the in-situ sensor in detecting high and low depth events. We wanted to determine whether it would be possible to use this as a temporary backup sensing modality if the in-situ sensor went offline.

We used between two and four visual data streams differing in terms of the algorithm used to determine if an event had occurred, based on the appearance or non-appearance of certain features. We had found one of the classifiers to perform very poorly on negative instances which formed the majority of the dataset, hence configurations of the visual streams were evaluated with and without this feature, where those without the inclusion of this feature generally performed much better. Overall we found the highest precision and recall values that can be achieved by at least one of the data streams to be very satisfactory for the high water level events. Performance was poorer for the February test data, even though we used a model trained using a combination of data from November, January and February. It may be the case that the models need to be improved for this particular time of year with better training data, and consideration of further features as input to

the SVM for modelling each of the depth features. The lower water level events achieved higher recall scores it seems at the expense of precision. However this was perhaps due to the small number of features that had to be classified correctly for this event compared to the high water event. Overall however we demonstrated the effectiveness of a camera as a back-up sensing modality in a river environmental monitoring network. The depth features could be effectively classified and related back to the in-situ readings and subsequently used for the detection of depth events. While it does not replace the use of the in-situ depth sensor in the manner it has been demonstrated in this context, it has successfully been demonstrated as a tool that can enhance the use of of an in-situ sensor network in the marine environment.

Secondly, we evaluated the use of satellite remote sensing data to complement an in-situ sensor network in Galway Bay. Our review on SST and ocean colour satellite data highlighted many issues and we subsequently chose to use a selection of SST and chlorophyll data analysis products combining data from a variety of sources in our study. The SST data products demonstrated mixed performance and it is apparent that certain data products are not suitable at all for use in the network. However three of the selected products produced excellent performance at high enough temporal scales to complement the use of an in-situ sensor, with these products appearing to have better performance at the Mid-Bay test site than Mace Head. The chlorophyll data products showed some interesting outputs and highlighted the need for a closer analysis looking at the percentage of missing data values. The GlobColour products are obviously not particularly suited to monitoring at this particular point, and the unreliability of the in-situ chlorophyll sensor subsequently rendered it difficult to come to any conclusions regarding the Ifremer product. This highlighted the issues with such products and the need for careful consideration where many alternative products may need to be investigated for their suitability to the application context. However overall there are many applications and benefits to using satellite data products in environmental monitoring networks and they can greatly improve and enhance their operation.

Thirdly, we evaluated the use of in-situ data models in a river environmental monitoring network. We developed and evaluated a variety of models for the prediction of depth, conductivity, dissolved oxygen and temperature, with many of these models demonstrating excellent performance in our initial evaluation. We evaluated different configurations of the three parameter models on unseen test data as these had performed very well in our initial evaluation. For the prediction of each parameter at least one model configuration reached a correlation of over 0.8. Considering the difficulties in modelling the dynamics at the Lee Maltings site, the performance obtained here demonstrates the powerful role these models may play in estimating missing data in a low cost manner with limited data availability and simplified input model scenarios. When we evaluated the performance of models for specific time periods, while it was apparent that very high performance could be achieved from various model configurations for the detection of each parameter, performance of specific models varied across months, and it was often difficult to predict how a model would perform on unseen test data based on its initial evaluation. This further reinforced the need for a real-time update tool for evaluating the behaviour of models over time in the network, allowing us to choose the most appropriate model at a specific point in time or during a specific event.

We evaluated the performance of these models from an applications perspective examining the percentage of outputs that are within a particular error margin of the in-situ sensor with many of the data models displaying a very promising performance that would be suitable for many application contexts. In the case study analysing their ability to detect high and low water events, overall the highest precision and recall values that can be achieved were very satisfactory for the high level event, with improvements perhaps desirable in January and February. Monthly models were used for predictions in these months as opposed to the generalised models, since they achieved better performance. However as previously outlined the weeks chosen for training the models were chosen arbitrarily from the particular month. It may be the case that data quality is more of an issue during these months since the

sensors are longer into the deployment and more susceptible to failure or producing faulty data. Hence a more focused training approach may be required. The low level event overall produced lower precision and recall scores from the models. However overall, it can be concluded that these models demonstrated themselves as effective and useful tools that can improve the effectiveness and operation of an in-situ sensor network in environmental monitoring applications.

Finally, we evaluated if context information from alternative modalities can be used to enhance the efficiency of an environmental monitoring network. Our investigation into the use of rainfall radar imagery data and water depth data to improve the operation of an in-situ chemical sensor produced a number of very interesting results. We found particular parts of the catchment to be more influential on the output of the model than others. Interestingly in many cases this corresponded to parts of the catchment that had been highlighted as being influential in our catchment description in Chapter 3. While the water depth data is the dominant influence on the model output, the rainfall data does appear to have some impact on the network despite its noisiness. Overall we found that the ANN performed very satisfactorily for the application context in question. We highlighted specific combinations of rainfall and water level lag times that appear to produce the best outputs. However it was also apparent that performance for specific time periods requires more analysis as our test sets were quite limited. Overall it can be concluded from this study that with limited training data, a system for controlling the sampling rate of the nutrient sensor using contextual information can be set up quickly and cost effectively at a deployment. It also has the potential of providing the basis for an early warning system for flood monitoring and can bring about great efficiencies to an environmental monitoring network.

Therefore based on the results of these four studies investigating the use of a camera, satellite remote sensing data, in-situ data models, and context information, we can successfully conclude that multiple sensing modalities enhance the use of an in-situ sensor network in the marine environment. However it was also appar-

ent from our analysis that each of these alternative data sources can be inherently unreliable and it is often hard to predict their performance. With many producing very promising results, an investigation into how to optimise their use in the network was carried out. This leads us to our conclusions in relation to the second hypothesis of this thesis. The second hypothesis states that a trust and reputation model adapted for use in a multi-modal sensor network will help to deal with the unreliability associated with the visual sensor streams and modelled outputs in the network and optimise their use by choosing the most reliable output at a particular point in time.

A trust and reputation framework from the literature known as RFSN was adapted for dealing with the unreliability associated with each of the in-situ data models and the visual sensor streams. The modified RFSN model was applied to three specific application scenarios to examine its resilience across three very different applications in environmental monitoring. We specifically wanted to evaluate the ability of the alternative data streams to replicate the output patterns of the in-situ sensor, and the model was hence tailored to this objective. We demonstrated the ability of the model to reflect the behaviours of data streams over time in relation to their ability to produce output patterns in line with the in-situ sensor node. The model also enabled us to pick up trends that are difficult to decipher otherwise and to determine the magnitude of difference in reliability in one model compared to another. However the model was not suitable for use in conditions where regular data values were not available. For example it was not suitable for use with the GlobColour data products since data was only intermittently available for the Mace Head and Mid-Bay sites throughout the year.

In the case study for detecting high and low water events, when we equipped the network with the trust and reputation framework it produced a very impressive performance. For the in-situ data models, despite not having been optimised for the detection of the events outlined, it resulted in an improvement in the output of the network in the majority of cases, with a significant increase in performance in

50% of the cases. The performance for the high level events was very satisfactory with an improvement in the output of the network in all months but one, with a significant increase in performance in 3 out of 5 of these months. Overall, the trust models seemed to perform best in months where a selection of some of the individual models were performing better. When we adopted a late-fusion approach whereby we combined the output of the top four trustworthy models as opposed to simply selecting the most trustworthy model, this resulted in an improvement in the performance of the trust and reputation framework in all cases. For the visual data streams the trust and reputation framework also performed very well. In all cases except for one the trust model was consistent with the highest score achieved by the best performing visual stream, for both events.

Therefore we can successfully conclude that a trust and reputation model helps to deal with the unreliability associated with the visual sensor streams and modelled outputs and optimises their benefits to the network. It provides a real-time update tool reflecting the behaviour of models over time and it has demonstrated itself as a useful tool in optimising the output of the network in the face of a number of unreliable sources. We have therefore successfully confirmed our two research hypotheses. The following section summarises the main contributions of this research. Subsequently we explore future avenues for research following the issues that have been raised while carrying out the studies in this thesis.

9.2.1 Research Contributions

The contributions of this research consist of the following:

- An investigation into the use of a low cost off-the shelf camera as a complementary sensing modality to an in-situ sensor network and the development and evaluation of models to classify the chosen parameter.

- A methodology for integrating pixel information from rainfall radar images and in-situ depth data into an ANN for predicting average freshwater levels

and an analysis of model outputs.

- The development and evaluation of models consisting of heterogeneous in-situ information to predict the value of alternative in-situ nodes in an in-situ sensor network where no redundant nodes are available at times of node failure.

- The adaption of the trust and reputation framework outlined in (Ganeriwal et al., 2008) to a multi-modal sensor network, an outline of its possible application scenarios and an examination of its use in these scenarios

- An evaluation of a multi-modal sensor network from an applications perspective and an evaluation of the performance of a trust and reputation framework in a particular application scenario at producing the most reliable output from multiple unreliable sensor streams in the network.

9.3 Future Work and Reflections

There are a wide range of exciting and novel opportunities for monitoring and characterising our marine environment with the help of a wide variety of sensing platforms. The findings of the research carried out in the context of this thesis have generally been very positive towards the use of multi-modal platforms, and have also opened up a whole range of possibilities with regards to future avenues of research. Procurement of knowledge across a range of disciplines and integration of a wide variety of technologies seems to be a re-occurring theme among many global and European led initiatives e.g. GMES[1], and recently funded European projects e.g. Seventh Framework Programme (FP7)[2]. Looking at the individual technologies evaluated within the context of this research, there are number of exciting opportunities for further research into their use in the context of operational environmental monitoring networks:

[1]Global Monitoring for Environment and Security, www.gmes.info
[2]$http://cordis.europa.eu/fp7/projects_en.html$

- **Camera Sensor Networks in a River Environment:** Whilst off-the-shelf low cost cameras might not provide the same precision that an in-situ sensor can, there are a large number of application opportunities. They have the benefit of being remote from the environment and so are not subject to the same issues as in-situ sensors such as biofouling. They are not affected by extreme events such as flooding, unlike in-situ sensors which may become unoperational when they are needed most. Essentially they act like a set of eyes on the water and there are numerous application opportunities. While in our study we were limited due to issues related to infrastructure and the types of sensors deployed as part of the in-situ sensor network, there are numerous avenues for future research exploring the use of multiple cameras at different points along the river. Subsequent research has already begun within CLARITY to develop a low cost self-powered mobile platform that could be deployed at any point along the river.

  While we evaluated the estimation of depth from the camera since it could be linked in with the in-situ sensor network, our visual sensing system also undertook the detection of other image features such as objects floating on the water, boats, water turbulence etc. However there are a wide range of further opportunities in this regard. For example, if there are further in-situ sensors such as flow, turbidity, etc at the test site, this provides a whole new set of possibilities for providing surrogate measurements from the camera for a range of potential parameters. At the Lee Maltings site we were limited with the features that could be linked up to the in-situ sensor network, but this may not be the case in future deployments or test sites. There are also other potential features or landmarks that could be evaluated with a specific application in mind, such as alerting for possible flooding.

- **In-Situ Data Models:** The performance of the in-situ data models was very impressive considering the number of complex processes in operation at the Lee Maltings which are difficult to account for and to model. We aimed to

investigate what we could achieve using simple input model scenarios with limited data, so that these could provide a form of redundancy in the network without the need for any additional technology or data in the case of intermittent node failure. The performance obtained here demonstrates the powerful role these models may play in estimating missing data in a low cost manner, and has provided a grounds for future research into further optimising their outputs. A longer deployment would allow a better development and evaluation of seasonal models where more data would also allow the quality of the data to be considered before being selected for training purposes. It would also be very interesting to evaluate the use of such models in a site without the dynamics of the Lee Maltings and to compare the performance achieved at both sites.

- **Rainfall Radar Information as a Contextual Data Source:** Our investigation into the use of rainfall radar and in-situ depth data in an artificial neural network with a view to capturing a catchment's response to rainfall and predicting freshwater levels at the Lee Maltings, raised some very interesting issues worth further investigation. These especially relate to the effects of different parts of the catchment on the output of the model or response times or lag times producing the best outputs. A larger dataset with a longer deployment would also enable further investigation into the performance of the model during specific time periods and the evaluation of seasonal models if required. Further studies could also be carried out in relation to the performance of the model during flood events and an investigation into its use as an early warning system.

- **Satellite Remote Sensing Data:** Satellite remote sensing data is a powerful tool for better understanding our environment and it will play a significant role in the future into the investigation of issues such as climate change. Its use in an operational monitoring network has significant potential. However

looking into the future, it appears that an integrated approach offers more potential especially in areas greatly affected by cloud cover. This in itself opens up a whole new area of research opportunities. If we can develop products using satellite remote sensing data in the more difficult regions, they can subsequently be promoted as a tool that can be suitably developed for any region. The particular satellite data products that we choose to incorporate into our studies however must also be carefully chosen within the specific context. For example from our analysis it is apparent that certain products are not particularly suited to monitoring the two test sites in Galway Bay. With the range of satellite data products becoming available through services such as MyOcean [3], it would be interesting to pursue this analysis further with a varying set of products. These products could also be further evaluated in a wide range of application contexts such as predictive systems, rapid detection and response systems, etc.

- **Trust and Reputation Framework:** The successful use of a trust and reputation framework for optimising the use of a variety of these alternative data streams to the network has demonstrated the benefits that the novel use of such a framework can bring to an environmental monitoring network. It has also opened the door to a whole new range of research opportunities. For example we defined a specific notion of reliability with certain assumptions, this could be adapted and optimised for use in other application scenarios. There are also other issues that require further analysis. For example further investigations may consider how much misbehaviour a node needs to exhibit before it is deemed more untrustworthy than another node or the actual trust value of the node may be considered as opposed to simply the most trustworthy. There are also possible opportunities for the integration of such a model into other frameworks such as the OCG[4]'s Semantic Web. Future opportunities

[3] *http* : *www.myocean.eu*

[4] Open Geospatial Consortium, *http* : *//www.opengeospatial.org/*

may lie in the further development of ontologies integrating such a framework for finding the most reliable data sources within a specific context.

Overall, in evaluating our research questions and hypotheses we have produced promising results that provide grounds for further investigation into the use of innovative methods for optimising our environmental sensor networks. It has also opened the door to further opportunities for the exploration of multi-modal systems and their potential applications with a view towards achieving the vision of large scale detection and response systems in our environment.

Glossary

AATSR Advanced Along-Track Scanning Radiometer.

ADEOS ADvanced Earth Observation Satellite.

ALI Advanced Land Imager.

AMSR Advanced Microwave Scanning Radiometer.

ATSR Along-Track Scanning Radiometer.

AVHRR Advanced Very High Resolution Radiometer.

CERSAT Centre ERS d'Archivage et de Traitement.

CNES Centre National d'Études Spatiales.

CNSA China National Space Administration.

COCTS Chinese Ocean Colour and Temperature Scanner.

COMS Communication Ocean and Meteorological Satellite.

CZCS Coastal Zone Colour Scanner.

CZI Coastal Zone Imager.

ERS European Remote Sensing Satellite.

ESA European Space Agency.

ETM+ Enhanced Thematic Mapper Plus.

GHRSST-PP Global Ocean Data Assimilation Experiment High-resolution Sea-Surface Temperature Pilot Project.

GOCI Geostationary Ocean Color Imager.

IFREMER Institut francais de recherche pour l'exploitation de la mer.

ISRO Indian Space Research Organisation.

KARI Korea Aerospace Research Institute.

KORDI Korea Ocean Research and Development.

MERIS MEdium Resolution Imaging Spectrometer.

MODIS Moderate Resolution Imaging Spectroradiometer.

MOS Modular Optoelectronic Scanner.

MSG Meteosat Second Generation.

NASA National Aeronautics and Space Administration.

NOAA National Oceanic and Atmospheric Administration.

NPP NPOESS Preparatory Project.

OCM Ocean Colour Monitor.

OCTS Ocean Colour and Temperature Scanner.

POLDER POLarization and Directionality of the Earth's Reflectance.

SeaWiFS Sea-viewing Wide Field-of-view Sensor.

SEVIRI Spinning Enhanced Visible and Infrared Imager.

SPOT Système Probatoire d'Observation de la Terre.

TMI TRMM Microwave Imager.

TRMM Tropical Rainfall Measuring Mission.

VIIRS Visible Infrared Imager Radiometer.

Appendices

Appendix A

Publications

We have published a number of works which have directly contributed to this thesis. These are as follows:

A.1 Papers

1. **Integrating Multiple Sensor Modalities for Environmental Monitoring of Marine Locations**

   E. O'Connor, A.F. Smeaton, N.E. O'Connor, D. Diamond

   In: ACM SenSys 2008 - 6th ACM Conference on Embedded Networked Sensor Systems, 5-7 November 2008, Raleigh, NC, USA.

2. **Views From the Coalface: Chemo-Sensors, Sensor Networks and the Semantic Sensor Web**

   J. Hayes, E. O'Connor, J. Cleary, H. Kolar, R. McCarthy, R. Tynan, G. O'Hare, A.F. Smeaton, N.E. O'Connor, D. Diamond

   In: SemSensWeb 2009 - International Workshop on the Semantic Sensor Web, 1 June 2009, Heraklion Crete, Greece.

3. **Environmental Monitoring of Galway Bay: Fusing Data from Remote and In-situ Sources**

   E. O'Connor, J. Hayes, A.F. Smeaton, N.E. O'Connor, D. Diamond

In: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IX, SPIE Europe Remote Sensing, 31 August - 3 September 2009, Berlin, Germany.

4. **River Water-Level Estimation Using Visual Sensing**

   E. O'Connor, E. O'Connor, C O'Conaire, A.F. Smeaton, N.E. O'Connor, D. Diamond

   In IEEE EuroSSC 2009 - 4th European Conference on Smart Sensing and Context, 16-18 September 2009, Guildford, U.K.

5. **Short-term rainfall nowcasting: using rainfall radar imaging**

   P. Wang, A.F. Smeaton, S. Lao, E. O'Connor, Y Ling and N.E. O'Connor

   In: EuroGraphics Ireland 2009 - 9th Irish Workshop on Computer Graphics, 11 December, 2009, Dublin, Ireland.

6. **Image processing for smart browsing of oceancolor data products and subsequent incorporation into a multi-modal sensing framework**

   E. O'Connor, J. Hayes, C. O'Conaire, A.F. Smeaton, N.E. O'Connor, D. Diamond

   In: RSPSoc Remote Sensing and Photogrammetry Society Annual Conference with Irish Earth Observation Symposium, 1-3 September 2010, Cork, Ireland.

7. **Image processing for smarting browsing of ocean colour data products: investigating algal blooms**

   J Hayes, E O'Connor, K.T. Lau, A. F. Smeaton, N.E. O'Connor, D. Diamond

   In: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IX, SPIE Europe Remote Sensing, 20 September - 23 September 2010, Toulouse, France.

8. **A multi-modal event detection system for river and coastal marine monitoring**

E. O'Connor, A.F. Smeaton, N.E. O'Connor

In: IEEE Oceans 2011, 6-9 June 2011, Santander, Spain.

A.2 Presentations

1. **The Application of Multi-Modal Sensor Networks to the Monitoring of Coastal and Inland Marine Environments**

   E. O'Connor, J. Hayes, A.F. Smeaton, N.E. O'Connor

   In: 3rd Annual Irish Earth Observation Symposium, 12th - 13th November 2009, Geological Survey of Ireland, Dublin.

2. **A reputation and trust based multi-modal sensor network for environmental monitoring**

   E. O'Connor, A.F. Smeaton, N.E. O'Connor, D. Diamond

   In: In: Environ 2010: 20th Irish Environmental Researchers' Colloquium, 17th - 19th February 2010, Limerick, Ireland

Appendix B

Technologies Developed as Part of This Research

We have published a number of technologies which have directly contributed to this thesis. These are as follows:

- **Visual Data Analysis Toolkit 1**

  This toolkit developed in $C\sharp$ was developed to provide a number of functionalities. Firstly it was developed to enable the visualisation of the images and the nearest in-situ sensor reading that corresponds to the time that image was captured. This enables the identification of features that correspond to the in-situ sensor readings. Secondly it was developed to carry out an initial investigation into a range of image processing techniques to extract information from the images. For example classification of an image through finding its closest match in an image database, carrying out background subtraction to detect change in the image, carrying out various thresholding techniques for the detection of various features, and the detecting of features such as objects floating on the water, boats passing.

- **Visual Data Analysis Toolkit 2**

  Following the investigation carried out through using the first visual data analysis toolkit, it was decided to build specific software for the extraction of

features for the detection of the four depth features outlined in Chapter 4 of this research. This was developed in Matlab (Version R2009A). This software extracts the relevant features from relevant parts of the image for subsequent incorporation into an SVM at a later stage.

- **Feature Set Development Software**

  Each of the classification and prediction scenarios (i.e. SVM classification for visual features, ANN prediction for freshwater levels, and regression tree prediction for the in-situ parameters) all required the development of a significant number of feature sets for training and testing the models and evaluating various issues e.g. the most important visual features for describing the depth features, the effect of different parts of the catchment on model output, the best combination of lag times, the best combination of in-situ parameters for the prediction of an alternative parameter, etc. Software was developed in Java for developing feature sets.

- **Satellite Image Analysis System**

  A satellite image analysis System was developed for the efficient browsing, searching and analysis of MODIS chlorophyll data. This system is described in O'Connor et al. (2010). The image processing was carried out in Matlab (Version R2009A) and the user interface was developed using $C\sharp$. This system is based on searching and analysing the data using the JPEG summary image of the source data. It allows the user to view trends over time from specific regions of Ireland and to search satellite data based on specific events.It also allows the user to find similar images to an image of interest in the database. This technology has been passed over for use to the Marine Institute.

- **Satellite Data Extraction Software** Software was developed using the Beam API[1] for extracting data HRDDS, GlobColour and Ifremer data products from the relevant pixels.

[1] http://www.brockmann-consult.de/cms/web/beam/

- **Trust and Reputation Framework** Software was developed to apply the modified RFSN model to the three applications scenarios described in the thesis and calculate trust values. Incorporated into this system was also software for carrying out the experiments where the trust and reputation framework was applied to a network for detecting depth events described in Chapter 8.

