The development of a Novel Geosmin sensor and WSN platform for the deployment of sensors in water catchment areas

Samuel Peter Ryecroft

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

This research programme was carried out in collaboration with United Utilities

January 2021

Abstract

Water is essential to our daily lives, we consume water in many ways, from drinking it to washing and bathing ourselves. In many developed countries such as the United Kingdom however these standards are ensured by water suppliers who processes and treat water before it enters the drinking water supply. This process involves the monitoring of water catchment areas used by treatment works which is currently monitored through manual methods only enabling the monitoring of small areas of large catchment areas. This thesis presents work that could enable the remote real time monitoring of water catchment sites.

This thesis presents work showing the impact of conductivity on radio frequency communication distances and showing real world communication distances in the Leeds Liverpool canal of up to 7 meters at data rates of 1.2 kbps The real world experiments showed that lower baud rates could achieved larger communication distances, with further work undertaken at a water catchment site where communication distances of 17 meters at data rates of 1.2 kbps achieved with direct communication and showing data transmission across the air water boundary.

This thesis also explores the development of microwave sensors for the detection of Geosmin and Alphacypermethrin in water samples, both of which are common contaminates within water catchment areas. This thesis applies machine learning to microwave data sets to identify frequencies sensitive to contaminates at levels as low as 5 ng/l for Geosmin and 25mg/m^2 for Alphacypermethrin, enabling the future development of a microwave sensor targeted at the detection of these contaminates.

Abbreviations

ADC	Analog to Digital Converter
AES	Advanced Encryption Standard
ANN	Artificial Neural Network
ASK	Amplitude Shift Keying
AUV	Autonomous Underwater Vehicle
bps	Bits Per Second
BSK	Binary Shift Keying
CH	Cluster Head
CPU	Central Processing Unit
CRAN	Comprehensive R Archive Network
CRC	Cyclic Redundancy Check
CSV	Comma Separated Value
DAC	Digital to Analog Converter
dB	Decibel
ELISA	Enzyme Linked Immunosorbent Assay
EM	Electro Magnetic
FSK	Frequency Shift Keying
GB	Gigabyte
GBM	Gradient Boosted Model
GFSK	Gaussian Frequency Shift Keying
GMSK	Gaussian Minimum Shift Keying
I2C	Inter-Integrated Circuit
IC	Integrated Circuit
IIoT	Industrial Internet of Things
IoT	Internet of Things
ISM	Industrial Science Medical
JSON	JavaScript Object Notation
Kbps	Kilobits per second
KNN	K Nearest Neighbor
LEACH	Low Energy Adaptive Clustering Hierarchy
LED	Light Emitting Diode
Mbps	Mega Bits Per second
OOK	On Off Keying
PCB	Printed Circuit Board
PEGASIS	Power Efficient Gathering in Sensor Information Systems
RAM	Random Access Memory
RF	Radio Frequency
RF	Random Forest
ROM	Read Only Memory
RSSI	Received Signal Strength Indication
SD	Secure Digital
SMA	SubMiniature version A
SPI	Serial Peripheral interface
SVM	Support Vector Machine
TEEN	Threshold sensitive Energy Efficient
UART	Universal Asynchronous Receiver-transmitter
USART	Universal Synchronous / Asynchronous Receiver-transmitter
UWSN	Underwater Wireless Sensor Network
VCO	Voltage Controlled Oscillator
VNA	Vector Network Analyzer
WHO	World Health Organization
WSN	Wireless Sensor Network

Table of Tables

Table of Figures

Table of contents

1 Introduction

1.1 Problem statement

Surface water covers around 70% of our planet, filling our oceans lakes and rivers; according to the World Health Organization (WHO) in 2017 only 71% of the global population used a water source that was located on the premises, available and free from contamination [1].

In 2019 the World Bank published a report into the growing issue of water quality describing it as the invisible water crisis. The report identified that the most important priority in tackling water quality issues was monitoring systems, calling for overlapping systems such as remote sensing and machine learning to add layers of verification [2].

Water contamination is a growing issue around the world; with water being a finite resource, more and more research is focusing towards the monitoring of large bodies of water. One form of water contamination of increasing concern is pharmaceutical contamination [3, 4], entering water from sources including the improper disposal of pharmaceuticals, human waste product and manufacturing plants [5] and is a global issue with investigations carried out across the world [6] including the USA [7], Sweden [8] and Germany [6, 9].

Pharmaceutical contamination is not the only form of contamination for water, with growing levels of heavy metal contamination within water being a growing cause for concern, with contaminants often entering water supplies in areas involved in mining activities [10]. Contamination of water with heavy metals has been reported across the world including Korea [11], England and Wales [12, 13].

Farming activities have also been identified as introducing contaminants into water bodies, introducing contaminants into water supplies from runoff and leaching into the ground [14, 15]. Contaminants introduced can have a significant impact on ecology and human health, nitrogen and phosphates can be introduced from farm run-off leading to eutrophication of water supplies leading to a decline in water quality, limiting its use in drinking water [16].

The WHO has published recommend guidelines for drinking water quality, setting out the maximum recommended levels of a range of common contaminants [17]. Individual governments have introduced their own legislation to enforce water quality; within the UK water supplied to consumers is governed by legislation [18] which covers a range of contaminants including nitrates, phosphates and pesticides. Within the UK water is supplied by private companies after the privatisation of water through the Water Act 1989 [19] which sold the 10 regional water authorities established by the previous Water Act 1973 [20], with water companies responsible for ensuring that supplies meet required standards.

Water suppliers are responsible for ensuring that water supplied to the end consumer is suitable and in line with the appropriate legislation, failure to abide by legal standards can lead to fines and other penalties. One of the most important parts of ensuring that water meets standards is to monitor the water intake sources such as reservoirs; understanding the quality of the water intake informs the treatment works of what contaminants need to be removed and in what quantities.

Presently utility companies use manual sampling methods to take samples from reservoirs and other intake sources, often these samples are taken from single points at the edge of the catchment area, limiting the insight that these samples provide. The samples are then taken to labs where they are processed and the relevant quantities of contaminants are identified, once this process is completed the treatment works can take action to adjust the water treatment process being applied based on the water intake quality.

Current processes pose two key issues, the first is that samples only represent a small area of what can be considerable bodies of water, so the samples do not reflect the water quality across the entire catchment area. The second issue is the time and effort required to take samples and process them, which introduces a delay between changes in the catchment area occurring and the changes being detected and acted upon.

Developments within wireless and sensor technologies have helped to develop a revolution of internet connected devices and sensor technologies capable of producing and relaying data in real time over geographically distributed areas. These developments have helped to enable a range of sensor networks that are able to collect data across large areas with applications such as air quality monitoring [21], connected smart farming [22] and smart street lighting management systems [23], which are all applications of IoT (Internet of Things).

The limitations that current methods of monitoring of catchment sites have, of limited samples and time delay, could be avoided through the use of a distributed sensor network, equipped with sensors to measure and quantify contaminants. These sensor readings could be taken across the whole catchment site through a wireless sensor network where data could be processed by treatment works operatives and acted upon in real time. Such an approach would enable a reactive approach to be taken to contaminants and provide a more reliable method of monitoring the intake water quality.

A real-time water quality monitoring system, capable of being deployed for prolonged periods would enable a better monitoring approach for water catchment sites, enabling larger data sets to be collected which could then be processed to potentially predict contamination incidents.

1.2 Novelty

A wireless sensor-networking platform for the long-term deployment of sensors targeted at water quality monitoring with a sensor targeted at the detection of Geosmin contamination of levels, usable by industry.

1.3 Research Aim and Objectives

1.1.1 Aim

The aim of this research project is to develop a system capable of the real-time detection of contaminants, focusing on Geosmin within large water bodies such as reservoirs using a dispersed multi-hop sensor network.

1.1.2 Objectives

- Research environmental monitoring methods within the UK including contaminants, sensor technologies, data analysis using machine learning and underwater communication.
- Research considerations for using radio frequency in an underwater environment.
- Create a multi-hop networking platform, capable of functioning in an underwater environment using RF communications and evaluate the performance of the platform.
- Create a microwave-based sensor capable of detecting Geosmin, which could be integrated into the deployed wireless sensor-networking platform.

1.4 Overview

Chapter 2 examines environmental monitoring within the water industry. The chapter covers the legislation within the UK and the contaminants, and the maximum allowable concentrations allowed. The chapter explores sensor technologies within the water industry, examining a range of sensors that are used within the industry. Chapter 2 also covers machine learning approaches to processing sensor data.

Chapter 3 examines in detail the considerations for radio frequency in underwater wireless sensor networks. The chapter covers antenna designs including examples of antenna designs that have been used in previous works. Chapter 3 also examines the propagation of electromagnetic waves and how loss can occur. The chapter also explores modulation techniques before covering digital communication data rates, examining factors that impact on the obtainable rates of communication. The chapter also covers routing approaches in wireless sensor networks as well as security considerations within wireless sensor networks.

Chapter 4 presents experimental works undertaken in the implementation of an underwater wireless sensor-networking platform using radio frequency. The chapter presents experimental work examining the role that conductivity plays in signal attenuation. The chapter presents experimental work into underwater communications in a real-world environment to create a point-to-point link, the experiments examine how data rates can impact upon communication distances. The chapter presents further real-world experiments at a water catchment site using a small multi-hop sensor network to increase communication distances. Finally, this chapter presents simulation results to examine the performance of the sensor-networking platform.

Chapter 5 presents an analysis approach to identifying frequencies that could be used to detect contaminants using microwave spectroscopy. The chapter presents a two-step approach to identifying frequencies that could be used to detect contaminants. The chapter presents the application of the outlined process to two data sets for two contaminants with microwave spectroscopy.

Chapter 6 concludes the thesis, recapping on the important facts of each chapter and how they relate to the works undertaken. The chapter identifies the contributions to knowledge and how they have been disseminated. Finally, the chapter identifies future work that should be undertaken to expand the works presented within this thesis further.

2 Environmental monitoring

2.1 Introduction

Water is used in many aspects of life, from drinking and food preparation to bathing and washing; ensuring that we have a safe source of water is a vital issue and which many people in the United Kingdom take for granted. The quality of water plays an important role in ensuring consumer confidence in the safety of the water that they drink. The World Health Organisation (WHO) has published guidelines on safe levels for a variety of common water contaminants [1], the guidelines cover a range of contaminants including nitrates, heavy metals and pesticides.

Contaminants within the guidelines set out by WHO such as nitrates have known human health risks, with contaminants such as heavy metals causing acute heavy metal poisoning as well as other health issues including miscarriage and premature birth in women[24]. There are other contaminants within WHO guidelines such as pesticides which are also known to have adverse effects on human health including some being known carcinogens[25].

WHO guidelines are not directly enforceable in a country, instead legislation must be in place to enforce any legal limits. The United Kingdom (UK) enforces its own limits of contaminants the legislation in The Water Supply (Water Quality) Regulations [18], the legislation sets out contaminants that must be monitored by water suppliers and the maximum acceptable levels. Should water companies fail to meet these requirements then penalties may be enforced by the regulator.

Over the years a wide range of sensor technologies have been deployed to monitor water quality metrics[26-28]. However, these sensors in many instances are not suitable for a continuous monitoring approach either due to the reliance on equipment that must be used in a lab-based setting, or the requirement of reagents. Where sensors have been developed that are suitable for continuous monitoring however there is a lack of a communication platform to support their deployment in an underwater environment, making it impossible to fully exploit these developments in sensor technologies, with previous attempts relying on communication systems that are not submerged such as the SmartCoast system[29].

This chapter explores environmental monitoring, examining the quality levels that water suppliers must meet after water treatment. This chapter also examines the current sensing technologies available for contaminants for Geosmin and pesticides, two contaminants that are of importance to water suppliers. Finally, this chapter explores underwater communication methods that could be used to form a network of sensors to monitor intake areas for water treatment works such as reservoirs, enabling detailed data capture across an entire catchment site.

2.2 Contaminants in water

2.2.1 Introduction

With the potential risks that poor water quality can have on human health, many countries set out legislation to define legal limits on the amount of contaminants that can be present in water for human consumption The UK has legislation[18] that covers a number of contaminants within the legislation which defines levels of contaminants that water companies must not breach in their supplies to consumers.

It is important for water companies to comply to the levels set out in legislation to ensure compliance with the law and ensure that they do not endanger the health of the consumers that they supply. It is important for water companies to understand the quality of the water within the catchment area of a water treatment works so that they ensure that the correct processes are applied so that contaminants are suitably removed to ensure compliance with the levels set out in legislation.

Each contaminant within the legislation has a prescribed maximum allowable quantity that can be found and the point at which water must comply to the prescribed level. Some contaminants such as coliform bacteria have strict levels of none being present, with the point of compliance being both at the water intake point of the reservoir and at the consumer's tap. Other contaminants such as nitrates have a prescribed level of 50 mg/l which must be complied with at the point of the consumer's tap. Table 2-1 provides a full list of contaminants that are regulated under the legislation in The Water Supply (Water Quality) Regulations [18].

Table 2-1Maximim allowable quantities of contaminants allowed by law in UK water supplies

Contaminant	Maximum allowable quantity	Units of measurement	Point of compliance
Enterococci	0	number/100ml	Reservoirs and treatment works
Escherichia coli	0	number/100ml	Reservoirs and treatment works
Coliform bacteria	0	number/100ml	Reservoirs and treatment works
Acrylamide	0.10	µg/1	
Antimony	5.0	µgSb/1	Consumer's taps
Arsenic	10	µgAs/1	Consumer's taps
Benzene	1.0	µg/1	Consumer's taps
Benzo(a)pyrene	0.010	µg/1	Consumer's taps
Boron	1.0	mgB/1	Consumer's taps
Bromate	10	µgBrO3/1	Consumer's taps
Cadmium	5.0	µgCd/1	Consumer's taps
Chromium	50	µgCr/1	Consumer's taps
Copper	2.0	mgCu/1	Consumer's taps
Cyanide	50	µgCN/1	Consumer's taps
1, 2 dichloroethane	3.0	µg/1	Consumer's taps
Epichlorohydrin	0.10	µg/1	
Fluoride	1.5	mgF/1	Consumer's taps
Lead	10	µgPb/1	Consumer's taps
Mercury	1.0	µgHg/1	Consumer's taps
Nickel	20	µgNi/1	Consumer's taps
Nitrate	50	mgNO3/1	Consumer's taps
Nitrite	0.50 0.10	mgNO2/1	Consumer's taps Treatment works
Pesticides	0.030	µg/1	Consumer's taps
Aldrin Dieldrin Heptachlor Heptachlor epoxide	0.030	µg1	Consumer's taps

Other pesticides	0.10	µg/l	Consumer’s taps
Pesticides: total	0.50	µg/1	Consumer’s taps
Polycyclic aromatic hydrocarbon	0.10	µg/1	Consumer’s taps
Selenium	10	µgSe/1	Consumer’s taps
Tetrachloroethene and Trichloroethene	10	µg/l	Consumer’s taps
Trihalomethanes: Total	100	µg/1	Consumer’s taps
Vinyl chloride	0.50	µg/1	
Aluminium	100	µgA1/1	Consumer’s taps
Colour	20	mg/1 Pt/Co	Consumer’s taps
Iron	200	µgFe/1	Consumer’s taps
Manganese	50	µgMn/1	Consumer’s taps
Odour	Acceptable to consumers and no abnormal change		Consumer’s taps
Sodium	200	mgNa/1	Consumer’s taps
Taste	Acceptable to consumers and no abnormal change		Consumer’s taps
Tetrachloromethane	3	µg/1	Consumer’s taps
Turbidity	4	NTU	

Two issues are Odour and Taste, these are not specified as contaminants but rather as properties of the water itself, for this reason no quantity is set other than them being acceptable to the end consumers and have no abnormal change. One main cause of taste and odour sources in water is Geosmin [30], a natural occurring organic compound that is produced by the breakdown of organic material such as algae in water catchment sites.

Another common family of contaminants that are regulated through legislation are heavy metals such as Lead, Chromium, Iron, Copper, Manganese, Cadmium, Mercury, Nickle and Arsenic. These contaminants can all lead to heavy metal poisoning, each contaminant has different symptoms but all include abdominal pain and nausea, other symptoms can include Hart abnormalities, kidney and liver damage, brain dysfunction and miscarriage or premature labour in women[24].

Nitrates are another common contaminant within water catchment sites that needs to be treated and removed before water can be supplied to the end consumer. Nitrates can be introduced to water catchment sites in a variety of ways including the run-off from farms during heavy rain. Nitrates can be closely linked to an increase in the concentrations of Geosmin within water. Increased nitrate levels within water allow for a growth in algal blooms within catchment sites, which leads to an increase in odour and taste complaints from a rise in Geosmin levels[31].

Pesticides form another group of contaminants regulated by legislation, these contaminants include specifically named pesticides including Aldrin, Dieldrin, Heptachlor and Heptachlor epoxide. Legislation sets limits on the specified pesticides as well as any other pesticides with a limit on the total quantity of pesticides. Pesticide contaminants can have a range of impacts on human health. Pesticides have several other potential impacts on human health including dermatological, gastrointestinal and neurological impacts as well as carcinogenic, respiratory reproductive and endocrine effects [25].

2.2.2 Geosmin

Geosmin in the water supply

Geosmin is an odour causing contaminant that can occur in large bodies of water, Geosmin is often introduced into reservoirs and lakes through the decomposition of organic matter such as algae broken down by microorganisms[32]. As previously highlighted nitrate contamination can lead ultimately at an increase in geosmin contamination of water bodies. Geosmin is a significant challenging contaminant to detect as humans can detect geosmin at extremely low concentrations of just 4 ng/l[33].

UK regulations do not specify any allowable level of geosmin contamination that is acceptable either at the consumers’ taps or in natural bodies of water, however in both cases legislation is clear that water should be of an acceptable odour. Geosmin has a distinctive earthy aroma and gives beetroots their distinctive smell and is the cause of the earthy smell that often occurs after rain also known as petrichor.

Geosmin does not pose a direct impact to human health, however taste and odour in drinking water is the only way an end consumer can determining the safety of tap water[34]. Odour complaints are one of the leading causes of complaints to water suppliers with 17,000 contacts being made to companies regarding odour issues in 2018[35].

Geosmin can be removed from water supplies with several treatment methods available, one commonly used method is the use of Powdered Activated Carbon (PAC), the process is normally only used during an outbreak of odours and takes place before the filtration stage of the water treatment process.

Detection methods

There are several methods of detecting geosmin levels in water, one such method is the use of solid-phase micro extract and gas chromatography proposed McCallum et al[36]. The methodology set out by McCallum provided a detection level of 1 ng/l and a good linearity over a range of 5 ng/l to 40 ng/l. Other work by S Keita et al[37] showed the use of gas chromatography with solid phase micro extraction to detect Geosmin at levels of 0.6 pg/ml.

Another detection method proposed by G. S. Braga et al[38] developed an electronic tongue to detect geosmin. The sensors used conducting polymer sensors to detect geosmin at a concentration of 25 ng/l up to 300 ng/l. The threshold of 25 ng/l is significantly below the needs of industry in order to successfully detect at the same levels as humans.

M. Son et al[39] developed a real-time sensor using a bioelectronic nose for the real-time detection of geosmin. The sensor used cloned human olfactory receptors sensitive to geosmin to replicate a human nose. The results show that the developed sensor was capable of detecting geosmin at concentrations as low as 10 ng/l making the sensor suitable for the needs of industry which demands detection of low levels.

The bromine reaction is another method which can be applied to the detection of geosmin first used by T P Hensarling and Susan K Waage[40] in 1990. The approach is a modified version of the TJ reaction[40], adding bromine to the water sample shows a colour change when geosmin is present in the test sample. The results present a simple method for the determination of the presence of geosmin within a water sample but they do not propose this as a method for quantifying the levels of geosmin within a test sample.

Enzyme-linked Immunosorbent Assay (ELISA) uses antibody-antigen binding properties to produce a signal that can be measured. ELISA methods have been applied to the detection of geosmin in previous works, for instance in Chung et al[41] , the work shows that ELISA could be used to detect geosmin down to levels of 1000 ng/l but significantly above the levels at which humans are capable of detecting geosmin. Other work presented by S W Kang et al[42] also used gold seeded magnetic nanoparticles with ELISA to detect geosmin at concentrations of 2000 ng/l ± 800 ng/l.

K Sato and S Tanaka[43] presented a chemiluminescence reaction for the detection of organic materials in water including geosmin. The work used a luminol oxidation reaction to identify organic material in the water over the course of 20 minutes. The detection method was shown to be capable of detecting at levels of just 0.055 ng/l in pure water making this method unsuitable for the detection and quantification of geosmin in water which may contain other organic material that could influence the reaction.

2.2.3 Pesticides

Pesticides in the water supply

Pesticides pose a real and significant threat to water supplies, often entering water catchment sites through run-off during direct application, drifting droplets run off during heavy rain[44]. The legislation in the UK is clear and prescribes a number of pesticides and maximum levels that are permitted within the water supplied to consumers, additionally legislation defines a total maximum allowable quantity of 50 mg/l for all pesticides which must also be adhered to in addition to any other limits set out.

Pesticides pose serious risks to human health if consumed, pesticides such as heptachlor being classified by the International agency for research on cancer as possibly carcinogenic to humans (Group 2B)[45]. Other risks from exposure to pesticides can include acute poisoning, headaches, dizziness and impaired vision[46, 47].

Pesticides can be removed from water taken in by a treatment works, using processes such as Granulated Activated Carbon filters in which the pesticides stick to small particles of coal or charcoal within the filters. Pesticides can also be removed through reverse osmosis, although this process can remove minerals from water which can impact on the taste of the supplied water.

Detection methods

Many research works have been undertaken to investigate methods of detection and quantification of pesticides in water. S Berijani et al[48] presented the use of dispersive liquid-liquid microextraction with gas chromatography-flame photometric detection for detecting organophosphorus pesticides in water. The results showed that the processes could detect organophosphorus contamination at levels as low as 0.003 -0.02 ug/l with a short extraction time of under 3 minutes.

Other work presented by S Suwansard et al[49] developed research into a semi-disposable reactor biosensor for detecting carbamate. The sensor used an enzyme reactor connected to a flow cell with an ion analyser. The sensor presented has a limited number of uses before the enzyme reactor must be replaced, the sensor presented could detect carbamate in spiked well-water with readings of 2.0 ppm and 8.0 ppm.

Work by A A Boyd-Boland et al [50] presented the simultaneous determination of 60 pesticides in water through the use of solid-phase microextraction and gas chromatography-mass spectrometry. The work used a mixture of pesticides with samples from each of the organonitrogen, organochlorine and organophosphorus classes of pesticides. The work examined two types of fibres, polyacrylate and polydimethylsiloxane-coated fibres from sample mixtures between the range of 0.1 – 100 ug/l. The research concluded that this approach could be used to simultaneously detect 60 pesticides over the range of 0.1 to 100 ug/l.

Work by F Ahmadi et al examined the detection of organophosphorus pesticides in water using single drop microextraction and gas chromatography-flame photometric detection. The work examined 12 organophosphorus pesticides in water. The work highlighted that in most cases the limit of detection was between 0.001 ug/l and 0.005 ug/l.

2.3 Microwave sensors

2.3.1 Principles of operation

Microwave spectroscopy uses electromagnetic waves within the microwave frequency and examines how the properties of the wave change as it passes through a substance. There are several parameters that can be used within microwave spectroscopy including the signal loss as it passes through a material as well as the change in phase. The changes to the electromagnetic wave's properties are based upon the permeability and permittivity of the material under test and will differ depending on the frequency of the wave used.

Microwave spectroscopy examines the interaction of magnetic waves within the microwave frequency of between 300 MHz and 300 GHz to create a unique spectrum response which varies based on the properties of the material under test such as the permittivity of the substance.

There are two core components that define the permittivity of a substance, the dielectric constant and the dielectric loss[51]. The dielectric constant or relative permittivity is equal to the ratio of a capacitor filled with the given substance to the capacitance of an identical capacitor in a vacuum without the dielectric substance[52]. The relative permittivity represents the storage of energy and the reduction in an electromagnetic wave velocity when the wave passes through a substance.

The dielectric loss of a material is the loss of energy that goes into heating the substance in a varying electric field[53]. The dielectric loss within microwave spectroscopy is the loss that the wave experiences as it propagates through the substance. The magnitude of the electromagnetic wave reduces due to the oscillation of molecules[54].

The measurements taken normally vary with changes to the permittivity and the conductivity of the substance under test, with measurements changing depending on the frequency of the electromagnetic wave that is applied to the test substance. Changes to the substance under test will lead to changes in the readings taken, enabling the differentiation of test substances based upon the response.

There are two key types of measurement used in microwave spectroscopy, the first is S_{11} also known as Return loss and S_{21} also known as transmission loss. Figure 2-1 shows the difference between an S11 measurement and an S21 measurement, the type of measurement that can be captured depends on the equipment used, with antennas most commonly used to only look at reflected power while other devices such as waveguides and resonant cavities can be used to take either measurement.