Appendix C

Graphs for the Analysis of an Artificial Neural Network incorporating In-Situ Depth Data and Rainfall Radar Data for Predicting Freshwater Levels

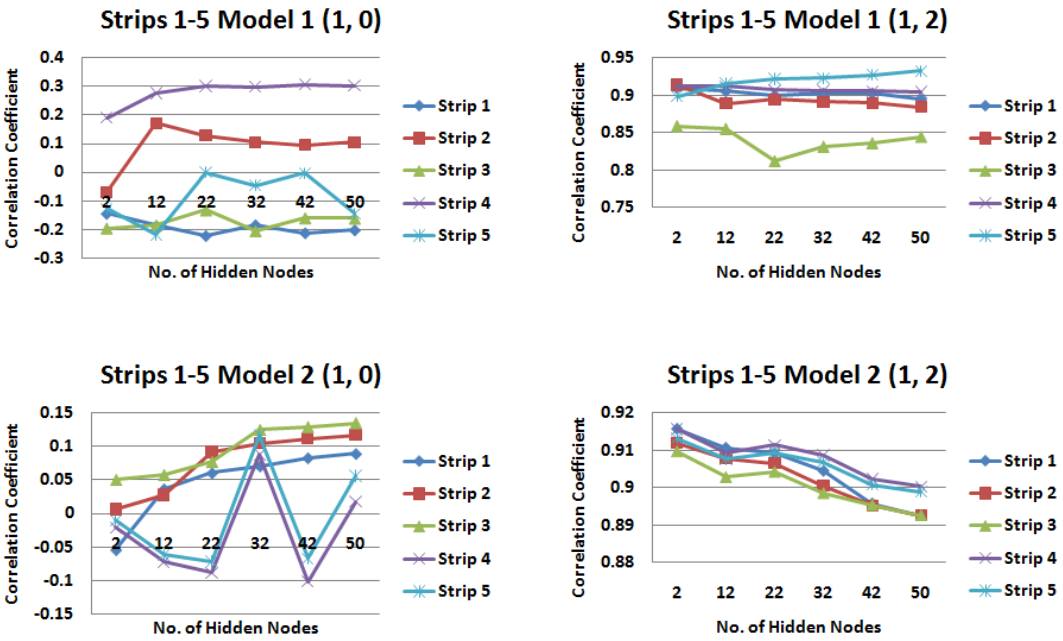

Figure C.1: Correlations for input models 1 and 2 for 1 day rainfall information and 0 and 2 days water level information for each strip of the catchment.

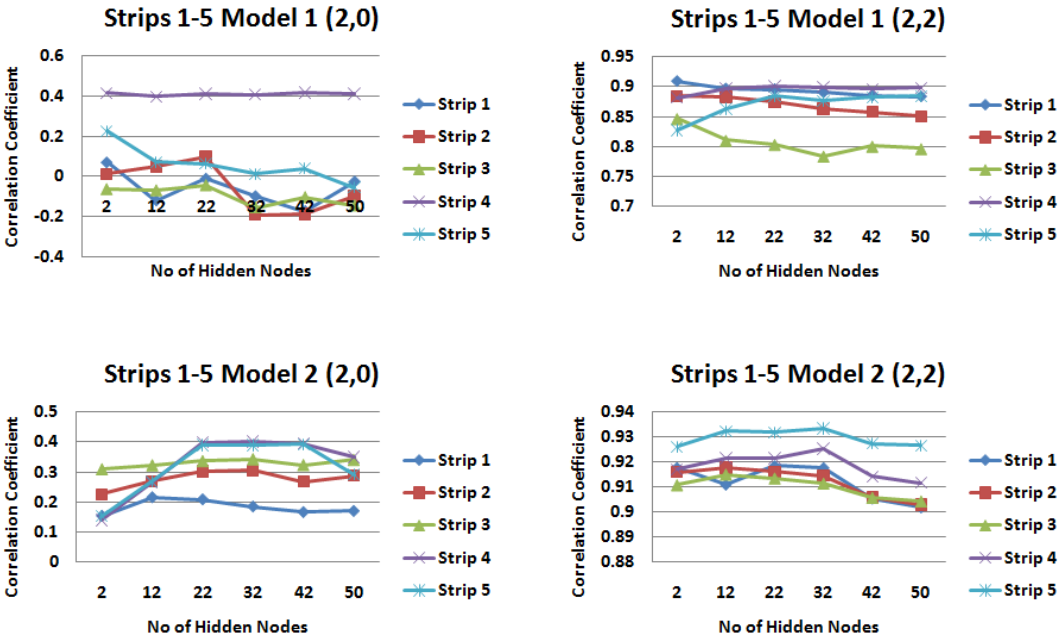

Figure C.2: Correlations for input models 1 and 2 for 2 days rainfall information and 0 and 2 days water level information for each strip of the catchment.

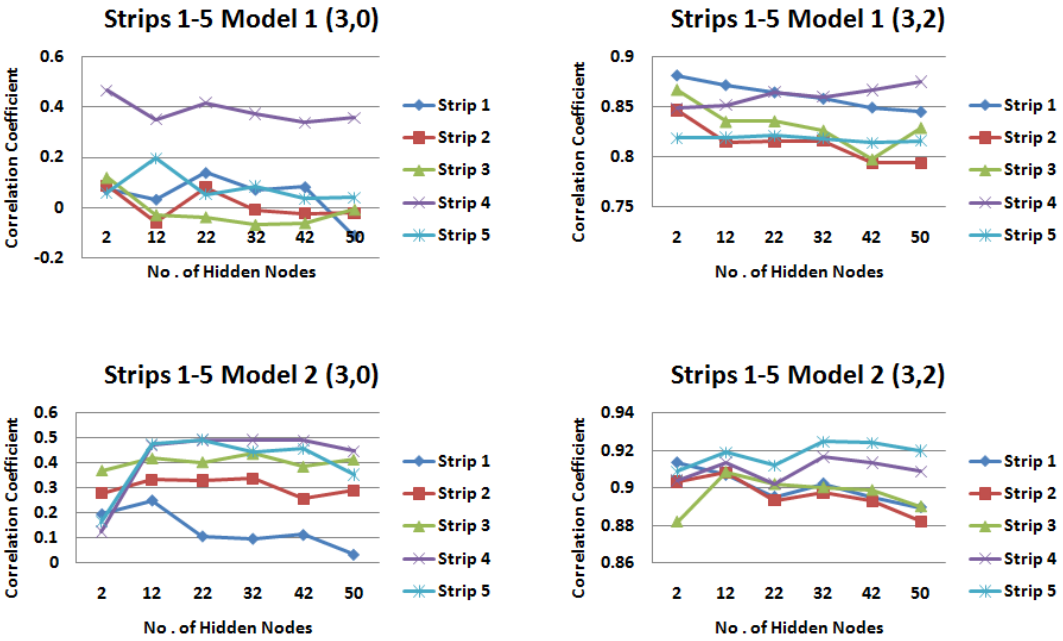

Figure C.3: Correlations for input models 1 and 2 for 3 days rainfall information and 0 and 2 days water level information for each strip of the catchment.

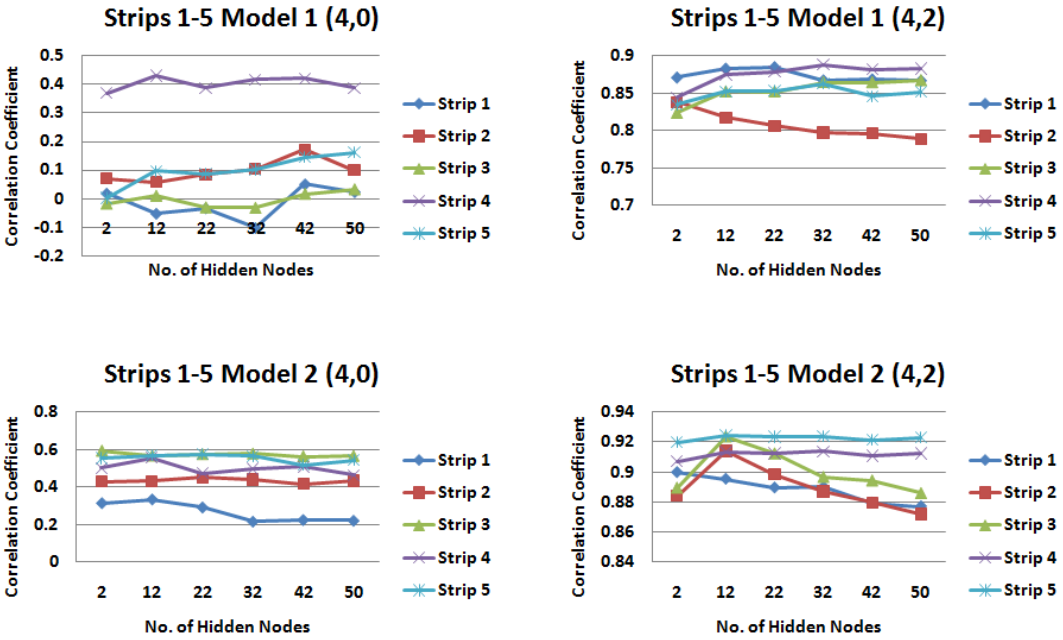

Figure C.4: Correlations for input models 1 and 2 for 4 days rainfall information and 0 and 2 days water level information for each strip of the catchment.

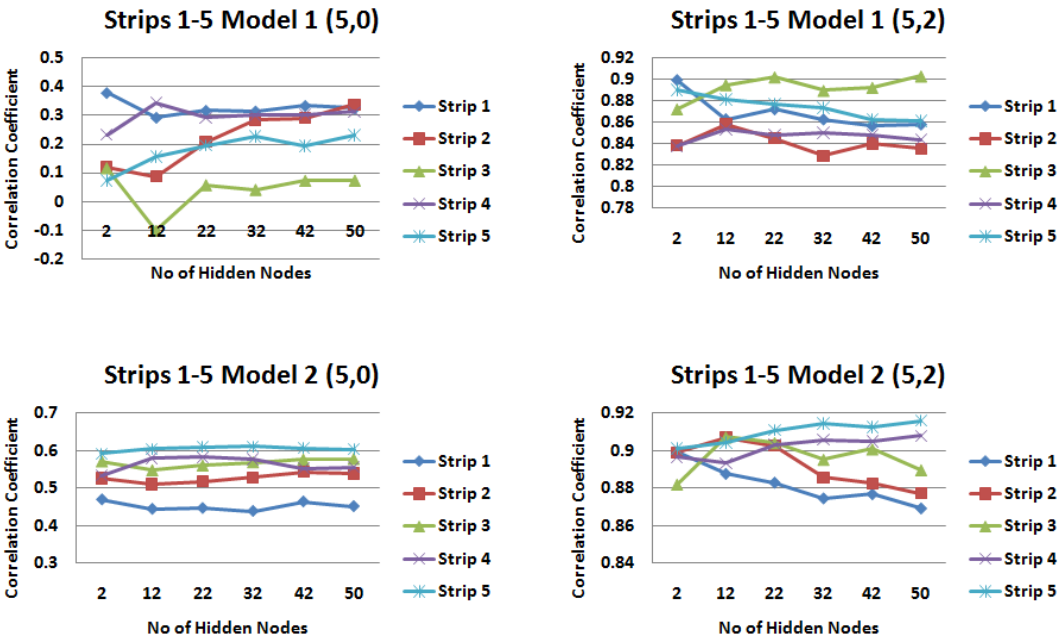

Figure C.5: Correlations for input models 1 and 2 for 5 days rainfall information and 0 and 2 days water level information for each strip of the catchment.

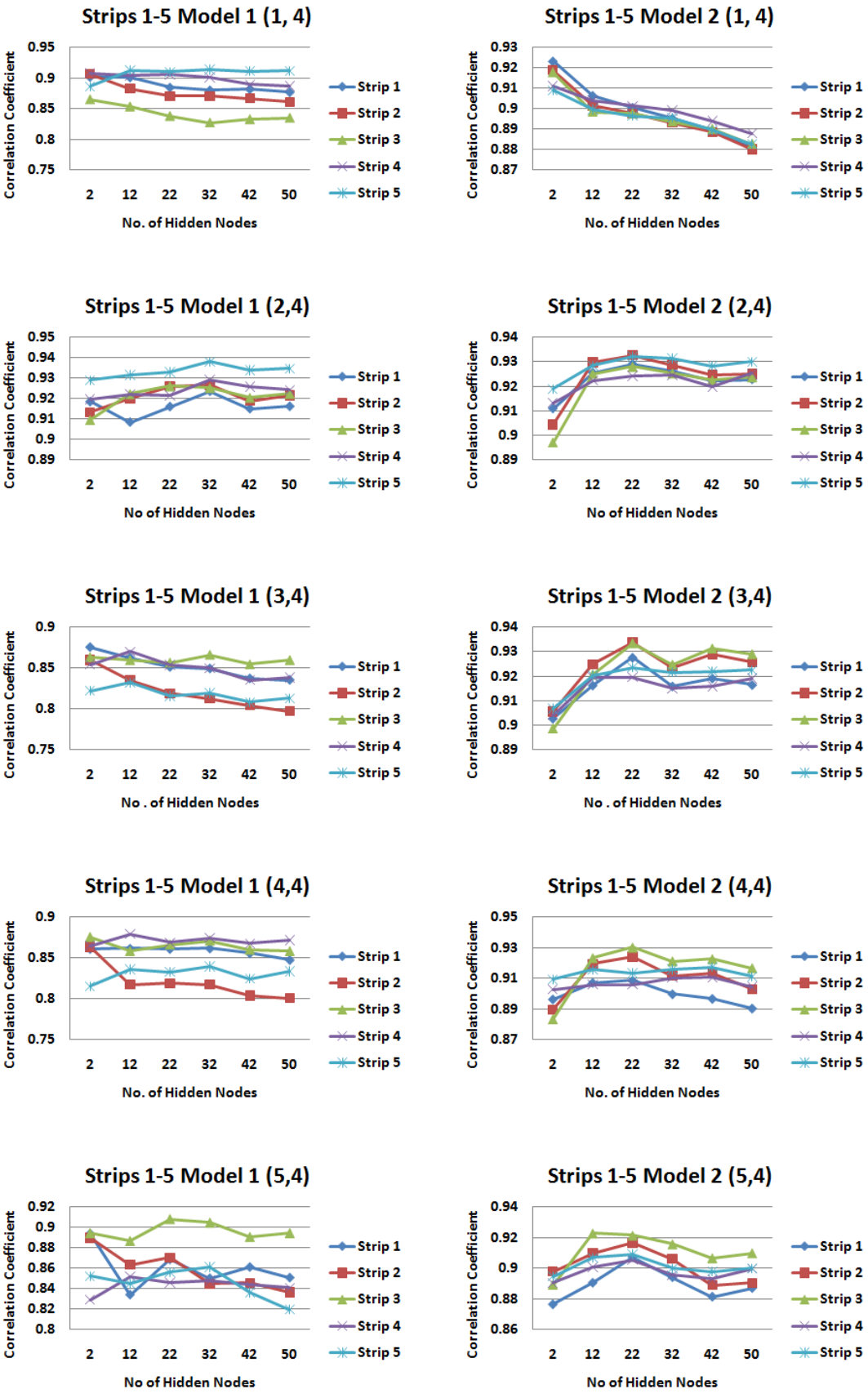

Figure C.6: Correlations for input models 1 and 2 for 1-5 days rainfall information and 4 days water level information for each strip of the catchment.

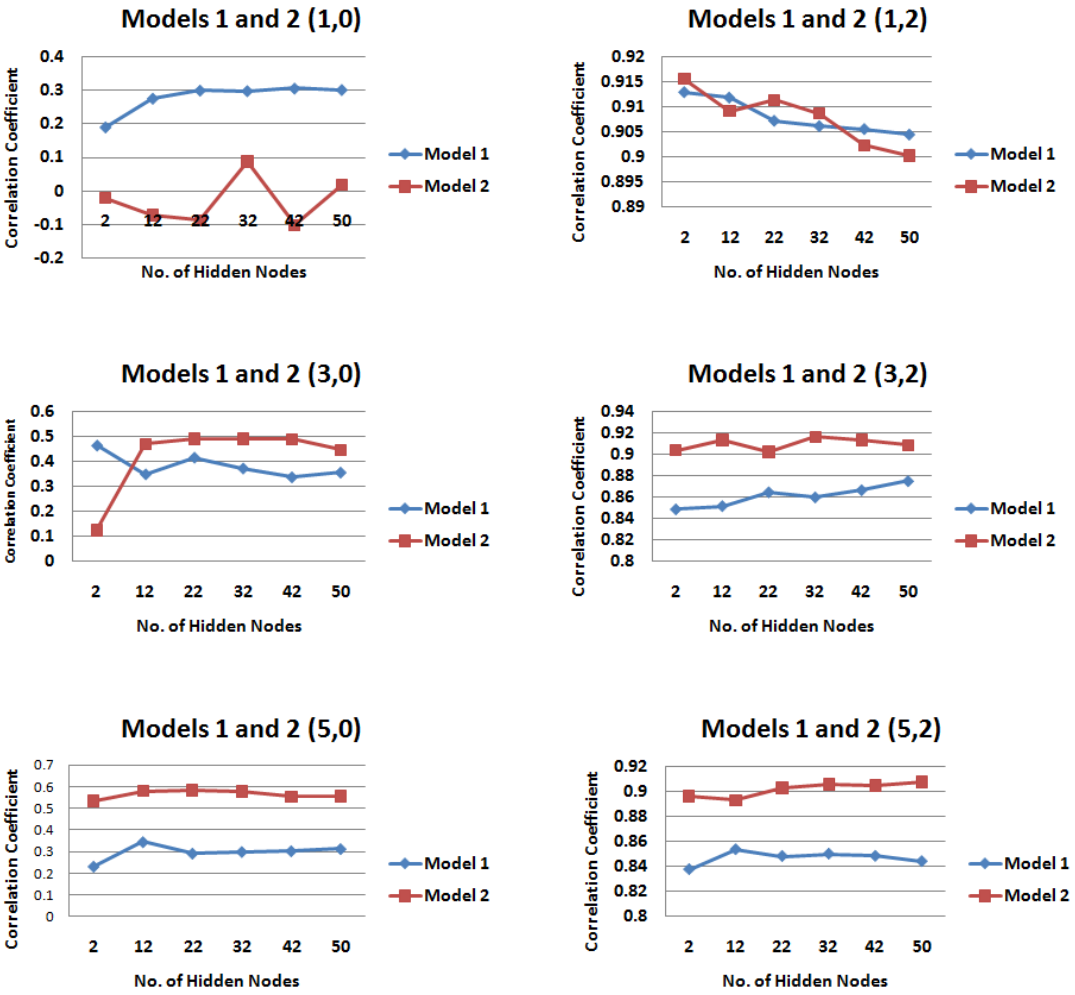

Figure C.7: Correlation coefficients for input models 1 and 2 for 1,3 and 5 days of rainfall information and 0 and 2 days water level information for strip 4 of the catchment, demonstrating the input model that produces the highest correlation for each scenario.

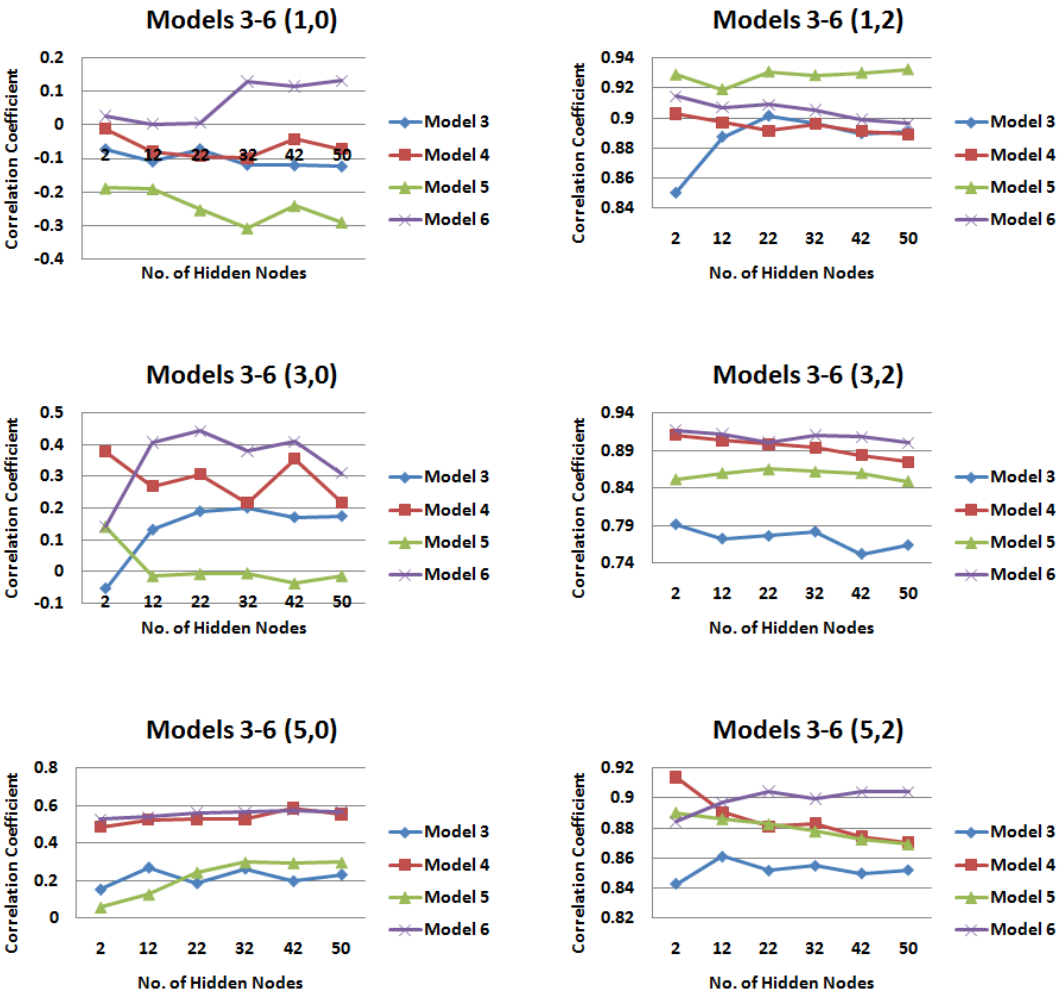

Figure C.8: Correlation coefficients for input models 3-6 for 1,3 and 5 days of rainfall information and 0 and 2 days water level information, demonstrating the input model that produces the highest correlation for each scenario.