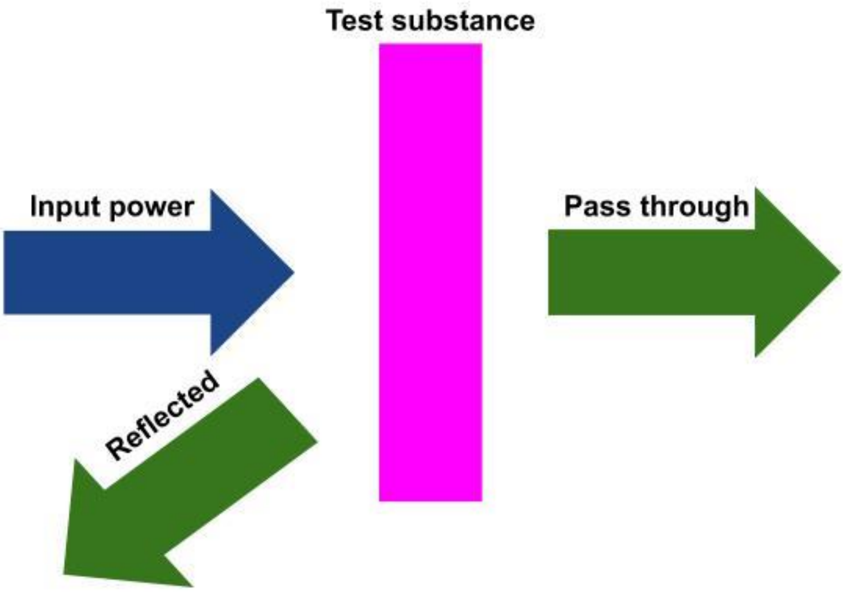

Figure 2-1 An example of reflected and pass through power

Readings within microwave spectroscopy consist of two values, the first is a magnitude reading. The magnitude reading represents the power of the observed microwave, this varies based upon the initial output power of the microwave source. Magnitude readings represent the power that is lost from the observed microwave as it passes through the substance under test, the power lost as the wave passes through the substance is related to the permittivity and permeability of the substance as shown by Equation 2-1.

Equation 2-1 Attenuation loss based upon the permittivity and permeability of a substance

$$a = \omega \left\{ \frac{\mu_0 \epsilon_0}{2} \left[\sqrt{1 + \left(\frac{\sigma}{\omega \epsilon_0} \right)^2} - 1 \right] \right\}^{\frac{1}{2}}$$

ϵ_0 Permittivity

μ_0 Permeability

ω Angular frequency

α Attenuation constant

σ Conductivity

Phase readings are the second form of readings taken within microwave spectroscopy; a phase reading represents the change in the speed of the wave as it passes through a substance. A phase reading is based on the difference in phase between the outgoing wave and the observed wave. The reading changes for two reasons, the first is a constant phase change due to the time taken for the wave to propagate and be observed, this should remain constant within an experimental set-up.

The second reason for difference in the phase of the observed microwave is due to the change in speed of the wave as it passes through the substance under test. The frequency of a wave can be derived using Equation 2-2 using the speed of light in a vacuum as a constant, however depending of the permittivity and permeability of the substance under test the speed of light will vary. The speed of the wave alters based on Equation 2-3 which shows the link between the speed of the wave and the permittivity and permeability of a substance under test.

Equation 2-2 Derivation of the frequency of a wave based upon the speed of light and wavelength

$$f = \frac{c}{\lambda}$$

f Frequency

c Speed of light

λ Wavelength

Equation 2-3 A derivation of frequency based on the permittivity and permeability of the transmission medium

$$f = \frac{\frac{1}{\sqrt{\epsilon_0 \mu_0}}}{\lambda}$$

c Speed of light

ϵ_0 Permittivity

μ_0 Permeability

F Frequency

λ Wavelength

Microwave sensors use a variety of designs to transmit and receive the electromagnetic waves. These include interdigitated antennas, aperture antennas such as a horn antenna and planar antennas. Antennas can only be used to capture return loss readings. The design of the antenna used plays a critical role in the sensor as antennas are sensitive to specific frequencies. Some antenna designs can cover small frequency ranges such as an interdigitated antenna or a planar antenna while other designs such as aperture antennas or horn antennas can be a broad band design and therefore sensitive to a much larger range of frequencies.

Microwave waveguides are another method that can be used to capture readings in microwave spectroscopy. Waveguides are often used in microwave circuits as an alternative to cables such as coaxial cables, they function by effectively directing the wave form the source to the receiver unlike a cable where the transmission is through conduction. Dimensions of waveguides are based upon the target wavelength that is to be transmitted through the guide, with lower frequencies using larger sized waveguides due to the longer wave lengths.

Microwave resonant cavities are a further way that microwave spectroscopy readings can be captured. A microwave resonant cavity is an enclosed space with a conducting surface in which microwaves can be maintained. The biggest advantage of a resonant microwave cavity is the significant reduction in power loss at the resonant frequency of the cavity. A resonant cavity can be used to capture either S_{11} or S_{12} readings.

2.3.2 Advantages

Microwave spectroscopy offers a significant advantage of being a non-destructive method that does not require any reagents or produce a waste product as a result of readings being taken. This advantage is of particular benefit in a platform for monitoring water quality, which may be deployed for prolonged periods of time where disposing of waste products and keeping a supply of reagents can cause issues.

Microwave spectroscopy-based sensors can be refined and developed to be small and portable. The advantage of being able to miniaturise the required circuitry makes it an ideal sensor technology for use on a sensor platform where space is at a premium.

The use of microwave spectroscopy offers the potential to overcome issues such as the build-up of biofilms on sensors over time. By adapting the sensor design and data processing it is feasible to stop the build-up of film interfering with the readings.

2.3.3 Disadvantages

Another disadvantage of microwave spectroscopy sensors is the frequencies that are sensitive to the desired contaminants may be licenced and cannot be used in a developed sensor due to arising legal issues. It may be possible to overcome this issue through using frequencies that are freely available if suitable or through getting the appropriate approvals for the frequencies that are required. While this does not constrain academic research, it is an important consideration in the development of commercial products.

Microwave based sensors are also sensitive to a variety of external factors including pressure, temperature and humidity; to develop a sensor that is suitable for all these environments it is important to account for these factors and how they impact on the raw data.

2.3.4 Applications of microwave-based sensors

Microwave based sensors are not new in the area of water quality monitoring, previous research presented by K. Zhang et al[55] has examined the use of a microwave sensor array targeted at detecting a range of common water contaminants. The research presented examined the use of the microwave frequency range between 1GHz and 10GHz measuring the sensor's responses to the following contaminants: Nitrate, Phosphate, Ammonium, Lead, Mercury, Chromium, PH levels Sodium Chloride and Dissolved Oxygen. The sensor measured the difference in resonance frequencies and the differences of the corresponding S_{21} values. The results showed that a microwave sensor array could be used to identify the studied contaminants with different levels of each examined contaminant detectable.

Other work presented by R Mirzavand et al [56] presented a zero-power microwave sensor for the real time assessment of water quality. The presented sensor harvested energy from a transmitting base station to detect humic acid to simulate organic compounds. The sensor presented used two coaxial probes, one in a test solution and one in a calibration solution.

Work presented by I Frau[57] et al examined the use of electromagnetic wave sensors for real time monitoring of lead pollution in water samples. The work examined two methods, a microwave resonant cavity and an Interdigitated electrode sensor. The results observed that lead concentrations could be detected using microwave sensors with variations in signal amplitude being observed for varying concentrations.

Work presented by A Azmi[58] examined the application of coating materials onto electromagnetic sensing arrays to detect water contamination. The work examined two coatings a N-Methyl-2-pyrrolidone membrane and an acrylic coating, the results showed that the membrane improved the sensitivity of the sensor by 338% when compared with the acrylic coating. The results showed that the sensor could detect variations in concentrations of contaminants such as nitrates and phosphates with both coatings.

2.3.5 Summary

Microwave-based sensors offer some distinct advantages that are vital for a sensor targeted at deployment in an underwater wireless sensor network. The fact that the sensor is capable of functioning without the use of reagents or producing a by-product allowing deployments to be limited only by power resource. The ability of microwave sensors to also overcome issues surrounding the build-up of biological material such as algae makes it a promising candidate, with other sensors such as electrodes and optical sensors being severely impacted by this issue.

One of the greatest concerns with microwave-based sensors is the use of the RF spectrum, which is an issue if large sections of the spectrum are required for sensing purposes. If a limited range of frequencies are required for operation of the sensor then this becomes much less of an obstacle. This issue becomes even further reduced where a microwave cavity or wave guide is used due to the electromagnetic waves being contained and not radiating out of the sensor.

While developing microwave-based circuitry can be a challenging task it is not impossible to develop an appropriate sensor board. It is important that the design stage of the board is well thought through with consideration given to all aspects of the layout of the board as well as ensuring that there are no sources of interference that could impact on the readings.

Microwave sensors have already been developed targeting a range of substances in water, with results indicating that the use of these sensors is feasible to detect contaminants such as nitrates and heavy metals. The demonstration of functional sensors lends support to the use the of microwave sensors to detect contaminants within water.

One of the largest potential issues of the long-term deployment of sensors is the build-up of biological material on the sensor surface. Through a combination of sensor design and data analysis it is possible to overcome this issue, allowing the emitted microwaves to pass through the film and interact with the test substance.

2.4 Analysis approaches to sensor data using machine learning

2.4.1 Introduction

The previous section covers a variety of approaches of collecting raw data, however, to derive information and knowledge from the raw data it must first be analysed. Over the years machine learning has grown, accelerated by the quantity and variety of data available, much of which is collected through IoT devices and sensors. Machine learning has previously been applied to a variety of data sets, providing valuable insights into a broad range of areas.

Microwave spectroscopy produces sizable data sets, with a single reading consisting of thousands of data points. With such large data sets it can be difficult to identify patterns and changes as in many cases the changes between substances can be subtle, this is especially true when examining variations in the same substances at different concentrations. Figure 2-2 shows an example of a partial data set, the full data set consists of readings taken at 4,000 frequencies with each reading consisting of a magnitude, phase reading as well as a real and an imaginary component. While such variations can be seen in the plot there are many more subtle variations that cannot be easily seen through visual inspection.

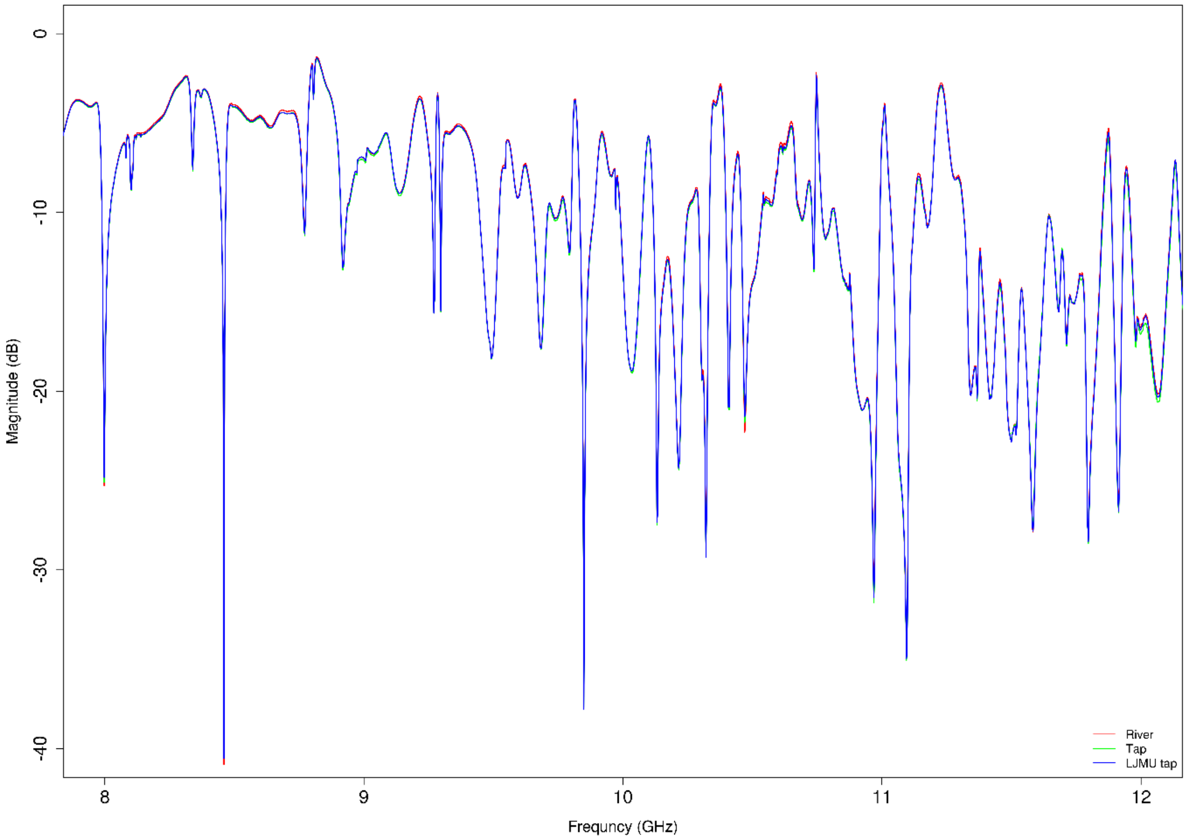

Figure 2-2 An example of three readings in a microwave data set

Machine learning offers the ability to process these large and complex data sets and extract information from them, using feature selection techniques combined with classification models frequencies that are sensitive to the contaminants. Based upon the frequencies identified a specific sensor can be developed that can then be used in convocation with a wireless communication platform.

The machine learning approaches examined do not need to be run upon the communication platform, instead the platform can be used to relay the required readings from the sensor. The data can then be communicated through the network to a sensor node connected to wider network such as the internet, the data can then be processed and categorised off the platform where computational power and energy resources are not constrained.

The presented data set in Figure 2-2 shows a significant amount of variation within the region below -60 dB, at this low power level there is an increase in the amount of noise within the signal. It is also important to consider that many microwave power detectors available are limited to a range of power readings up to around -30 dB making magnitude readings with higher powers preferable to use.

2.4.2 Machine learning approaches

Machine learning can broadly be grouped into two categories of supervised learning and unsupervised learning. Supervised learning can only be used with data sets that include labels however unsupervised learning can be applied to both labelled and non-labelled data sets. Supervised learning techniques are typically applied to classification problems where a model can process an unlabelled input and provided a classification for the unknown input.

Unsupervised learning approaches focus on clustering, feature learning and density estimation. Unsupervised learning is often used during the exploratory analysis of data to help identify a structure and build the foundations of initial hypotheses. Unsupervised learning can also be applied to dimensionality reduction to learn relationships between individual features allowing a reduction in the number of features needed to represent the data.

Given the fact that during the development of a sensor, the data required would have to be captured, it is possible to apply either approach as data can be labelled when captured with the appropriate classification. Given that the data used will be labelled it is feasible to apply supervised learning used for classification problems as the data sets that will be used will fit into discrete classes.

The data set produced during classification will collect readings from samples of a range of concentrations, due to experimental limitations these concentrations will fit into distinct classes, rather than being able to collect discrete data. For that reason, a supervised classification approach is more appropriate than a supervised regression as the data available will only fit into a limited number of classes.

Treating the initial problem as a classification problem supports the use of supervised learning as the most appropriate approach. Using classification allows the assessment of the potential feasibility of detecting contaminants and concentrations based on microwave spectroscopy, as well as allowing the extraction of features most useful to the models which can then be used to develop a targeted sensor for the contaminant.

2.4.3 Information Gain

Information gain is an example of a feature filtering algorithm, the algorithm analyses each feature considering the mutual information that the feature contains measuring the ability of each feature to predict the classification value; information gain filters are quick to apply to feature sets due to the fact that they only consider each feature on its own and not in combination with other features.

2.4.4 Support Vector Machine

Support vector machines[59] (SVM) are a popular machine learning approach with a wide range of applications including forecasting of freight volume[60] and finger print recognition[61]. SVM uses a hyperplane to separate the two classes, the SVM training process optimises the location of this hyperplane for each classification model by placing the hyperplane to provide as much separation between the two classes as possible. The optimised hyperplane can then be used to produce classifications for unlabelled data points, the data is processed to classify which side of the hyperplane the data point lies on for each classification.

Support Vector Machines are an example of a binary classifier that is only capable of placing a prediction into one of two classes. The binary nature of SVMs are overcome within a multi-class problem by using multiple SVMs, using one per classification known as a One-Vs-All approach where one class and all other classes are used. Variations have been proposed to SVM to enable the support of multiple classes[62-64] however it is argued that a tuned One-Vs-All approach can perform as well as other approaches[65] and are widely believed to be easier to work with.

2.4.5 K-Nearest Neighbour

K-Nearest Neighbour (KNN) is a relatively simple machine learning approach, the approach makes no assumption about the underlying data. KNN does not require training like other algorithms such as SVM, RF and GBM. KNN makes a prediction based on the whole training data set which means that the full training data set must be provided to the algorithm at run time. KNN has been used in a range of applications including text mining[66] and stock price prediction[67].

KNN examines an unknown data point against the entire training set provided to the model. The algorithm calculates the nearest neighbours to the unknown point within the training set, the number of neighbours examined is known as the K value. The classification of the K nearest neighbours within the training set are then used to classify the unknown point with a majority rule system being used.

2.4.6 Random Forest

Random Forest[68] (RF) is an example of an ensemble approach in which several weak learners, in this case multiple decision trees, are used together to generate an overall predicted value. RF has been used previously in a range of works including the detection of Android malware[69] and with remote sensing and geographic data[70] .

Random Forest uses a collection of decision trees which are built on the entire provided data set, the forest consists of a collection of decision trees. Each decision tree within the forest uses a random selection of observations within the training set to build multiple decision trees to form the forest. Each decision tree within the forest is then used to generate a classification, each decision tree output is then used to produce an overall classification from the forest.

2.4.7 Gradient Boosted Machine

Gradient Boosted Machine[71-73] (GBM) involves another example of an ensemble method where a weak learner is improved through the combination of multiple learners. GBMs have been used in a wide range of works, including mobility prediction and analysis in smart cities[74] and bankruptcy predictions[75].

GBM uses a weak learner, in this case a regression tree, the trees are added to a model one at a time with existing trees in the model remaining unchanged. The output of the newly added tree is then added to the output sequence of the previous trees to correct or improve the final output of the model, the process continues until a fixed number of trees have been added, training stops once the loss reaches a threshold or no further improvements are made during validation.

2.4.8 Applications of machine learning microwave spectroscopy

Machine learning techniques have been applied previously to data generated by microwave sensors. One such example is work presented by S Cashman et al [76], the work examined the detection of nitrate levels in water, the work used four samples including a sample of deionised water. The work presented used two machine learning methods A logistic regress and a multi-layer perceptron with both approaches performing well in separating samples which contained nitrate and control samples which did not, with both models achieving a classification accuracy of 96%.

Other research work presented by L Harrison et al[77] examined the use of a microwave sensor array combined with machine learning for material identification. The work attempted to identify materials including cardboard, wood and plastic. The work processed the sample data using a KNN outlier detection algorithm to remove outliers within the data set. The work then applied a decision-tree-based support vector machine to classify the three categories combining the benefits of efficiency of decision trees with the accuracy of SVM. The results presented showed an accuracy of between 79.8% and 86.4% accuracy in the classification of materials.

Work undertaken by M Soltani and M Omid[78] examined the use of machine learning with dielectric spectroscopy to detect the freshness of poultry eggs. The work applied four machine learning techniques to the sensor data collected: Artificial Neural Networks (ANN), Support Vector Machines (SVM), Bayesian Network (BN) and Decision tree. The results enabled improvements in the speed of models by reducing the size of the data set required while providing high levels of accuracy with BN, ANN and SVM providing an accuracy of 100%.

Other work presented by A Ali et al [79] examined the detection of cracks in metallic surfaces using a wave guide using two ANN models and an SVM model in a majority voting system. The results show that the developed system was capable of an accuracy rate of over 99% for the classification of cracks.

2.4.9 Features within microwave spectroscopy

There are two key features that are captured through microwave spectroscopy, the first feature is magnitude. Depending upon the type of reading taken the magnitude represents the power that is reflected (S_{11}) or transmitted through the substance (S_{12}).

The second important feature captured using microwave spectroscopy is phase, the phase reading is based upon the change of the transmitted wave compared to the one that is observed at measurement.

There are other indirect features that could be derived between data sets, one such feature is the phenomenon of peak shift. These shifts can often be observed between readings of substances at different concentrations and are caused by changes in the dielectric constant of a material. The change in dielectric constant causes a change in wave velocity which can then been observed as changes in the location of a peak or trough. Little investigation has been undertaken in using peak shift as a feature set in the application of machine learning within microwave spectroscopy data sets.

2.4.10 The case for machine learning in microwave data sets

There is great potential in applying machine learning to microwave spectroscopy data sets, offering the ability to remove large quantities of information and identifying frequencies that could be used in the detection of target substances.

Previous works have examined the use of machine learning with microwave spectroscopy data sets with significant successes including the work presented by S Cashman et al [76] which was able to determine concentrations of nitrates in water using an artificial neural network.

Feature filtering methods such as information gain enables large feature spaces such as the ones collected through microwave spectroscopy with a significant reduction in computational time required to process. Feature filtering is often used as a first stage in feature selection processes to help quickly reduce the size of feature spaces before applying more computationally costly approaches such as wrapper feature selection.

Because data will be collected specifically for the purpose of analysis it is possible to have considerable control of the data set allowing the data to be labelled during collection enabling the use of either supervised or unsupervised learning. The proposed methodology could make use of supervised learning enabling the approach to take full advantage of the labelled data sets provided.

Using supervised learning allows for the use of classification methods for discrete data sets or regression approaches for continuous data sets. Due to the limited number of quantities which will be able to be captured in a lab-based environment only a limited number of discrete classes will be collected, rather than a range of readings across a wide spectrum, this limits the ability to develop a regression model as there will only be limited discrete data readings.

The aim of the application of machine learning techniques is to identify frequencies that could be used to detect the concentration of specified substances. For this reason, discrete data in a limited number of classes should provide a sufficient amount of data to identify frequency ranges that are sensitive to a specific substance with the classification a by-product of the processes to help support the feasibility of using each frequency.

There are a range of features within a microwave spectroscopy data set that can be used, with magnitude being the most common feature used in other academic works. Microwave spectroscopy also offers phase readings which could be used as an alternative. There is also the potential of deriving a peak-shift feature set, identifying the location of peaks and troughs within the data set, using one class as a baseline and calculating the shift of each peak in subsequent classes.

2.4.11 Feasibility of using machine learning in microwave spectroscopy

To establish the feasibility of using microwave spectroscopy with water samples, initial experimentation was undertaken using three water samples collected from three locations: A local River, Home Tap and Tap at LJMU. The samples were inserted into a resonant microwave cavity connected to a VNA which was in turn connected to a laptop with LabVIEW software set up to collect data from the VNA using a pass through measurement.

Each sample was measured ten times, for each reading the sample was removed from the cavity before being agitated and rotated to a new position. The sample was rotated to ensure that the readings were not impacted by the surface of the centrifuge tube.

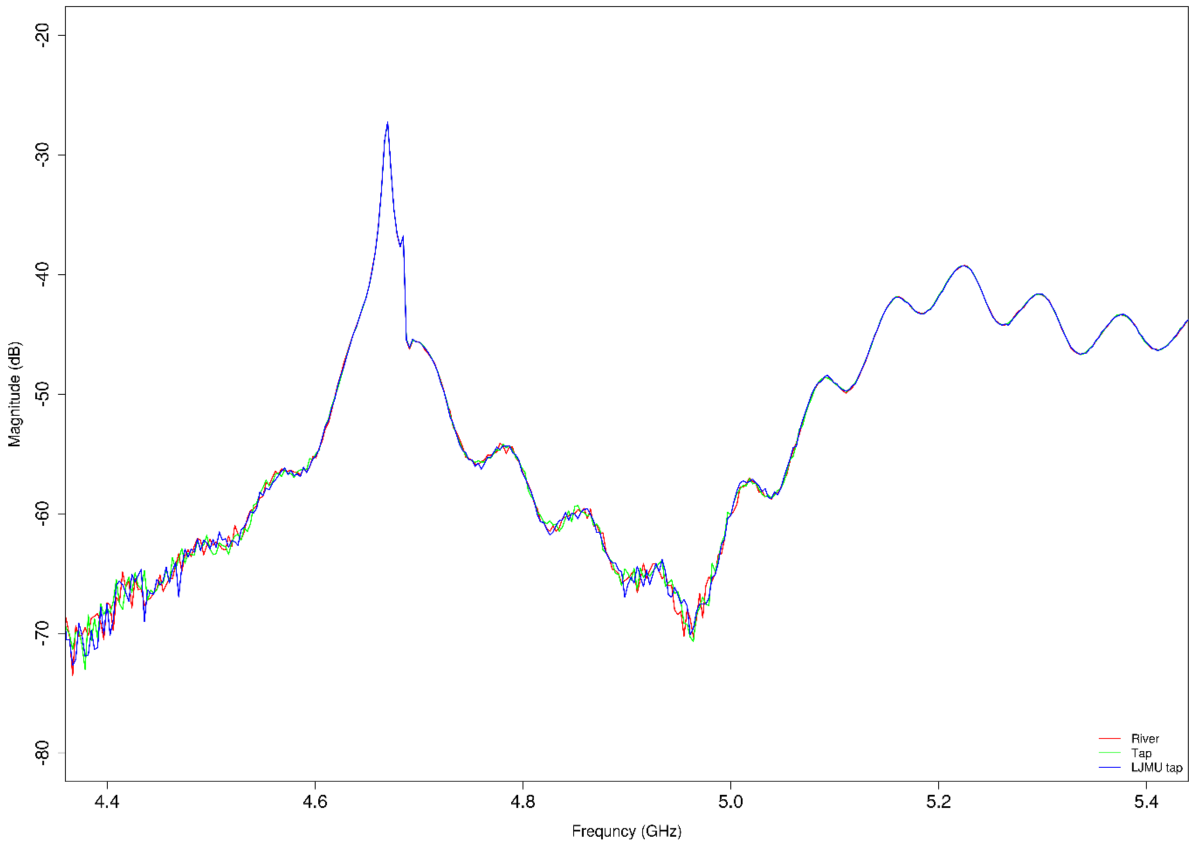

Figure 2-3 A plot of water sample responses to the 4.5 GHz to 5.5 GHz frequency range

Figure 2-3 shows a frequency plot of the samples taken within the frequency range of 4.5 GHz to 5.5 GHz. The plot shows several important points that must be considered when processing microwave spectroscopy data. Firstly, there is a clear peak for all three samples within the 4.6 GHz frequency range, as this is an S_{12} measurement this shows a large amount of the transmitted power is passing through the substance and is not being absorbed.