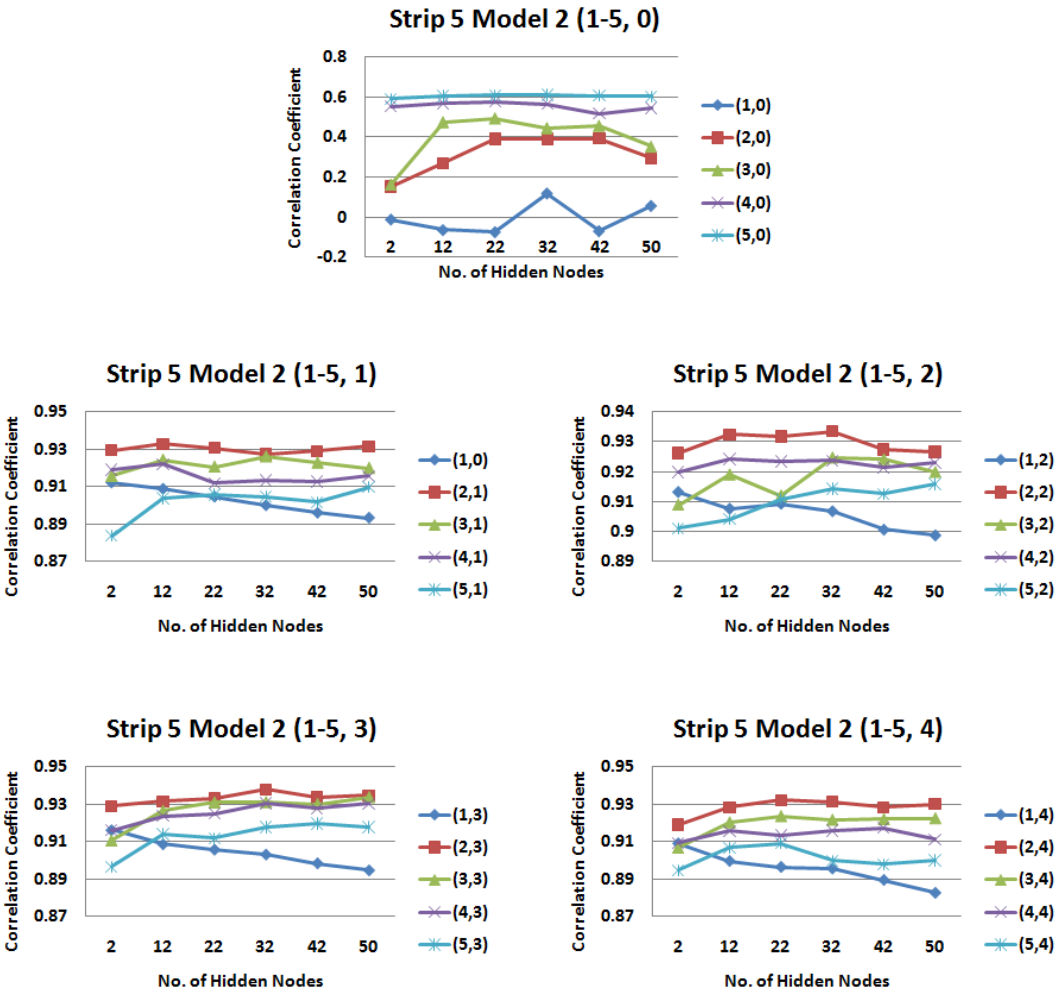

Figure C.9: Correlation coefficients for input model 2 for 1 - 5 days of rainfall information and 0 - 4 days water level information for strip 5 of the catchment - demonstrating the effect of rainfall and water level information on the model

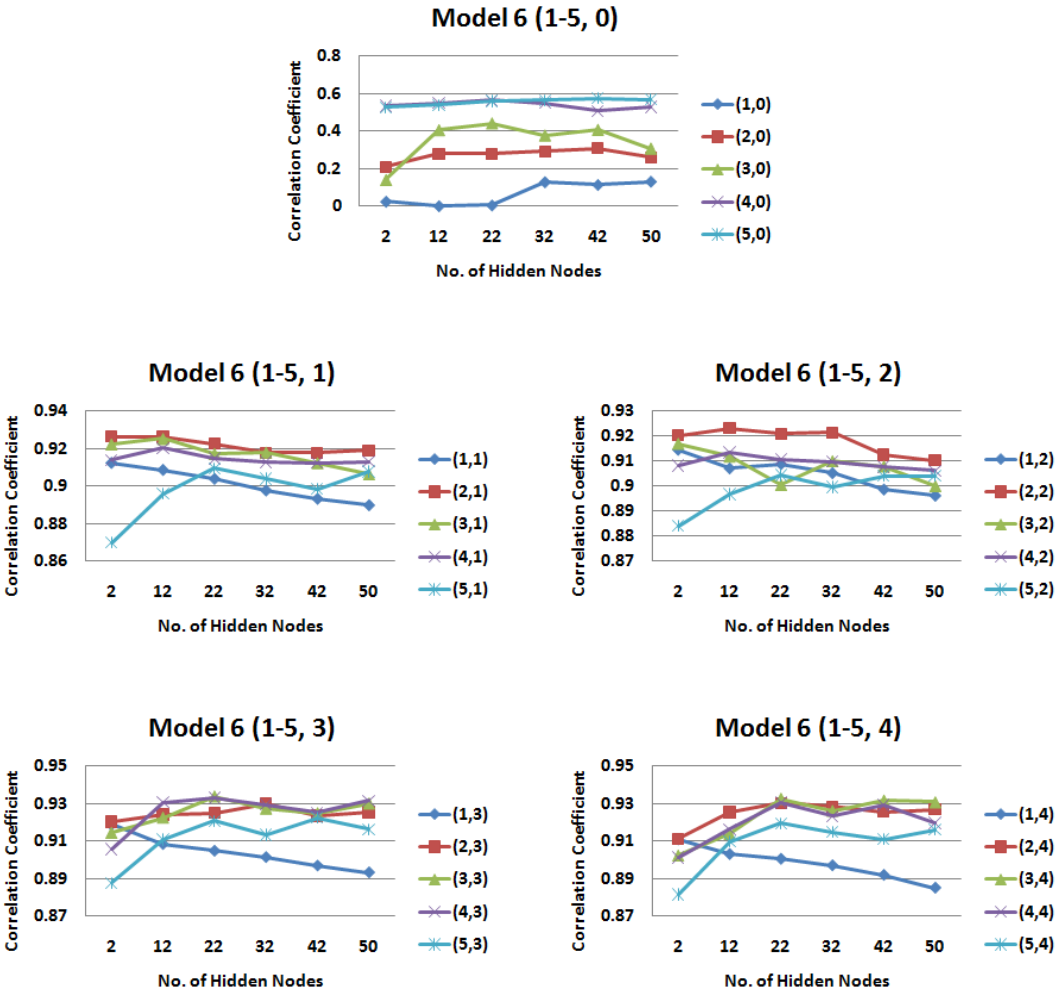

Figure C.10: Correlation coefficients for input model 6 for 1 - 5 days of rainfall information and 0 - 4 days water level information - demonstrating the effect of rainfall and water level information on the model

Appendix D

Application of RFSN to Models for Dissolved Oxygen and Temperature Prediction

D.1 Dissolved Oxygen at the Lee Maltings

For the prediction of dissolved oxygen the generalised model versions of M1-M7 were used for predictions in June and July. The monthly model versions of M1-M7 was used for predictions in May, January and February. In particular for January and February, the generalised model appeared to produce quite high MAE's for certain models. However Table 7.2 demonstrates that the range of these test sets differed somewhat more than other months to the generalised dataset.

The generalised models use 10,000 training instances, the May and February monthly models use 2,000 training instances and the January models range from 488 to 2570 training instances (M1, M4, M5, M7 - 488, M2, M6 - 2531, M3 - 2570). The number of test instances in each month are - May - 2006, June - 2304, July - 1954, January - 2089, February - 2108. The performance of the models on the test data is shown in Table D.1. Overall the results are generally poorer than that achieved for depth and conductivity. It may be the case that there are more complexities

involved in the prediction of these parameters compared to that of conductivity and depth at the Lee Maltings site. It would be interesting to develop models for an alternative site without the dynamics of the Lee Maltings in order to try and determine the influences on these parameters. This may provide an indication on how the modelling could be improved with limited input data.

Future studies may subsequently involve investigating the possible

The use of alternative data values for the prediction of these parameters

may require more complex input parameters or an increased number of input parameters.

May	CC	MAE	RMSE
M1	0.3714	1.1324	1.4124
M2	−0.2335	1.4548	1.8027
M3	0.1011	1.2575	1.5956
M4	0.2435	1.2599	1.5501
M5	0.5225	0.9909	1.2285
M6	0.1295	1.3043	1.6423
M7	0.4332	1.0603	1.3184

June	CC	MAE	RMSE
M1	0.513	1.3534	1.5852
M2	0.1343	2.4235	2.7759
M3	0.0179	2.1252	2.4892
M4	0.4899	1.4366	1.6715
M5	0.2847	1.4075	1.7235
M6	0.3808	2.0574	2.36
M7	0.4206	1.5209	1.7745

July	CC	MAE	RMSE
M1	0.6191	1.2958	1.5844
M2	−0.0571	2.1543	2.7522
M3	−0.1082	1.8556	2.4176
M4	0.7245	1.1451	1.4233
M5	0.5746	1.287	1.562
M6	−0.1622	1.8775	2.4767
M7	0.6391	1.286	1.5717

Jan	CC	MAE	RMSE
M1	−0.3127	2.4237	2.8355
M2	0.126	2.4286	2.7901
M3	0.457	1.5017	1.7344
M4	−0.1746	2.468	2.8841
M5	−0.1107	2.1347	2.5095
M6	0.4394	1.5071	1.7614
M7	0.0537	2.1904	2.5526

Feb	CC	MAE	RMSE
M1	−0.0032	0.6928	0.8298
M2	0.4347	0.6067	0.7348
M3	−0.1228	0.7064	0.8406
M4	0.152	0.6908	0.8129
M5	−0.1694	0.7369	0.8709
M6	0.3142	0.6567	0.7807
M7	−0.0433	0.7189	0.8452

Table D.1: Performance of each of the models when evaluated in the corresponding test data for the prediction of dissolved oxygen

D.1.1 Application of the Trust and Reputation-Based framework

Similar to the system applied for the previous two in-situ parameters (i.e. conductivity and depth), at each epoch the Watchdog assigns a level of cooperation to a model based on the deviation of its output from the output value of the sensor (See Equation 7.3 on page 187). Due to the fact the range of the dataset can vary more from month to month than depth or conductivity, the deviation of the error value is normalised to a range of [0-1] using a different range depending on the model used. Since May, January and February predictions were carried out using monthly models, the range of the training dataset for the model in question was used for carrying out the normalisation by the *Watchdog*. For May test data, the error measures were normalised by dividing by the max-min range of the May training set - [9.9-2.4]. For January test data the max-min range of [13.59-8.63] is used and for February test data the max-min range of [11.71-7.97] is used. June and July predictions were carried out using the generalised model, hence the error values were normalised to a range of [0-1] by dividing by the range of the full training set [14.8-2.05]. Similar to previous, this value is then subtracted from 1. Again the *Reputation* component uses this output of the *Watchdog* to update the reputation of the model (Equation 7.1 on page 177) and there is the subsequent calculation of trust (Equation 7.2 on page 177).

Normalising the monthly models with the range of the full training set will not generally change the patterns found in terms of the reliability of the models. However it will affect the magnitude of reliability as well as the overall height of trust values. For example, in February dividing an error by the range of Feb training set will result in a much harsher cooperation metric than if dividing by the range of the range of the full dataset. Figures D.3 and D.2 demonstrates the trust values when normalisation using the range of the full dataset is used and normalisation using the range of the February training set is used. The prediction outputs for the models

for this time period are shown in Figure D.1. While overall patterns don't generally change, the magnitude of difference between models and the height of trust values achieved varies. Normalisation using the range of the February training set results in a greater magnitude of differences between the models and overall lower trust values. The level of trust values may be important if using an application whereby a thresholding approach was being implemented as to whether or not to consider the value of a model.

D.1.2 Results

The output of each of the models for the first two weeks in July along with the actual values output by the DO sensor are shown in Figures D.4 and D.5. The associated trust values of each of the models at each epoch is shown in Figure D.6.

Similar to the scenerio with depth and conductivity, the trust values generally reflect the performance of the model in relation to predicting the values output by the sensor. Of the single parameter models (M1-M3), it is clear while none of the models accurately represent the phenomenon at the beginning of the first week that M1 (Conductivity) appears to produce a pattern more in line with the DO sensor than the other two models. For the remainder of the first week and for the second week, it also seems to be producing patterns and outputs that replicate that of the DO sensor more than the other models. This is reflected in the trust values whereby M1 is deemed a much more reliable model than M2 or M3. However it can also be seen there is a slight dip in its trust values at the beginning of week one reflecting its inability to pick up the event taking place during this time period. However its trust values increase after this.

A similar scenario is seen for the two and three parameter models (M4-M7), however in this case it is M4 (Cond-Depth), M5 (Cond-Temp) and M7 (Cond-Temp-Depth) that are the most accurate in replicating the activities of the DO sensor and M6 (Depth-Temp) which performs the poorest. These patterns seen in relation to the trust values are again representative of the performance metrics of M1-M7 on

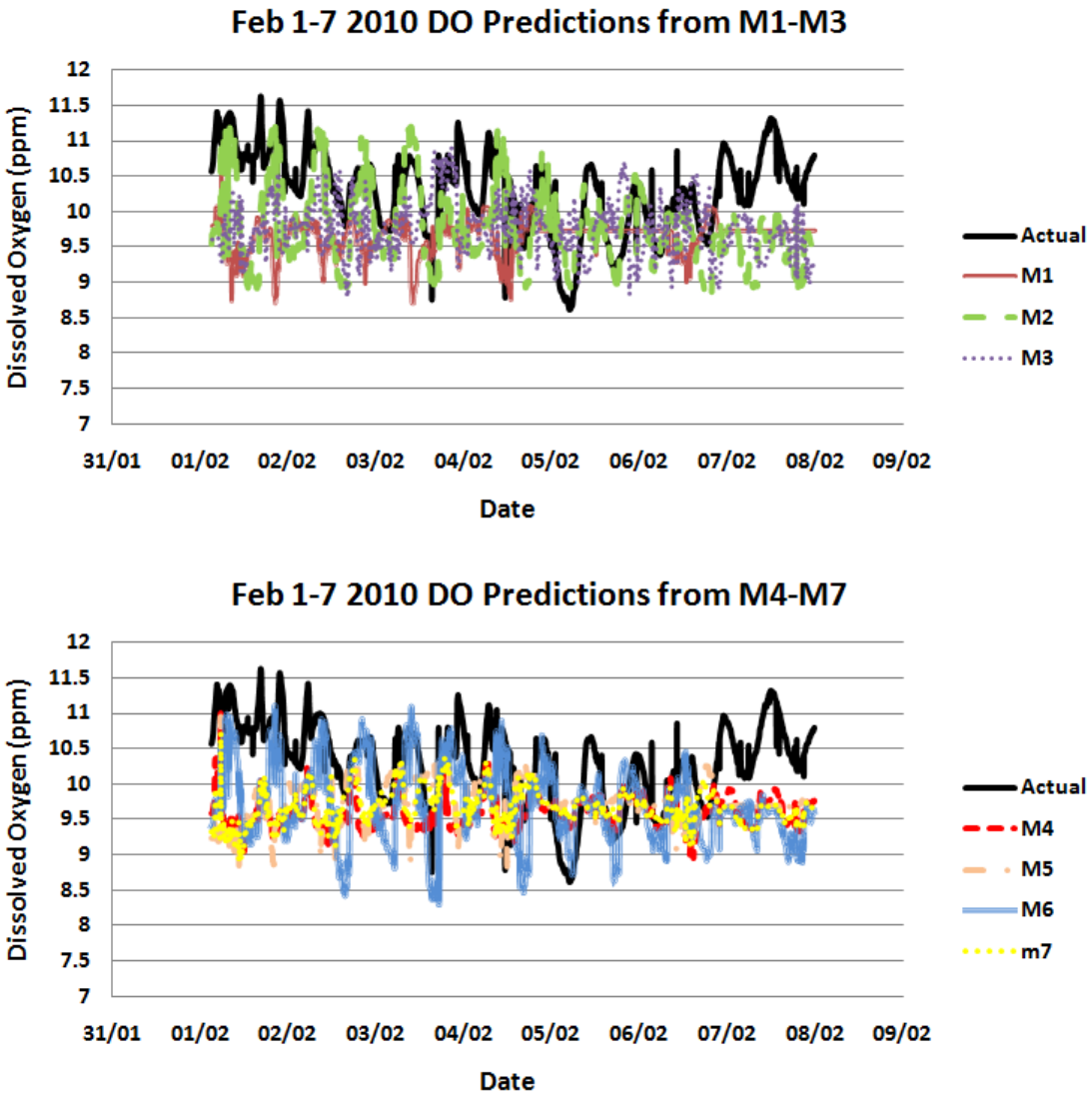

Figure D.1: Graphs showing the dissolved oxygen values output by M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) compared to the actual values output by the dissolved oxygen sensor for Feb 1-7 2010

the test data - with M1, M4, M5 and M7 clearly producing better performances in terms of CC, MAE and RMSE compared to the other models. M4 outperforms M1, M5 and M7 with a CC of 0.7245, a MAE of 1.1451, and a RMSE of 1.4233. This is reflected in the trust and reputation framework where it has the highest trust values at all times except for a period in the second half of week one.

Sometimes it can be seem from plotting the outputs of the each of the models

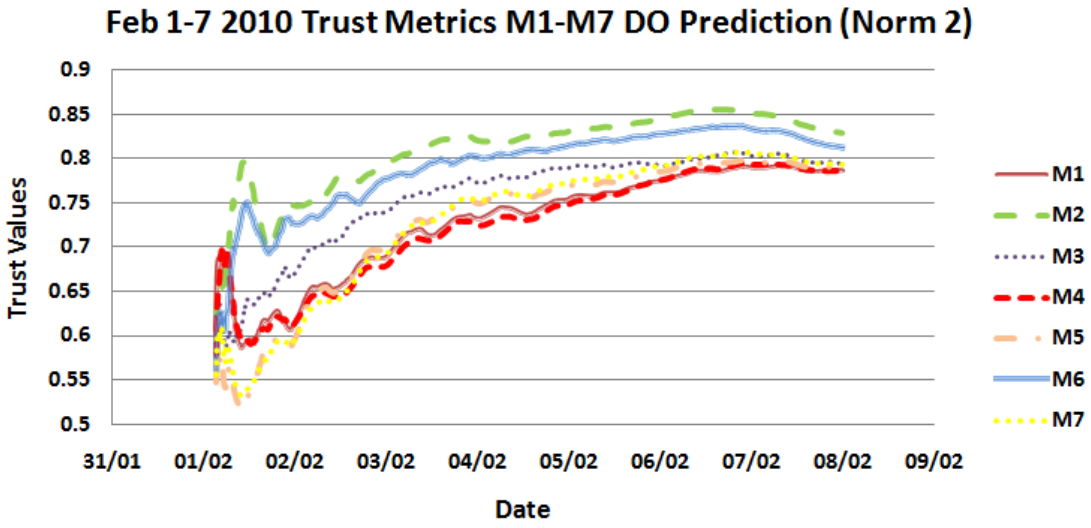

Figure D.2: Graphs showing the trust values for M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) for predicting dissolved oxygen from Feb 1-7 2010 when normalisation using the range of the February training set is used

which particular models should generally have the higher trust values in the trust framework. However this framework can also tell you the magnitude of unreliability of one model in replicating the activities of the sensor in comparison to another. In Figure D.1 we see a scenario whereby it is very difficult to determine which model is most reliable from a plot of the outputs. This is the output of the models for DO prediction for the first week in February. This is reflected in the output of the trust model whereby the trust values for each of the models are close but it does show differences in the level of performance of the models which is difficult from simply looking at a plot of the outputs (See Figure D.2).

Best Output

Here a similar evaluation is carried out as to that for Conductivity and Depth prediction. The results of this evaluation can be seen in Table D.2. The TrustModel (TM) produces very similar outputs to the best performing model(s) on test data for each month (*May*- M5, *June*- M1, *July*- M4, *Jan*- M3, *Feb*- M2).

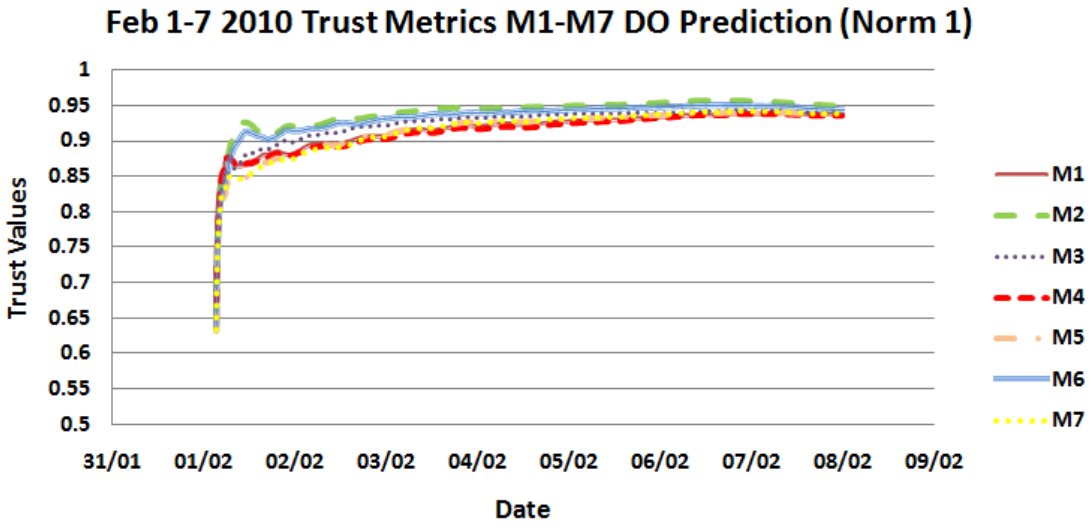

Figure D.3: Graphs showing the trust values for M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) for predicting dissolved oxygen from Feb 1-7 2010 when normalisation using the range of the full training set is used

	CC	MAE	RMSE
May	0.5211	0.9906	1.2351
June	0.4898	1.3473	1.6009
July	0.7025	1.1423	1.4247
Jan	0.4416	1.5011	1.7499
Feb	0.4360	0.6065	0.7342

Table D.2: Performance metrics when the output of the most trustworthy model is selected at each epoch for DO prediction

This outlines the ability of TM to capture the best performing nodes even in times of noisy outputs. As previously outlined the performance of the models here is generally lower than that achieved by the models for conductivity or depth. Figure D.7 shows the values from the dissolved oxygen sensor plotted alongside the predicted dissolved oxygen values when the value from the most trustworthy model is chosen at each time epoch, for the last two weeks in June.