The plot also shows that for a large proportion of the data the readings follow each other indicating that these frequencies are not sensitive to any of the substances which differentiate each of the samples. The readings do however differ within several frequency ranges indicating that there are differences between each of the samples tested, this is to be expected as the tap water samples were taken from different sites on different days. It is also likely that river water will contain additional substances which are not present in tap water, it is also likely that tap water will contain some substances which are not present in river water such as processing chemicals used to purify water for human consumption.

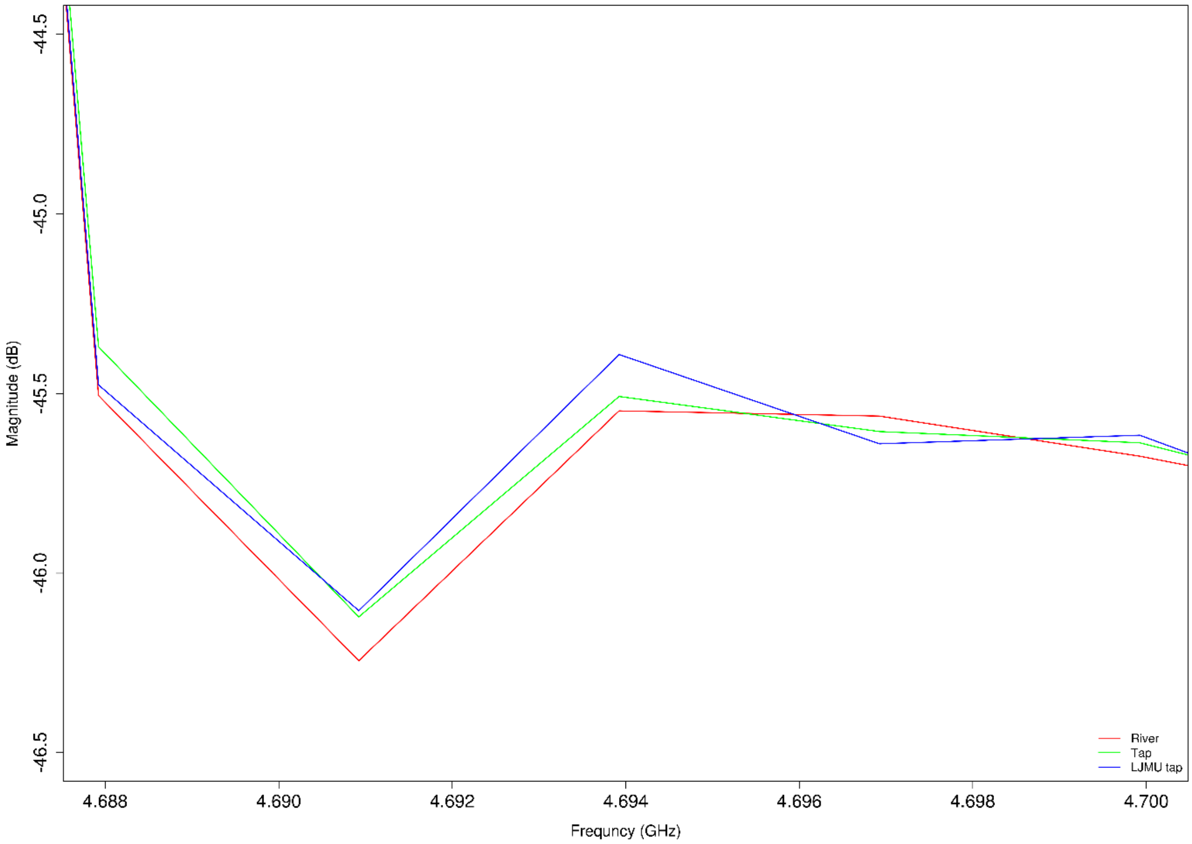

Figure 2-4 A focused plot of between 4.688 GHz to 4.695 GHz

Figure 2-4 shows a focused plot within the 4.68 GHz to 4.69 GHz frequency range, the plot shows a deviation between the samples. These differences are small but consistent across the averages within the data set. Given that a significant proportion of the readings agree with each other it is likely these changes are due to differences in the substances within the water.

Using a simple information gain filter, it is possible to identify features within the data set that could be used to identify potential features that differentiate the three water samples. Figure 2-4 shows the top 5 ranking features using a simple information gain filter to identify frequencies likely to be usable in the identification of the three substances.

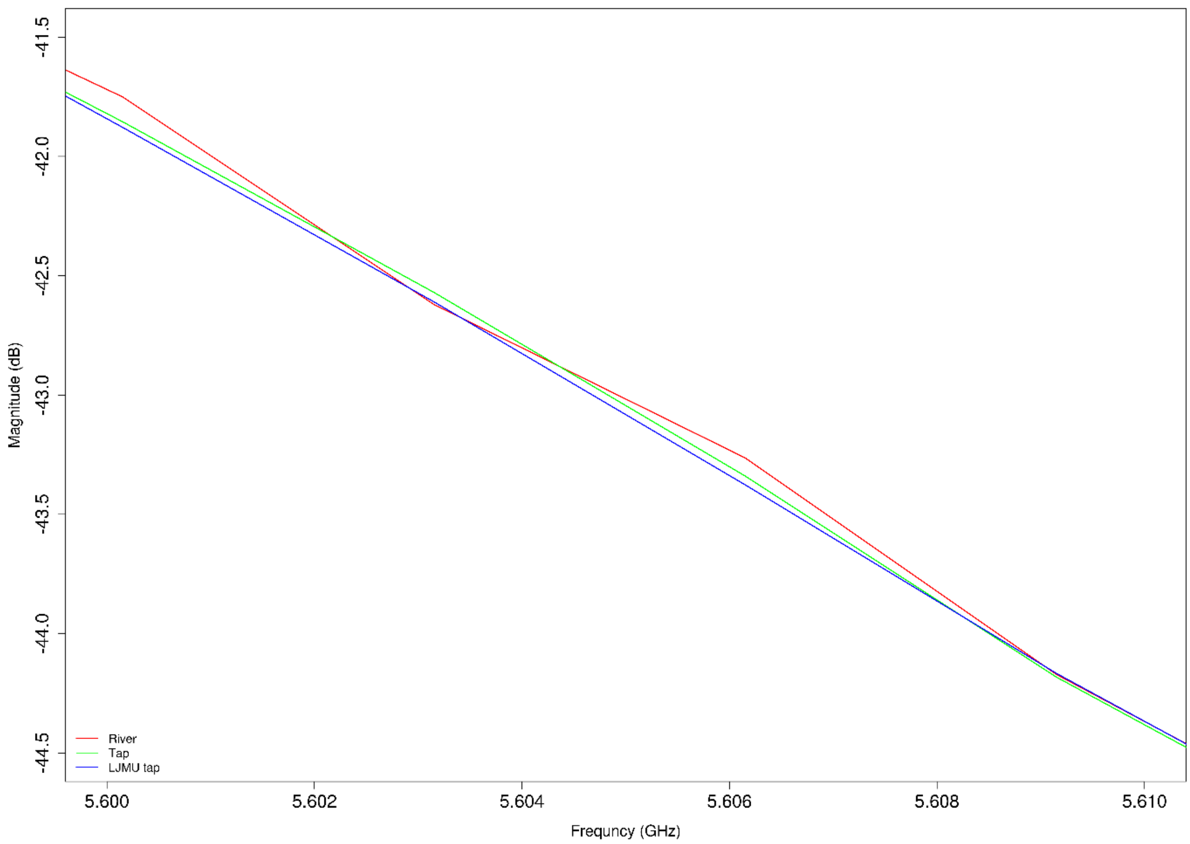

Figure 2-5 A plot of the 5.268 GHz to 2.280 GHz frequncy range

Figure 2-5 Shows a plot of features identified using a simple information gain filter applied to identify frequencies, ignoring regions where the magnitude levels fall -60 dB the following features were identified using an information gain filter. Table 2-2 shows the frequencies identified through the application of an information gain filter within the identified region.

Table 2-2 Features identified within the 4.70 GHz region

Information gain filter frequencies (Hz)
4700925440
4703426048
4705926656
4708427264
4704676352
4709677568

Results suggest that feature engineering techniques such as information gain filters can be applied to microwave spectroscopy data. The frequencies identified through this approach show a clear difference between the data sets collected. Further work needs to be undertaken on a range of microwave spectroscopy data sets investigating the use of approaches such as information gain filters as well as more advanced supervised learning techniques to assess the potential to use machine learning to identify appropriate frequencies within a microwave spectroscopy data set.

3 Communication in underwater sensor networks

3.1 Communication methods in underwater sensor networks

3.1.1 Introduction

Communication methods form the fundamental backbone of WSNs and UWSNs, they allow for network nodes to relay data between network nodes allowing data to pass through to key network nodes. Communication methods can have a large impact on how the sensor networks operate, what routing protocols can be used on the network, how far nodes can be separated and how quickly data can be transmitted through the network.

This section investigates common communication methods within both conventional WSNs as well as UWSNs, the advantages and drawbacks of these communication methods, this fundamental principle of operation and achievable data rates and communication distances that each communication method offers.

3.1.2 Optical

Introduction

Optical communications in UWSNs offer high speed point-to-point communication links using light as the communication method. Optical communication has been an established method of communication, using pulses of light to encode messages.

Principles of operation

Optical communications use a focused light source to communicate, encoding digital data into a series of light pulses, light can be passed through a variety of mediums such as air, fibreoptic and water, to name some of the mediums used in optical communication. A receiving optical modem can then be used to receive the light pulses from the transmitting modem and decode the pulses back into the original digital data.

Optical communications depend on two key components, a light source used to transmit a signal, and a photoelectric sensor to detect the transmitted light. Common light sources for optical communications including LEDs are capable of a high level of radiance in a small form factor and can be modulated quickly to encode data. The second light source commonly used is a laser light source, capable of providing a source usable over larger distances and providing higher data rates than those of LEDs, due to lasers providing a coherent light, enabling higher data rates.

Photo sensors are used in optical communication to detect the light provided by the transmitter light based on the principles of the photoelectric effect, in most cases a fast response photodiode is used. The photodiode responds to the light signals of the transmitter and outputs the detected pulses from the transmitter. The output of the photodiode can then be demodulated to reproduce the transmitter data.

Advantages

Due to the speed at which light travels, optical communications offer unparalleled speed of communications in the right circumstances. However, these conditions are not always achievable, especially in uncontrolled environments such as air and water. Commercial products offer communication link speeds of up to 10 Mbps in the case of the BlueComm 200[80].

Optical communication links can form point-to-point links over significant distances with commercial products offering communication distances of 150 M with products such as the BlueComm 200[80]. However, these distances depend heavily upon the optical qualities of the water within which they deploy, which can have a significant impact on the real-world communication distances.

Drawbacks

Communication depends on the clear transmission of light from one point to another, this raises several issues within underwater environments. Due to the need to transmit light from one point to the other a direct path is required for the light to be sent between modems. This can severely limit the usefulness of optical communications within a UWSn. Requiring a direct path and alignment between the transmitter and receiver, failure to correctly align the transmitter and receiver can have a significant impact on the achievable data rates and can cause the link to fail completely. The requirement of alignment also means that sources of vibration could cause the nodes to become misaligned causing data rate drops or link failure over time.

Due to the requirement of having line of sight with each communicating node an optical modem is only able to be used to communicate with a single other node at any one time, meaning that only a point-to-point link can be formed. While a point-to-point network can be formed using optical communications, these point-to-point links form a significant weakness as multiple point to point links must be established if network redundancy is required.

The reliance on using optical communication also requires that any medium that the light uses to transmit data has suitable optical characteristics, in uncontrolled underwater environments this is not always achievable, with discolouration and particles such as suspended solids impacting on reliability, data rates and achievable distances.

The ability to only form point-to-point networks with optical communications also impacts on the resilience of networks created, point-to-point links limit the network topologies that can be formed; network topologies such as mesh networks, popular within WSNs, cannot be easily formed with point-to-point links, requiring multiple optical modems to form a mesh topology. Many remaining network topologies are less resilient to network node failures which are a common occurrence within WSNs when network nodes often leave the network due to a variety of factors such as depletion of power, damage or physical removal.

State of the art

Research work presented by F Hanson and S Radic[81] examined the use of optical communications at speeds of 1Gbit/s in an underwater environment in controlled conditions. The research used a 1064 nm laser diode supplied with 10 mW of power. The work was undertaken in a 2 m tube filled with simulated sea water. The results showed that the system could be used to achieve 1 Gbit/s speeds over distances of 2 m in sea water conditions.

N Farr et al [81] demonstrated the use of optical communications in an underwater environment using a 470 nm light source , the work carried out experimental works in a 15 m pool, with the signal reflected with mirrors multiple times to create an effective transmission distance of 91 m. Real world experimental works were undertaken at a vertical separation of 10 m and was able to achieve a OOK modulated signal of 10 MHz.

Other research work presented by J B Snow et al [82] showed the use of pulse modulated communications using a green laser beam at 514 nm to communicate in lab based environments as well as a salt water environment. The results presented were limited by the instruments used and were able to achieve 50 Mbit communication speeds at distances of 9 m in seawater but were limited by the modulation and detection equipment. The results suggested that data rates of over 100 Mbit/s could be achieved over 10 attenuation lengths, while a reduction in data rates could achieve double.

D Anguita et al[83] presented preliminary simulation and experimental results for an optical wireless underwater communication system for AUVs. The work concluded that there are many important factors that must be considered including the water turbidity, the work also identified that ambient noise from daylight would play a part in communications in shallow waters. The simulated results projected that a 140 mw could have an effective range of over 20 m at rates of 1.8 Mbps.

Other recent work published by J Wang et al[84] presented an optical wireless communication system using a 520 nm laser diode with data rates of 500 Mbps at distances of 100 m. The works presented include experimental work undertaken in a large water tank, the work used mirrors to repeatedly reflect the laser output to increase the total distance to 100 m. The work concluded that while distances of 100 m were demonstrated it would be possible to achieve distances of 146 m at data rates of 500 Mbps.

3.1.3 Acoustic

Introduction

Acoustic communication is a popular method for underwater communication, acoustic communications operate in a similar way to conventional RF communication using lower frequencies than RF communications impacting on the achievable data rates but allowing for long distance communication to be achievable in an underwater environment.

Principles of operation

Acoustic communication normally uses a hydrophone which is used to produce a sound wave within the sound wave spectrum close to or within the acoustic range of the sound wave spectrum 20 Hz through to 20 KHz though some commercial products are able to produce waves within the infrasound and acoustic sound spectrum range of the sound spectrum. Hydrophones can also be designed to use the piezoelectric transducer to produce a sound wave to also transmit data, though this can also be a separate transducer specifically designed to output a sound wave at a frequency range maximising performance.

Hydrophones are microphones specifically designed to match the acoustic impedance of water, while a conventional microphone could be used, however due to the acoustic characteristics of sound travelling through water this approach is significantly less effective than a purpose-built hydrophone. Most hydrophones are built in the same way as microphones, using a piezoelectric transducer that produces electrical potential when subjected to pressure change from a sound wave.

Data is modulated using similar approaches to those applied within RF communications with some modulation methods being more commonly used due to the behaviour of sound waves in underwater environments and this is covered in Chapter 3.

Sound waves in an underwater environment act as a pressure wave, in underwater environments this pressure wave is capable of travelling faster and further than in air, with the speed of sound in water being 1498 m/s in distilled water[85] compared to 343 m/s in air[86].

Advantages

Acoustic communication offers many advantages in an underwater environment, one of the most impressive advantages of acoustic communication in an underwater environment is the communication distances that are achievable, due to the way in which sound waves propagate through water allowing for significant communication distances to be achieved compared to other communication methods.

Acoustic communications also reduce the complexity with network set-up, while some products do require some form of alignment many offer a significant degree of flexibility allowing the hydrophone to be placed in the general direction of the incoming transmission.

Acoustic communications further allow for multiple network nodes to use the same communication channel, allowing for nodes to be able to communicate with multiple nodes without the need for multiple links to be formed in the case of optical communications.

Acoustic communication offers significant communication systems with commercial products offering communication distances of between 2 Km and 6 Km in the case of the Teledyne marine 900 series[87]. This is significantly higher than other communication methods available such as optical and radio frequency.

Drawbacks

Acoustic communication does suffer from some disadvantages, particularly with communication data rates that can be achieved[88] compared to other communication methods. Due to the wavelength of sound waves used in acoustic communication these rates can severely impact on the achievable data rates. Commercial products such as the Teledyne marine 900[87] series offers data rates of between 80 bps and 15,360 bps.

Acoustic communications also suffer in shallow water deployments, due to an increase in the number of echoes produced from sound wave reflections from the surface of the water and the bottom of the deployment zone such as a riverbed or ocean floor. The echoes cause an increase in the amount of noise impacting on communication distances and requiring additional signal processing techniques to reliably communicate[89] such as equalization, spread spectrum signalling and error coding.

Acoustic communication techniques do not readily allow for communication across water air boundaries, this can add complications during implementation with sink nodes that might require a communication link above ground such as a 3/4G link or a WiFi connection still requiring an underwater deployment to retrieve data from the rest of the network.

Acoustic communications do not function well out of water, while this is not normally an issue in areas of deployment, in tidal areas it is possible that some or all nodes may at times be out of the water; during these periods it is likely that nodes will be unable to communicate with one another using acoustic communications, and would likely have to fall back to methods such as RF communications during these periods.

Acoustic communication also introduces additional noise into a marine environment, many animals within marine environments depend on hearing for a variety of reasons including, hunting, detecting predators and reproduction. Studies have shown that additional noise within marine environments can cause issues, including damage to the hearing of animals and upsetting the predator-prey balance[90].

State of the art

Work presented by X Tu et al[91] examined the use of prefix-free frequency domain equalization for underwater acoustic signal carrier transmissions. The work examined the proposed method through experimental works at four distances between 250 m and 1500 m in the Gulf of Mexico. The work used a frequency of 85 KHz at a data rate of 22.7 kbps, the field trials demonstrated the effectiveness of the TR-FDE scheme developed outperforming existing approaches such as overlapping FDE.

Other work put forward by S Zheng et al[92] evaluated the use of acoustic networks for autonomous underwater vehicles in confined spaces. The work used a carrier frequency of 16 KHz with a data rate of 63.49 bps. The work used 6 nodes to form an underwater acoustic network to establish an acoustic link to an AUV in a confined space, data was relayed from the network nodes to a control node through the use of an RF wireless link using an antenna located above the surface of the water.

Work by A K Morozov and J C Preisig[93] presented work related to multi-carrier modulation. The work showed that the proposed system offers performance improvements over traditional feedback equalizer. The experimental work showed data transmission through a shallow-water channel at a depth of 12 meters with communication distances of 2 km at data rates of 2.5 Kbs.

Work by A Abdi and H Guo [94] presented a multichannel receiver for underwater wireless communication. The work presented simulations of the proposed multi-channel receiver module using a vector sensor receiver rather than a conventional pressure sensor. The simulation results showed that the proposed vector shows a significant improvement over the signal pressure sensor, the simulated results were based on a separation distance of 5 Km at a bit rate of 2.4 Kbs using the 12 kHz frequency.

3.1.4 Radio Frequency

Introduction

Radio Frequency communication is a common communication approach in conventional wireless sensor networks. Using modulated radio and microwave carrier waves to transmit data in a similar approach to acoustic communication, Radio frequency communications depend on magnetic waves as carrier waves meaning the propagation of them in water can be a significant challenge, meaning that the use of RF in an underwater setting is uncommon due to the attenuation of magnetic waves in water.

Principles of operation

Radio Frequency uses a carrier wave within the Radio and Microwave of the electromagnetic spectrum. The carrier wave takes the form of a sinusoidal wave generated from a source such as a crystal oscillator. Radio Frequency communications consist of two types of fields, the first is the magnetic field, known as the H field or near field, these fields normally occur between antennas that are placed at one wavelength or less and have an inductive effect. The E field also referred to as the radiation field and is formed from an electric field. The magnetic near field component decays at a significant rate, and the near field region is considered to exist up to 1 wavelength before entering a ‘transition zone’ which extends from 1 wavelength to two wavelengths. Within the transition zone the effects of both near field and far-field are both important, within the transition zone the behaviour of the near field ceases to be important with the far field effects as the dominant interactions.

Radio Frequency transmitters use an oscillator source to produce an alternating current at the desired frequency, in most instances the oscillator produces a sinewave that is used as the carrier wave. The carrier wave is then used in conjunction with a modulator which is used to encode the desired data into the carrier wave. The modulated carrier wave is then amplified through an amplifier before being transmitted through an antenna.

Radio Frequency receivers use an antenna to receive a transmitted signal, the signal is then passed through an RF amplifier to amplify the weak received signal. The signal is then passed through a tuner circuit that extracts the desired frequency from the mix of signals that are received, the received signal is then passed through to a demodulator. Depending on the type of modulation technique used the demodulator can either take the form of a simple circuit or a more complicated IC that is capable of demodulating multiple modulation methods.

Advantages

Due to the higher frequencies involved in using RF communications there is the potential to provide significantly higher data rates than those offered from acoustic communications. Published works show that data rates of up to 11 Mbps were achieved by J Lloret [95] et al during experimental works using the 2.4 GHz at distances of 0.17 M.

RF communications also offer the potential to cross the water air boundary, this has the potential to enable the deployment of sensors both in air and water. This benefit could also enable submerged sensors to collect data which is then transmitted to a wireless sensor network deployed in air, enabling longer range communications.

Due to the nature of RF communications data transmissions are not impacted by factors such as environmental noise, unlike acoustic communications, or vibrations unlike optical communications, making RF an ideal candidate for applications such as oil and gas exploration where environmental noise and vibrations from drilling are not uncommon.

Unlike optical communications the optical quality of the water that the sensor is deployed in does not impact on RF communications. This benefit enables RF to operate in environments where there are high levels of turbidity or suspended solids in the water without impacting on communication performance.

Drawbacks

RF communications suffer from issues with attenuation of the signal when deployed within an underwater environment, this can have a significant and detrimental impact on the communication distances that can be achieved using RF in such an environment.

Little work has covered feasible communication distances that could be useful to an Underwater Wireless Sensor network, with most works being able to achieve distances measured in centimetres rather than metres. Research work by A. Abdou [96] has shown communication distances of 30 M can be achieved. Other research work using loop antennas has shown distances of 90 M in the case of work undertaken by A. Shaw [97] using highly directive antennas.

State of the art

Much focus has been given to underwater communication with RF, work by Lloret et al[95] examined the possibility of using the 2.4 GHz frequency. The work presented managed to achieve distances of up to 20 cm with data rates of 11 Mbps being achieved in controlled indoor environments

Work by Rubino et al[95, 98] examined the user of a ¼ wave length mono pole antenna for the transmission of a live video feed from an underwater UAV. The work used a transceiver module using the 868.3 MHz frequency with a power output of 25 mW using frequency shift keying modulation. The results showed communication at data rates of 4.9 Kb/s at distances of 5 m could be achieved with some packet loss.

Other work by A Shaw et al[97] investigated using low frequencies of 3.5 MHz, 4.7 MHz and 5 MHz combined with directive loop antennas. The research was able to achieve significant communication distances in real-world experiments of up to 90 m in seawater when appropriately aligned at data rates of 50 kbs. The work used highly tuned antennas that would be unlikely to function correctly if exposed to air rather than water.

A A Abdou et al[96, 99] has presented multiple works focused on communication in an underwater environment using radio frequency. The works presented used a bowtie antenna designed for use at the 433 MHz frequency. The results showed that in real-world experimentations in the Leeds Liverpool Canal it was shown that a 433 MHz signal generated with a signal generator and transmitted though the designed antenna could be detected at distances of up to 30 m.

Work presented by B Kelley and K Naishdham[100] presented work using broadband underwater antennas based on a loop antenna design and coated with TiO_2 paint to help match the impedance between the antenna and the seawater. The works presented combined with previous work by B F Bush et al[100, 101] shows that RF communications using the 5 MHz frequency cross the water-air boundary enabling longer range propagation in shallow deployments possibly enabling distances of up to 100 metres to be achieved.

3.1.5 Comparison of methods

There are a range of communication methodologies used within underwater wireless sensor networks, with a wider variety of communication methodologies used than in conventional wireless sensor networks. Due to the challenges associated with radio frequency communications in underwater wireless sensor networks, acoustic and optical communications have dominated in the commercial sector with a variety of products available using acoustic and optical communications.

Table 3-1 shows a summary of the key features between the three communication methods discussed, highlighting distance, data rates, environmental considerations, and supported network topologies and gives a summary of the advantages and disadvantages of each communication technology.