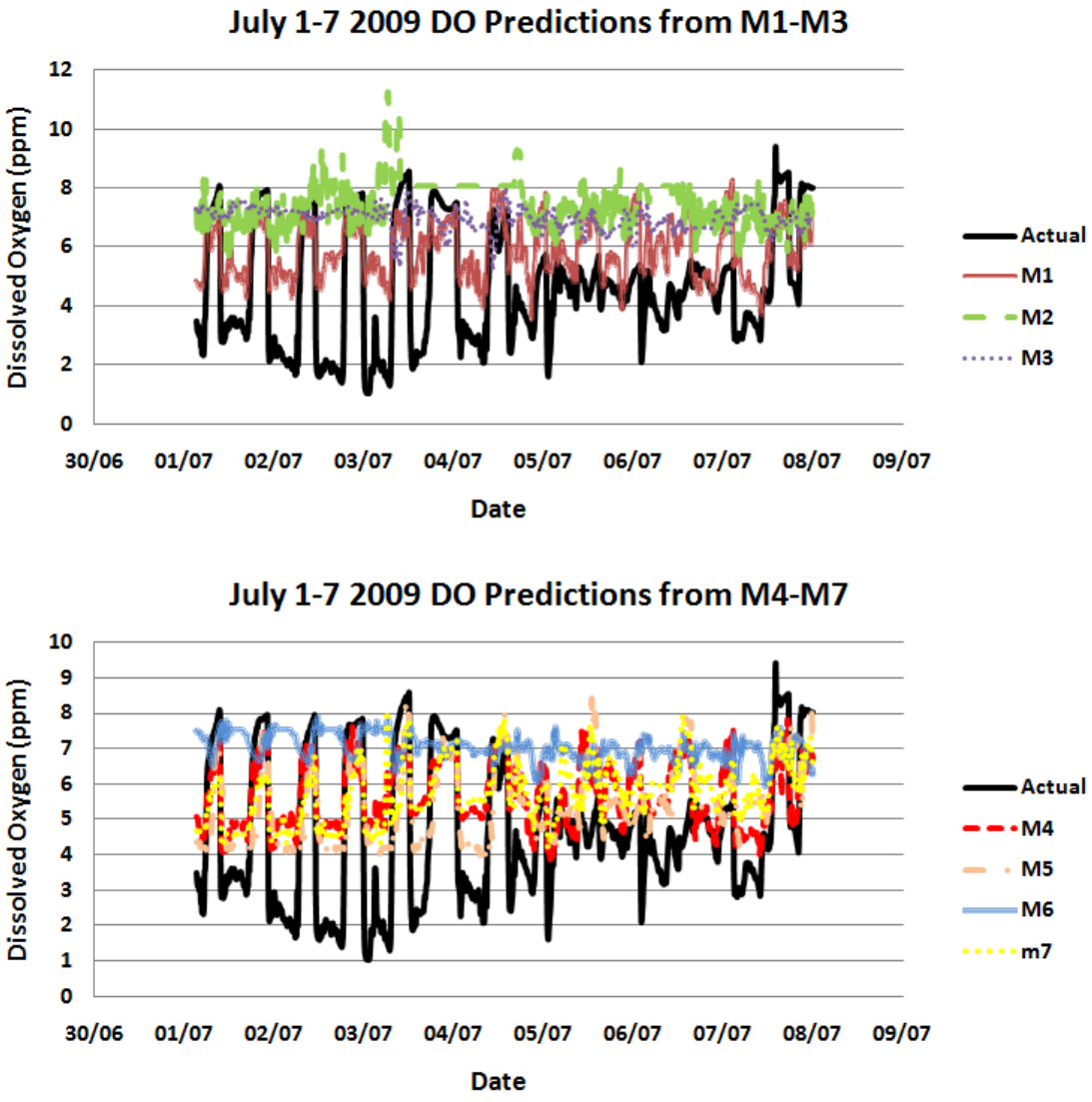

Figure D.4: Graphs showing the dissolved oxygen values output by M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) compared to the actual values output by the dissolved oxygen sensor for July 1-7 2009

D.2 Temperature at the Lee Maltings

For the prediction of temperature the generalised model versions of M1-M7 were used for predictions in June. The monthly model versions of M1-M7 was used for predictions in May, July, January and February. Similar to the case with dissolved oxygen, the generalised model appeared to produce quite high MAE's for certain models in January and February.

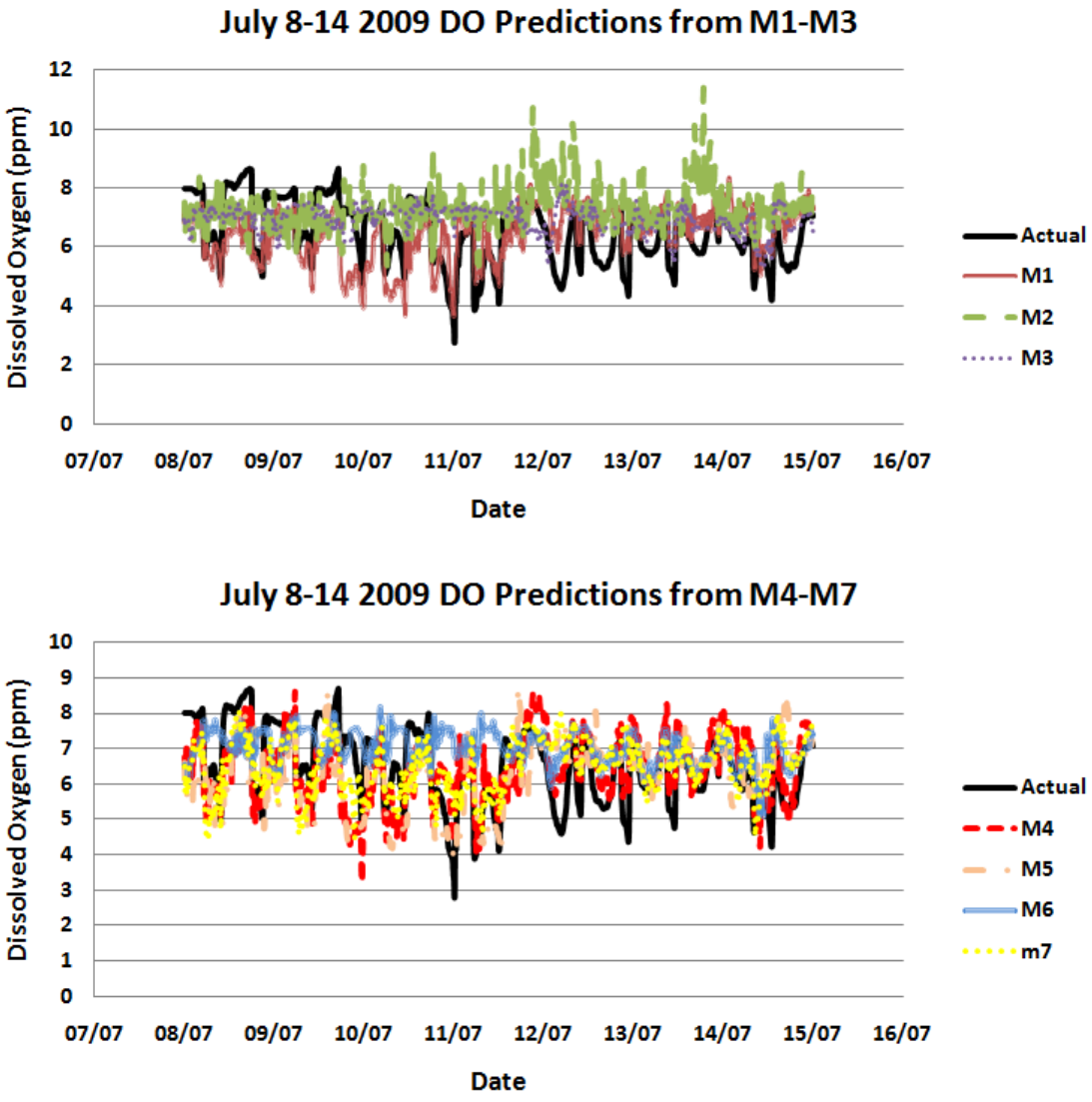

Figure D.5: Graphs showing the dissolved oxygen values output by M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) compared to the actual values output by the dissolved oxygen sensor for July 8-14 2009

Again the generalised model uses 10,000 training instances, the May, July and February monthly models use 2,000 training instances and the January models range from 488 to 2551 training instances (M1, M4, M5, M7 - 488, M2 - 2551, M3, M6 - 2531). The number of test instances in each month are - May - 2006, June - 2305, July - 1956, January - 2089, February - 2108. The performance of the models on the test data is shown in Table D.3. The May June and July results are generally

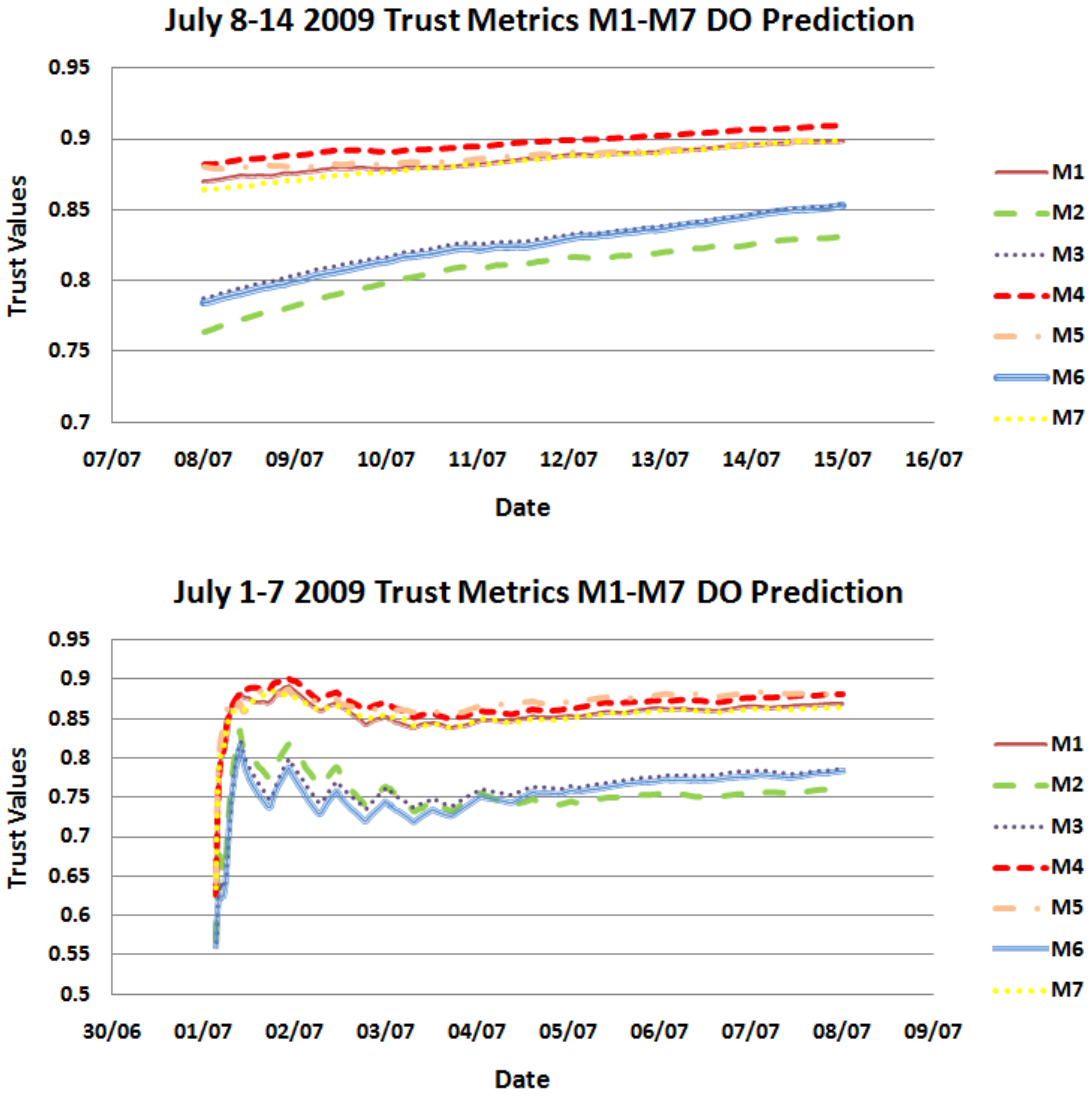

Figure D.6: Graphs showing the trust values for M1-M3 (Cond, Depth, Temp) and M4-M7 (Cond-Depth, Cond-Temp, Depth-Temp, Cond-Temp-Depth) for predicting dissolved oxygen from July 1-14 2009

poorer than that achieved for depth and conductivity.

D.2.1 Application of the Trust and Reputation-Based framework

Again at each epoch the Watchdog assigns a level of cooperation to a model based on the deviation of its output from the output value of the sensor (See Equation 7.3

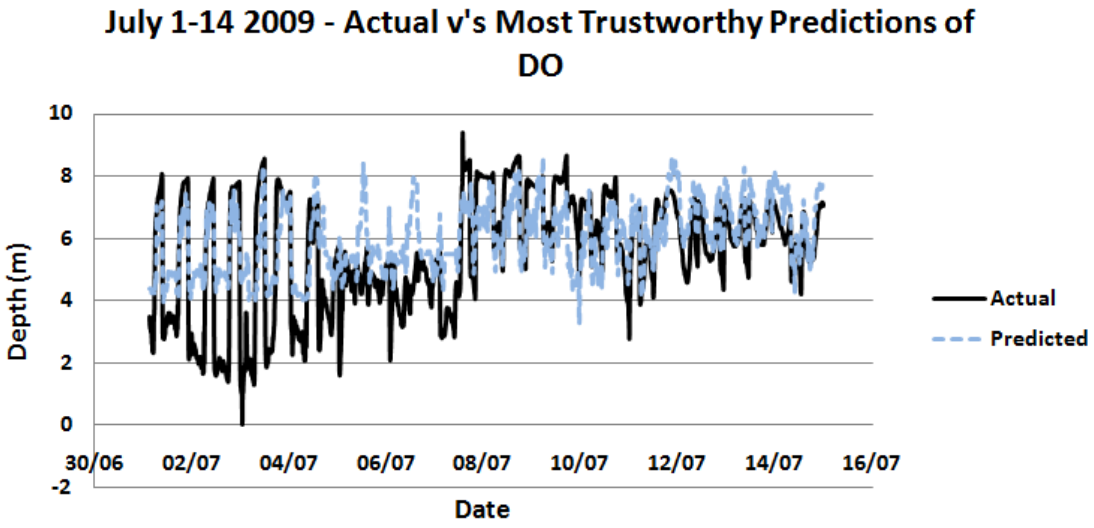

Figure D.7: Actual v's predicted values for dissolved oxygen when the value from the most trustworthy model is chosen at each time epoch for July 1-14 2009

on page 187). Similarly to the case with dissolved oxygen, the range of the dataset can vary greatly from month to month as can be seen in 7.2. In this case monthly models were used for all predictions except June. Thus for May test data, the error measures were normalised by dividing by the max-min range of the May training set - [17.12-11.89]. For July test data, the range of [18.56-15.03] is used. For January test data the max-min range of [10.89-3.26] is used and for February test data the max-min range of [7.91-5.01] is used. Then for June test data, where a generalised model is used for carrying out predictions - a range of [18.56-3.26] - is used. Similar to previous, these values are then subtracted from 1. The same methodology is applied as before for updating reputation and trust values (See Equation 7.1 on page 177 and Equation 7.2 on page 177).

D.2.2 Results

The output of each of the models for the first two weeks in January along with the actual values output by the DO sensor are shown in Figures D.8 and D.9. The associated trust values of each of the models at each epoch is shown in Figure D.10.

May	CC	MAE	RMSE
M1	0.3775	1.14	1.3339
M2	0.0776	1.0545	1.2364
M3	0.3759	1.1171	1.2828
M4	0.3456	1.1898	1.4
M5	0.434	1.1639	1.3261
M6	0.3389	1.1828	1.344
M7	0.4315	1.1321	1.2847

June	CC	MAE	RMSE
M1	0.2466	1.6425	1.8228
M2	−0.0091	1.9302	2.2597
M3	0.172	3.5561	3.8212
M4	0.1215	1.778	2.0979
M5	0.2942	1.6482	1.8215
M6	−0.0109	1.9796	2.253
M7	0.1104	1.8413	2.1409

July	CC	MAE	RMSE
M1	0.2162	1.0685	1.3026
M2	0.1422	1.398	1.5885
M3	0.2134	1.0451	1.2067
M4	0.2566	1.2128	1.4489
M5	0.2167	1.0898	1.3108
M6	0.1858	1.2481	1.4537
M7	0.2603	1.2154	1.4464

Jan	CC	MAE	RMSE
M1	0.5741	2.6865	2.8242
M2	0.1808	0.9273	1.1888
M3	0.0701	1.9394	2.2132
M4	0.773	2.3116	2.4193
M5	0.5488	2.8167	2.9511
M6	0.219	0.8929	1.1559
M7	0.7684	2.3154	2.424

Feb	CC	MAE	RMSE
M1	0.6057	0.4335	0.6006
M2	−0.1081	0.6716	0.8266
M3	0.5234	0.5172	0.6247
M4	0.5829	0.4487	0.6039
M5	0.6376	0.4289	0.5859
M6	0.2767	0.599	0.7275
M7	0.5981	0.4499	0.5945

Table D.3: Performance of each of the models when evaluated in the corresponding test data for the prediction of temperature

It is clear from looking at the output of the models that for the single parameter models (M1-M3) that M2(DO) is clearly most in line with the outputs of the temperature sensor compared to the other two models. For the two or three parameter models (M4-M7), it appears to be M6 (DO-Depth). The flat line of outputs by M1 (Cond), M5 (Cond-Depth) and M7 (Cond-Depth-DO), is most likely due to missing data in the inputs to the model. It is also interesting to note that while M2 and M6 appear to coincide better with the outputs of the temperature sensor, they are not really picking up the oscillations taking place at the beginning of week 2. M1, M4 (Cond-DO), M5 and M7 appear to correlate much better with with these oscillations however at a higher scale. However, the manner in which the cooperation values are calculated will render M2 and M6 more reliable in this scenario. An alternative methodology in calculating these value would produce a different result.

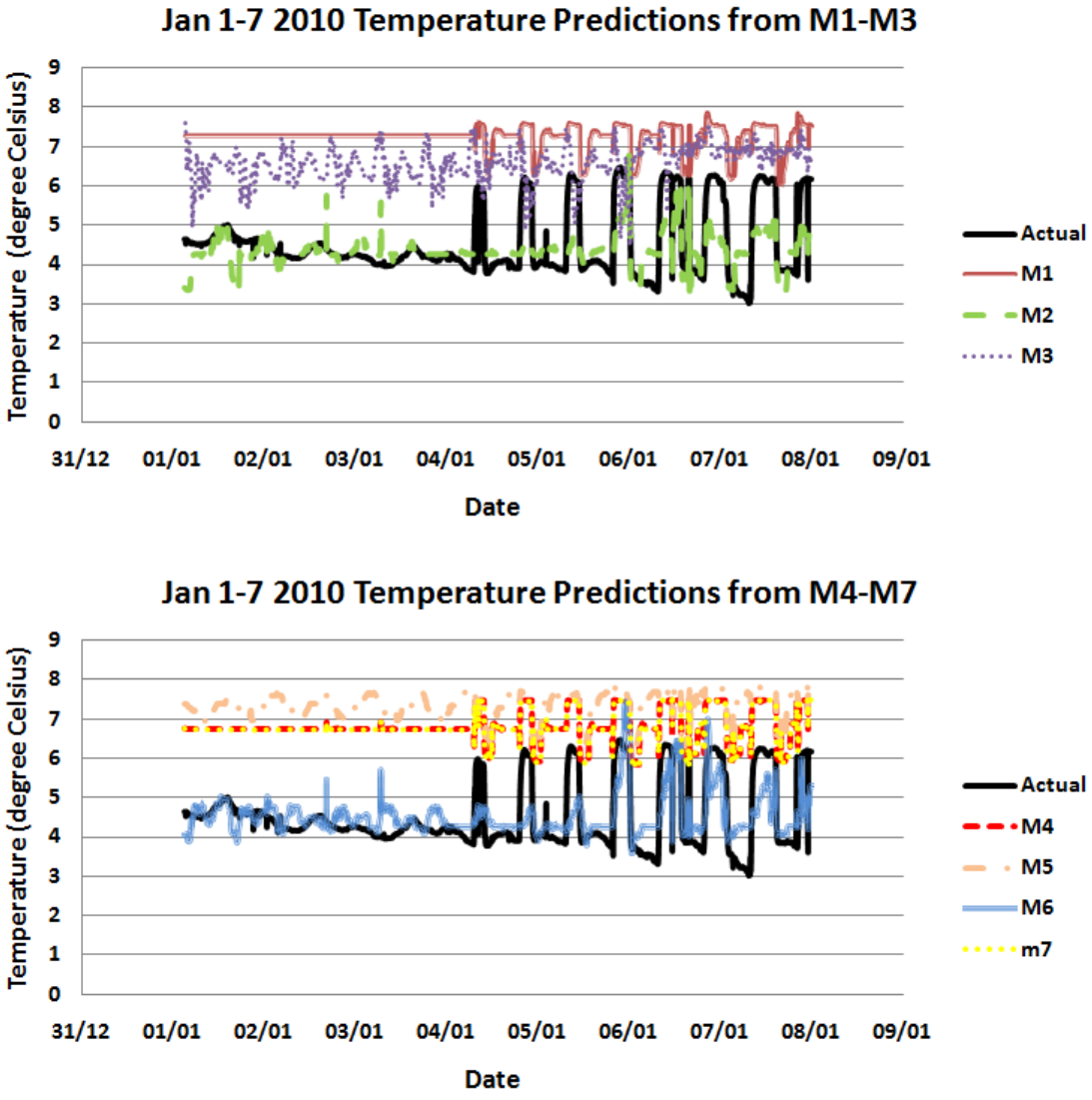

Figure D.8: Graphs showing the dissolved oxygen values output by M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Depth, DO-Depth, Cond-Depth-DO) compared to the actual values output by the dissolved temperature sensor for Jan 1-7 2010

There is a slight dip in the trust values of M1 towards the end of the second week, which is probably caused by the flat line output. This can also be seen for M1 along with M4, M5 and M7 at the beginning of the first week where similar outputs are seen for each. There is also a dip in trust values for M3, which doesn't produce a flat line of outputs, however it is still producing outputs further from the curve produced by the temperature sensor than it did in the first half of of the week.

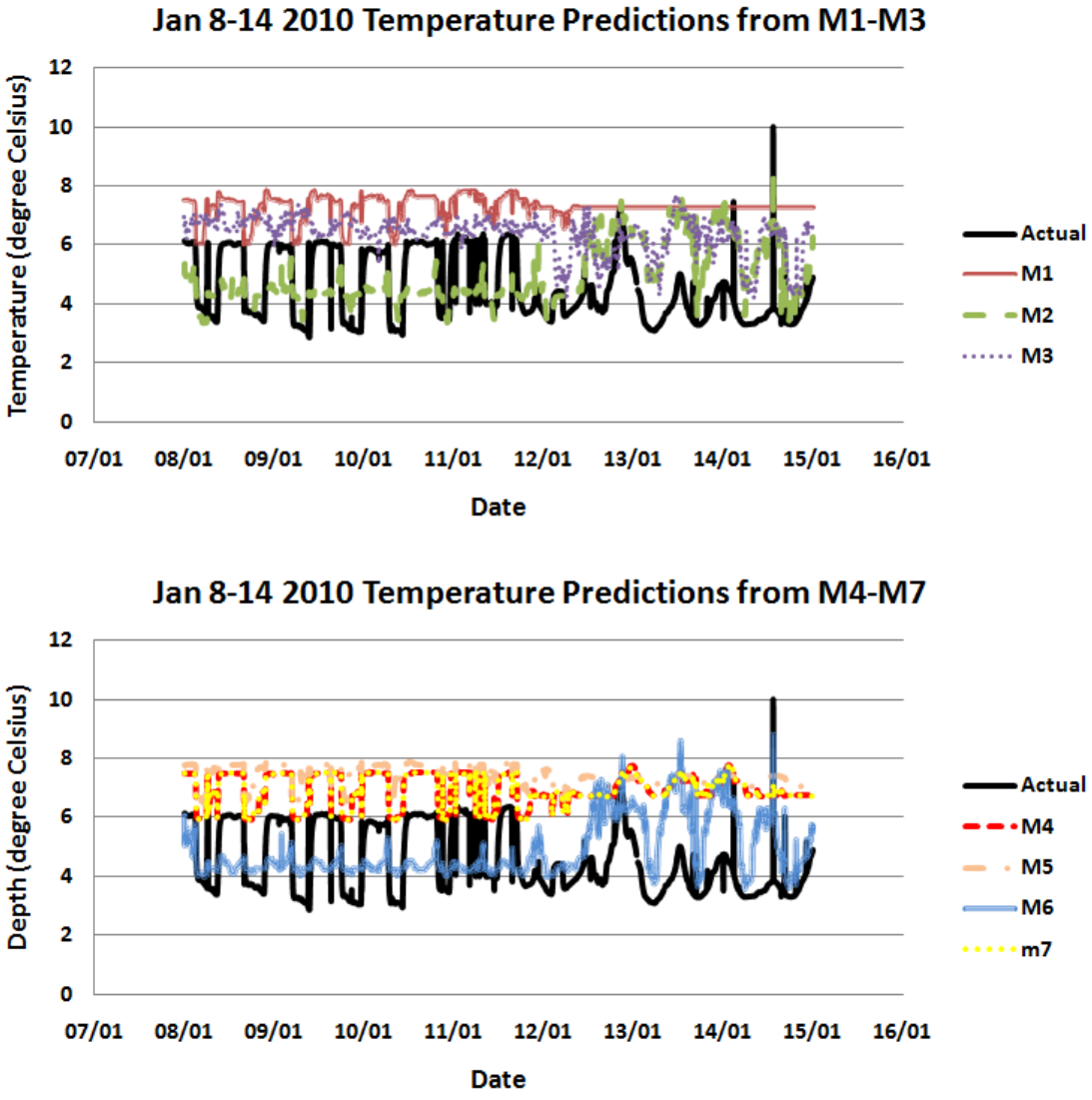

Figure D.9: Graphs showing the dissolved oxygen values output by M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Depth, DO-Depth, Cond-Depth-DO) compared to the actual values output by the dissolved temperature sensor for Jan 8-14 2010

The inability of M2 and M6 to pick up the oscillations at the beginning of the second week is coincided with a reduction in trust values during this time period. These trust values continue to decrease as even though it it replicating the pattern produced by the temperature sensor much better towards the end of the week, it it producing values that are higher than the values of the temperature sensor. It is interesting to note that these trends coincide with measures of MAE produced by

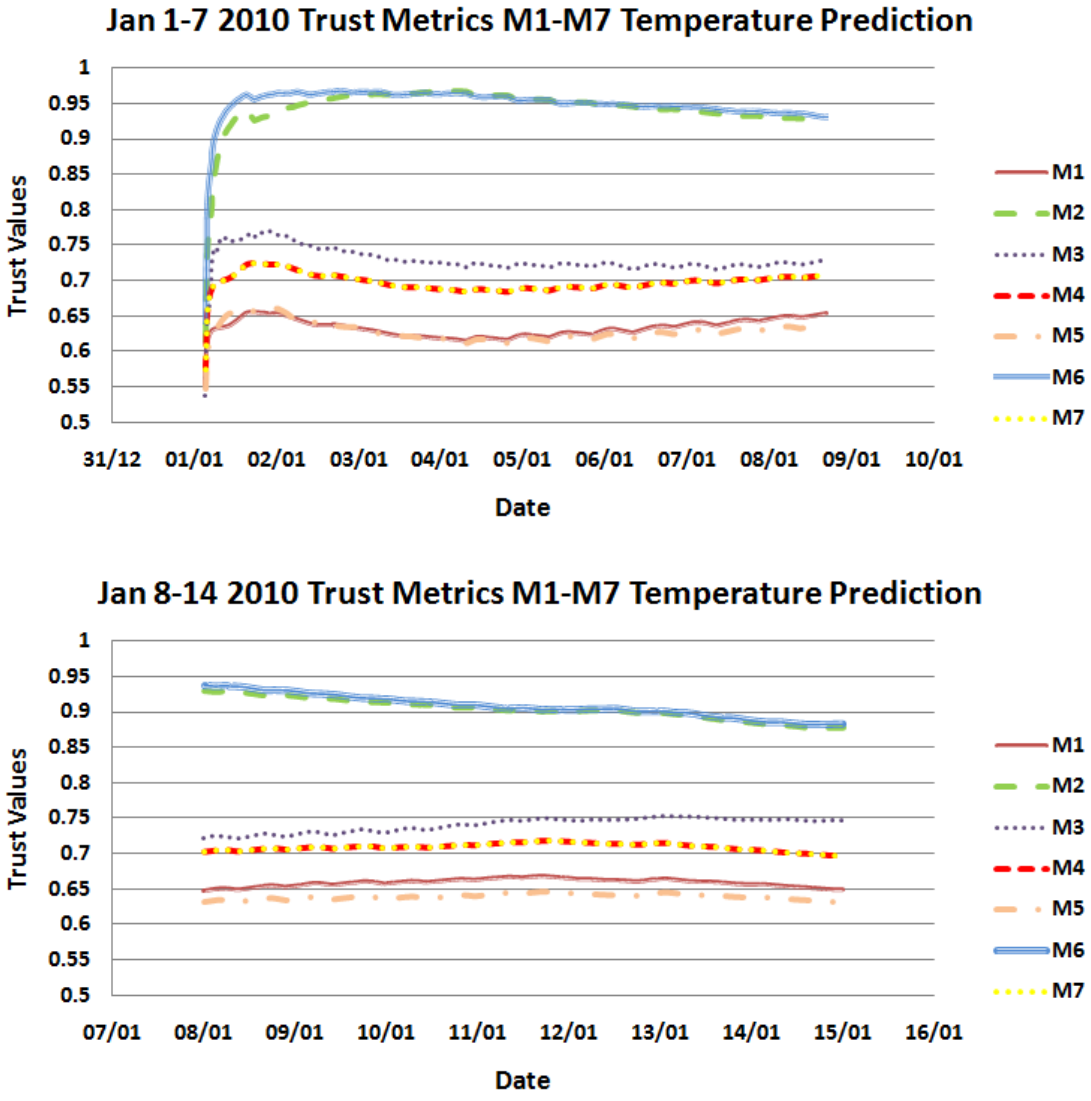

Figure D.10: Graphs showing the trust values for M1-M3 (Cond, DO, Temp) and M4-M7 (Cond-DO, Cond-Depth, DO-Depth, Cond-Depth-DO) for predicting temperature from Jan 1-14 2010

M1-M7 on the test data. From table D.3 it can be seen that while M1, M4, M5 and M7 produce by far the highest correlations on the test data which is apparent from the analysis above, they produce much higher MAE's than M2 and M6 which have the by far the lowest MAE's of all the models and are deemed to be the most trustworthy by the reputation and trust model. Hence the reputation and trust framework is accurately reflecting the type of method that is being implemented to determine cooperativeness or reliability i.e. deviation from the output value of the

temperature sensor.

Again it is apparent that the trust and and reputation model allows us to pick up trends which are difficult to decipher from simply plotting the data. From Figure D.10 it is clear that M2 and M6 are the most reliable in terms of outputting values close to that of the temperature sensor. It is difficult to tell from plotting the outputs which of these are most reliable but from the trust values it appears that M6 has a slight edge. From looking at Figures D.8 and D.9, it appears that M3 would be deemed the next reliable with M5 and M4 the least reliable. However it is not really clear. However these assumptions are confirmed by the output of the trust and reputation model. It also allows us to see the difference in the magnitude of trust between models. For example it reflects that M2 and M6 are similar in reliability, whereas a large difference can be seen between their trust values and the trust values of the other models.

Best Output

	CC	MAE	RMSE
May	0.2220	1.0534	1.2485
June	0.3364	1.6411	1.8331
July	0.2133	1.0451	1.2067
Jan	0.2168	0.8927	1.1569
Feb	0.6233	0.4279	0.5745

Table D.4: Performance metrics when the output of the most trustworthy model is selected at each epoch for temperature prediction

Here a similar evaluation is carried out as to that for conductivity, depth and DO prediction. This results are shown in Table D.4. Here we are seeing cases where we have models which produce high CC's but also high MAE's. However it is notable here in that TM (TrustModel) matches the best performing models in terms of MAE's as opposed to CC's while sometimes improving correlation.

In terms of MAE and RMSE the best performing model for May is M2, although it has a poor correlation of 0.0776. Other models are reaching correlations of 0.43,

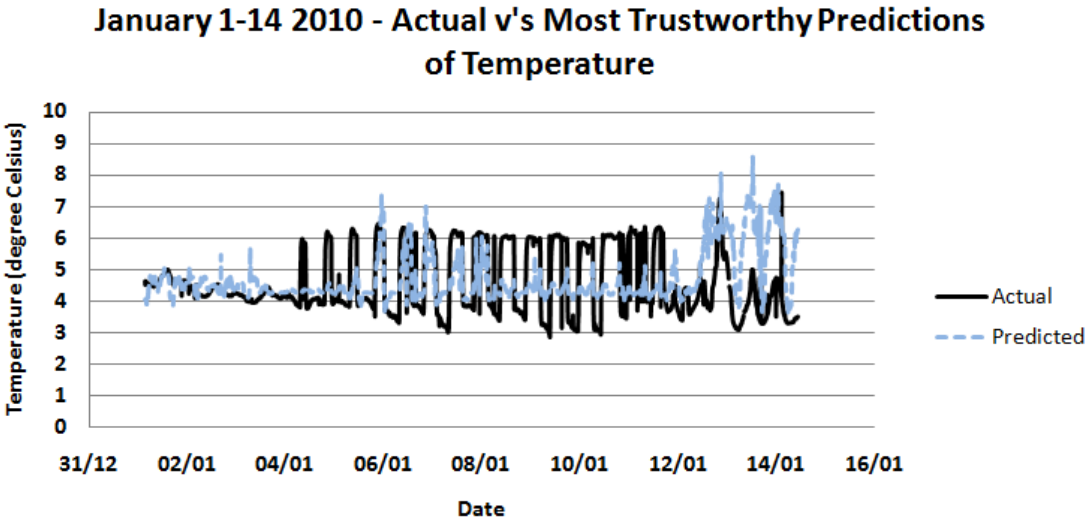

Figure D.11: Actual v's predicted values for temperature when the value from the most trustworthy model is chosen at each time epoch for Jan 1-14 2010

but with slightly higher MAE's and RMSE's. TM produces an output with a CC of 0.222 and a MAE and RMSE almost identical to that of M2, at 1.0534 and 1.2485. For June test data, TM produces an output which has an MAE very slightly lower than the lowest MAE produced by any of the models, with an increased correlation of 0.3364. In June the best performing models are M1 and M5 with CC's of 0.2466 and 0.2942, and MAE's of 1.6425 and 1.6482. For July, TM matches the model with the lowest MAE and RMSE (M3). M1, M4 and M7 produce slightly higher CC's but also slightly higher MAE's and RMSE's. As one would expect from the analysis carried out on the test data from January above, TM produces a very slightly lower MAE to the lowest achieved by any of the models on test data from January, and a very similar CC and RMSE. For February, TM achieves a very similar output to M5, the best performing model for February test data, with a marginally lower MAE and RMSE. Figure D.11 shows values from the temperature sensor plotted alongside the predicted temperature values when the value from the most trustworthy model is chosen at each time epoch, for the first two weeks in January.

Appendix E

Threshold Margins for Models and Remote Sensing Data Outputs

E.1 Modelled Ouputs at the River Lee

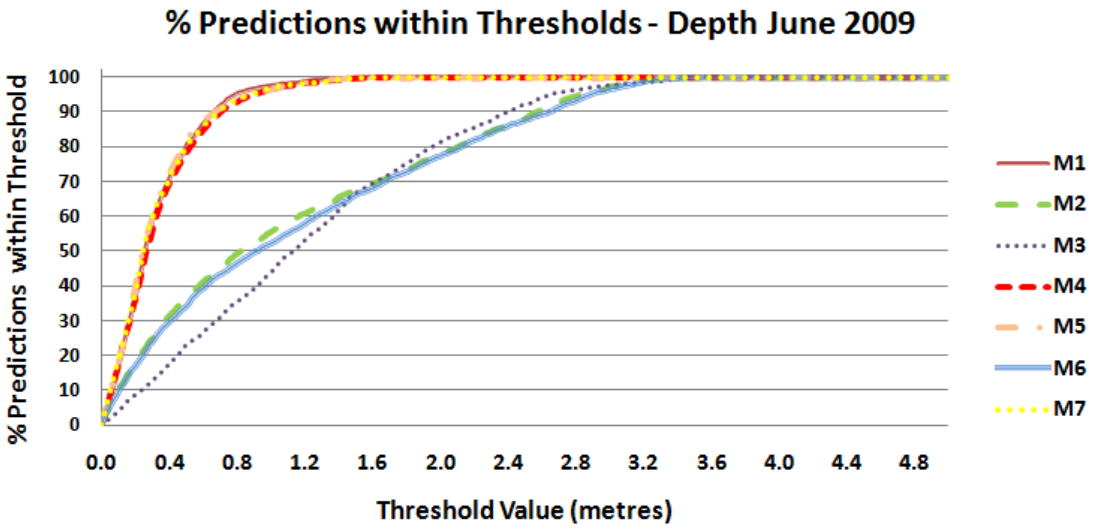

Figure E.1: Models 1-7 for predicting depth June 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

E.2 HRDDS SST Analysis Products

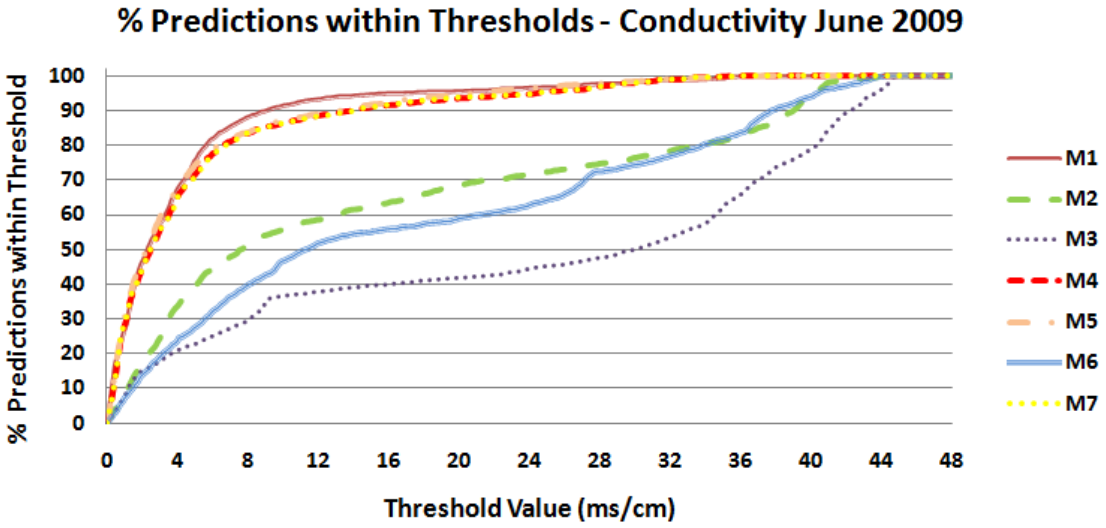

Figure E.2: Models 1-7 for predicting conductivity June 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

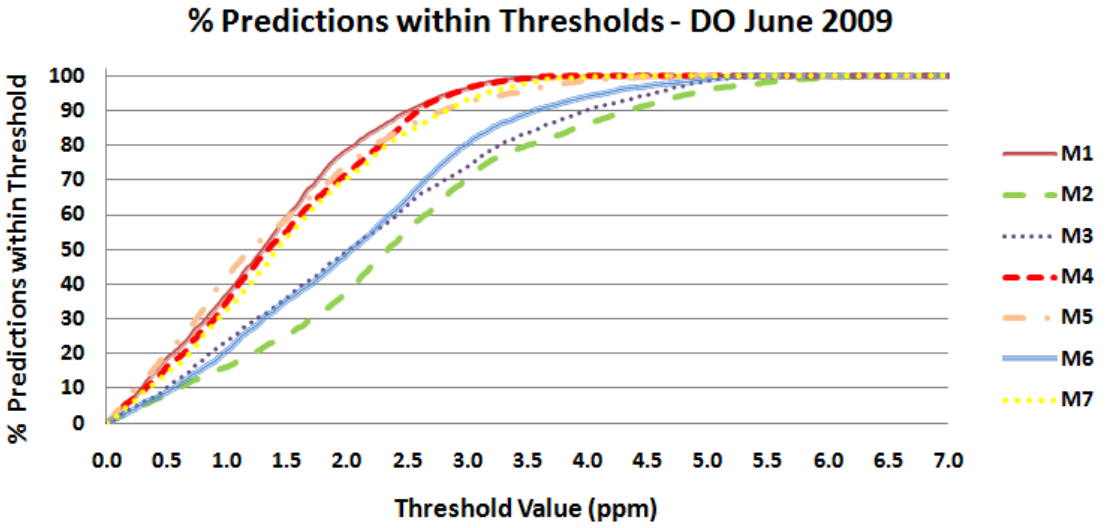

Figure E.3: Models 1-7 for predicting dissolved oxygen June 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

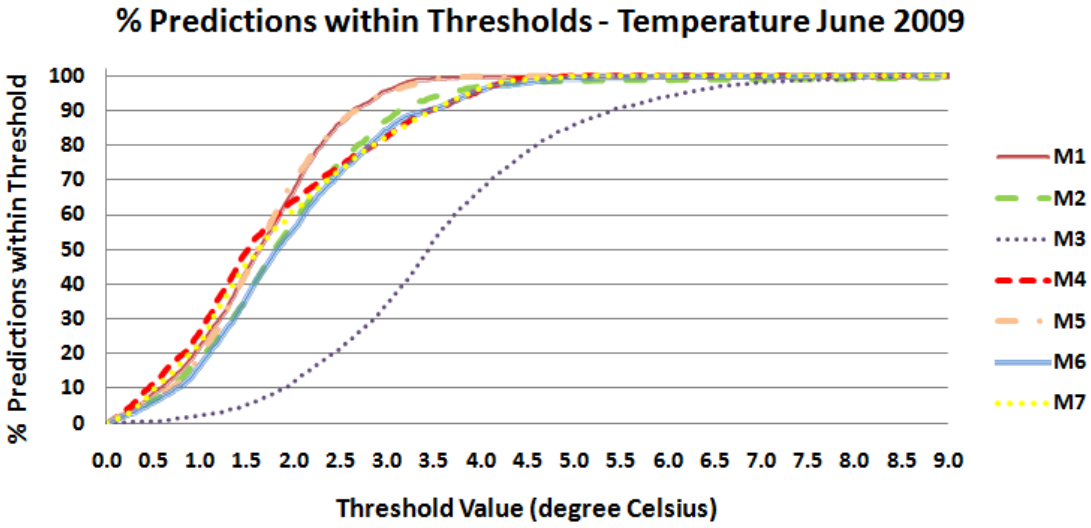

Figure E.4: Models 1-7 for predicting temperature June 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

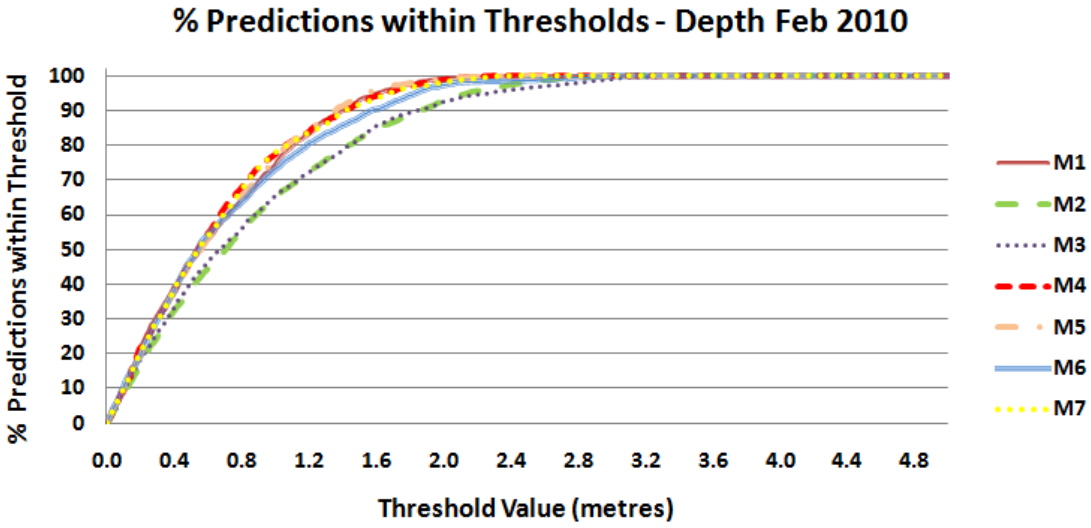

Figure E.5: Models 1-7 for predicting depth February 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

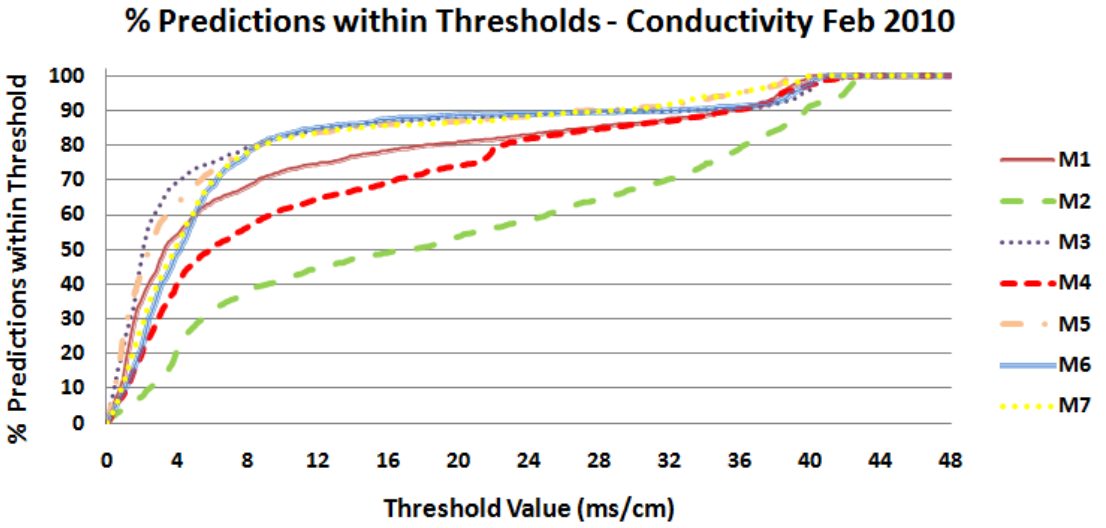

Figure E.6: Models 1-7 for predicting conductivity February 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