Table 3-1 A comparison of features of Optical, Acoustic and Radio Frequency communications in underwater environments

	Optical	**Acoustic**	**Radio Frequency**
Data rates	10 mbps [80]	80 bps – 15360 bps [87]	50 kbs [97] 4.8 kbs [98] 11mbps [95]
Communication distance	150 m [80]	2 km – 6km [87]	90 m [97] 100 m [100, 101] 30 m [96, 99] 5 m [98] 0.2 m [95]
Topology	Point to Point	Broadcast	Broadcast
Environmental considerations	Data rates impacted by suspended solids and optical clarity	Concerns over introduction of noise into marine environments	Conductivity of water
Alignment	full alignment	Some	Some
Impacted by clarity of water	Yes	No	No
Concerns over impact on wildlife	No	Yes	No
Impacted by conductivity	No	No	Yes

Optical communications meet a clear need in underwater wireless communications, providing high speed point-to-point links at a considerable distance when alignment and optical conditions permit. Acoustic communications also play an invaluable role where long-distance communications at significantly lower data rates are required. Existing communication methods leave space for a communication method that provides moderate data rates at short to medium ranges, RF offers the potential to fill this gap, offering the potential for no alignment and a broadcast topology without reliance on the optical quality of water.

RF communication presents potential advantages over existing technologies such as optical communications which demand significant alignment and can be impacted by the quality of the water that the system is deployed in. Similarly, RF communication also presents advantages over Acoustic communications including being usable in shallow environments, being able to cross the air-water boundary and offering potentially higher data rates.

3.2 Antenna design

3.2.1 Introduction

Antenna design plays an important role in radio frequency communications, antenna designs impact on factors such as directionality and the ranges of frequencies that the antenna will function at, for example broadband antenna designs allowing for a wider frequency range. Antenna designs are specific to the designed frequencies with measurements based on the target frequencies.

3.2.2 Dipole antenna

Dipole antennas are the simplest form of antenna design used in radio frequency communication, dipole antennas are usually split in the centre and fed from a balanced transmission line.

A common dipole antenna design is a half wave dipole antenna; a half wave dipole is formed of a conductor with a length of half of the wave length for the target frequency[102]. A half wave dipole antenna is divided in the middle, meaning either side forms approximately ¼ of the wavelength. Figure 3-1shows an example of a basic half wave dipole antenna.

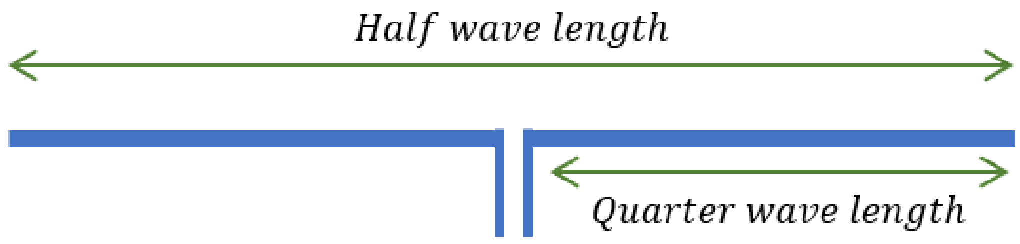

Figure 3-1 An example of a half wave dipole antenna design

Another example of a dipole antenna is a short dipole antenna, this type of antenna has a length significantly less than that of the wavelength of the frequency targeted normally with a length of less than one-half wavelength. These forms of dipole antennas are used where using a standard one-half wavelength dipole antenna would not be practical.

As a short dipole antenna is shorter than a resonant antenna such as the one-half wave length dipole antenna, a short dipole has increased capacitive reactance at the feed points; to counteract this a load coil or other form of matched load network is used with these forms of antenna. The loading coil counteracts the increased capacitive resistance of the antenna, stopping power being reflected into the transmission line. Figure 3-2 shows a short dipole antenna with load coils positioned at the base of the antenna.

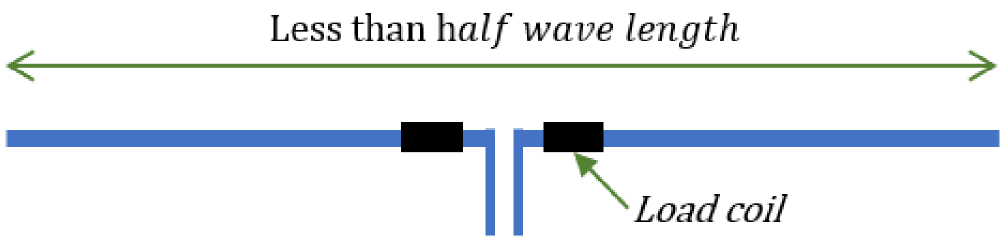

Figure 3-2 An example of a short dipole antenna with load coils

One-half wavelength dipole antennas and short dipole antennas have an omnidirectional radiation pattern in the plane that is perpendicular to the axis of the conductor used for the antenna. Due to the design of both the one-half wavelength dipole antenna and the short dipole antenna both antennas have a low bandwidth.

Research by O. Aboderin et al[103] developed two dipole antenna designs targeted at radio frequency communications in underwater wireless sensor networks. Both antennas used a parasitic element with a dipole antenna. The antennas designed and presented achieved a bandwidth of up to 70 MHz between 10 MHz and 80 MHz The antennas used different parasitic elements, with the first using an ordinary straight parasitic element, the second antenna used a ‘coily’ parasitic element, the results form experimental and simulated work showed that the second antenna performed slightly better than the conventional straight parasitic element.

Investigations undertaken by S. I. Inácio et al[104] examined the use of dipole antenna designed for underwater communications. The work used two antennas with a length of 50 cm and a thickness of 3 mm. The work used the 38 MHz frequency range, the results showed that simulated results matched well with the supporting experimental work, the results obtained show a likely operational range of 5.5 m in fresh water.

3.2.3 Monopole antenna

A monopole antenna normally consists of a slight rod-shaped conductor mounted perpendicular over a ground plane.

Whip antennas are a common form of monopole antennas and used for a variety of applications including car radios, cordless phones as well as base station antennas. Like one-half dipole antennas the pole of the antenna is normally based on the wavelength using a quarter of the wavelength of the target frequency.

Work by Z. M. Loni et al[105] investigated the use of a floating monopole antenna for ocean monitoring applications using the 433 MHz frequency range. The work used a simple monopole antenna for communication in sea water, the results show that a propagation distance of approximately 4 m was achieved using the submerged monopole antenna.

3.2.4 Biconical antenna

Biconical antennas share some of the characteristics of dipole antennas however they provide a wider bandwidth coverage, this factor makes them particularly useful as it can enable the antenna to function in and out of water. Conical antennas improve on some aspects of basic dipole antennas, improving on bandwidth due to the element structure used in these forms of antennas. Conical antennas are also commonly known as a Butterfly, Bowtie and Bifin antenna.

A number of publications have been based on the investigation of using these types of antenna, one example is work published by E. A. Karagianni et al[106]. The work developed a bowtie antenna designed for use in seawater, the antennas were designed to function using the 2.4 GHz and 5 GHz frequencies, the same frequencies used for Wireless Local Area Network (WLAN) commonly referred to as WiFi as set out in the IEEE 802.11-2016 [107] standards. The work showed that the designed antenna functioned in both pure water and simulated seawater with 3.1% salinity. The experimental work presented shows functionality at short ranges of 15.2 cm, with a power measurement of -60 dBm in pure water and a power measure of - 64 dBm to - 71 dBm in simulated seawater, in comparison to a power measurement of – 30 dBm in air.

Research works produced by A. A. Abdou et al[96] examined the use of a bowtie antenna targeted at the 433 MHz frequency range. The work examined and designed an antenna suitable for use in underwater wireless sensor networks. The work examined the propagation of electromagnetic waves in underwater environments designing a bowtie antenna with these factors in mind. The antenna designed was simulated and supported with supplementary experimental works. The results show that the antenna, the simulated and experimental works were both in good agreement, the results showed that the antenna had a loss of 40 dB at separation distances of 45 cm.

Additional research carried out by A. A. Abdou et al[108] further examined the antenna developed previously in field trials. Experimentation undertaken used a voltage-controlled oscillator to output a constant 433 MHz signal using a bowtie antenna. A spectrum analyser was used to measure the signal strength across a range of tested distances. The experimental works examined signal decay form a starting distance of 0 m up to 30 m, the results showed that distances of up to 30 m could be achieved with the transmitted signal being received at – 110 dBm. The work identified that transceivers with a high sensitivity would still be capable of receiving this signal level.

Work by K. N. Alvertos et al[109] examined the use of a bowtie antenna for use in the 2.4 GHz and surrounding frequencies used by WLAN commonly referred to as WiFi, the work designed and tested an antenna targeted for use in underwater environments. The work conducted simulations on the antenna design before moving forward with supporting experimental works in controlled conditions. The authors concluded that the antenna design tested was able to support communication distances of up to 60 cm.

3.2.5 Loop antenna

Loop antennas can be formed in a variety of shapes including circular, rectangular, square or hexagonal. Like dipole antennas Loop antennas are resonant antennas and have a length based on the wavelength of the frequency that the antenna is designed for. Loop antennas offer a more compact form than other antennas such as a one-half wavelength dipole antenna. As well as a standard loop antenna there is also a small loop antenna, like a short dipole antenna this antenna is a smaller variant of the larger full-sized one-half wavelength antenna.

Loop antennas have proved to be a popular area of research for underwater communications with work such as that published by A. Shaw et al[97] which examined the use of loop antennas in seawater. The experimental works used two loop antennas using three frequencies: 5 MHz, 4.5 MHz and 3.7 MHz at distances of up to 90 metres. The work showed that by using the 5 MHz frequency, it would be theoretically feasible to transmit at data rates of up to 500 kb/s. The work showed that most signals attenuate quickly but showed that communication could be achieved at distances of up to 90 meters.

Other experimental works by S. I. Inácio et al[110] have examined the use of a loop antenna design in an underwater environment. The work simulated the antenna design used based on using a frequency of 39 MHz before proceeding with supporting experimental works. The results from simulation and experimental works were in good agreement. The works showed that at 5 metres of separation, a power reading of – 80 dB was recorded. The works also highlighted the change in the orientation of the radiation pattern of the antennas between fresh and sea water, showing the need for conductivity to be accounted for at the design stage of the antenna.

Work by O Aboderin et al[111] exemplified the use of a loop antenna with ground plane for underwater communication. The work presented simulation work for the proposed antenna and showed that the directivities of the antenna were higher than the test J-pole antenna used in comparison. The work presented was supported with experimental works which examined the antenna design in freshwater, the experimental works showed the antenna exhibits wideband communication and high levels of directivity within the examined frequency range of 0 MHz to 200 MHz.

Other work by A Massaccesi and Paola Pirinoli[112] examined a range of loop antenna designs for use in fresh water and seawater, with simulations for fresh water using the 433 MHz frequency and 10 KHz for seawater. The simulation works highlighted that antennas of the same electrical length behave differently in fresh water and salt water, but it is possible to find different solutions with similar behaviours in the two environments.

3.2.6 Conclusion

Based on the existing works within the field the proposed experimental works through the works presented will use a bowtie antenna design. The work will use this antenna design for several reasons, the first is due to the omnidirectional design of the antenna, enabling the antenna to be used without the need for a high level of alignment between sensor nodes. Not only does the lack of need for alignment enable easy deployment and set-up of sensor nodes but also enables sensor nodes to communicate with multiple nodes, forming a broadcast network topology.

The second reason the works will use a bowtie antenna is due to the broadband nature of the bowtie antenna. The wide bandwidth nature of the bowtie antenna allows it to effectively function in air and underwater environments, enabling a mix of sensor nodes to be used both submerged in water and deployed in air.

A further reason for using a bowtie antenna is due to the existing supporting literature, with the works produced by A A Abdou [96, 108, 113] supporting the use of a bowtie antenna design at 433 MHz. Other works such as E. A. Karagianni et al[106] and by K. N. Alvertos et al[109] have also shown the use of a bowtie antenna design for underwater communications at 2.4 GHz and 5 GHz.

The bowtie antenna design presented in previous works has been presented to potentially offer significant communication distances of 60 cm in the case of work produced by K. N. Alvertos et al[109] and more significantly 30 metres in real-world environments presented in work undertaken by A. A. Abdou et al[108], showing that significant distances can be achieved using this antenna design.

The bowtie antenna has been shown to function in several publications with initial fieldwork being undertaken as part of previous research. The antenna offers a wideband communication that allows for functionality in and out of water. The bowtie antenna also offers an omnidirectional radiation pattern reducing the need for alignment between sensor nodes, enabling a broadcast network topology and reducing deployment complexity.

3.3 Propagation of electromagnetic waves

3.3.1 Introduction

Propagation of electromagnetic (EM) waves poses a challenge within water, with factors such as attenuation of signal vastly increased compared to in air. The speed of EM waves within water is also reduced, impacting on the frequency and attenuation of signal.

Signal attenuation is an important consideration in the application of RF in a wireless sensor network for underwater communication, as the more attenuation a signal suffers the weaker the signal becomes. The weaker the signal becomes the harder it is for a radio transceiver to identify a signal against background noise, transceiver modules are also only sensitive to signals up to a certain strength, beyond this the transceiver is unlikely to be able to function correctly.

The impact that an underwater environment has on frequency is also an important consideration. This is due to signal frequencies changing based on the permeability and permittivity of the environment that is deployed within. It is possible that an antenna that is sensitive to a frequency in air will not be sensitive to the same frequency in water due to the changes in permeability and permittivity of the medium.

3.3.2 Frequency

Frequency is an important consideration for an underwater environment, it is important to understand the influence that the permeability and permittivity play in the frequency of a transmitted radio wave. In order to ensure that communications function in an underwater environment the antenna used must account for this change of signal frequency, either through a wide band design that supports a wide range of frequencies or through the use of a tuned antenna to the appropriate frequency within an underwater environment.

Based on equation 3-1 for the calculation of frequency the speed of the wave must be accounted for. In air the speed of the wave can be assumed to be the speed of light in a vacuum, however in the case of EM propagation in water this is not the case, since the speed of light is dependent upon the dielectric properties of the medium it travels through as shown by Equation 3-2.

The frequency of a wave can be calculated based on its wavelength and the speed of light constant shown in equation 3-1. In most cases the speed of light used is based upon the speed of light in a vacuum.

3-1 Calculation of frequency using the speed of light and wave length

$$f = \frac{c}{\lambda}$$

f Frequency
c Speed of light
λ Wave length

Maxwell's equation shows that the permittivity of the material that the electromagnetic wave travels through impacts on the speed at which the wave travels. Using a derivation of Maxwell's equation allows the calculation of the speed of light[114, 115] for a given permittivity using the equation shown in 3-2. Using equation 3-2 the equation presented in 3-1 can be expended to equation 3-3.

3-2 Derivation of the speed of light using Maxwell's equations

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

c Speed of light
ϵ_0 Permittivity
μ_0 Permeability

3-3 Derivation of frequency using Maxwell's equations for the speed of light

$$f = \frac{\frac{1}{\sqrt{\epsilon_0 \mu_0}}}{\lambda}$$

c Speed of light

ϵ_0 Permittivity

μ_0 Permeability

f Frequency

λ Wave length

3.3.3 Attenuation

Attenuation is an important concern that can severely limit the effectiveness of communications for an underwater wireless sensor network using radio frequency. Two forms of attenuation to consider are attenuation over distance and the attenuation suffered from refractive loss, as a signal passes from one medium to another.

Loss over distance

One of the most important considerations for the use of radio frequency in an underwater wireless sensor network is the attenuation of signal that occurs over the distance transmitted. All RF signals suffer from some form of attenuation as they travel through a medium. The attenuation of signal is related to the permeability and the permittivity of the medium.

Electromagnetic waves suffer from significant attenuation when in water with it being shown[116] that radio frequency attenuation suffers significant losses in sea water. The work showed the significant attenuation experienced in sea water with an increase of attenuation from 3.5 dBm at 10 kHz to 11 dBm at 100 kHz.

The attenuation constant can be calculated using Equation 3-4[117, 118] the equation shows the link between permittivity, conductivity and frequency on the attenuation of a Radio Frequency signal, an increase in frequency or conductivity will increase the attenuation.

Equation 3-4 Calculation for signal attenuation over distance based on permittivity and permeability

$$\alpha = \omega\left\{\frac{\mu\epsilon}{2}\left[\sqrt{1+\left(\frac{\sigma}{\omega\epsilon}\right)^2}-1\right]\right\}^{\frac{1}{2}}$$

α Attenuation

ϵ Permittivity

μ Permeability

ω Angular frequency

Refractive loss

Electromagnetic waves suffer additional loss when changing from one medium to another, this is known as refractive loss. This loss is introduced on top of the attenuation loss of a signal, the value of the loss can be calculated in the formula given in Equation 3-5 [117, 118].

Equation 3-5 Formula for the calculation of refractive loss

$$a = 10\log_{10}\left(|T|^2 Re\left(\frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{\sqrt{\frac{\mu_1}{\epsilon_0\bar{\epsilon}}}}\right)\right)$$

ϵ_0 Permittivity

μ_0 Permeability

$\bar{\epsilon}$ Relative permittivity

T Transmission coefficient

Refractive loss is an important factor to consider as it plays an important role within the propagation path that a signal travels. In some cases, the amount of refractive loss that occurs for example when crossing the water-air boundary and vice versa plus the loss experienced in air is less than the signal would suffer if it passed solely through the water.

3.3.4 Propagation path

Within an underwater environment it is important to consider the propagation path that the signal will take. This will often depend upon the depth at which the communications take place. In many instances it is possible that the propagating path is through the exiting of water and transmitting through air before re-entering the water. This will be the case when the loss experienced through attenuation through only water is greater than the sum of the loss from refractive loss and attenuation loss in air and water.

There are examples of this being demonstrated within experimental works including that presented by I I Smolyaniov et al[119] who examined how depth impacted on communication distances in seawater; the work used a 50 MHz using impedance matched antennas. The work examined depths between 0 m and 2.5 m and distances of 0 to 5 m with both simulated and experimental works, the works showed that as depth increased the communication distance decreased. The work concluded that in seawater received signals were dominated by surface electromagnetic waves.

Additional works by A Palmeiro[120] highlight this issue as well, the work discusses that in cases where two divers are 1 Km apart and submerged 2 m an RF signal will experience less loss than predicted if the signal passes through water. The work identifies that instead, the water crosses the air-water boundary before propagating through air and re-entering the water.

3.4 Modulation techniques

3.4.1 Introduction

Carrier wave modulation is fundamental to communication methods such as acoustics and RF, the modulation approach allows for data to be encoded using the carrier wave, in turn allowing the signal to be decoded by receiving nodes. This section covers some popular carrier wave encoding approaches that are used, covering the modulation approach, the advantages and the drawbacks of the modulation schemes.

3.4.2 Amplitude Shift Keying

Amplitude Shift Keying (ASK) two binary values are represented by two amplitudes, in most cases one of the amplitudes used is zero. This allows the presence of the carrier signal to represent the second binary value. ASK modulation is commonly used in fibre-optic communications.

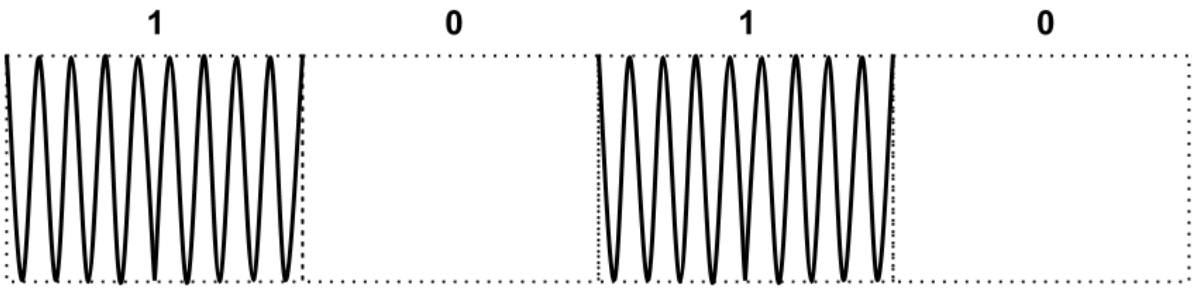

Figure 3-3 An example of OOK modulation for the binary value of 1010

OOK uses the presence of a carrier wave and the lack of presence of a carrier wave to encode digital data, this differs from ASK where a lower amplitude can be used for a low signal rather than the complete lack of presence of a signal. Figure 3-3 shows an example of binary data encoded using an OOK scheme with the digital overlaid over a representation of the signal.

Data encoded using ASK is trivial to decode, assuming a sample is taken at twice the frequency rate the signal can be decoded using a simple lowpass filter. The most simplistic form of amplitude modulation is On Off Keying (OOK).

3.4.3 Frequency Shift Keying

Frequency Shift Keying (FSK) is a modulation scheme in which the frequency of the carrier wave is discreetly altered to encode data into the carrier wave. The simplest form of FSK is Binary Frequency Shift Keying (BFSK), in this modulation scheme a pair of frequencies are used to encode and transmit binary data. Figure 3-4 shows an example of a modulated carrier wave using BFSK.

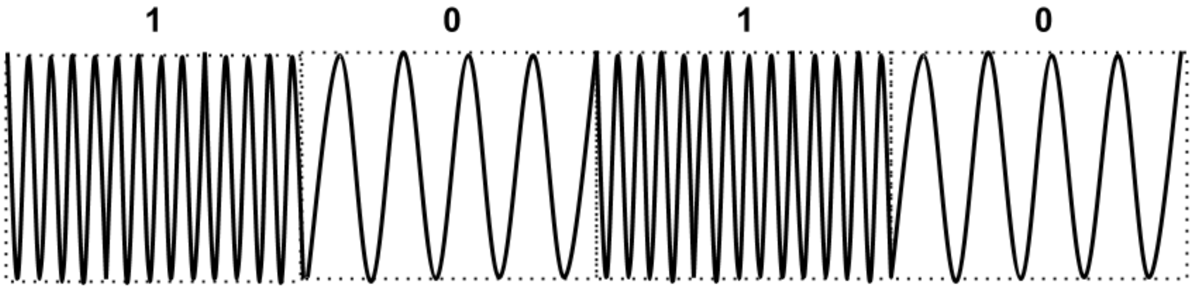

Figure 3-4 An example of FSK modulation for the binary value of 1010

Data encoded using BFSK can be decoded with relative ease without significant amounts of computational power using the Goertzel Algorithm[121] which can be implemented even on devices with limited computational power such as microcontrollers. The approach measures a single frequency component.

3.5 Digital communication data rates

3.5.1 Introduction

Communication rates play an important factor in any form of computer network, from networks such as the internet through to wireless sensor networks. This section explores the limitations of wireless sensor network communication baud rates which are impacted by factors such as frequency and the signal to noise ratio.

3.5.2 Nyquist theorem

Nyquist's theorem [122] is an important cornerstone within digital signal processing, not just limited to being used to calculate network data rates, the theorem plays a role in other computer science areas such as the digitisation of sound.

Nyquist's theorem provides a formula to calculate the maximum amount of data that can be encoded into a given frequency. Equation 3-6 shows Nyquist's theorem, while the value given by this is a theoretical maximum that can be encoded, it shows that the frequency used has a direct impact on the amount of data that can be transmitted.

Equation 3-6 Nyquist's theorem

$$C = 2B\ log_2 L \tag{1}$$

C = channel capacity (bits per second)

B = Bandwidth of the channel (Hz)

L = number of signal levels used to represent data

3.5.3 Shannon Hartley theorem

The Shannon Hartley[123] theorem adapts and expands Nyquist's theorem further and develops work by Hartley. The Shannon Hartley theorem enables the calculation of a maximum theoretical baud rate based upon frequency and signal to noise ratio. The equation given by the Shannon Hartley theorem is shown in Equation 3-7.

Equation 3-7 Shannon Hartley equation

$$C = B \, log_2(1 + SNR) \tag{1}$$

C = channel capacity (bits per second)
B = Bandwidth of the channel (Hz)
SNR = Signal to Noise ratio

The Shannon Hartley theorem identifies that as the signal to noise ratio increases the maximum baud rate must decrease. This is an important consideration where there may be an increased amount of signal to noise ratio such as in built-up areas with other wireless devices or where the transmission signal is weaker; for example, in cases where there is a large distance between the transmitter and receiver.

The theorem shows that there is a link between the baud rate and the maximum communication distance that can be achieved, showing that distances can be increased by sacrificing data communication speeds to reduce the signal to noise ratio.

3.5.4 Data rates

There is an important distinction within data communications between baud rates and bit rates. Baud rates represent symbols per second, symbols in the case of digital communications are the number of bits that can be carried in each signal unit. The baud rate is then calculated as the number of signal units that can be transmitted per second.

Bit rates within digital communications relate to the number of bits that can be communicated per second, this value can be based on the baud rate multiplied by the number of bits that a signal unit carries. In cases where the number of bits that a single signal unit can carry is one, the baud rate will be the same as the bit rate.

The work of Shannon Hartley shows that the theoretical maximum data communication rate is influenced by the signal to noise ratio, with a higher amount of noise within an environment leading to a decrease in the maximum theoretical communication link speed. Based upon this we can infer that reducing the data rate enables operation in higher noise environments or where there is a weaker signal, for example at further distances.

A practical example of this is in the case of the 802.11 family of protocols better known as WiFi. The 802.11 specifications include link speed negotiation between devices to enable devices to connect where there is a low signal to noise ratio, either due to a weak signal attributed to a further separation distance between devices or where these is an increase in signal interference from other devices.

The approach taken by the 802.11 specifications enables a balance between clients in favourable conditions of a high signal to noise ratio while ensuring that clients at further distances, or in higher noise environments are also able to communicate. This approach to balancing data rates within wireless networks could be applied to an underwater wireless sensor network, enabling sensor nodes to use lower data rates to achieve further communication distances, while allowing sensor nodes that are more closely located to communicate at higher data rates.

3.6 Routing protocols

Routing protocols form a vital part of WSNs, enabling data to traverse the network from one sensor node to another and have been explored in detail within a range of works published on the subject. Routing protocols have an impact on factors of WSNs such as node power consumption, network resilience and the ease of deployment of network nodes. This section covers a range of routing protocols used within WSNs, how they can impact on factors such as ease of deployment, power consumption and resilience to network node losses.

3.6.1 Low Energy Adaptive Clustering Hierarchy

Low Energy Adaptive Clustering Hierarchy (LEACH) is a popular routing protocol for wireless sensor networks, using a clustering approach to distribute power consumption across nodes maximising overall network lifespan by attempting to distributing the workload of the network evenly across all network nodes. LEACH has formed the basis on many modified variants that add additional functionality such as multi hop relay support, this section covers the original LEACH approach, with later sections covering some popular variants of the LEACH protocol.

Principles of operation

The LEACH protocol consists of two key phases, the first is a cluster head (CH) selection phase, during this stage cluster heads are elected and set up with nodes selecting the most appropriate CH. The second phase is the steady phase in which data is relayed though the network, the network remains in the second phase until the end of the steady phase when the cluster head selection phase begins again, and cluster heads are elected.

Cluster Head selection

The Cluster Head selection phase of LEACH allows for cluster heads to be rotated, this process allows for the more power demanding task of long-range data transmission and aggregation of data to be offloaded to designated network nodes and rotated to different nodes, sharing the burden and balancing power demanding activities throughout the network.

The LEACH cluster selection process consists of three stages: advertisement, cluster-set-up and schedule creation. The first stage involves a node deciding on if it should put itself forward as a CH, the probability of a node advertising itself as a CH. Each node generates a random number between 0 and 1 and evaluates if the number is greater than a threshold calculated based upon the number of rounds that have passed and the number of desired cluster heads. The original approach works on the assumption that all nodes begin with the same amount of energy although variations of the Cluster Head threshold equation have been developed.

Each identified CH transmits an advertisement message to other nodes on the network, all nodes transmit the advertisement message at the same level of power. During this phase all nodes must keep their receivers to receive cluster head advertisements, one completed non-cluster head node selects the cluster which it will join for the next round based on the received signal strength of the advertisement messages with the highest signal strength being selected, in cases of a tie for which cluster head to use a random cluster head is selected.

Once each node has selected a Cluster Head to belong to, it informs the Cluster Head node that it will be a member of the cluster. During this stage all cluster head receivers must remain on. Once all nodes have registered with Cluster Heads each cluster head forms a schedule based on the number of nodes in the cluster and is broadcast to all members of the cluster.

Steady Phase

The steady phase of operation can begin once a schedule is put in place. Assuming nodes will always have data to send they can transmit data to the Cluster Head during the allocated transmission time. Transmissions to the Cluster Head can use the minimum amount of energy required for the Cluster Head to receive the transmission, based on the received signal strength for the original advertisement message. The steady phase of the LEACH algorithm continues until the start of the next round when new cluster heads are elected.

Advantages

The LEACH algorithm allows for energy intensive data transmissions to be spread across all nodes, making improvements to the overall energy efficiency of the network by maximising the lifetime of the network. Using adaptations in other works for the definition of the threshold allows for factors such as remaining energy to be factored into the selection of a Cluster Head, providing further opportunities to ensure power consumption is distributed across the network nodes as effectively as possible.

The LEACH algorithm lends itself to allowing data aggregation before sending data onwards since each Cluster Head receives data from nodes for which it is the nearest Cluster Head. This means that there is the opportunity to reduce the amount of data transmitted over long ranges by aggregating data that is collected from the nearby area.

The LEACH algorithm reduces the amount of network set-up required, allowing for the formation of an ad hoc network with no knowledge required of other nodes on the network. This also allows for new nodes to be added to the network throughout its lifetime, though an appropriate approach for the derivation of a threshold for Cluster Head selection based on residual power would need to be used to ensure that the network continued to distribute power throughout the network.

The LEACH approach manages node deaths throughout the network, in cases where a Cluster Head dies during the steady stage the orphaned nodes will be reincluded into the network upon the start of the next round; in cases where a node that is not a Cluster Head dies the network will continue to operate normally with the scheduled slot for the dead node going unused. Upon the next round the network will disregard the dead node and will not schedule for it.

LEACH has formed the basis for several adaptations which add additional functionality to the approach. MS-LEACH [124] and M-level LEACH [79] are two examples that add support for multi-hop routing allowing for an increase in network range by routing packets through multiple clusters to reach the target sink node. Other adaptations such as LEACH-C [125] which uses a high energy base station to finish the choice of Cluster Heads improving the chances of nodes with higher residual energy being elected as Cluster Heads.

LEACH is an active area of research being undertaken into variations that add additional features and improve on power consumption. LEACH forms the basis of other routing approaches such as TEEN [126] which uses LEACH as the basis of its approach to communication within wireless sensor networks.

Drawbacks

The original LEACH protocol suffers from issues, such as the lack of multi-hop support reducing the area that the original protocol can be used to cover. Adaptations such as MS-LEACH and M-LEACH have been developed to resolve this gap within the LEACH protocol and provide approaches that allow for a multi-hop network to be formed.

The original protocol also does not consider the energy levels of nodes to set the probability threshold of a node becoming a Cluster Head. This can have an impact on the overall network lifetime, with some nodes dying considerably earlier than others, this issue is resolved with adaptations such as LEACH-C.

LEACH and derivatives of it focus on a singular direction of data transmission, with no provision for communication from the sink node to nodes within the network. In most cases this is not an issue but, in some cases, it may be desirable to allow for bidirectional communication between nodes and the sink nodes in instances where commands wish to be sent to a node or configuration parameters need to be altered. It would also be desirable as such a facility would allow for over-the-air updates to be pushed to nodes on the network.

3.6.2 Power-Efficient Gathering in Sensor Information Systems

Power-Efficient Gathering in Sensor Information Systems [127] (PEGASIS) is a commonly used routing protocol that can offer significant improvements in prolonging network lifetimes compared to the LEACH protocol. Applying a chain-based approach with a singular head node, this approach means that each network node need only to forward data on to the next network node within the chain reducing the amount of data each network node has to communicate.

Principles of operation

PEGASIS takes a chain-based approach to routing packets though a network of nodes, each node has prior knowledge of the network. The protocol uses a chain-based approach with each node apart from the starting and ending node receiving a new packet; the receiving node then fuses its additional data to the data received before forwarding on to the next network node.

PEGASIS uses either a greedy algorithm graph traversal algorithm to form a chain that traverses all nodes within the sensor network or, alternatively the work suggests that a chain can also be computed by the sink node and transmitted to all the sensor nodes on the network. The chain is computed before the first round of communication takes place.

Advantages

The PEGASIS protocol enables a significant reduction in the amount of data that is directly communicated to the sink node of the sensor network. The approach opts to fuse data together, allowing a single packet to pass through all sensor nodes within the network. This approach allows each sensor to contribute its own data to the packet before passing the packet on to the next sensor node in the chain. This approach allows for all nodes to relay data using a signal transmission without the need for the transmission to be repeated by intermediary nodes to reach the sink node.

The PEGASIS approach allows for adaptations to the network, enabling all sensors to communicate data. The approach mitigates the need for multi-hop routing for a sensor node to transmit data to the sink node by making multi-hop routing a first-class citizen and the basis for the protocol. The chain-based approach allows for the protocol to react to changes in sensor node locations, enabling the sensor network to automatically configure itself once deployed and cope with the loss of sensor nodes.

Drawbacks

The PEGASIS protocol makes heavy use of data fusion throughout the routing process, while this allows for a significant reduction of data transmitted it also reduces the amount of data that is received and significantly reduces the granularity of the data received. Data fusion might include the use of processes such as mean, max and average values for data, however this limits the data analysis that can be done.

The loss of granularity within the data fusion process of PEGISIS also makes the capture of region-specific data more difficult using the protocol. It would be possible for sensors to be allocated a region and use a sensor packet that allows for the fusion of multiple regions of data, though this adds additional complexity to the sensor network and would cause some sensor nodes to transmit a larger than required data packet.

3.6.3 Threshold Sensitive Energy Efficient Sensor Network Protocol

Threshold Sensitive Energy Efficient Sensor Network Protocol [126] (TEEN) is a commonly used routing protocol that can offer significant improvements in network node lifetimes over the conventional LEACH protocol implementation by applying a threshold check to sensor data collected, only relaying sensor readings when sensor data meets certain criteria requirements. The protocol uses the LEACH protocol as a basis for routing packets between network nodes but adds a thresholding approach to sensor readings to reduce the data transmitted though the network.

Principles of operation

The TEEN routing protocol uses the same system for routing messages through a network as LEACH, electing cluster heads and relaying data through the cluster head to the sink node. The TEEN protocol adds an additional layer on top the of the LEACH protocol adding a threshold at which data from a sensor node must be transmitted to a cluster head.

The protocol uses a Hard Threshold (H_T), this threshold represents the threshold beyond which the sensor node must transmit to the cluster head. The protocol also uses a second Soft Threshold (S_T) which is used; this represents the minimum change at which the sensor node must transmit data to the cluster head.

A sensor node using the TEEN routing approach continually takes readings from the environment, the reading taken is evaluated against the H_T value defined, if the sensed data exceeds this threshold, data is transmitted to the cluster head and the sensor value (SV) stored in memory.

The sensor node will continue to take sensor readings but will not transmit unless the new reading exceeds the value of H_T and the new reading is less than $SV - S_T$ or greater than $SV + S_T$. If both conditions become true, the sensor will transmit the new value and the newly transmitted value will become the new value for SV.

Advantages

TEEN is built upon the LEACH protocol, which has been developed over the years with a range of alternatives which adapt and improve the original protocol by supporting features such as multi-hop routing. Being built upon LEACH provides TEEN with the benefit of these previous works and developments.

TEEN enables a reactive approach to be taken with sensing applications, data is relayed at regular intervals or when there is a significant change that needs to be known. This has applications in many sensor-networking areas including environmental and water quality monitoring where knowing that thresholds have been exceeded has advantages so that action can be taken as soon as possible.

Drawbacks

TEEN requires that sensors take regular readings, regardless of whether that reading will be transmitted, while this is not as power consuming as taking a reading and transmitting it there is still an associated power consumption cost that must be accounted for. A system that implements a TEEN routing approach will always consume more power than the same implementation using the same underlying LEACH implementation as more sensor readings will be taken.

The Hard Threshold and Soft Threshold within the TEEN protocol play a significant role in how many times the sensor node will transmit additional data beyond the amount that would be transmitted using the equivalent LEACH implementation. It is important that these values are set appropriately so that data is not needlessly transmitted. An incorrect or misjudged value set for either H_T or S_T could lead to a significant reduction in the lifetime of the sensor network due to the increased number of transmissions made within the network.

3.6.4 Controlled flooding

Controlled flooding routing is a simple but effective routing approach that provides a simple implementation which enables multi-hop routing. The approach ensures messages are communicated as far as possible by having each network node retransmit the message it has just received if it has not previously transmitted that message, stopping endless loops of the same message being transmitted across the network.

Principles of operation

Controlled flooding [128] is a simplistic protocol where a sensor node transmits a message with a payload including a unique message ID. Each sensor node that receives the packet checks the message ID and compares it to a list of message IDs that it has retransmitted. If the sensor node has not retransmitted the message the ID of the message is stored, and the message is retransmitted.

Controlled flooding builds upon the uncontrolled flooding routing approach, but significantly improves on one of the major weaknesses of the uncontrolled flooding approach where messages end up being transmitted endlessly between two nodes. This is because in uncontrolled flooding, messages are not checked to see if they have been previously retransmitted, but instead are automatically retransmitted regardless. This can in many cases cause close sensor nodes to repeat the same message between one another until one of the sensor nodes runs out of power.

Advantages

A controlled flooding routing approach allows for a simple implementation which can be implemented with a trivial amount of effort. The implementation requires a method of acquiring a unique message ID such as system clock and a method for storing and evaluating previous message IDs that have recently been transmitted to ensure the sensor node does not transmit the message more than once.

Controlled flooding ensures that a message propagates through to all reachable sensor nodes, including the sink node. This gives the message the best chance of reaching the sink node as in all cases where there is a viable route between the transmitting node and the sink node the message will be received eventually through the network.

Controlled flooding as a routing method enables the sensor network to handle situations such as the changing of locations of sensor nodes as well as the loss of sensor nodes. Because no node requires any knowledge of other nodes within the network there is no impact on the overall network

Drawbacks

One of the major flaws in controlled routing is the disregard that the protocol gives for power efficiency. The protocol gives the best chance of a message being received by the sink node but at the cost of propagating the message through to all sensor nodes within the network, many of which would not need to receive or broadcast the message for it to reach its desired destination, wasting significant amounts of energy with unnecessary transmissions.

The controlled flooding approach to routing also reduces the number of messages that can be transmitted through the network since every sensor node in the network plays a role in the transmission of each message. This increases the number of transmissions that are used for each sensor reading, reducing the time available for other sensors to transmit readings.

3.6.5 Comparison

There are a range of routing protocols available within wireless sensor networks, the routing protocols examined depend on being able to form links with multiple nodes to ensure the network is resilient to node failures.

The LEACH protocol is a well-established hierarchical routing approach, with adaptations such as M-LEACH to enable multi-hop routing within the network. LEACH rotates the more energy demanding role of cluster head amongst sensor nodes to balance the power consumption of nodes within the network to optimise network lifetimes.

The TEEN routing protocol builds upon the LEACH protocol, evaluating sensor readings on a regular basis to enable a more reactive sensor reading reporting approach, allowing changes in conditions to be more quickly relayed to the sink node than in the case of LEACH on its own. The TEEN protocol needs to have appropriate threshold values set to ensure that readings are only transmitted when appropriate, if not, the overall network lifetime can be significantly impacted.

The PEGISIS protocol offers an alternative to LEACH which TEEN is ultimately built upon. PEGISIS uses data fusion to reduce the amount of data that needs to be transmitted by each sensor node. The approach requires that a sensor node receives a message, fuses its data into the received packet before forwarding the packet onto the next sensor node in the chain. PEGIS offers improvements of network lifetimes over LEACH[129] at the cost of data granularity due to the data fusion process.

Controlled flooding is a very simple routing approach, that enables a simple implementation at the expense of power efficiency. Due to sensor nodes having no required knowledge of other sensor nodes within the network the protocol is flexible and able to easily cope with changes in the locations of sensor nodes and with the death of sensor nodes without any additional overhead. Controlled flooding does have a significant overhead of unnecessary transmissions due to all sensor nodes receiving and then retransmitting each data packet.

3.7 Summary

The chapter has identified and examined the key communication methods used within an underwater wireless sensor network, examining important factors for communication including alignment, This chapter has examined considerations that are important to the creation of an underwater wireless sensor-networking platform. highlighting the impacts of antenna design, examining existing implementations of underwater communications using a variety of designs covering the data rates and distances achieved by the works.

The chapter has also explored the way in which electromagnetic waves propagate in an underwater environment. The section shows the importance of ensuring that changes in frequency due to the change in permittivity and permeability are accounted for to ensure the antenna functions properly. The section has also examined the loss experienced by electromagnetic waves due to attenuation and refractive loss.

The chapter has also examined modulation techniques that can be applied to digital communications to encode binary data into a carrier wave. The section also examined factors that impact on communications including how frequency can impact the maximum achievable data rates. The chapter also identified the role that signal to noise plays in the maximum achievable data rate and how lowering the baud rate can positively impact on the achieved communication distances.

The chapter has also covered routing protocols commonly used in conventional wireless sensor networks, examining the fundamental principles of the protocol as well as the advantages and disadvantages of them. The section showed that there are several ways routing protocols can be used to improve the reliability of the network as a whole and ensure the maximum network lifetime.

4 Underwater wireless sensor networks using radio frequency

4.1 Introduction

This chapter presents the use of radio frequency to form an Underwater Wireless Sensor Network (UWSN), examining factors such as conductivity and baud rate's impact on communication distances. This chapter presents a first implementation of a UWSN using a simple broadcast routing protocol to route packets across multiple network nodes.

4.2 Considerations for processor units

This section covers considerations for microcontrollers with a microcontroller having the CPU, RAM, ROM and other peripherals to operate. This section will examine a range of properties that are important in selection of a microcontroller, this section will only examine microcontrollers which are designed for specific tasks and not microprocessors which are designed for unspecific tasks.

Power consumption is an important factor for microcontrollers, as this will have a direct impact on the overall power consumption of the platform. Many microcontrollers offer a range of power saving features including low power sleep states and reducing the clock speed of the processor to reduce power usage.

Technology maturity is another important consideration, during development the ease of access to development and support materials is often related to the maturity of the technology in use, with established technologies offering a range of supporting materials.

Connectivity is vital to ensure that the developed platform can connect to a variety of sensors, with a range of communication interfaces available including SPI, I2C, Serial. A microcontroller providing a range of connectivity options enables a wider range of sensors to be supported enabling a more flexible platform.

Cost is an important factor in the selection of a microcontroller, for a sensor platform to be viable it must be cost-effective to deploy in large numbers and in conditions where not all sensor nodes may be recovered.

Complexity of the overall circuity required to support the processor should also be considered, not only does this help to reduce the overall cost for a sensor node but also reduces the complexity of the required circuit design.

4.2.1 Atmel 328

The Atmel 328 microcontroller is popular within industry and with hobbyists with support from clock speeds of up to 20 MHz at 5 V although most common configurations use an 8 MHz clock speed at 3.3 V or 16 MHz at 5V. The Atmel 328 offers a range of power saving modes enabling current consumption of 40 µA in sleep state and 5.2 mA when active.

The Atmel 328p offers a range of connectivity including support for one SPI, one I2C and a UART interface. The Ateml 328p microcontroller also has support for 23 programmable Input Output lines.

The Atmel 328p has a wide range of development materials available both directly from the manufacturer as well as from the wider community. The Atmel 328p is also used on a wide variety of development boards including the Arduino Uno as shown in Figure 4-1 and the Adafruit Metro enabling rapid development of prototypes.

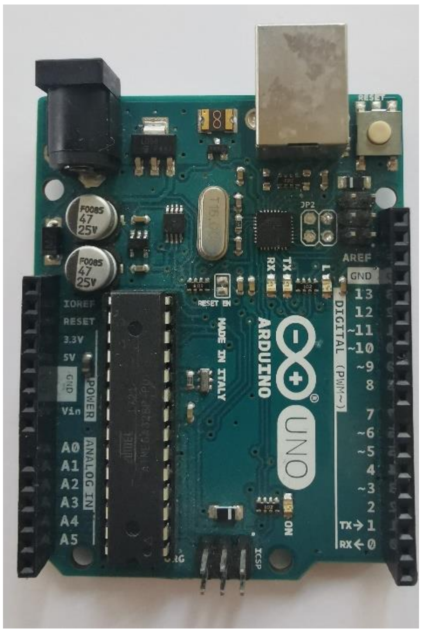

Figure 4-1 Arduino Uno development board

4.2.2 Microchip PIC24F

The Microchip PIC24F is a 16 bit microcontroller supporting speeds of up to 32 MHz with support for 32Kb of SRAM. The PIC24F operates with a supply voltage of between 2.0V – 3.6 V with a current draw of 6.3 mA when active and 1.16 mA sleep state and 300 nA in a sleep state.

The Microchip PIC24F offers a range of connectivity options including support for three SPI interfaces, 2 I2C interfaces and 4 UART interfaces. The PIC24F also provides support for up to 85 I/O pins including 24 analogue inputs with 12-bit resolution.

The PIC24F is popular within the Microchip community, there are a range of forums and example code. Microchip also provide supporting documentation as well as a range of development boards such as the PIC24F Curiosity development board shown in Figure 4-2 available to enable rapid prototyping.

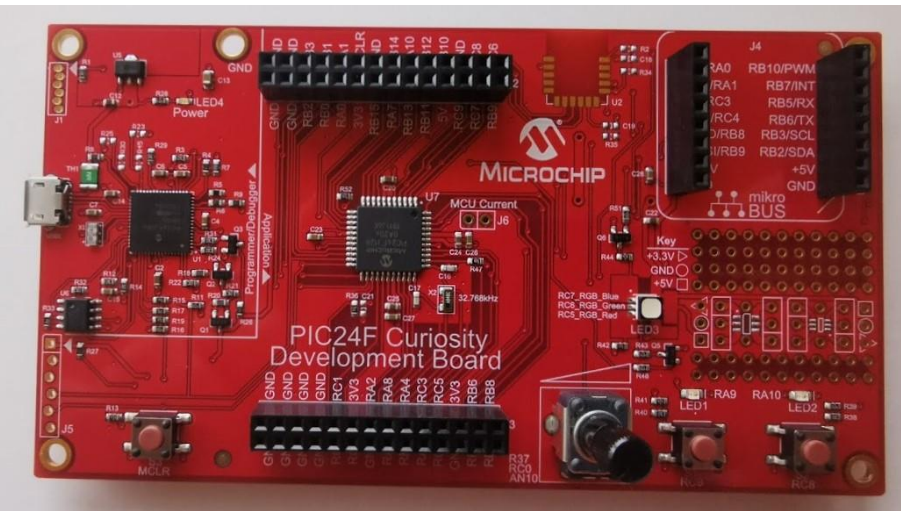

Figure 4-2 A PIC24F Curiosity development board

4.2.3 STM32F0

The STM32F0 is a 32-bit entry level microcontroller with a range of variances including low-voltage and value variants. The F0 series offers a range of clock speeds up to 48 MHz with 8 Kbytes of SRAM. The STM32F0 supports voltages of between 2.0 V – 3.6 V with a typical sleep current consumption of 2.8 mA at a clock speed of 48 MHz and a current of 12.0 mA at 48 MHz when executing code from memory.

The STM32F0 series supports a range of connectivity including support for up to I2C interfaces with support for fast mode. The microcontroller also supports two USART with support for one ISO7816 interface. The controller also provides support for two SPI interfaces.

The STM32F0 processor series is a widely supported microcontroller, used in a range of connected devices and a wide range of materials available online. The STM32F0 also has a variety of development boards such as the STM32F072B-DISCO shown in Figure 4-3 which can be used to create initial prototypes before continuing to develop further iterations.

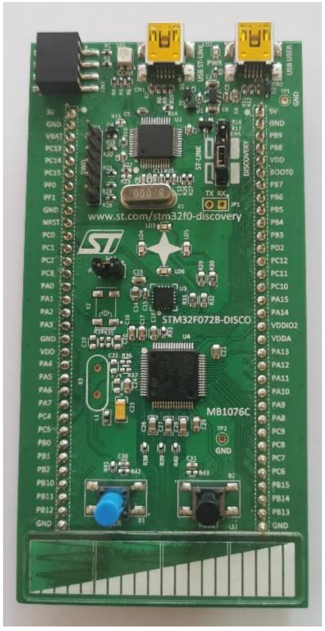

Figure 4-3 STM32F077B-DISCO development board

4.2.4 Intel Edison

The Intel Edison is a computer module shown in Figure 4-4 targeted for use in the internet of things and wearable devices. The board uses a duel core Intel Atom processor at 500 MHz combined with an intel Quark microcontroller running at 100 MHz combined with 1 GB of RAM. The module also includes a 4 GB of internal storage and has integrated Bluetooth and WiFi integration.

The Intel Edison module offers a range of interfaces including a dedicated SD card interface, two UART controllers, 2 I2C controllers, 1 SPI controller with 2 chip selects and a I2S controller. The module also supports an additional 14 general purpose input and output pins.

The Intel Edison was discontinued in September 2017, while there remains a community surrounding the Intel Edison they are no longer manufactured and will not be developed further. There is however a wide range of example code and documentation that still supports development using this platform. Figure 4-4 shows an example of the Intel Edison evaluation board.

Figure 4-4 Intel Edison development board

4.2.5 Conclusion

There are a variety of microcontrollers all offering a range of features including low power usage and a range of connectivity options. Table 4-1 shows a comparison of key features for each of the discussed development boards. The microcontrollers examined offer their own distinct features with products such as the Microchip PIC24F offering low sleep currents while other microcontrollers such as the STM32F0 offer increased clock speeds.

Table 4-1 A comparison of features of potential development boards

Microcontroller	Sleep current	Wake current	Clock speed	SPI	I2C	UART	Cost (£)
Atmel 328P	1.16 mA	5.2 mA	20 MHz	Y	Y	Y	1.59
Microchip PIC24F	300 nA	1.16 mA	32 MHz	Y	Y	Y	1.68
STM32F0	2.8 mA	12.0 mA	48 MHz	Y	Y	Y	1.43
Intel Edison	N.A	N.A	1 GHz	Y	Y	Y	39.92

Given the intended use of the microcontroller there is no need for a high clock speed due to the limited amount of processing that each sensor node will undertake, with any high levels of data processing being completed once data is transmitted to its final location. This means that relatively low clock speed of several MHz is more than enough for the intended purposes.

Documentation and support are important factors with products such as the Intel Edison being less useful due to the lack of ongoing support from the manufacture and limited community support compared to other offerings such as the STM32F0 and Atmel 328P.

Connectivity is important to ensure that the developed platform can communicate with a range of devices, allowing the usage of the platform to be expanded to include additional sensors. All the examined microcontrollers support a range of communication methods including SPI, I2C and UART which are commonly used.

The Atmel 328P offers a reasonable clock speed for the intended application while providing all the required connectivity options. The cost of the Ateml 328P also makes the microcontroller an attractive option, allowing for the development of a cost-effective solution. The Atmel 328P also has large support both from the manufacture and with a supportive community, providing a range of example applications and development boards to get started with.