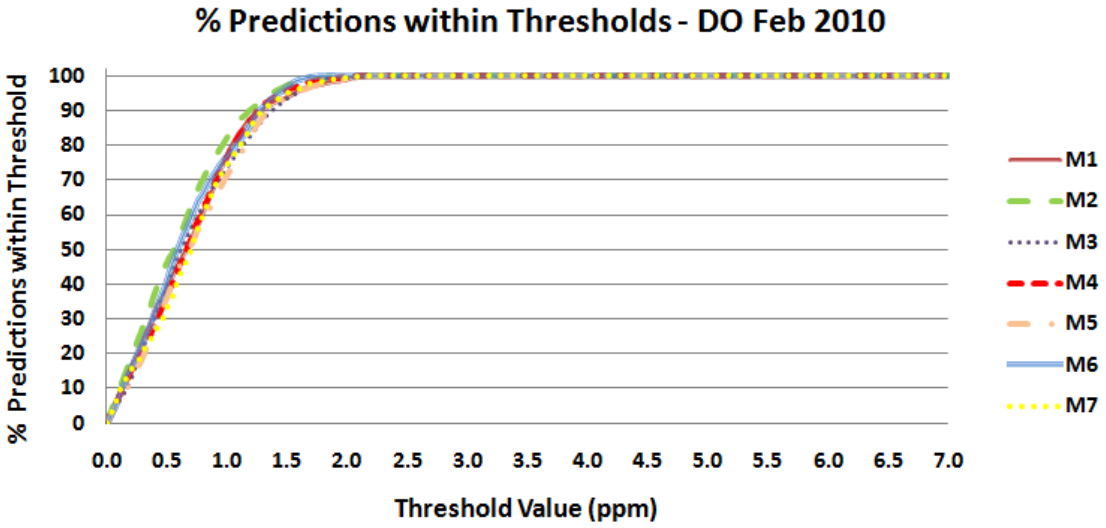

Figure E.7: Models 1-7 for predicting dissolved oxygen February 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

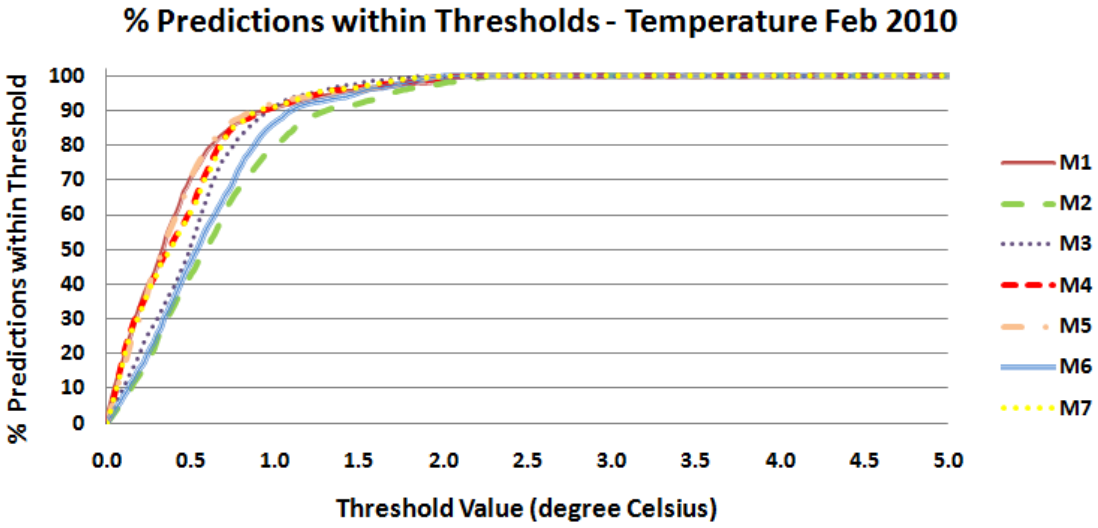

Figure E.8: Models 1-7 for predicting temperature February 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

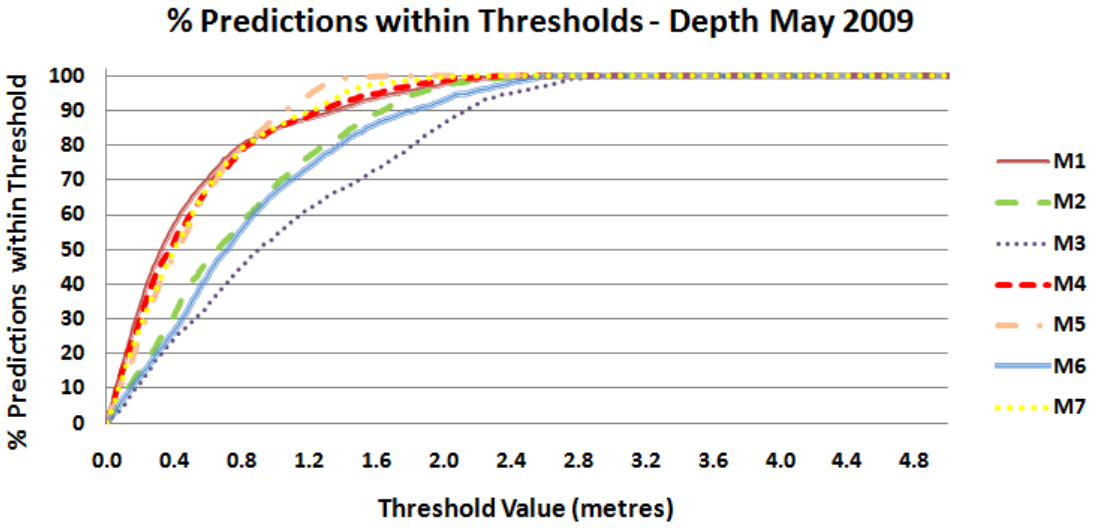

Figure E.9: Models 1-7 for predicting depth May 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

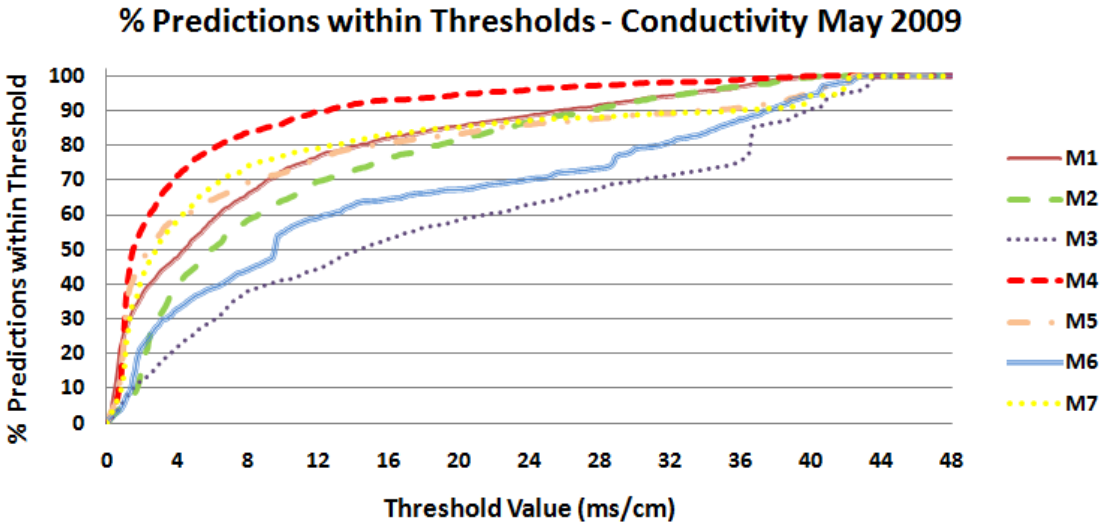

Figure E.10: Models 1-7 for predicting conductivity May 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

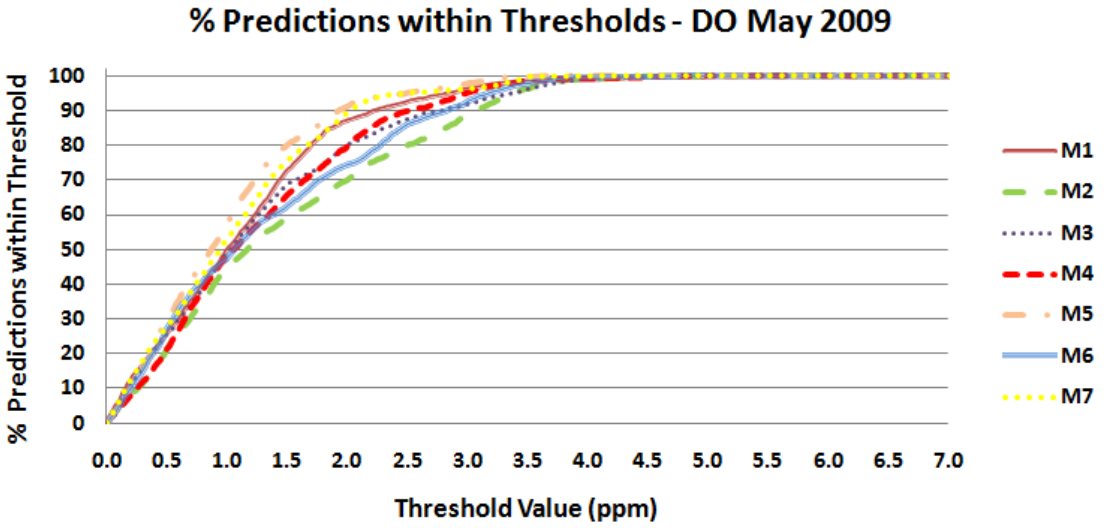

Figure E.11: Models 1-7 for predicting dissolved oxygen May 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

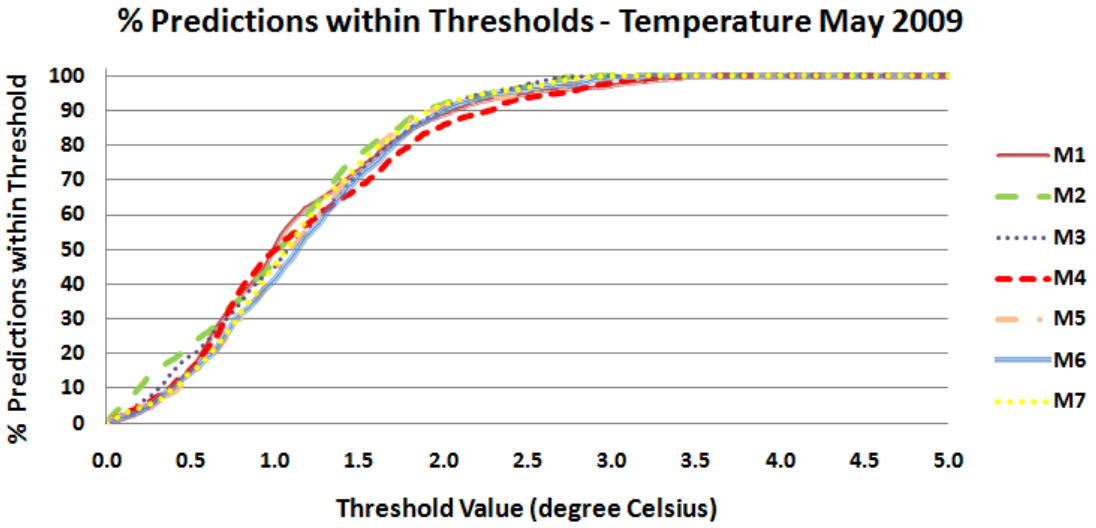

Figure E.12: Models 1-7 for predicting temperature May 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

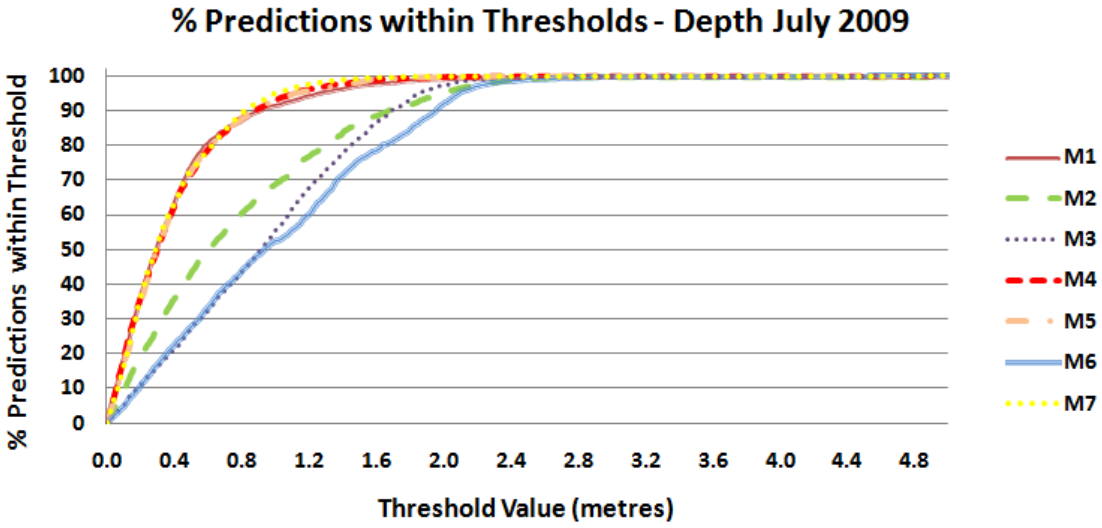

Figure E.13: Models 1-7 for predicting depth July 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

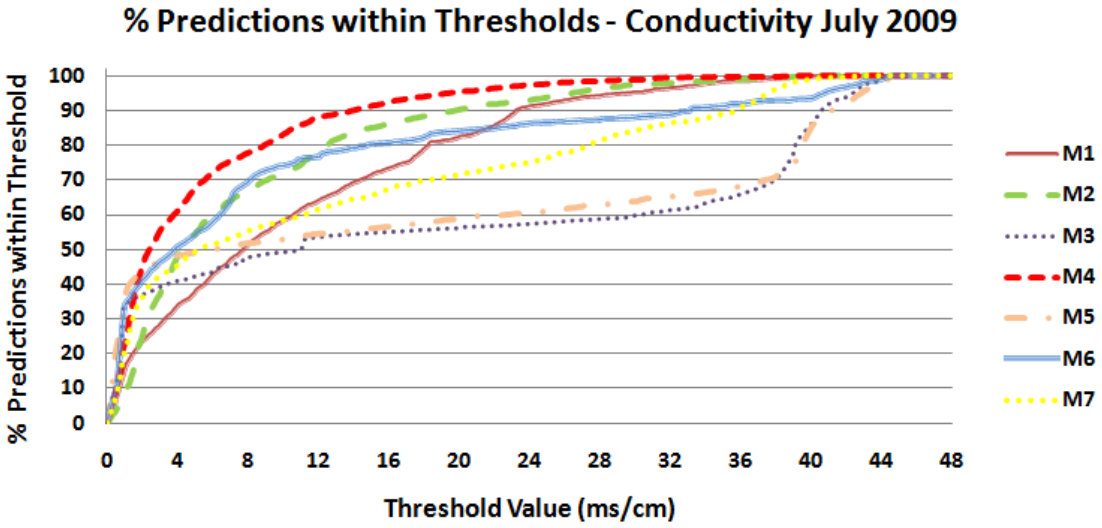

Figure E.14: Models 1-7 for predicting conductivity July 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

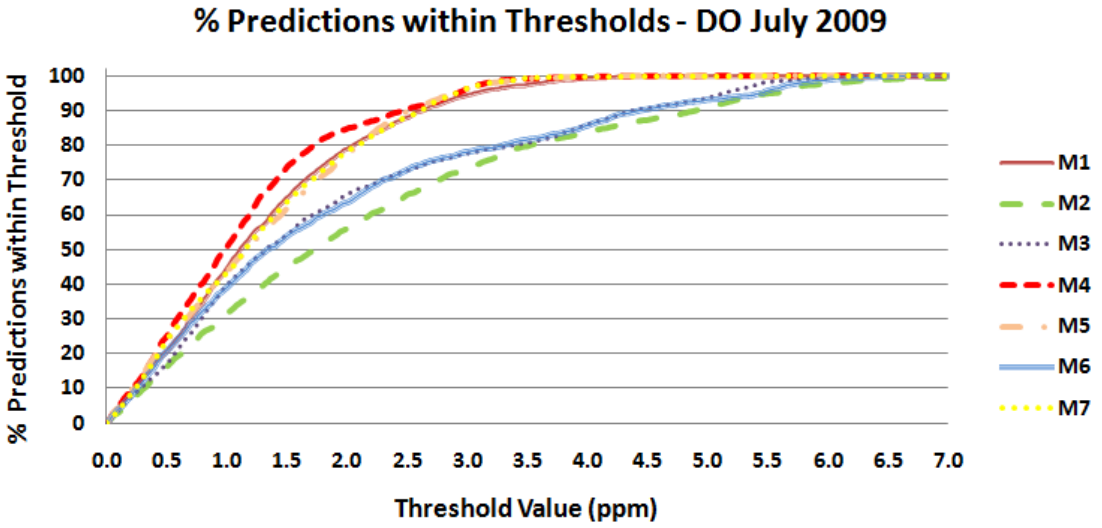

Figure E.15: Models 1-7 for predicting dissolved oxygen July 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

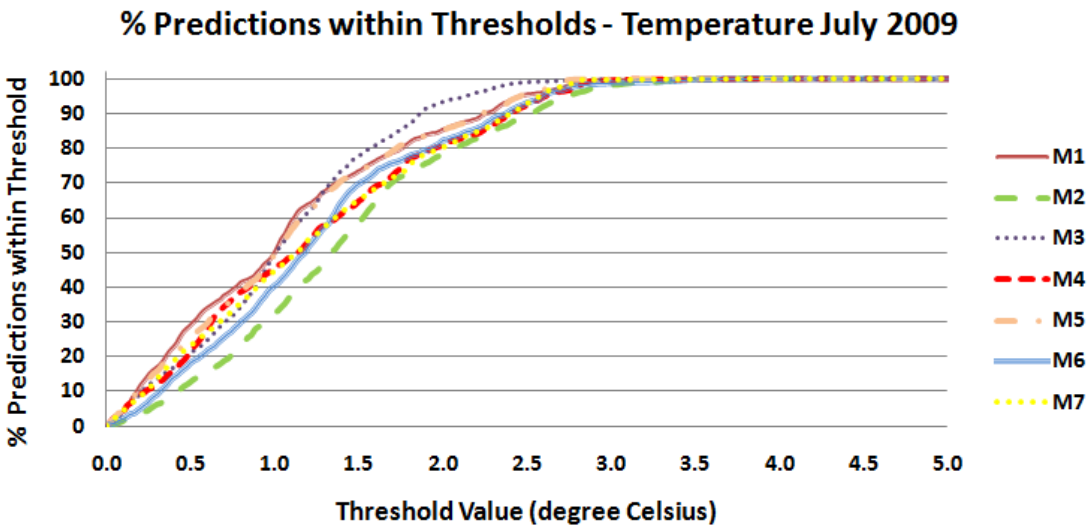

Figure E.16: Models 1-7 for predicting temperature July 2009 - Percentage of predictions within various thresholds of the in-situ sensor readings.