Given the connectivity offered by the Atmel 328P enables a range of sensors and other devices including a transceiver to be connected. The Atmel 328p also offers low power consumption figures during wake and sleep states enabling a power efficient platform. The Atmel 328P is also popular with developers and provides a range of support materials and development platforms enabling rapid prototyping. The features and functionality offered by the Atmel 328P make it an excellent option to develop an underwater wireless sensor-networking platform.

4.3 Routing approach

Previous chapters have shown that there are a range of routing approaches available for use within wireless sensor networks. One protocol discussed was LEACH and derivatives, LEACH is an effective clustering protocol that enables power efficient routing in large scale wireless sensor networks, due to the clustering approach used within the protocol the effects of LEACH are significantly enhanced in larger scale networks for longer term deployments. There is also a certain amount of complexity involved in the implementation of the LEACH protocol ensuring that cluster heads are correctly rotated, and all nodes join a cluster.

PEGISIS is another routing protocol used within wireless sensor networks, the protocol is another example of using multiple nodes within the network to relay a message to the sink node of the network. As with LEACH the implementation of PEGISIS has a significantly higher level of complexity to implement compared to other routing approaches.

Controlled flooding is another protocol used within wireless sensor networks, an improved version of flooding in which every node resends every message it receives which can cause messages to be sent repeatedly throughout the network. Controlled flooding ensures that each node only retransmits a message once, ensuring messages are not repeatedly transmitted. Controlled flooding allows a simple implementation and ensures that messages are transmitted to all sensor nodes within the network ensuring that where a path from the sensor node to the sink node exists the message with be received by the sink node.

Given the initial applications intended to use a small number of sensor nodes more complex routing protocols provide little benefit; due to the number of network nodes and the distribution of them it is unlikely that the benefits of protocols such as LEACH or PEGISIS will be an advantage. The additional complexity of these protocols could also inhibit experiments by introducing more points of failure.

TEEN is another protocol used within wireless sensor networks; the approach is based upon the LEACH protocol but also applies a threshold for when readings are sent. As well as the issues highlighted with the LEACH and PEGISIS protocols TEEN suffers from the additional issue of not continually transmitting data, while this approach is useful in a deployed wireless sensor network it inhibits the ability to perform testing as readings would have to change on a regular basis for data to be transmitted. This factor introduces an element of doubt into if the network is functioning correctly or not.

Given the simplicity of implementing controlled flooding compared to other approaches such as TEEN, LEACH and PEGISIS it makes controlled flooding an attractive approach for initial development, with a simpler routing approach ensuring the implementation is functional. The use of controlled flooding could be enhanced using a similar approach to only transmit data when it falls within a given threshold or when data has changed by a specified amount.

The advantage of controlled flooding being a simple approach to implement, combined with the fact that other approaches examined, such as LEACH, PEGISIS and TEEN, only show true advantages in prolonging network lifetimes due to the size of the network, make controlled flooding a strong option. Using controlled flooding enables a simple implementation of a mesh wireless sensor network, reducing the change for issues during initial testing. Controlled flooding is the best suited to developing a wireless sensor network, while enabling an alternative to be implemented at a later point in development.

4.4 Antenna design

Antenna design plays an important role in how signals will be transmitted and received, 3.2 Antenna design covers a number of antenna designs that have been used in previous experimental works within UWSNs . The experimental works undertaken used a bowtie antenna, this design was selected due to the benefits of functionality in and out of water, due to the ultrawide bandwidth supported by the antenna, as well as the omnidirectional design reducing the need for alignment.

Experimental works used a bowtie antenna based on the design presented by A. Abdou et al[96], Figure 4-5 provides a design of the antenna used with measurements in mm. The presented antenna is designed to operate at the 433 MHz ISM (Industrial Scientific and Medical frequency). The antenna design offers advantages over other designs including a wideband design that enables it to function in air and water.

The bowtie antenna wideband design allows the antenna to function over a wider range of frequencies. This enables the antenna to operate in both air and water. As shown by Maxwell's equations the frequency of a signal will differ between air and water based on the permeability and permittivity. The wide band enables the antenna to function in water with a range of conductivities which impact on the permeability and permittivity of the water.

The bowtie antenna design has an omnidirectional radiation pattern reducing the need for alignment between deployed sensor nodes in a deployment environment. This omnidirectional design also enables a single sensor node to communicate with multiple sensor nodes enabling a mesh network topology, rather than just point-to-point communication links. A mesh network topology enables the network to become more resilient helping to eliminate single points of failure.

The antenna design selected has also been shown to work in water, with previous experimental works in real-world environments achieving distances of 30m. The combination of advantages of the bowtie antenna design are its wide bandwidth nature, the omni directional design with the specific designs used by A. Abdou et al[96] which uses the 433 MHz ISM frequency and it has been tested in real-world environments.

The bowtie antenna was selected for several reasons, firstly it has been shown that the bowtie antenna can operate at the 433 MHz frequency with distances of 30m showing a signal strength that could be viable for communication. Additionally, the bowtie antenna design is a broadband antenna design, allowing it to operate across a wider frequency range, this enables the antenna to function effectively in and out of water. The size of the antenna plays an important role, due to the design of the bowtie antenna it is possible to create a shorter antenna comparted to the dipole equivalent antenna. A further reason for the use of a bowtie antenna design is due to the omni-directional nature of the antenna, making it ideal for forming a broadcast network.

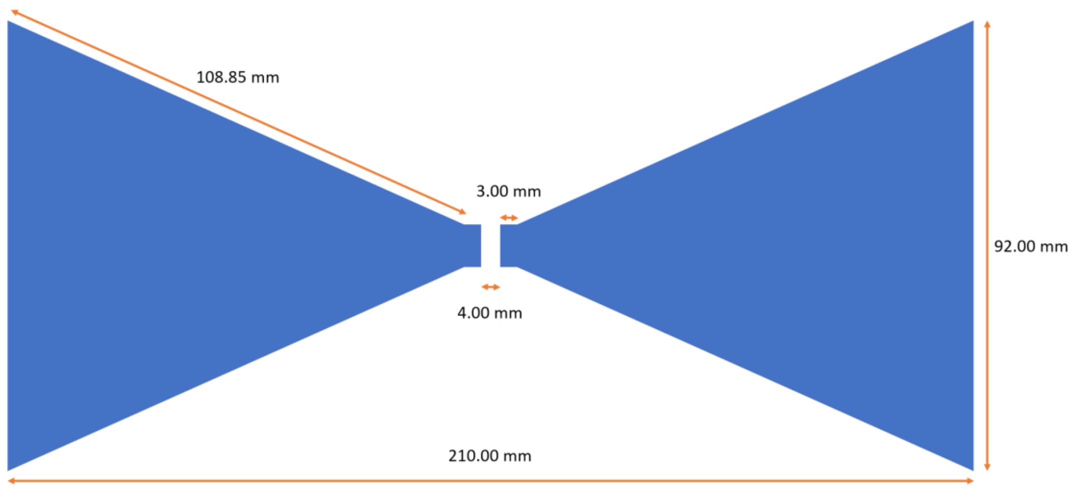

Figure 4-5 Design and measurements of the bowtie antenna used throughout experimental works based on previous works by A Abdou et al

The antennas used throughout the experimental works and based on the designs presented in Figure 4-5 were created using a chemical etching process on FR4 board with thickness of 1.5 mm, a layer of copper with a thickness of 35 µm was used for the planes of the antenna. The antennas were coated with a top layer of epoxy resin approximately 0.8 mm thick to protect the surface of the antenna from abrasion and damage and to ensure that the planes of the antenna were not able to short. Figure 4-6 shows the layers and thicknesses of the antenna used. The antenna was created in Autodesk Eagle and manufactured using a chemical photo etching process.

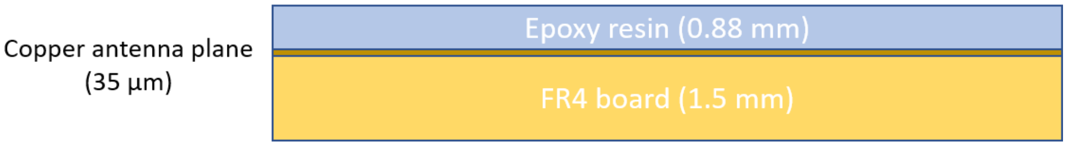

Figure 4-6 A representation of the layers within a cross section of the antenna used during experimental works

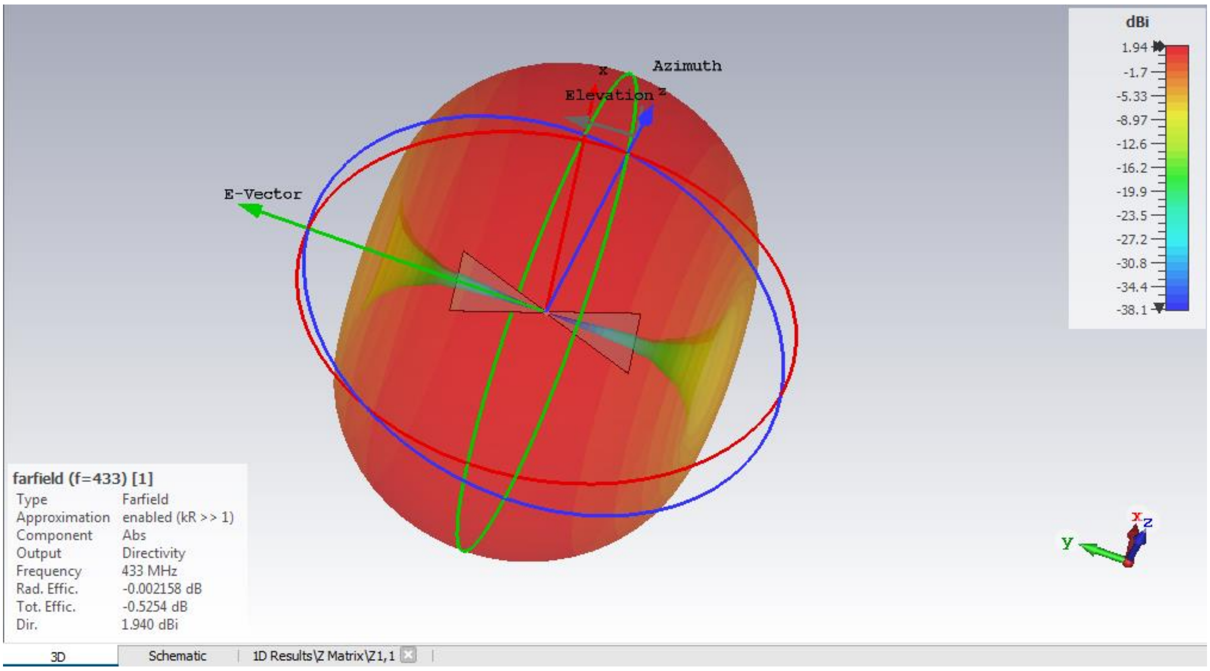

Figure 4-7 A simulation of the bowtie antenna's radiation pattern

Figure 4-7 Shows the simulated radiation pattern of the bowtie antenna presented in Figure 4-5. The radiation pattern shows a strong radiation pattern in all directions, this radiation pattern allows the signal to radiate well in all directions removing the need for alignment between sensor nodes.

4.5 Communication circuit design

Experimental works used a common hardware platform based upon a number of iterations of the platform presented in Figure 4-8, the platform used an Atmel 328p micro controller. The microcontroller was used to implement software to control the various other components of the platform including sensor readings, data transmission using the platforms transceiver and providing serial output for data collection.

The Atmel 328p micro controller was selected for a number of reasons, the first is the flexibility of the controller itself, it offers support for I2C, SPI, Serial communications and UART . Using the Atmel328P as a base allowed for the platform to develop enabling the easy integration of additional peripherals as and when required.

A **B** **C**

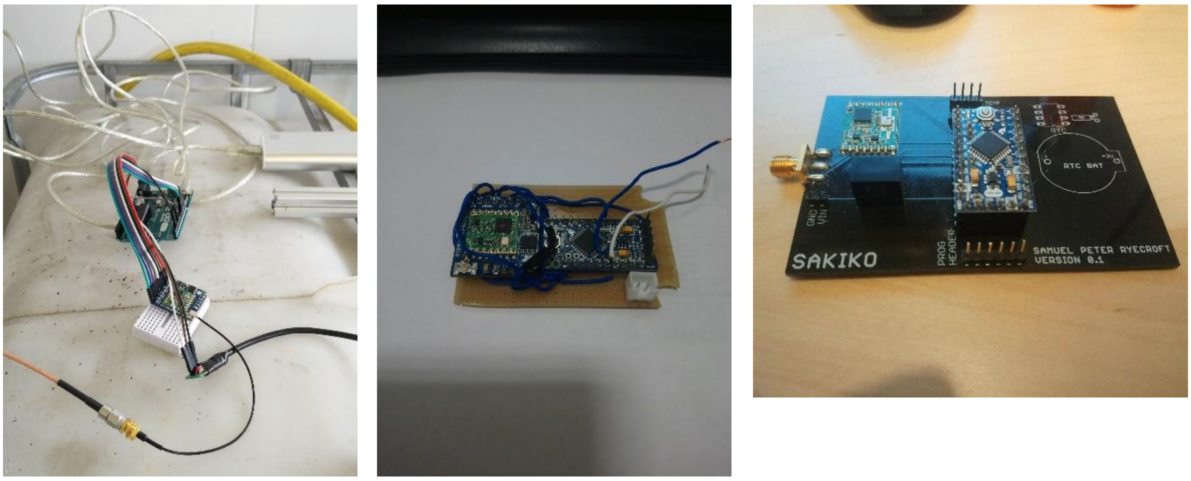

Figure 4-8 Three generations of prototype communication platform: A) First generation solderless breadboard prototype, B) Second solder breadboard prototype, C) Third printed circuit board prototype

The Atmel328P micro controller also offers a range of power saving features and modes, enabling a range of sleep states to be used by the micro controller to conserve power. The conservation of power is vital for any wireless sensor platform to ensure the maximum possible lifetime, for large amounts of the deployment the platform will not be active, and it is therefore beneficial to be able to use hardware sleep modes to reduce power consumption as much as possible.

The platform used a HopeRF RFM69HW transceiver module for use with the 433 MHz ISM frequency, the transceiver module is capable of a transmission power output of +20 dBm and a sensitivity of down to -120 dBm. The transceiver module was connected to the SPI bus of the microcontroller to enable interaction between the microcontroller and the transceiver.

The first generation of the platform was used for initial feasibility experimentation ensuring that the hardware was fully compatible and that the associated libraries were functional and enabled interaction between the transceiver and the microcontroller. The second generation of the platform was used for further testing and was developed to ensure that the connected components did not come loose during movement, storage and during testing, this soldered breadboard prototype was used to test additional sensors that could be connected to the platform by enabling some flexibility to the platform while keeping key components securely connected. The final printed circuit board prototype was developed once all the hardware components including sensors had been identified and selected. This enabled several prototype boards to be produced while remaining consistent ensuring that antenna connections were as similar as possible across all prototypes.

The transceiver supports a variety of modulation techniques: FSK (Frequency Shift Keying), GFSK (Gaussian Frequency Shift Keying), MSK (Minimum Shift Keying), GMSK (Gaussian Minimum Shift Keying) and OOK (On Off Keying) enabling communication speeds of 300 kb/s using more advanced modulation techniques. OOK is limited in data rates to 25 kb/s.

The HopeRF module includes a packet mode enabling automatic processing of tasks such as preamble and sync word generation, CRC calculation and checking, address filtering and the handling of AES encryption and decryption through dedicated hardware.

The platform was connected to two sensors using the I^2C bus of the microcontroller. The first sensor was a TSYS01 temperature sensor which enables temperature readings to be taken at a resolution of 24 bits with an accuracy of ±0.1°C. The second sensor was a MS5837-30BA depth sensor capable of taking depth readings of up to 300 M with a resolution of 2 mm. The sensors were used to collect real-time data to transmit to other sensor nodes.

4.6 Radio frequency in conductive environments

4.6.1 Introduction

Initial experimentation was undertaken in a controlled lab environment to establish the impact on signal strength at distances of between 10 cm and 50 cm at a variety of conductivities. The experimentation examined how these factors impact on how signal strength decays in a variety of conductivities and how these factors are likely to impact in real-world environments.

4.6.2 Method

The experiment used two bowtie antennas, one antenna was connected to the developed experimental hardware platform described in 4.5 Communication circuit design. The platform was programmed to transmit a continuous signal at the 433 MHz frequency. A second bowtie antenna was connected to a Hameg HMS 3000 spectrum Analyzer set-up to monitor the 433 MHz frequency.

Both antennas were suspended from a wooden rod suspended above a water tank measuring 50 cm by 40 cm by 60 cm. The tank was filled with water so that the suspended antennas were fully submerged. The first antenna connected to the experimental platform was placed 5 cm away from the edge of the tank with the antenna facing forward. The second antenna connected to the spectrum analyser was placed at a predefined distance away from the transmitting antenna. Figure 4-9 shows a bird's eye view of the experimental set-up used.

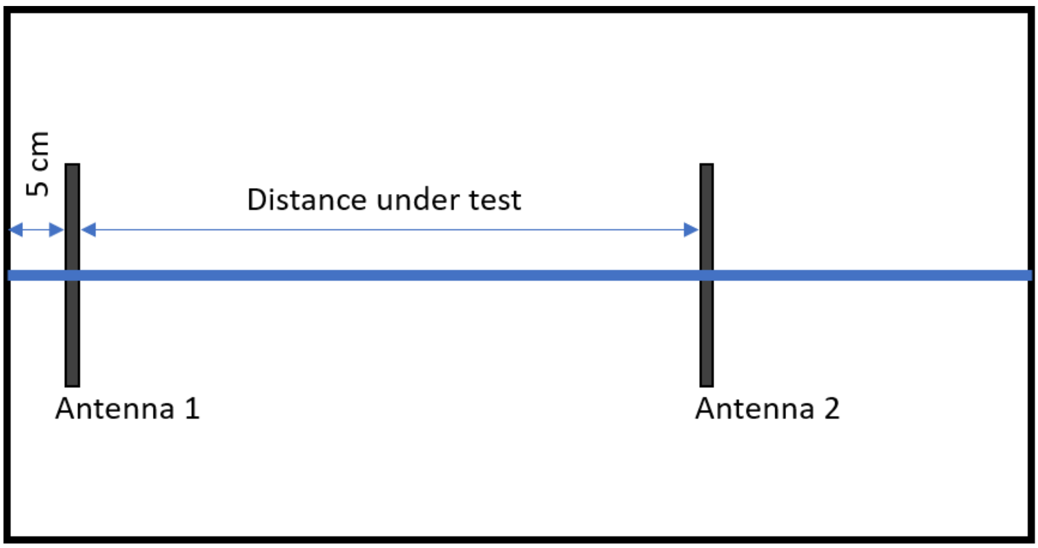

Figure 4-9 A bird's eye view of the water tank setup with the antennas in position

The conductivity of the water within the tank was increased using sea salt, while sea salt is not likely to be present in large quantities in reservoirs, it is present in seawater and brackish water and is the main cause of the increased conductivity of these environments. The use of sea salts does not have a large impact on the overall PH of water due to the extremely weak basic nature of the CL^- bond.

The water conductivity was measured using a Hanna Instruments 98188 conductivity meter, the initial tank water had a conductivity of 0.2 ms/cm. The conductivity of the water within the tank was raised using sea salts. The conductivity of a 1 molar solution of sea salt was used to calculate the required quality of sea salts to raise the conductivity of 1 litre of water by 0.1 ms/cm. The amount of sea salt required was then used to calculate the amount of salt required to raise the water tank to the required conductivity based on the amount of water in the tank.

For each reading taken the second antenna connected to the spectrum analyser was moved along the wooden rod to the desired distance. The antennas and tank were then left to settle for a minimum of 30 seconds. The spectrum analyser was then observed, and a signal strength reading taken. The antenna was then moved to the next desired distance and left to settle before the next reading was taken. This process was repeated 10 times for each distance for each conductivity rating.

Readings were taken for distances of 10 cm to 50 cm in 5 cm intervals, readings were taken from between 10 cm to 50 cm due to the dimensions of tank used in experimental works, with 5 cm of separation being used between readings to enable a range of readings to be obtained in a limited space while providing enough separation between readings.

Readings were taken for the following conductivities: 0.4 ms/cm, 0.6 ms/cm, 0.8 ms/cm, 1.0 ms/cm and 5.2 ms/cm. The conductivities used represented potential conductivities of surface and waste water with an additional higher conductivity of 5.2 ms/cm being used to represent the potential conductivity of sea water to examine how much of an impact high conductivities would have on potential communication distances.

4.6.3 Results and Discussion

Table 4-2 contains the mean average results from experimentation for each distance and conductivity tested. The results within the table indicate that the signal strength generally decreases as distance increases. The results also indicate a general decrease in signal strength as conductivity increases. The results show significant additional losses being observed in the case of testing at a conductivity of 5.2 ms/cm.

Table 4-2 A table of mean averages for the distances and conductivities under test

Distance (cm)	0.4 ms/cm	0.6 ms/cm	0.8 ms/cm	1.0 ms/cm	5.2 ms/cm
10	-27.38	-24.34	-25.15	-24.90	-38.74
15	-25.68	-22.06	-23.50	-23.70	-33.61
20	-28.33	-26.14	-28.09	-28.95	-46.27
25	-30.00	-28.45	-29.44	-29.97	-47.64
30	-33.87	-30.26	-32.20	-34.13	-51.57
35	-41.40	-34.21	-35.51	-36.22	-48.50
40	-41.53	-38.99	-38.01	-41.87	-50.58
45	-41.38	-44.44	-42.10	-43.53	-51.70
50	-46.02	-44.98	-43.87	-44.48	-50.69

The results indicate an increase in attenuation of the signal as the conductivity of the water increases. This agrees with the literature which shows the relationship between conductivity of the medium where good conductors lead to an increase in attenuation of signal.

The results and experimental set-up suggest that it is likely that the results are observations of near field power rather than far field power, with near field power often observed at distances of up to 1 wavelength. Near field power decays at a significantly increased rate compared to far field power.

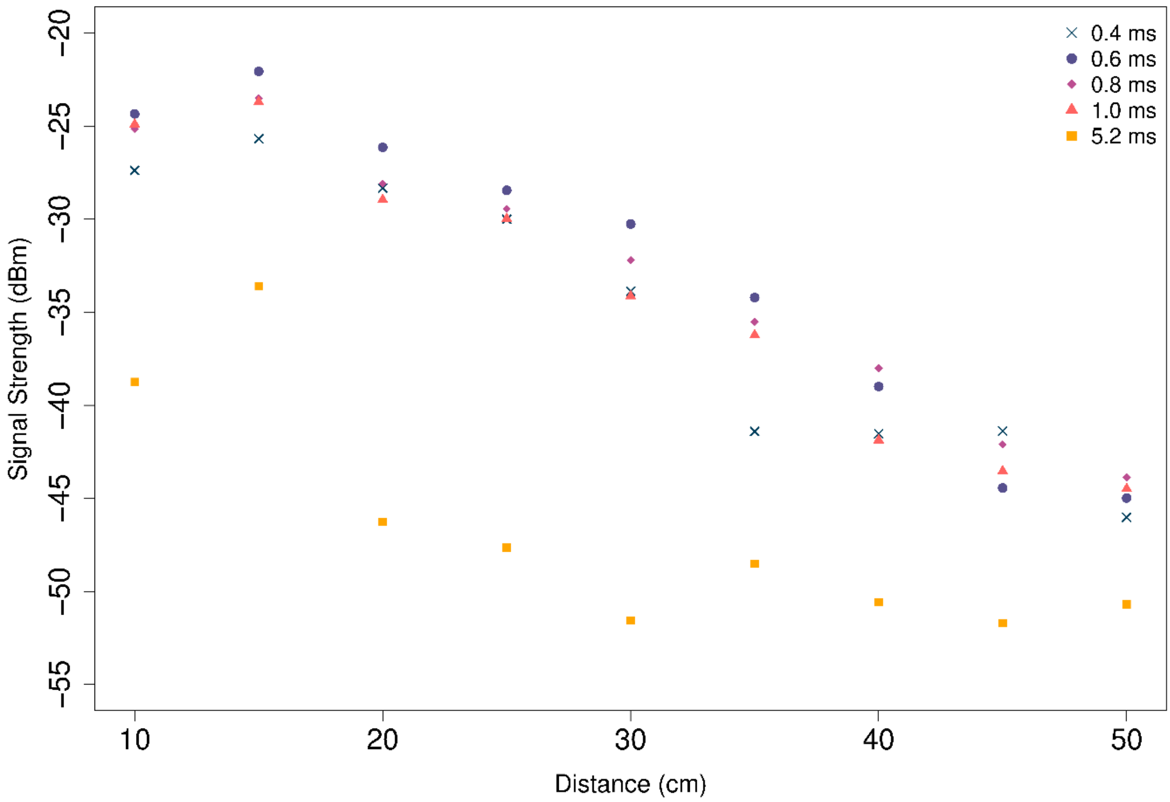

Figure 4-10 A plot of results from experimental works of signal strength against distance

Figure 4-10 shows a plot of the average data collected form lab-based experimentation, the data presented had a maximum standard deviation of +/- 3 dBm and in most cases was within +/- 1 dBm. The plot shows a plot of signal strength against distance for each of the conductivities tested.

The results show that signals at 10 cm are weaker than at 15 cm, the signal strength then drops again at 20 cm. There are several possibilities for this result, one explanation is the geometry of the tank could result in either a weaker signal being detected at the 10 cm reading point due to reflected waves from the end of the tank interfering with the signal at this distance, meaning that the readings taken at 15 cm are a true reflection of signal strength.