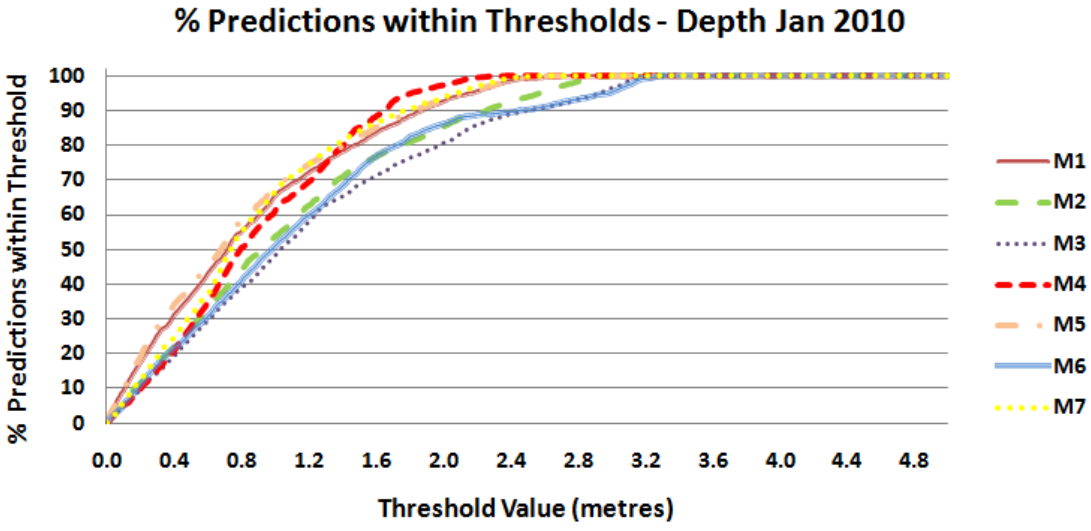

Figure E.17: Models 1-7 for predicting depth January 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

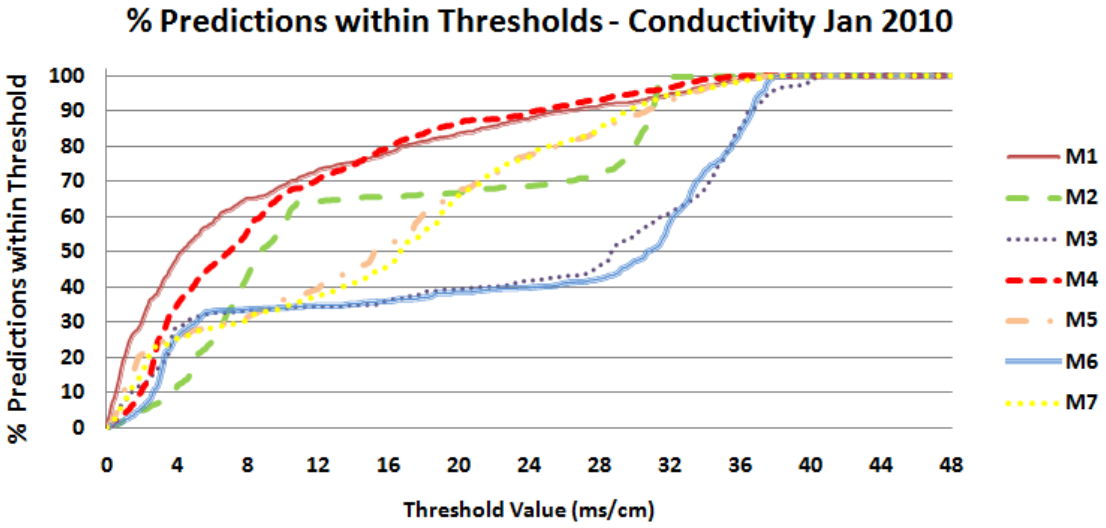

Figure E.18: Models 1-7 for predicting conductivity January 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

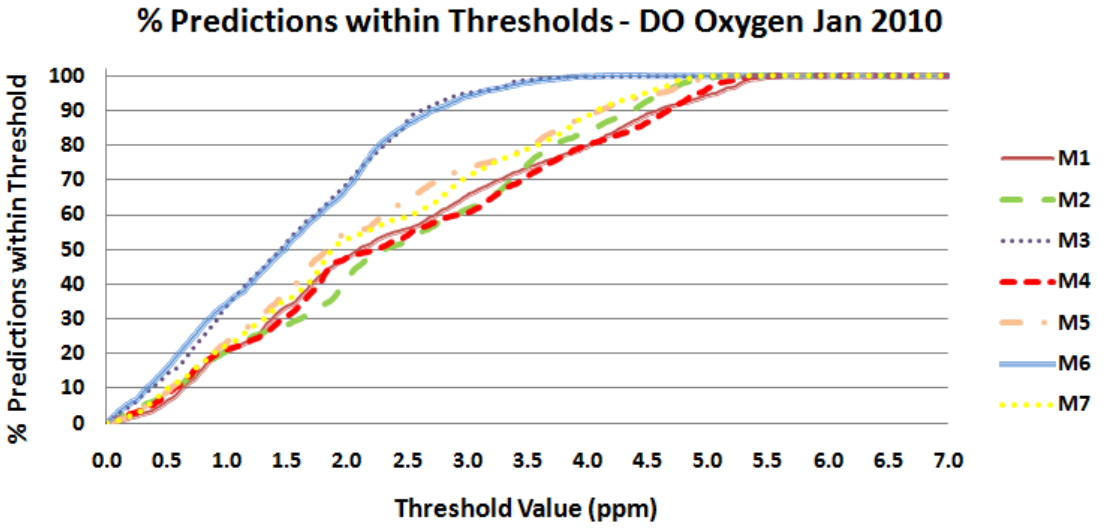

Figure E.19: Models 1-7 for predicting dissolved oxygen January 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

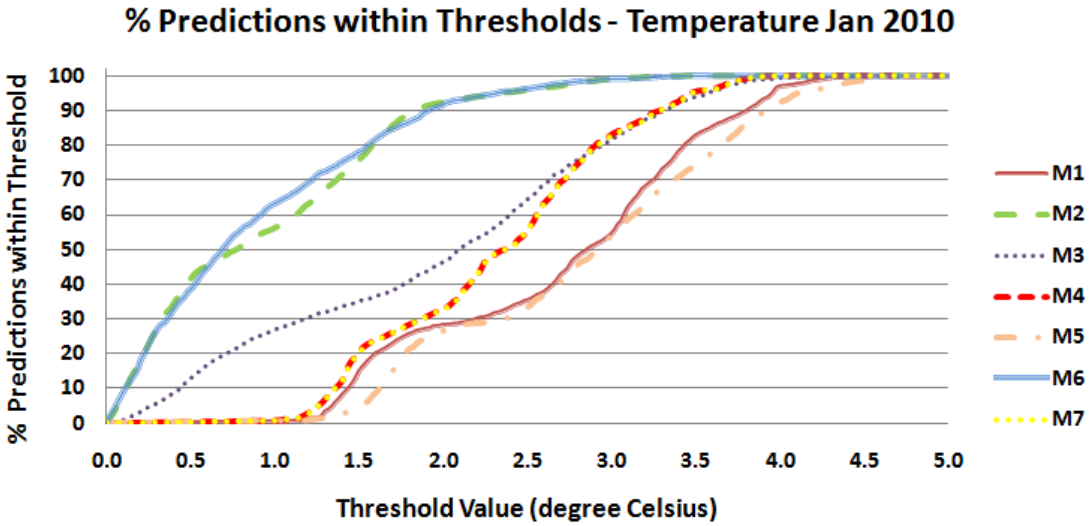

Figure E.20: Models 1-7 for predicting temperature January 2010 - Percentage of predictions within various thresholds of the in-situ sensor readings.

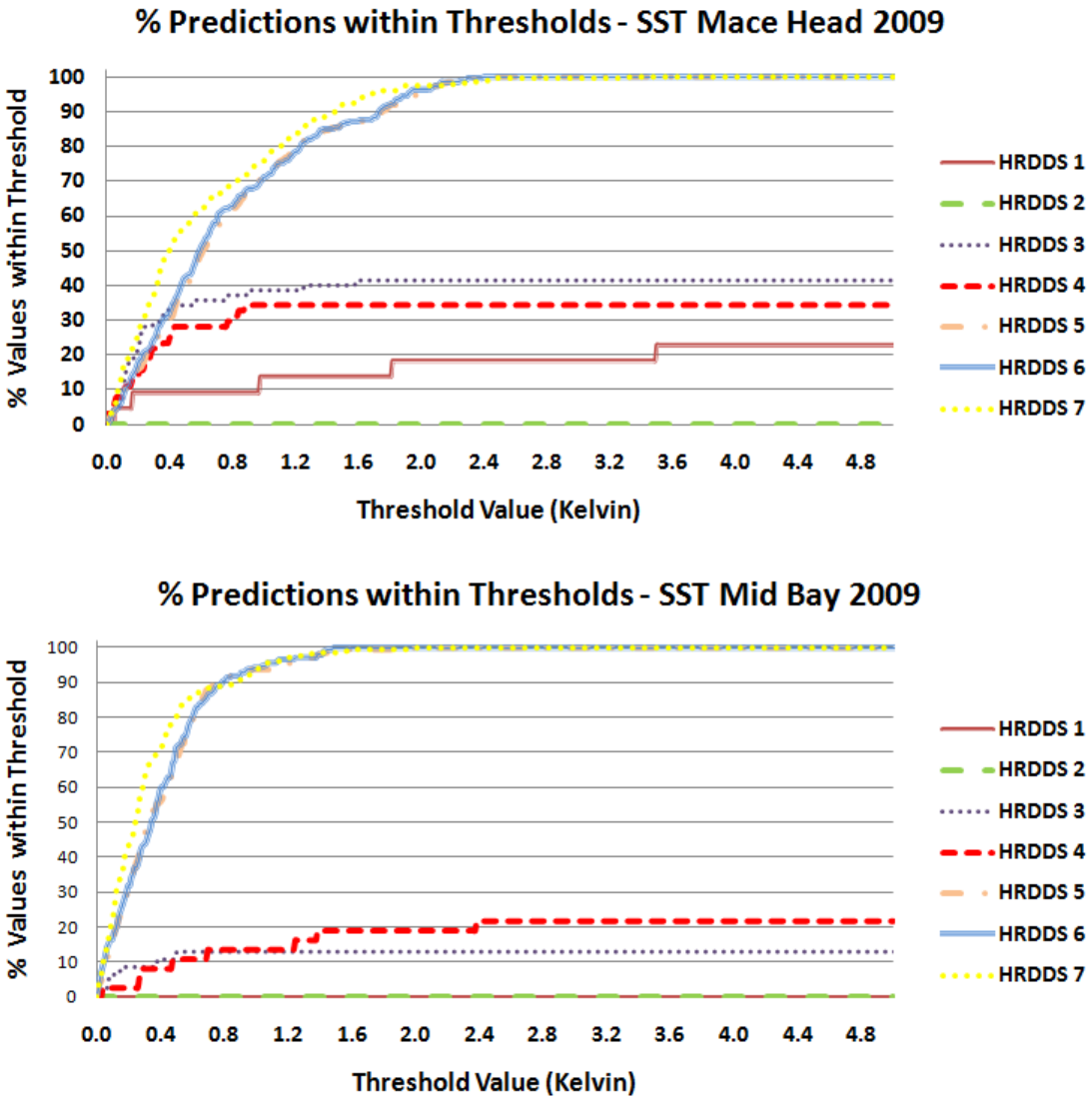

Figure E.21: HRDDS SST products for 2009 at Mace Head and Mid-Bay - The percentage of values of values within various thresholds of the in-situ sensor reading.

Bibliography

Aarninkhof, S. G. J., Turner, I. L., Dronkers, T. D. T., Caljouw, M., and Nipius, L. (2005). Nearshore subtidal bathymetry from time-exposure video images. *Geophysical Research*, 110 (C6).

Abdul-Rahman, A. and Hailes, S. (2000). Supporting trust in virtual communities. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*.

Aberer, K. and Despotovic, Z. (2001). Managing trust in a peer-2-peer information system. In *CIKM '01: Proceedings of the Tenth International Conference on Information and Knowledge Management*, pages 310–317, New York, NY, USA. ACM.

ACRI-ST (2008). GlobColour Final Report. `http://www.globcolour.info/docs/GlobColour_FR_v1.0.pdf`. [Online; accessed 24-June-2011].

ACRI-ST (2011). MyOcean GlobColour Products User Guide. `http://www.globcolour.info/CDR_Docs/MyOcean_ACRI_PUG.pdf`. [Online; accessed 23-May-2011].

Ahmadian, S., Ko, T., Coe, S., Rahimi, M., Soatto, S., Hamilton, M., and Estrin, D. (2007). A vision system to infer avian nesting behavior. Technical report, University of California, Los Angeles.

Aivaloglou, E., Gritzalis, S., and Slkianis, C. (2008). Trust establishment in sensor networks: Behaviour-based, certificate-based and a combinational approach. *International Journal System of Systems Engineering*, 1(1-2):128–148.

Akyildiz, I., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). A survey on sensor networks. *Communications Magazine, IEEE*, 40(8):102–114.

Albaladejo, C., Sanchez, P., Iborra, A., Soto, F., Lopez, J., and Torres, R. (2010). Wireless sensor networks for oceanographic monitoring: A systematic review. *Sensors*, 10(7):6948–6968.

Amin, R., Zhou, J., Gilerson, A., Gross, B., and Moshary, F. (2009). Novel optical techniques for detecting and classifying toxic dinoflagellate karenia brevis blooms using satellite imagery. *Optics Express*, 19:9126–9144.

Anderson, D., Gilbert, P., and Burkholder, J. (2002). Harmful algal blooms and eutrophication: Nutrient sources, composition, and consequences. *Estuaries*, 25(4b):704–726.

Areerachakul, S., Junsawang, P., and Pomsathit, A. (2011). Prediction of dissolved oxygen using artificial neural network. In *2011 International Conference on Computer Communication and Management*.

Areerachakul, S. and Sanguansintukul, S. (2010). Classification and Regression Trees and MLP Neural Network to Classify Water Quality of Canals in Bangkok, Thailand. *International Journal of Intelligent Computing Research (IJICR)*, 1(1-2).

Astoreca, R., Rousseau, V., Ruddick, K., Knechciak, C., Mol, B. V., Parent, J.-Y., and Lanceolot, C. (2009). Development and application of an algorithm for detecting Phaeocystis globosa blooms in the Case 2 Southern North Sea waters. *Journal of Plankton Research*, 31:287–300.

Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books, New York.

Baban, S. M. (1995). The Use of Landsat Imagery to Map Fluvial Sediment Discharge into Coastal Waters. *Marine Geology*, 123:263–270.

Barker, K., Mazeran, C., Lerebourg, C., Bouvet, M., Antoine, D., Ondrusek, M., Zibordi, G., and Lavender, S. (2008). MERMAID: The MERis MAtchup In-situ Database. In *MERIS and (A)ATSR User Workshop*, ESA, Esrin.

Beggs, H., Zhong, A., Warren, G., Alves, O., Brassington, G., and Pugh, T. (2011). RAMSSA - An operational, high-resolution, Regional Australian Multi-Sensor Sea surface temperature Analysis over the Australian region. *Australian Meteorological and Oceanographic Journal*, 61:1–22.

Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Blockeel, H., Deroski, S., and Grbovi, J. (1999). Simultaneous Prediction of Multiple Chemical Parameters of River Water Quality with TILDE. In Zytkow, J. and Rauch, J., editors, *Principles of Data Mining and Knowledge Discovery*, volume 1704 of *Lecture Notes in Computer Science*, pages 32–40. Springer Berlin / Heidelberg.

Boole, G. (1854). *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*. Originally published by Macmillan, London. Reprint by Dover, 1958.

Boukerch, A., Xu, L., and El-Khatib, K. (2007). Trust-based security for wireless ad-hoc and sensor networks. *Computer Communications*, 30:2413–2427.

Bowden, M., Sequiera, M., Krog, J. P., Gravesen, P., and Diamond, D. (2002). Analysis of river water samples utilizing a prototype industrial sensing system for phosphorpus based on micro system technology. *Analyst*, 4(5):767–771.

Bradley., E., Toomey, M., Still, C., and Roberts, D. (2010). Multi-Scale Sensor Fusion With an Online Application: Integrating GOES, MODIS, and Webcam Imagery for Environmental Monitoring. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(4):497 –506.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.

Breunig, M., Kriegel, H., Ng, R., and Sander, J. (2000). LOF: Identifying density-based local outliers. In *ACM SIGMOD International Conference on Management of Data*.

Brodsky, L. (2008). Supervised crop classification from middle-resolution multitemporal images. In *MERIS and (A)ATSR User Workshop 2008*, ESA, ESRIN.

Brown, C. W., Connor, L. N., Lillibridge, J. L., Nalli, N. R., and Legeckis., R. V. (2007). *Remote Sensing of Coastal Aquatic Environments*, chapter 2, An introduction to satellite sensors, observations and techniques, pages 21–49. Springer.

Buchegger, S. and Boudec, J. (2003a). The effect of rumor spreading in reputation systems for mobile ad-hoc networks. In *Proceedings of WiOpt 03: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*.

Buchegger, S. and Boudec, J.-Y. L. (2002). Performance analysis of the confidant protocol. In *MobiHoc '02: Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*, pages 226–236, New York, NY, USA. ACM.

Buchegger, S. and Boudec, J. Y. L. (2003b). Coping with false accusations in misbehavior reputation systems for mobile ad-hoc networks. Technical report, Tech. rep. IC/2003/31, EPFL-DI-ICCA.

Buchegger, S. and Boudec, J.-Y. L. (2003c). A robust reputation system for mobile ad-hoc networks. Technical report, Technical Report IC/2003/50, EPFL-IC-LCA.

Buchegger, S. and Boudec, J.-Y. L. (2004). A robust reputation system for peer-to-peer and mobile ad-hoc networks. In *In Workshop on Economics of Peer-to-Peer Systems.*

Byrne, R. and Diamond, D. (2006). Chemo/bio-sensor networks. *Nature Materials*, 5:421–424.

Cannizzaroa, J. P., Cardera, K. L., Chena, F. R., Heilb, C. A., and Vargoa, G. A. (2008). A novel technique for detection of the toxic dinoflagellate, Karenia brevis, in the Gulf of Mexico from remotely sensed ocean color data. *Contintental Shelf Research*, 28(1):137–158.

Cardell-Oliver, R., Smettem, K., Kranz, M., and Mayer, K. (2004). Field testing a wireless sensor network for reactive environmental monitoring [soil moisture measurement]. In *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 7 – 12.

Cerpa, A., Elson, J., Estrin, D., and Girod, L. (2001). Habitat monitoring: application driver for wireless communications technology. In *In ACM SIGCOMM Workshop on Data Communications in Latin America and the Caribbean.*

Chang, C.-C. and Lin., C.-J. (2011). LIBSVM: A Library For Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27):1–27.

Chelton, D. and Wentz, F. (2005). Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. *Bulletin of the Americal Meteorological Society*, 86:1097–11.

Chen, H., Wu, H., Hu, J., and Gao, C. (2008). Agent-based trust management model for wireless sensor networks. In *MUE '08: Second International Conference on Multimedia and Ubiquitous Engineering*, pages 150 –154.

Chen, M. and Singh, J. P. (2001). Computing and using reputations for internet ratings. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 154–162, New York, NY, USA. ACM.

Chickadel, C. C., Holman, R. A., and Freilich, M. H. (2003). An optical technique for the measurement of longshore currents. *Geophysical Research*, 108 (C11):3364.

Chong, C.-Y. and Kumar, S. (2003). Sensor networks: Evolution, opportunities, and challenges. *Proceedinga of the IEEE*, 91(8):1247–1256.

Coulibaly, P., Anctil, F., and Bobee, B. (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology*, 230:244–257.

Crosby, G., Pissinou, N., and Gadze, J. (2006). A framework for trust-based cluster head election in wireless sensor networks. In *In DSSNS '06: Second IEEE Workshop on Dependability and Security in Sensor Networks and Systems*, pages 10 pp. –22.

Darecki, M. and Stramski, D. (2004). An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment*, 89(3):326–350.

Dareckia, M., Weeks, A., Saganb, S., Kowalczukb, P., and Kaczmarekb, S. (2003). Optical characteristics of two contrasting Case 2 waters and their influence on remote sensing algorithms. *Contintental Shelf Research 23*, 23:237–250.

Davidson, M., Koningsveld, M. V., de Kruif, A., Rawson, J., Holman, R., Lamberti, A., Medina, R., Kroon, A., and Aarninkhof, S. (2007). The CoastView project: Developing video-derived Coastal State Indicators in support of coastal zone management. *Coastal Engineering*, 54(6-7):463 – 475.

Dawson, C., Harpham, C., Wilby, R., and Chen, Y. (2002). Evaluation of Artificial Neural Network Techniques for Flow Forecasting in the River Yangtze, China. *Hydrology and Earth System Sciences*, 6(4):619–626.

Dawson, C. and Wilby, R. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1):80–108.

de Vos, N. and Rientjes, T. (2005). Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. *Hydrol. Earth Syst. Sci.*, 9:111–126.

Dellarocas, C. (2000). Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *EC '00: Proceedings of the 2nd ACM conference on Electronic commerce*, pages 150–157, New York, NY, USA. ACM.

DEPLOY (2010). Deploy: A technology demonstration. `http://www.deploy.ie`. [Online; accessed 01-Aug-2011].

Diamond, D. (2004). Internet scale sensing. *Analytical Chemistry*, 76(15):278A–286A.

Diamond, D., Coyle, S., Scarmagnani, S., and Hayes, J. (2008a). Wireless sensor networks and chemo-/biosensing. *Chemical Reviews*, 108(2):652–679.

Diamond, D., Lau, K., Brady, S., and Cleary, J. (2008b). Integration of analytical measurements and wireless communications - current issues and future strategies. *Talanta*, 75(3):606–612.

Donlon, C., Minnett, P., Genetemann, C., Nightingale, T., Barton, I., Ward, B., and Murray, M. (2001). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *J. Climate*, 15:353–369.

Donlon, C., Robinson, I., Casey, K., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Pioll, J., Barton, I., Beggs, H., Poulter, D., Merchant, C., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R., Kawamura, H., and Rayner, N. (2007). The Global Ocean Data Assimilation Experiment High-resolution Sea Surface Temperature Pilot Project. *Bulletin of the American Meteorological Society*, 88(8):1197–1213.

Donohue, I., Styles, D., Coxon, C., and Irvine, K. (2005). Importance of spatial and temporal patterns for assessment of risk to diffuse nutrient emmissions to surface water. *Journal of Hydrology*, 304(1-4):183–192.

East, T. L. (2006). Development of a Decision Tree Model for the Prediction of the Limitation Potential of Phytoplankton in Lake Okeechobee, Florida, USA. *Archiv Fur Hydrobiologie*, 165(1):127–144.

Elliott, J. and May, L. (2008). Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. *Hydrobiologia*, 559:401–411.

Esaias, W. E., Abbott, M. R., Barton, I., Brown, O. B., Campbell, J. W., Carder, K. L., Clark, D. K., Evans, R. H., Hoge, F. E., Gordon, H. R., Balch, W. M., Letelier, R., , and Minnett, P. J. (1998). An Overview of MODIS Capabilities for Ocean Science Observations. *IEEE Transactions on Geoscience and Remote Sensing*, 36:1250–1265.

Estrin, D. (2007). Refections on wireless sensing systems: From ecosystems to human systems. In *IEEE Radio and Wireless Symposium*, Long Beach, CA.

European Space Agency (2006). MERIS Product Handbook. `http://envisat.esa.int/pub/ESA_DOC/ENVISAT/MERIS/meris.ProductHandbook.2_1.pdf`. [Online; accessed 30-Jun-2011].

Fay, C., Lau, K.-T., Beirne, S., Conaire, C. ., McGuinness, K., Corcoran, B., OConnor, N. E., Diamond, D., McGovern, S., Coleman, G., Shepherd, R., Alici, G., Spinks, G., and Wallace, G. (2010). Wireless aquatic navigator for detection and analysis (WANDA). *Sensors and Actuators B: Chemical*, 150(1):425–435.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Fernandez-Gago, M., Roman, R., and Lopez, J. (2007). A survey on the applicability of trust management systems for wireless sensor networks. In *In SECPerU '0: Third International Workshop on Security, Privacy and Trust in Pervasive and Ubiquitous Computing*, pages 25 –30.

Froidefond, J., Lahet, F., Hu, C., Doxoran, D., Guiral, D., Prost, M., and Ternon, J.-F. (2004). Mudflats and mud suspension obsevred from satellite data in French Guiana. *Marine Geology*, 208:153–168.

Ganeriwal, S., Balzano, L. K., and Srivastava, M. B. (2008). Reputation-based framework for high integrity sensor networks. *ACM Transactions on Sensor Networks*, 4(3):1–37.

Ganeriwal, S. and Srivastava, M. B. (2004). Reputation-based framework for high integrity sensor networks. In *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 66–77, New York, NY, USA. ACM.

Geological Survey of Ireland (2011a). Geological Survey of Ireland - INFOMAR Overview. `http://www.gsi.ie/Programmes/INFOMAR+Marine+Survey/INFOMAR+Overview.htm`. [Online; accessed 25-May-2011].

Geological Survey of Ireland (2011b). Irish Seabed Survey. `http://www.gsi.ie/Programmes/INFOMAR+Marine+Survey/INFOMAR+Overview.htm`. [Online; accessed 25-May-2011].

Gitelson, A. A., Dall'Olmo, G., Moses, W., Rundquist, D. C., Barrow, T., Fisher, T. R., Gurlin, D., and Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sensing of Environment*, 112:3582–3593.

Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., and Kleinman, J. E. (2004). Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies. *Journal of Experimental Marine Biology and Ecology*, 300(1-2):409–448.

Glenn, S. M., Dickey, T. D., Parker, B., and Boicourt, W. (2000). Long-term real-time coastal ocean observation networks. *Oceanography*, 13:24–34.

Goddijn-Murphy, L., Dailloux, D., White, M., and Bowers, D. (2009). Fundamentals of in Situ Digital Camera Methodology for Water Quality Monitoring of Coast and Ocean. *Sensors*, 9(7):5825–5843.