4.6.4 Conclusion

The results from lab-based experimentation supports the use of RF in underwater environments, the results indicate that there is a quicker loss of signal strength although this does begin to level off at larger distances.

The distances explored within these experiments represent very short distances, due to the nature of the tests in a water tank it is possible that the water tank geometry may play a part in the results found. It is likely that the results are observations of near field power rather than far field power.

The initial findings support the use of communication in underwater environments, the results also show the need for further experimentation in larger environments where larger distances of separation can be achieved.

4.7 Communication distances in Liverpool Leeds canal

4.7.1 Introduction

Initial lab-based experimentation presented in 4.6 Radio frequency in conductive environments showed the feasibility of short-range communications in a variety of conductivities. The experimental works highlighted the need for additional experimentation in larger environments. The experimentation also set out to test feasible data rates and the impact that data rates had on the communication distances achieved in real-world environments.

4.7.2 Method

Experimentation was carried out in a section of the Leeds Liverpool canal managed by the Canal and River Trust. The work used the experimental communication platform described in 4.5 Communication circuit design and bowtie antennas covered in 4.4 Antenna design.

One of the platforms and sensors was mounted in a watertight housing manufactured by Blue Robotics. A bowtie antenna was mounted to the outside of the housing, secured by cable ties, and connected to the platform though a cable penetrator into the housing. Power and control commands were sent to the platform using an umbilical cord connected to the platform and laptop computer. The housing was attached to a vinyl coated mushroom anchor to weigh the housing down, Figure 4-11 shows the first platform submerged and in position.

Figure 4-11 A submerged sensor node in the Leeds Liverpool canal

A second platform was connected directly to an antenna without a housing, the antenna was submerged by attaching it to a vinyl coated mushroom anchor to ensure that the antenna remained in position. The second platform was connected to a mobile phone to provide power to the platform and facilitate the configuration of the desired baud rate. Figure 4-12 Shows the experimental setup with the second node located with by the buoy.

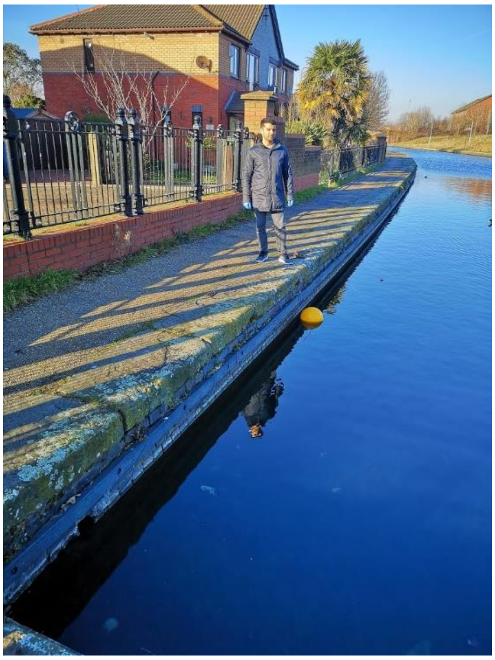

Figure 4-12 The experimental set-up used for taking readings during fieldwork in the Leeds Liverpool canal

The first platform was used to transmit data from the connected sensors to the second platform node. The platform would first be set to a specified baud rate, once a baud rate was set the first platform was programmed to send a heartbeat message at intervals of 1 second. The heartbeat message consisted of the temperature and depth readings from the connected sensors and placed in a JSON encoded message to represent an example network communication payload.

The second platform was used to receive data from the first platform, the second platform would be set to receive at the same baud rate as the first platform. The software would then check for newly received messages, if a new message had been received the software would output the JSON payload as well as the RSSI for the received packet.

Immediately before undertaking the fieldwork the conductivity of the water was measured as 0.51 mS/cm using a Hanna Instruments HI933000 conductivity meter. Throughout experimental works the temperature was captured using the TSYS01 temperature sensors integrated into the platform with an average temperature of 6.1 $^{\circ}$C.

During experimentation the first platform was placed at one end of the testing area, the platform remained in position throughout the experiment. The first platform was then set to the desired baud rate for the experiment. The second platform was then placed at 1-metre intervals away from the first platform.

The second platform was set to the correct data rate, messages were then saved in a data capture file. The second platform was left for 30 seconds after which readings were taking. The platform was then moved at 1-metre intervals until the heartbeat message was no longer received. A 1-meter interval was used as provided a significant unit of progression while still demonstrating the differences in communication distances between the two data rates used within the experimental works. The process was repeated three times for each distance and data rate tested.

The experiment undertaken used two data rates, a higher data rate of 25 kbps which was selected as it is representative of communication rates that can be used in air without issue at reasonable distances. A second lower data rate of 1.2 kbps was also used to examine the impact that a slower data rate would have on communication distances, this data rate was selected as it remains comparable to data rates achieved by acoustic communication in underwater communications and remains a useful data rate within wireless sensor networks.

4.7.3 Results and Discussion

Table 4-3 shows the mean average readings for each of the two data rates at the distances tested. The table shows that there is a clear and significant difference between the distances achievable with a data rate of 1.2 kbps achieving an additional 2 metres distance of separation functioning at up to 7 metres away, compared to 25 kbps which was capable of communication at distances of up to 5 metres.

Table 4-3 Mean average of signal strength for 1.2 kbps and 25 kbps data rates and distance

Distance (m)	1.2 kbps average	25 kbps average	Overall average
1	49.00	46.00	47.50
2	69.33	72.00	70.67
3	87.33	88.33	87.83
4	87.67	94.00	90.83
5	107.00	106.33	106.67
6	109.67	N.A	N.A
7	110.33	N.A	N.A

The results indicate a broadly similar rate of loss in received signal strength which is to be expected. These results indicate that the reduction is due to the difficulty in being able to distinguish the transmitted signal from background noise when higher data rates are used, this indicates that the results follow the principle of the Shannon-Hartley theorem[123] which states that as signal strength decreases the baud rate must also be decreased.

The results indicate that baud rates have a significant impact on the maximum distance that can be achieved for point-to-point communications using radio frequency in an underwater environment. The results show that using a 1.2 kbps data rate enabled a 40% increase in communication distances achievable compared to 25 kbps.

The results indicate that a trade-off can be made between communication distance and the link speed that can be established. This indicates that an approach similar to link speed negotiation in WiFi could be used to enable an effective solution that would enable links to dynamically adapt to conditions enabling the best link speed to be used between sensor nodes within an underwater wireless sensor network.

The work establishes that communications can be achieved using the 433 MHz frequency to transmit data in an underwater environment. The experiments highlight that 25 kbps links can be established at distances of up to 5 metres. The experiments also show that further distances of up to 7 metres can be established with a link speed of 1.2 kbps.

Communication distances of 7 metres offer significant improvements over previous works such as the work produced by Lloret et al[130] which achieved significantly higher data rates of 11 mbps although this was only achieved at distances of up to 20 cm. The works were able to achieve significantly improved communication speeds over the distances achieved in these experimental works, the use of the 2.4 GHz frequency enabled high speed data rates but at the cost of communication distances.

Other works such as that produced by Rubino et al[98]. used the 868.3 MHz frequency in a controlled water tank with a ¼ λ monopole antenna design. The results established an effective communication rate of 3.9 KB/s using the 868.3 MHz frequency, in comparison, the maximum data rate used in this study was equivalent to 3.125 KB/s.

There are existing technologies in underwater communications such as the Bluecom 200 series[80] which offers data rates of 5 Mbps at distances of up to 5 metres. The results presented show significantly reduced data rates but provide slightly improved distances. The proposed use of RF does overcome some of the issues of using optical communications including the need for alignment between nodes, the lack of support for communication with multiple nodes and the reliance on optical properties of water within the deployment environment.

Other commercial products for underwater communications such as the Teledyne marine 900[87] series provide significantly better communication distances of between 2 Km and 6 Km. The data rates presented within these results show a significant improvement over the data rates offered by products such as the Teledyne marine 900 series which provides data rates of between 140 bps and 15,260 bps in comparison to the baud rates of 1,200 bps to 25,000 bps achieved in experimental works presented.

It is possible that transmission distances could be further improved with alterations to the hardware platform used during experimentation, either through the amplification of the output signal or the use of a high sensitivity receiver module. Alterations could also be made to improve distances achieved using a lower data rate, which has the potential to improve distances further.

4.7.4 Conclusion

The results from experimentation support the use of RF communication in real-world underwater environments, enabling data rates of 25 Kb/s at distances of 5 metres and at distances of 7 m with a data rate of 1.2 Kb/s. A communication distance of 7 metres could be applied to some applications of underwater wireless sensor networks with the opportunity to extend communications between sensor nodes as well as improve overall coverage using a multi-hop routing.

4.8 A wireless broadcast network in a reservoir

4.8.1 Introduction

Hurleston reservoir is a water catchment site for United Utilities with a treatment works located next to the site for the supply of drinking water The reservoir is located north-west of Nantwich, Cheshire, United Kingdom and is fed from the Llangollen Canal which is in turn fed from the River Dee at Llantysilio. The reservoir was enlarged in 1959 and has an estimated capacity of 85 million gallons of water, Figure 4-13 shows an image of Hurleston reservoir.

Figure 4-13 Hurleston reservoir

First field trials in the Leeds Liverpool canal supported the use of RF communications in underwater environments, explored in experimental works conducted in real-world environments. The fieldwork showed that communication distances were impacted by the baud rate used, following the Shannon Hartley principle which states that as signal strength decreases baud rates must also decrease. Previous lab based experimental works showed that conductivity can significantly impact on the attenuation of signals in water.

The results from previous works show that distances of up to 7 metres could be achieved in real-world environments. Based on these findings it is important to explore if RF communications could be used to create an underwater wireless sensor network using a multi hop approach to extend communication distances though using an intermediary sensor node.

4.8.2 Method

Experimentation was undertaken at Hurleston reservoir, a water catchment area for United Utilities. The experiment used three sensor nodes to form a three-node underwater wireless sensor network. The experiments made use of the platforms described in 4.5 Communication circuit design and the antennas coved in 4.4 Antenna design.

The first platform was placed in a static location at one edge of the reservoir, in water deep enough to ensure the platform was fully submerged. The sensors and circuitry were placed in watertight enclosure constructed from acrylic tube with aluminium end caps, the antenna was mounted on the outside of the enclosure with the face of the antenna facing into the reservoir. The circuitry was connected to a Raspberry Pi computer using an umbilical cable connected to the circuitry in the housing. The housing was attached to a vinyl coated mushroom anchor and a polyfoam buoy and submerged 50 cm below the water's surface.

The second platform was placed at the water's edge, at a point usually submerged when the reservoir is full. This sensor node was positioned to simulate a sensor node that has been exposed to air due to a drop in the water level in the reservoir. The platform with the attached sensors was housed in a watertight enclosure with aluminium end caps and was powered by a 12 V lead acid battery. Figure 4-14 shows the second platform in position. Figure 4-15 shows an example of the platform with the antenna and anchor attached.

Figure 4-14 Second sensor platform in position outside of water at Hurleston reservoir

Figure 4-15 An example of the sensor platform used during testing at Hurleston reservoir with the buoy and anchor attached

The third platform was placed at different locations in the reservoir to conduct further distance testing to establish the maximum communication distance in the new environment. The platform and sensors were housed in a third watertight enclosure connected to a polyfoam buoy and vinyl coated mushroom anchor.

The first and second platform were left in place for the entirety of the fieldwork, the third platform was moved to varying distances from the other sensor nodes moving away from an initial starting point of 4 metres away moving up to 19 metres away. The platform was left in position, readings were taken using the first sensor platform with data collected and output to the connected Raspberry Pi, the platform was left for 1 minute and 6 readings captured from the first sensor node, including the received signal strength, the original node the message was from, and the node the message had been transmitted from.

Based on the findings in previous works the experimental works presented in 4.7 Communication distances in Liverpool Leeds canal, the fieldwork carried out used data rate of 1.2 kb/s which provided longer range communications. The experimental works used a separation distance of 1 meter as with previous experiments due to the level of progression while enabling the identification of maximum communication distances achievable using RF communication.

The platforms were all loaded with the same software that created broadcast network, Figure 4-16 shows a flowchart of the software used during experimentation. The platform sent sensor readings at 10 second intervals and transmitted this to all sensor nodes within range, sensor nodes would then re-transmit this message if they had not yet ‘seen’ the message based on an incremental message ID. Due to the nature of the experimental works, power efficiency was not a factor that was in play, for this reason the software did not make use of power saving features within the software and hardware platforms.

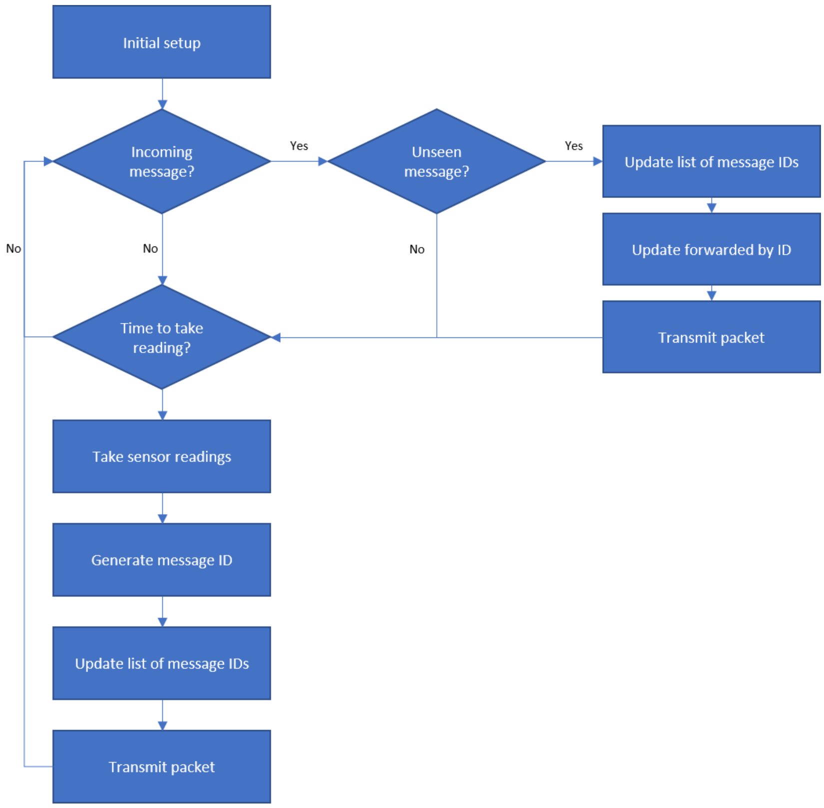

Figure 4-16 A flowchart of the software used during experimental works at Hurleston reservoir

4.8.3 Results and Discussion

Table 4-4 shows the results obtained during experimental works, the table shows the received signal strength when data was received directly between the first and third platform. The table also indicates which messages were received through the multi-hop networking.

Table 4-4 Results from field work undertaken at Hurleston reservoir

Distance	Received signal strength	Received signal strength Standard deviation	Platform forwarded
4	-90.17	0.69	False
5	-92.83	0.68	False
6	-100.34	0.47	False
7	-99.00	0.58	False
8	-105.67	1.57	False
9	-109.17	0.37	False
10	-101.17	0.90	False
11	-103.50	1.50	False
12	-106.67	1.25	False
13	-103.67	0.47	False
14	-107.34	1.70	False
15	-107.67	0.75	False
16	-106.17	0.37	False
17	-108.17	0.69	False
18	N.A	N.A	True
19	N.A	N.A	True

The results show a significant improvement in communication distances achieved in previous worked carried out as described in 4.7 Communication distances in Liverpool Leeds canal where communication distances of 7 metres were achieved in direct point-to-point communications. The results indicate a significant improvement of over 240% with distances of 17 metres being achieved during experimentation.

The results show that additional separation distance was achieved between the first and third platform by using multi-hop routing, this approach used the second sensor node as an intermediary. Better spacing of nodes could allow for further distances to be achieved, in cases where the intermediary node is directly between the two platforms it would be feasible to achieve communication distances of 34 metres using three platforms.

Table 4-5 Distances between platform 3 and 1 and 3 and 2

Distance between 3 and 1	**Distance between 3 and 2**
18	22.85
19	23.80

Table 4-5 shows how the distances between 3 and 1 and 3 and 2 differ due to the positioning of the platforms. The results show that the distance between 3 and 2 was greater than that between 3 and 1. This indicates that signals can travel further when one platform is not submerged in water. This indicates that the signal is likely propagating through the air; as the signal does not need to cross the air-water boundary twice the signal is able to travel further. Work by I. I Smolyaninov et al[119] has shown that signals will cross the water-air boundary and propagate through air.

4.8.4 Conclusion

The results highlight the importance of environment within deploying a UWSN using RF communications, with previous works demonstrating communication distances of 7 metres, these results show a significant improvement of over 240% with communication distances of 17 metres being achieved.

The results presented from the fieldwork highlight the benefit of RF communications being able to cross the air-water boundary, with the second sensor platform being placed out of the water being able to communicate with platform 3 (the sink node) and with the first platform to relay messages to platform 3 when platform 1 and platform 3 were unable to communicate with one another.

The works show a simple implementation of a multi-hop underwater wireless sensor network that could be used to extend communication distances in an extended underwater wireless sensor network. This simplistic example shows a three-node network that applies a simple but effective controlled flooding routing approach, more advanced applications would most likely take advantage of more complex routing protocols to improve power efficiency of network nodes.

4.9 Simulations of a wireless broadcast network

4.9.1 Introduction

An important factor in wireless sensor networks is the lifetime for which the network will operate, this is even more important in cases of underwater sensor networks where sensor nodes are more difficult to access and therefore maintain sensor nodes.

Many factors can play a part in network lifetimes including the firmware running on the sensor node, the number of sensor nodes within the network, the type of routing protocol used and the deployment area size. This section presents simulations based on the hardware platform used in experimental works, presenting the potential network lifetime considering factors such as communication intervals and the number of deployed network nodes.

4.9.2 Methodology

The simulation runs were based on an adapted version of firmware designed for power conservation, the firmware designed is shown as a finite state machine shown in Figure 4-17. The states are sleep, transmit and receive. The diagram shows the paths to each of the states that the sensor can enter with the sensor always eventually entering the sleep state until a new message is received.

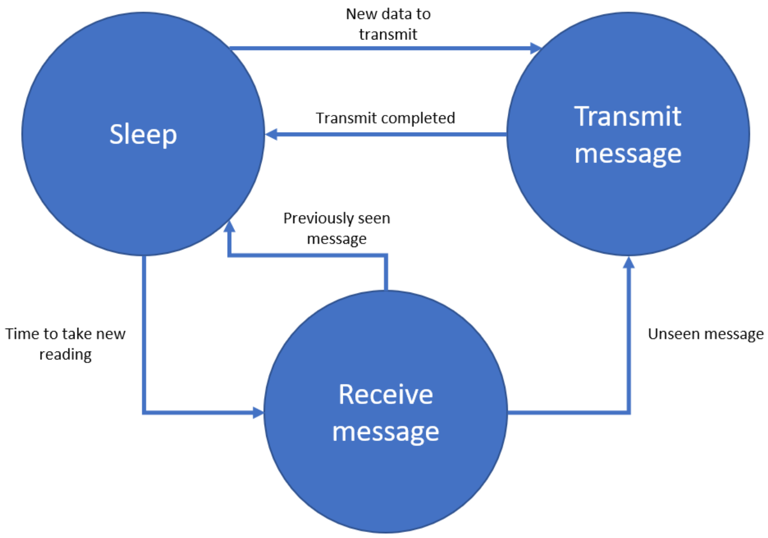

Figure 4-17A finite state machine of the states that the simulated sensor can be in

To allow for accurate simulations of network lifetimes, readings were taken of the current consumption of the platform for each of the three states. Current readings were taken using a Keithley DMM7510 bench top multi-meter. Figure 4-18 shows the bench top multi meter connected to one of the developed prototypes and the data captured is presented in Table 4-6. The current consumption figures used were based on a 12V source.

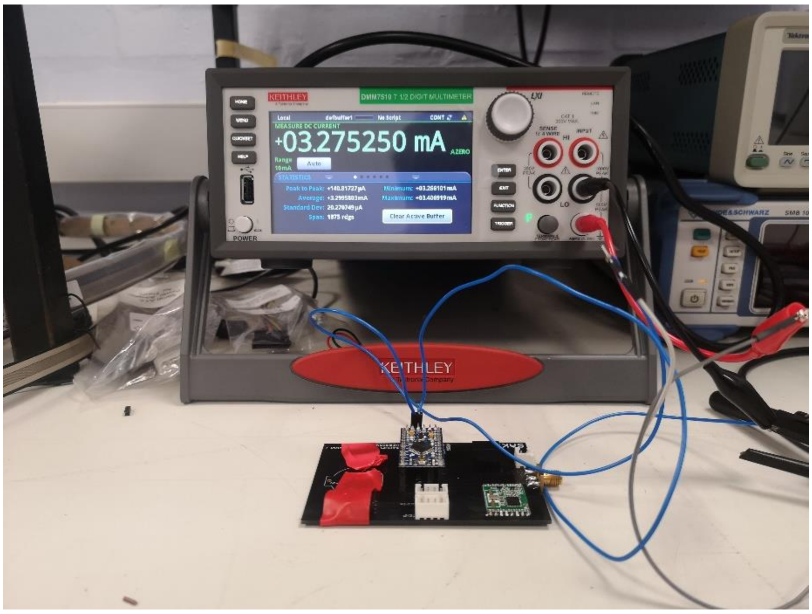

Figure 4-18 Experimental setup used to capture current draw of the developed platform

Table 4-6 Power consumption measurements of the developed platform

State	Current consumption at 12V (mA)
Sleep	1.99
Transmit	26.97
Receive	8.47

To calculate the energy consumption of each state the duration of the state must also be considered, the duration of the sleep state varies based on the amount of time that the node is set to sleep for between transmission intervals shown in Table 4-7.

Table 4-7 Durations and calculated energy usage for sleep states

Use case	Duration (s)	Energy usage (j)
One transmission per minuet	60	1.43
One transmission per hour	3600	85.97
One transmission per 12 hours	43200	1031.62
One transmission per 24 hours	86400	2063.23
One transmission per 168 hours (1 week)	604800	14442.62

The duration of transmission and receive states are based upon the size of the network packet being transmitted, Table 4-8 shows the duration of the transmission and receive states for a range of packet sizes.

Table 4-8 Duration and energy usage for transmitting and receiving data for a range of packet sizes

Packet size (Bytes)	Duration	Transmit energy usage (j)	Receive energy usage (j)
8	0.045	0.015	0.005
16	0.167	0.054	0.017
32	0.252	0.082	0.026
64	0.463	0.150	0.047
128	0.988	0.320	0.100

The energy consumption for each state was calculated using the equation shown in 4-1 Equation with the calculated values shown in the respective tables.

4-1 Equation for calculating energy in joules based on time, current and voltage

$$\boldsymbol{E = T \times I \times V}$$

E = energy (J))
T = time (S)
I =Current (A)
V = Voltage (V)

Simulations were conducted based on a range of node densities and communication intervals; network node densities used ranged from 100 to 500 giving a good spread of possible deployment densities. Three transmission intervals were simulated: daily, twice daily and hourly which could be of use to industry for continual monitoring and higher resolution monitoring for contamination events. In total 15 simulation scenarios were run.

Each simulation scenario was run 2,000 times, with a mean average being taken for the number of node deaths at the end of each round across all the simulation runs. The simulations assumed a 20 Ah capacity lead acid battery that supplies a constant voltage of 12 V for simplicity. The simulation space used was 260 metres by 360 metres, based upon the size of Hurleston reservoir where previous testing was undertaken.

The simulations assumed a network node communication distance of 15 metres, based on distances achieved in the second round of field trials undertaken, with a slight reduction made to compensate for the increase in noise introduced from having additional sensor network nodes. The simulations assumed a packet size of 64 bytes, this was selected as it allows for a considerable amount of data to be relayed through the network without becoming a drain on resources allowing a realistic simulation. A summary of the simulation is presented in Table 4-9 which remained constant for all simulation scenarios examined.

Table 4-9 Summary of common simulation parameters used across all simulation scenarios

Parameter	Value
Simulation space width (m)	260
Simulation space height (m)	360
Simulations run	2000
Node communication range (m)	15
Battery voltage (V)	12
Battery amp hour rating (Ah)	20
Battery charge (J)	864000
Packet size (Bytes)	64

Simulations were run using a piece of software specifically created for these simulations. The simulations used a controlled flooding routing approach, the same approach shown in previous works. Each simulation run started with nodes configured in random locations, each node would then transmit a message that would then travel through the network being received by nodes and then retransmitted if required.

4.9.3 Results and discussion

The results for each simulation scenario can be seen in Table 4-10 with the number of days before the first node death, 25% of nodes died, 50% of nodes died, 75% of nodes died and 100% of nodes died for each of the 15 scenarios examined.

Table 4-10 Results of the simulations for each tested scenario and the day on which the first node, 25% of nodes, 50% of nodes and 100% consumed all power resource

Nodes	Transmission interval	First node death (days)	25% nodes dead (days)	50% nodes dead (days)	75% nodes dead (days)	100% nodes dead (days)
100	Hourly	406	415	417	418	418
200	Hourly	371	406	413	416	418
300	Hourly	278	376	398	411	418
400	Hourly	148	274	331	389	418
500	Hourly	89	160	207	265	418
100	Twice daily	418	418	419	419	419
200	Twice daily	414	418	419	419	419
300	Twice daily	400	415	417	418	419
400	Twice daily	363	397	409	416	419
500	Twice daily	320	359	372	388	419
100	Daily	419	419	419	419	419
200	Daily	417	419	419	419	419
300	Daily	410	417	418	419	419
400	Daily	391	406	411	417	419
500	Daily	363	386	393	402	419

The results show that in the cases of daily and twice-daily transmission of data there is little difference in the network lifetime that can be expected, this is because in these instances the sleep state plays a significant role in the amount of power being consumed which remains constant regardless of transmission intervals.