Gons, H. J., Rijkeboer, M., and Ruddick, K. G. (2002). A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *Journal of Plankton Research*, 24(9):947–951.

Govindaraju, R. and Rao, A., editors (2000). *Artifical Neural Networks in Hydrology*. Kluwer, Dordrecht, Netherlands.

Graham, E. A., Yuen, E. M., Robertson, G. F., Kaiser, W. J., Hamilton, M. P., and Rundel, P. W. (2009). Budburst and leaf area expansion measured with a novel mobile camera system and simple color thresholding. *Environmental and Experimental Botany*, 65(2-3):238 – 244.

Gregg, W. W. and Woodward, R. H. (1998). Improvements in Coverage Frequency of Ocean Color: Combining Data from SeaWiFS and MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 36:1350–1353.

Guan, L. and Kawamura, H. (2004). Merging satellite infrared and microwave SSTs: Methodology and evaluation of the new SST. *J. Oceanography*, 60:905–912.

Guralnik, V. and Srivastava, J. (1999). Event detection from time series data. In *Proceedings of the 5th Int'l. Conference on Knowledge Discovery and Data Mining*, pages 33–42.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA Data Mining Software: An Update. *ACM SIGKDD Explorations Newsletter*, 11(1).

Hansen, L. B., Srensen, J. V. T., Erichsen, A. C., Kronborg, M.-B., Larsen, J., and Kaas, H. (2008). Operational data assimilation and marine forecasting based on ENVISAT data covering Danish waters. In *MERIS and (A)ATSR User Workshop*, ESA, Esrin.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. MacMillan: New York.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. London: Prentice Hall, 2nd edition.

Hellweger, F., Schlosser, P., Lall, U., and Weissel, J. (2004). Use of satellite imagery for water quality studies in New York Harbour. *Estuarine, Coastal and Shelf Science*, 61:437–448.

Holman, R. and Chickadel, C. (2004). Optical Remote Sensing Estimates of the Incident Wave Angle Field During NCEX . In *Proceedings of the 29th International Conference on Coastal Engineering*, ASCE, Lisbon.

Holman, R. and Stanley, J. (2007). The history and technical capabilities of Argus. *Coastal Engineering*, 54(6-7):477 – 491.

Hooker, S. and McClain, C. (2000). The calibration and validation of SeaWiFS data. *Progress in Oceanography*, 45:427–465.

Hsu, K., Gupta, H., and Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*, 31(10):2517–2530.

Hu, X., Gao, X., Wan, J., and Xiong, N. (2011). Trust Index Based Fault Tolerant Multiple Event Localization Algorithm for WSNs. *Sensors*, 11:6555–6574.

INFOMAR (2011a). Infomar- Galway Bay. `http://www.infomar.ie/surveying/Bays/Galwayv1.php`. [Online; accessed 24-May-2011].

INFOMAR (2011b). Infomar- Integrated Mapping for the Sustainable Development of Ireland's Marine Resource. `http://www.infomar.ie`. [Online; accessed 24-May-2011].

Iwahashi, M., Udomsiri, S., Mai, Y., and Fukuma, S. (2006). Water Level Detection for River Surveillance utilizing JP2K Wavelet Transform. In *APCCAS '06: IEEE Asia Pacific Conference on Circuits and Systems*.

Johnson, K., Needoba, J., Riser, S., and Showers, W. (2007). Chemical sensor networks for the aquatic environment. *Chemical Reviews*, 107(2):623–640.

Jøsang, A. (2001). A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311.

Jøsang, A. and Ismail, R. (2002). The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, Bled, Slovenia.

Jøsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644.

Kahru, M. and Mitchell, B. (1998). Spectral reflectance and absorption of a massive red tide off southern california. *Journal of Geophysical Research*, 103(C10):601–621.

Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. (2003). The Eigentrust algorithm for reputation management in P2P networks. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 640–651, New York, NY, USA. ACM.

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., and Giering, R. (2008). Remote Sensing Input for regional to global CO2 flux modelling. In *MERIS and (A)ATSR User Workshop*, ESA, Esrin.

Krasniewski, M., Padma, V., Rabeler, B., Bagchi, S., and Hu, Y. (2005). TIBFIT: trust index based fault tolerance for arbitrary data faults in sensor networks. In *DSN '05: Proceedings of International Conference on Dependable Systems and Networks*.

Kutser, T., Pierson, D. C., Kallio, K. Y., Reinart, A., and Sobek, S. (2005). Mapping lake CDOM by satellite remote sensing. *Remote Sensing of Environment*, 94(4):535–540.

Kutsera, T., Liisa Metsamaaa, N. S., and Vahtmea, E. (2006). Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science*, 67(1-2):303–312.

Lachtermacher, G. and Fuller, J. (1994). Backpropagation in hydrological time series forecasting. In Hipel, K., McLeod, A., Panu, U., and Singh, V., editors, *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*. Kluwer Academic, Dordrecht.

Lawlor, A. (2010). DEPLOY - Smart Catchment Demonstration: Long-Term Deployment of Sensor Monitoring System. In *Sensor Systems for Environmental Monitoring*, Royal Society of Chemistry, Burlington House, London.

Lopez, J., Roman, R., Agudo, I., and Fernandez-Gago, C. (2010). Trust management systems for wireless sensor networks: Best practices. *Computer Communications*, 33(9):1086–1093.

Lorenc, A. (1981). A global three-dimensional multi-variate statistical interpolation scheme. *Monthly Weather Review*, 109:701–721.

Maier, H. and Dandy, G. (1996). The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, 32(4):1013–1022.

Maier, H. and Dandy, G. (1998). The effect of internal parameters and geometry on the performance of back-propagation neural networks: An empirical study. *Environmental Modelling and Software*, 13(2):193–209.

Maier, H. R. and Dandy, G. (2000). Neural networks for the prediction and forecasting of water resources variables; a review of modelling issues and applications. *Environmental Modelling and Software*, 15(1):101–124.

Mainstone, C. and Parr, W. (2002). Phosphorous in rivers - ecology and management. *Science of The Total Environment*, 282-283:25–47.

Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R., and Anderson, J. (2002). Wireless sensor networks for habitat monitoring. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97, New York, NY, USA. ACM.

Manov, D. V., Chang, G. C., and Dickey, T. D. (2004). Methods for reducing biofouling of moored optical sensors. *Journal of Atmospheric and Oceanic Technology*, 21(6):958–968.

Manz, A., Graber, N., and Widmer, H. (1990). Miniaturized total chemical analysis systems: A novel concept for chemical sensing. *Sensors and Actuators B: Chemical*, 1(1-6):244–248.

Marine Institute (2006a). Sea change (2007-2013) part ii: Marine foresight exercise for ireland. `http://www.marine.ie/home/Publications/Publications/SeaChangeStrategyandForesightPublications.htm`. [Online; accessed 17-May-2011].

Marine Institute (2006b). Sea Change: A Marine Knowledge, Research and Innovation Strategy for Ireland 2007-2013. `http://www.marine.ie/home/Publications/Publications/SeaChangeStrategyandForesightPublications.htm`. [Online; accessed 17-May-2011].

Marine Institute (2008). Sea Change: Annual Progress Report 2007. `http://www.marine.ie/home/Publications/Publications/SeaChangeStrategyandForesightPublications.htm`. [Online; accessed 18-May-2011].

Marine Institute (2010). Sea Change: Annual Progress Report 2008. `http://www.marine.ie/home/Publications/Publications/SeaChangeStrategyandForesightPublications.htm`. [Online; accessed 18-May-2011].

Marti, S., Giuli, T. J., Lai, K., and Baker, M. (2000). Mitigating routing misbehavior in mobile ad hoc networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 255–265, New York, NY, USA. ACM.

Matthews, M. W., Bernard, S., and Winter, K. (2010). Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sensing of Environment*, 114(9):2070–2087.

McGarrigle, M., Lucey, J., and Cinnide, M. . (2010). Water quality in ireland 2007-2009. `http://www.epa.ie/downloads/pubs/water/waterqua/name,30640,en.html`. [Online; accessed 01-May-2011].

Mertes, L. A. and Warrick, J. A. (2001). Meauring flood output from 100 coastal watersheds in California with field measurements and SeaWiFS. *Geology*, 29(7):659–662.

Michiardi, P. and Molva, R. (2002). Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In *Proceedings of the IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security*, pages 107–121, Deventer, The Netherlands. Kluwer, B.V.

Miller, R. L., Castillo, C. E. D., and McKee, B. A. (2007). *Remote Sensing of Coastal Aquatic Environments*. Springer.

Millie, D. F., Schofield, O. M., Kirkpatrick, G. J., Johnsen, G., Tester, P. A., and Vinyard, B. T. (1997). Detection of Harmful Algal Blooms Using Photopigments and Absorption Signatures: A Case Study of the Florida Red Tide Dinoflagellate, Gymnodinium Breve. *Limnology and Oceanography*, 42(5):1240–1251.

Morel, A. and Prieur, L. (1977). Analysis of variation in ocean color. *Limnology and Oceanography*, 22(4):709–722.

Moscetta, P., Sanfilippo, L., Savino, E., Allabashi, R., and Gunatilaka, A. (2009). Instrumentation for continuous monitoring in marine environments. In *OCEANS '09: MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*.

Muller-Karger, F. E., Hu, C., Andrfout, S., Varela, R., and Thunell, R. (2007). *Remote Sensing of Coastal Aquatic Environments*, chapter 5, The color of the coastal ocean and applications in the solution of research and management problems, pages 101–127. Springer.

Murray, M., Allen, M., Mutlow, C., Zavody, A., Jones, M., and Forrester, T. (1998). Actual and potential information in dual-view radiometric observations of sea surface temperature from ATSR. *Journal of Geophysical Research*, 103(C4):8153–8165.

Ni, K., Ramanathan, N., Chehade, M. N. H., Balzano, L., Nair, S., Zahedi, S., Kohler, E., Pottie, G., Hansen, M., and Srivastava, M. (2009). Sensor network data fault types. *ACM Transactions on Sensor Networks*, 5(3):1–29.

Nightingale, T., Srensen, K., and Folkestad, A. (2008). AATSR SST Validation with SISTeR over the Skagerrak. In *MERIS and (A)ATSR User Workshop 2008*, ESA, Esrin.

Noyes, E., Minnettb, P., Remediosa, J., Corletta, G., Gooda, S., and Llewellyn-Jones, D. (2006). The accuracy of the AATSR sea surface temperatures in the Caribbean. *Remote Sensing of Environment*, 101(1):38–51.

O'Connor, E., Conaire, C. ., Smeaton, A., O'Connor, N., and Diamond, D. (2009). River water-level estimation using visual sensing. In *EuroSSC '09: Proceedings of the 4th European Conference on Smart Sensing and Context*.

O'Connor, E., Hayes, J., O'Conaire, C., Smeaton, A., O'Connor, N., and Diamond, D. (2010). Image processing for smart browsing of oceancolor data products and subsequent incorporation into a multi-modal sensing framework. In *RSPSoc Remote Sensing and Photogrammetry Society Annual Conference with Irish Earth Observation Symposium*, Cork, Ireland.

O'Connor, E., Smeaton, A. F., O'Connor, N. E., and Diamond, D. (2008). Integrating multiple sensor modalities for environmental monitoring of marine locations. In *SenSys '08: Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 405–406, New York, NY, USA. ACM.

Office of Public Works (2008). Lee CFRAMS: Lee Catchment Flood Risk Assessment and Management Study, Hydrological Report April 2008. `http://www.leecframs.ie/downloads/documents/REP004_HydrologyReport.pdf`. [Online; accessed 01-May-2011].

O'Flynn, B., Martinez-Catala, R., Harte, S., O'Mathuna, C., Cleary, J., Slater, C., Regan, F., Diamond, D., and Murphy, H. (2007). Smartcoast: A wireless sensor network for water quality monitoring. In *SenseApp '07: Second IEEE International Workshop on Practical Issues in Building Sensor Network Applications*, pages 1853–1856.

O'Flynn, B., Regan, F., Lawlor, A., Wallace, J., Torres, J., and O'Mathuna, C. (2010). Experiences and recommendations in deploying a real-time, water quality monitoring system. *Measurement Science and Technology*, 21(12).

O'Reilly, J. and Maritorena, S. (1998). Ocean Color Chlorophyll Algorithms for SeaWiFS. *Journal of Geophysical Research*, 103(C11):24937 – 24953.

Patton, M. and Josang, A. (2004). Technologies for trust in electronic commerce. *Electronic Commerce Research*, 4(1-2):9–21.

Pedrazzani, D., Julia, Y. B., Bombn, R. V., and Centeno, G. F. (2008). Testing MERIS FR capabilities and limits to map forest classes in Spain. In *MERIS and (A)ATSR User Workshop 2008*, ESA, Esrin.

Pon, R., Kansal, A., Duo, L., L., M. R., Shirachi, Kaiser, W., Pottie, G., Srivastava, M., Sukhatme, G., and Estrin, D. (2005). Networked infomechanical systems (NIMS): next generation sensor networks for environmental monitoring. In *IEEE MTT-S International Microwave Symposium Digest*, page 4 pp.

Pote, D., Daniel, T., Nichols, D., Sharpley, A., Moore, P. J., Miller, D., and Edwards, D. (1999). Relationship between phosphorus levels in three ultisols and phosphorus concentrations in runoff. *Journal of Environmental Quality*, 28(1):170–175.

Probst, M. J. and Sneha, K. K. (2007). Statistical trust establishment in wireless sensor networks. In *Proceedings of the 13th International Conference on Parallel and Distributed Systems - Volume 01*, pages 1–8, Washington, DC, USA. IEEE Computer Society.

Rast, M., Bézy, J., and Bruzzi, S. (1999). The ESA Medium Resolution Imaging Spectrometer MERIS - a review of the intrument and its mission. *International Journal of Remote Sensing*, 20(9):1681–1702.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Resnick, P. and Zeckhauser, R. (2002). Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. The Economics of the Internet and E-Commerce. In Baye, M. R., editor, *Volume 11 of Advances in Applied Microeconomics*. Amsterdam, Elsevier Science.

Reyes, D., Iossifidis, D., Auroux, P., and Manz, A. (2002). Micro total analysis systems 2. Analytical standard operations and applications. *Analytical Chemistry*, 74(12):2637–2652.

Reynolds, R. and Smith, T. (1994). Improved global sea surface temperature analyses. *Journal of Climate*, 7:929–948.

Reynolds, R., Smith, T., Liu, C., Dudley, B., Chelton, K., and Schlax, M. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20(22):5473–5496.

Richardson, A., Jenkins, J., Braswell, B., Hollinger, D., Ollinger, S., , and Smith, M. (2007). Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Ecosystem Ecology*, 152(2):323–334.

Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Jenkins, J. P., and Ollinger, S. V. (2009). Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, 19(6):1417–1428.

Roman, R., Fernandez-Gago, M., Lopez, J., and Chen, H. (2009). Trust and reputation systems for wireless sensor networks. In *On Security and Privacy in Mobile and Wireless Networking*. Troubador Publishing, Leicester, U.K.

Ruddick, K., Park, Y., Rousseau, V., Borges, A., Astoreca, R., Lancelot, C., and Lacroix, G. (2008). Applications of the MERIS algal2 product in Belgian waters. In *MERIS and (A)ATSR User Workshop 2008*, ESA, Esrin.

Segaran, T. (2007). *Programming Collective Intelligence*. O'Reilly.

Sellner, K. G., Doucette, G. J., and Kirkpatrick, G. J. (2003). Harmful algal blooms: causes, impacts and detection. *J Ind Microbiol Biotechnol*, 30(7):383–406.

Shaghaghian, M. R. (2010). Prediction of Dissolved Oxygen in Rivers Using a Wang-Mendel Method Case Study of Au Sable River. *World Academy of Science, Engineering and Technology*, 62.

Shaikh, R., Jameel, H., Lee, S., Rajput, S., and Song, Y. J. (2006). Trust management problem in distributed wireless sensor networks. In *Proceedings of 12th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, 2006*, pages 411 –414.

Shamseldin, A. (1997). Application of a neural network technique to rainfall-runoff modelling. *Journal of Hydrology*, 199(3-4):272–294.

Simis, S. G., Peters, S. W., and Gons, H. J. (2005). Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limonology And Oceanography*, 50(1):237–245.

Slater, C., Cleary, J., Lau, K.-T., Snakenborg, D., Corcoran, B., Kutter, J., and Diamond, D. (2010). Validation of a fully autonomous phosphate analyser based on microfluidic lab-on-a-chip. *Water Science and Technology*, 61(7):1811–1818.

Smeaton, A. F. (2010). Protecting the bay: Weaving the strands to protect galway bay. In *WCIT 2010: 17th World Congress on Information Technology, Amsterdam, The Netherlands*.

Smith, V., Tilman, G., and Nekola, J. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ. Pollution*, 100(1-3):179–196.

Srinivasan, A., Teitelbaum, J., Liang, H., Wu, J., and Carde, M. (2006a). Reputation and trust-based systems for ad hoc and sensor networks. In *Algorithms and Protocols for Wireless Ad Hoc and Sensor Networks*. John Wiley & Sons.

Srinivasan, A., Teitelbaum, J., and Wu, J. (2006b). Distributed reputation based beacon trust system. In *DASC '06:2nd IEEE International Symposium on Dependable Autonomic and Secure Computing*.

Staehr, P. and Cullen, J. (2003). Detection of Karenia mikimotoi by spectral absorption signatures. *Journal of Plankton Research*, 25(10):1237–1249.

Stark, J. D., Donlon, C. C., Martin, M. J., and McCullock, M. E. (2007). OSTIA: An operation, high resolution, real time, global sea surface temperature analysis system. In *IEEE Oceans 07*, Aberdeen, Scotland.

Stefouli, M., Kouraev, A., and Charou, E. (2008). Monitoring water level and water quality of lakes Macro Prespa and Ochrid using radar altimetry and MERIS data. In *MERIS and (A)ATSR User Workshop 2008*, ESA, Esrin.

Stumpf, R., Culver, M., Tester, P., Tomlinson, M., Kirkpatrick, G., Pederson, B., Truby, E., Ransibrahmanakul, V., and Soracco, M. (2003). Monitoring Karenia brevis blooms in the Gulf of Mexico using satellite ocean color imagery and other data. *Harmful Algae*, 2(2):147–160.

Sullivan, T. and Regan, F. (2011). The characterization, replication and testing of dermal denticles of scyliorhinus canicula for physical mechanisms of biofouling prevention. *Bioinspiration and Biomimetics*, 6(4).

Szewczyk, R., Mainwaring, A., Polastre, J., and Anderson, J. (2004). An analysis of a large scale habitat monitoring application. In *SenSys '04: Processedings of the 2nd international conference on Embedded networked sensor systems*, pages 214–226, Ney York, NY, USA. ACM.

Thirumalaiah, K. and Deo, M. (1998). Real-time flood forecasting using neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 13:101–111.

Tolle, G., Polastre, J., Szewczyk, R., Culler, D., Turner, N., Tu, K., Burgess, S., Dawson, T., Buonadonna, P., Gay, D., and Hong, W. (2005). A macroscope in the redwoods. In *SenSys '05:Proceedings of the 3rd international conference on Embedded networked sensor systems*, pages 51–63, New York, NY, USA. ACM.

Vega, M., Pardo, R., Barrado, E., and Debn, L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Research*, 32(12):3581–3592.

Wang, P., Smeaton, A. F., Lao, S., O'Connor, E., Ling, Y., and O'Connor, N. E. (2009). Short-term rainfall nowcasting: using rainfall radar imaging. In *EuroGraphics Ireland 2009 - 9th Irish Workshop on Computer Graphics*, Dublin, Ireland.

Weiss, R. (1970). The solubility of nitrogen, oxygen and argon in water and seawater. *Deep Sea Research and Oceanographic Abstracts*, 17(4):721–735.

Wentz, F. J., Gentemann, C., Smith, D., and Chelton, D. (2000). Satellite measurements of sea surface temperature through clouds. *Science*, 288:847–850.

Werner-Allen, G., Lorincz, K., Johnson, J., Lees, J., and Welsh, M. (2006a). Fidelity and yield in a volcano monitoring sensor network. In *Proceedings of the 7th symposium on Operating systems design and implementation*, OSDI '06, pages 381–396, Berkeley, CA, USA. USENIX Association.

Werner-Allen, G., Lorinz, K., M, R., Marcillo, O., Johnson, J., Lees, J., and Welsh, M. (2006b). Deploying a wireless sensor network on an active volcano. *Internet Computing*, 10(2):18–25.

Whelan, A. and Regan, F. (2006). Antifouling strategies for marine and riverine sensors. *Journal of Emvironmental Monitoring*, 8(9):880–886.

Wimmer, W., Robinson, I., and Donlon, C. (2008). Validation of AATSR using the ISAR radiometer - Results since 2004 and a new approach for a match-up quality indicator. In *MERIS and (A)ATSR User Workshop 2008*, pages ESA, Esrin, 22-26 Sept.

Witten, I. H., Eibe, F., and Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, Burlington, MA, third edition.

Xiong, L. and Liu, L. (2003). A reputation-based trust model for peer-to-peer e-commerce communities. In *CEC '03. IEEE International Conference on E-Commerce*, pages 275–284.

Xiong, L. and Liu, L. (2004). PeerTrust: Supporting Reputation-Based Trust for Peer-to-Peer Electronic Communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):843–857.

Xu, X., Gao, X., Wan, J., and Xiong, N. (2011). Trust Index Based Fault Tolerant Multiple event Localization Algorithms for WSNs. *Sensors*, 11:6555–6574.

Yang, W., Matsushita, B., Chen, J., Fukushima, T., and Ma, R. (2010). An Enhanced Three-Band Index for Estimating Chlorophyll-a in Turbid Case-II Waters: Case Studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. *IEEE Geoscience and Remote Sensing Letters*, 7(4):655–659.

Yu, B. and Singh, M. (2000). A social mechanism of reputation management in electronic communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents IV, The Future of Information Agents in Cyberspace*, pages 154–165, London, UK. Springer-Verlag.

Zacharia, G., Moukas, A., and Maes, P. (2000). Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, 29(4):371 – 388.

Zealand, C., Burn, D., and Simonovic, S. (1999). Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*, 214(1-4):32–48.

Zhang, W., Das, S., and Liu, Y. (2006). A trust based framework for secure data aggregation in wireless sensor networks. In *SECON '06: IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, volume 1, pages 60 –69.

Zia, T. (2008). Reputation-based trust management in wireless sensor networks. In *ISSNIP '08: Intelligent Sensors, Sensor Networks and Information Processing*, pages 163 –166.

Zimba, P. V. and Gitelson, A. (2006). Remote estimation of chlorophyll concentration in hyper-eutrophic aquatic systems: Model tuning and accuracy optimization. *Aquaculture*, 256(1-4):272–286.

Zimmerman, P. (1995). *The Official PGP Users Guide*. MIT Press.