Figure 4-19 shows a plot of the number of nodes that were dead after each round. The results show that as expected there is a direct relationship between the number of nodes within the network and the rate at which they die off. This is due to the usage of a controlled flooding routing approach as the more neighbours within range that a sensor node has, the more times that sensor node will receive messages and retransmit them.

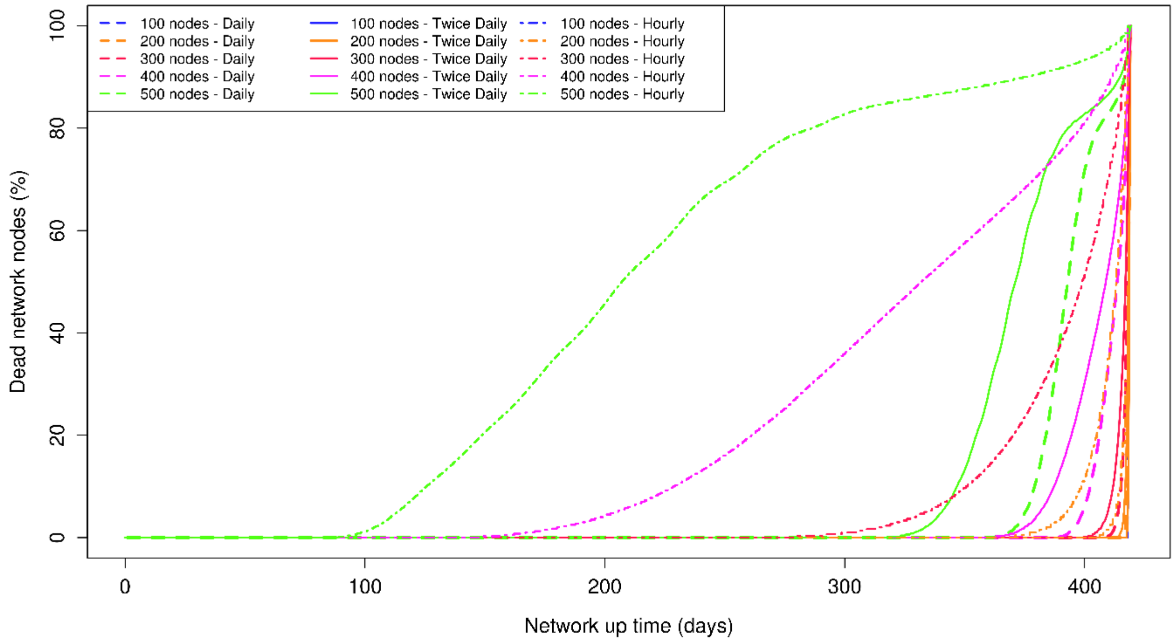

Figure 4-19 A graph of node deaths against the number of days

The results show that in most cases the rate of node deaths shows a sudden and significant increase in the number of node deaths occurring in the same time period. This can be attributed to the fact that a significant proportion of the power consumption within these nodes can be attributed to being in the sleep state, this sleep state can be considered a constant across all simulations as the sensor platforms consume the same amount of power in sleep state over the course of one day, regardless of the transmission rate.

4.9.4 Conclusion

The results from simulated works show that an underwater wireless sensor network could be deployed using the presented hardware and software while offering network lifetimes that would be significant for industrial applications for continual environmental monitoring.

The results presented show that a significant proportion of energy consumption is due to the amount consumed in the sleep state. It is possible that further reductions could be made to reduce this further either using more efficient sleep states, refinement in hardware platforms or a combination of these two factors.

While improvements in the sleep state would have a significant impact on the network lifetime it is possible that refinements within the routing approach could also lead to significant improvements in the overall network lifetime. These improvements could be achieved using approaches such as LEACH,[124] TEEN[126] or PEGISIS[127] which can offer significant improvements in energy consumption compared to a broadcast flooding approach.

4.10 Conclusion

The results from the experimental work undertaken highlight the possibility of using RF communications for underwater wireless sensor networks. The initial works undertaken in 4.6 Radio frequency in conductive environments show that conductivity plays an important factor in signal decay. These findings indicate that RF could be applied to underwater wireless sensor networks in low conductivity environments such as fresh water, canals and reservoirs where water is likely to be of a low conductivity.

The findings from initial lab-based works also identified that at close ranges it is likely that signals will be near field RF emissions rather than far field. This means that at short distances there is the possibility that the signal strength will significantly drop, this is due to the faster rate of decay of near field signals compared to far field emissions.

The experimentation conducted in the Leeds Liverpool canal and at Hurleston reservoir highlight the impact that environmental conditions can have on communication distances achieved and the maximum communication speeds that can be achieved, as the results from testing in the Liverpool Leeds canal highlighted the trade-off that is required between baud rates and communication distances.

The results from experimentations in 4.7 Communication distances in Liverpool Leeds canal suggest that a link speed negotiation protocol could be implemented within Underwater Wireless Sensor Network (UWSN). Such a protocol could allow for the balance between maximising the communication link speed that can be achieved while ensuring that all the required network nodes remain connected. The implementation of a link speed negotiation protocol would also have to consider how to optimise transmission power and ensure that a balance between these three factors can be achieved.

The results from fieldwork undertaken at Hurleston reservoir confirmed the possibility of using RF as a communication method for underwater wireless sensor networks, demonstrating point-to-point communication distances of 17 metres. This considerable increase in range between the two experimental works is likely due to environmental factors such as the geometry of the canal where the original testing was undertaken. Alternatively, it is possible that due to the location of the test site there was more interference from other devices on the same frequency.

The field work at Hurleston reservoir also indicated the ability to use a simple controlled flooding routing approach to routing network packets. Using this approach enabled communications to reach a sink node when peer-to-peer connections were not viable, this enabled additional distances to be reached, during experimentation this enabled an additional 2 metres of separation, however in practice such an approach has the potential of increasing the communication distance by a full 17 metres in a perfect scenario.

The simulation results show that with the platform used during experimental works, it would be possible to deploy an underwater wireless sensor network with 400 nodes capturing data daily with all nodes function for 391 days and 50% of sensor nodes remaining functional for 411 days.

The simulation results indicate that the largest factor that needs to be considered to progress the platform is the power consumption of the platform during the sleep state, sensor nodes spend significant periods of time in this sleep state making it vital to ensure that the sleep state of the platform is as efficient as possible. Small savings in energy consumption would significantly improve the lifetime of the network.

5 Feature selection in microwave data sets

5.1 Introduction

This chapter explores the use of machine learning feature selection in microwave spectroscopy data sets with a view to identifying frequencies that could be usable in the detection of specific substances. This chapter presents two example data sets that use a two-stage feature selection process to identify frequencies that could allow for the detection of contaminants at a range of frequencies.

5.2 Quantisation of peak-shift as a feature set

The focus of much of the previous work in the application of machine learning techniques to microwave spectroscopy data sets has focused on using magnitude readings. However, another important phenomenon that can be seen in microwave spectroscopy data sets is peak-shift, caused by a change in the microwave's velocity as it passes through a substance.

If a peak-shift can be quantised, then it could be used as a feature within machine learning to identify the concentration of a substance. A peak-shift feature could be used either as a stand-alone feature in a classification model or in combination with other features such as magnitude and phase.

One of the most important considerations when extracting a peak shift feature space is the distance maximum allowable distance between peaks before peaks are considered different peaks known as a window size. Using a large window size value for this feature space will significantly reduce the number of peaks that can be found within the feature space as only one peak can sit within the window size multiplied by two to represent the peak either side. Using too small a window size may lead to peaks being ignored within the feature space.

One of the largest challenges is detecting peaks across a range of readings, this is due to the fact that some readings may not contain the same peaks, it is also important to establish an appropriate method of detecting and grouping peaks together, while ensuring that the correct peaks are grouped as a set. An important tuning parameter to detect this peak shift is the maximum shift that is allowed before a peak is considered to be not found, too low and this value will cause too few peaks to be detected, while a large value will limit the number of features that are extracted from the data set.

5.2.1 Extraction of a peak-shift feature space

In order to identify a peak-shift feature space a baseline for peaks within a data set needs to be established, this can be achieved by using a specific class to establish the baseline of peak and trough locations within the data set. Using this baseline, it is then possible to determine the peak-shift for each class.

The size of a peaks-shift feature space can be controlled using the window size to limit the number of features that are extracted, with the larger the window size the fewer the features extracted. Using a window size that is too small may cause distortion within the features that are extracted that could impact upon the prediction accuracy.

The extraction of peaks was done through the use using a specifically developed TypeScript application. The extraction processes used a simple linear search to identify if each reading represented a peak or trough within the dataset. The algorithm checked readings to the left and right of the target reading checking to see if any values exceeded the value in the case of a peak or are lower than the value in the case of a trough.

The algorithm uses the provided window size to assess readings to the left and right of the selected reading, checking all frequencies that are within the window size range. The algorithm identifies if the reading is a peak or trough point by comparing the reading against others within the window region.

5.2.2 The application of a peak-shift feature space in machine learning

To examine the use of peak-shift within a microwave spectroscopy data set a small sample data set was collected of Geosmin concentrations and methanol control samples. The concentrations were: with readings taken using a Rohde and Schwarz ZVL13 VNAconnected to a horn antenna.

The data set examined consisted of three concentrations of alpha cypermethrin: 25 mg/m^2 400 mg/m^2 and 425 mg/m^2. The data set consisted of 150 readings with 50 readings for each concentration examined.

Table 5-1 shows the number of features that were extracted from the data set for each window size used, it shows a rapid change with the first three window sizes used before levelling off. Using large window sizes of 100 MHz and above significantly limits the number of features that are available. Using a window size of 2 MHz sill provides a significant number of features, without severely limiting the feature space available.

Table 5-1 Peak shift feature space size based upon window sized used

Window size (MHz)	Feature space size
2	1544
5	319
10	131
20	109
50	59
100	18
150	15
200	8

The approach taken applied four machine learning techniques: KNN, SVM, RF and GBM as well as an information gain filter. Unlike other approaches examined in these works the data used in the machine learning algorithms are not first reduced using an information gain filter due to the significantly reduced size of the peak-shift feature space compared to conventional microwave spectroscopy data sets.

The analysis identified a number of frequencies which present a peak shift that could be used to identify, Figure 5-1 shows an example of how a peak shift in magnitude readings could be used to identify between the classes within the data set. The plot shows the average magnitude of the three bases used within the dataset with a clear separation in the peak location between the three classes between 1.96 GHz and 1.98 GHz. The troughs in 1.98 GHz to 2.00 GHz also show a significant shift between the three classes that could be used for classification.

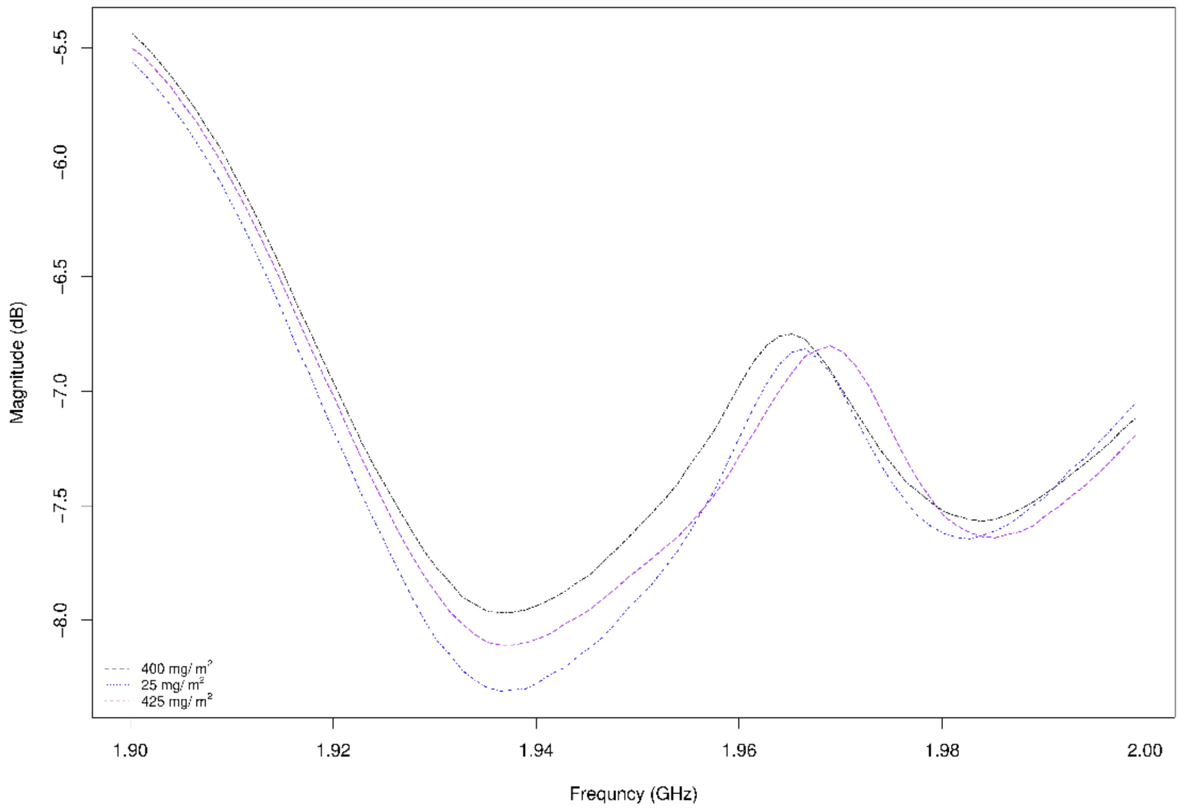

Figure 5-1 Peak shift identified through a peak shift feature space and machine learning within the 1.9 GHz to 2.0 GHz range

Another region of peak shift which was identified through the machine learning approach was the 1.7 GHz to 1.8 GHz region shown in Figure 5-2. The plot shows a clear trough shift between the 425 mg/m^2 and the other classes. The peak shift identified here is not a clear as the one shown in Figure 5-1.

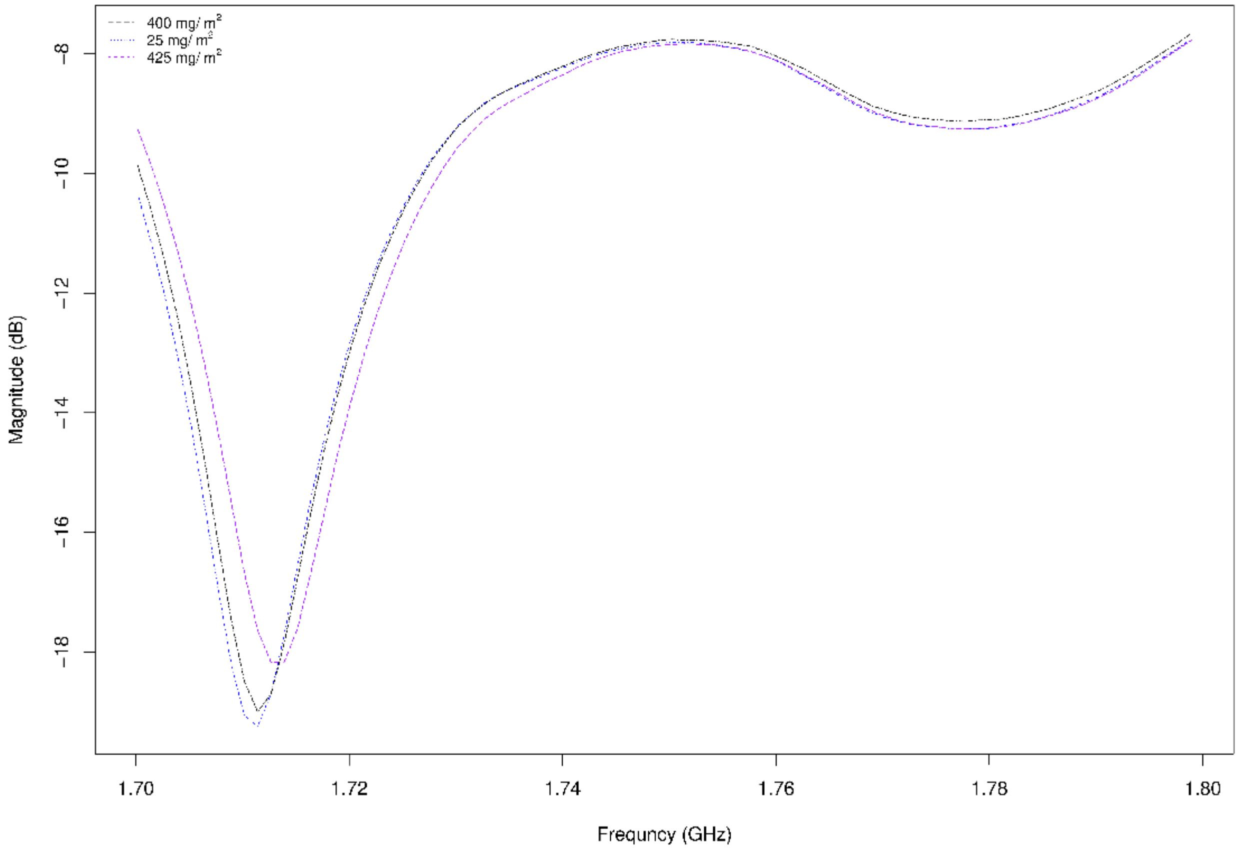

Figure 5-2 Peak shift identified through a peak shift feature space and machine learning within the 1.7 GHz to 1.8 GHz range

Figure 5-3 shows two further trough shifts identified through the application of machine learning to identify peak shifts that could be useful in a classification approach. The first peak within the region of 4.60 GHz to 4.62 GHz shows that there is a significant difference in troughs between the 425 mg/m^2 and the two remaining classes. The second trough shift between 4.68 GHz and 4.70 GHz shows separation between all three classes.

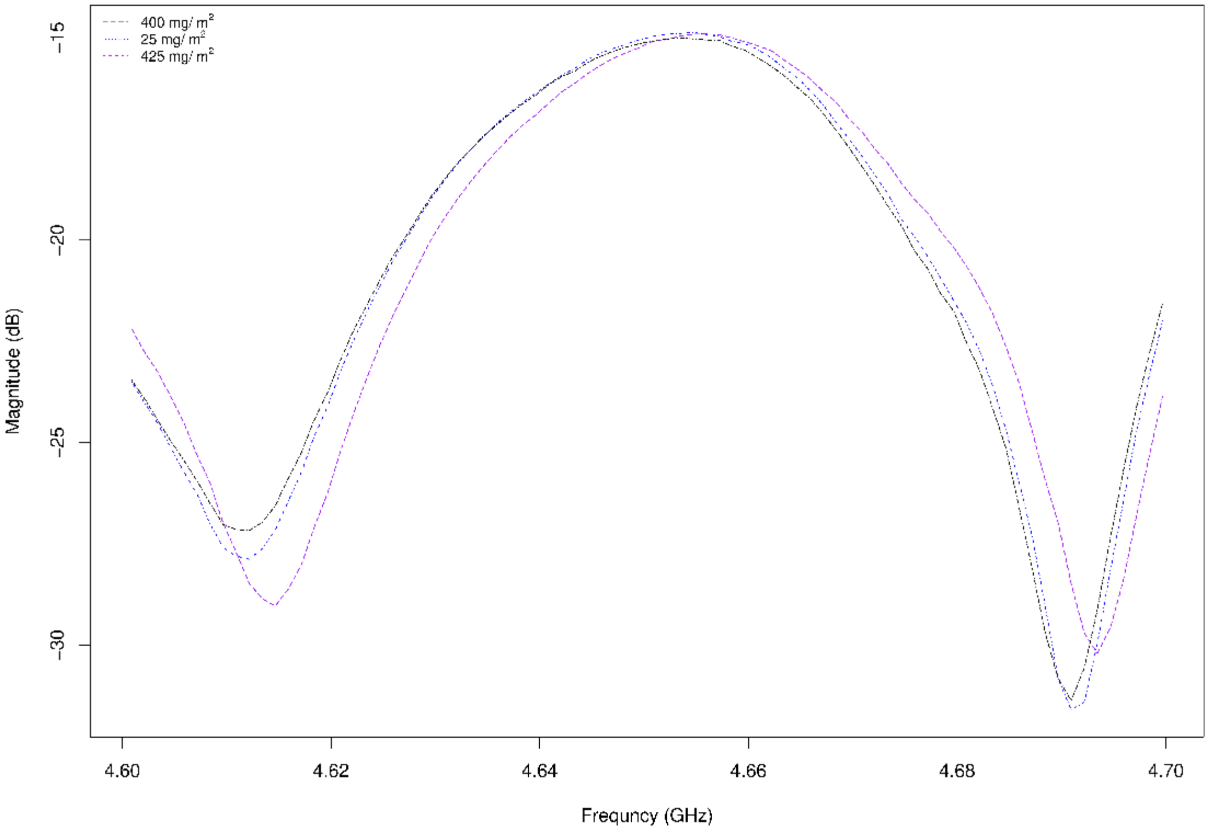

Figure 5-3 Peak shift identified through a peak shift feature space and machine learning within the 4.6 GHz to 4.7 GHz range

The largest advantage offered by using a peak shift feature space is the abstraction of the true magnitude value, rather than using the captured magnitude reading; the feature space depends upon the relative magnitudes. The feature space depends upon the location of a peak rather than the value of the reading itself.

There are however issues with using a peak shift feature space, the largest of which is the requirement to read several frequencies. The need to take readings for the larger number of frequencies means that any sensor board would require the ability to produce several microwave frequencies.

As well as issues surrounding the production of multiple microwave frequencies there is the added issue of an increased amount of data that must be transmitted or additional data pre-processing that must be undertaken to produce a reading. This is due to the need to use multiple frequencies that would have to be communicated to capture the peak location, or alternatively the sensor would have to analyse the captured data to identify the peak location and communicate the frequency at which the peak appeared.

5.3 Analysis approach for microwave spectroscopy data using magnitude

Data was analysed using the R programming language[131], using packages from CRAN (Comprehensive R Archive Network) the analysis used the reshape2[132] package for data transformation and shaping of the data from the loaded CSV format to the format used during analysis.

The analysis used a range of packages from CRAN for data processing including fselector[133], mlbench[134] and caret[135]. The work used these libraries due to being freely available for anyone to use. These packages also benefit from being used in other academic works and have been contributed to by multiple experts.

5.3.1 Analysis approach

Before analysis, data was loaded and pre-processed in the R programming language environment using the load csv functionality built into the language, during the loading the data was labelled with the appropriate class that the data pertained to.

Once a file was loaded it was transformed into a wide data format, with each column representing a frequency within the data frame rather than each row, this made it easier to manipulate and work with data during the processing. The data was reformed using the functionality provided by the reshape2 R library. Figure 5-4 shows an example of the shape of the data before and after the data transformation process.

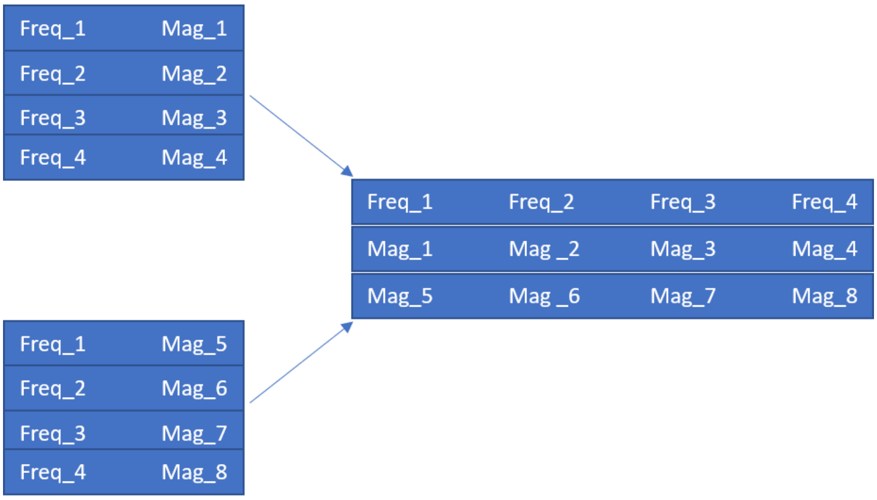

Figure 5-4 An example of the data transformation process

After transformation the data was bound with the previously processed data, after processing all files the data set was normalised. Normalisation desensitises the processes to the magnitude of numbers within the data set during the feature selection process and improves the numerical stability of the models used.

5.3.2 Initial dimensionality reduction

The second stage of the process reduced the dimensionality of the data set from a large feature space of 4,000 to a reduced number using feature filtering techniques. The use of feature filtering enables computationally quick analysis of the data set, allowing feature spaces to be quickly reduced in size, allowing more computationally costly techniques such as wrapper feature selection to be run on smaller feature spaces. There are several feature filtering options available including Information Gain, Correlation Coefficient score and Chi Squared test. The analysis approach opted for an information gain filter to firstly reduce the size of the feature space with an information gain filter being selected due to its speed.

In the case studies examined the implementation of the information gain filter used was from the fselector R library. The processes identified the top ranking 500 features, which were carried over to the next stage of the feature selection process. A boundary of 500 features was used after experimentation with higher and lower boundaries.

The initial dimensionality reduction stage enables the reduction of the feature space to 12.5% of its original size in both case studies, though this reduction appears sizeable it is important to note that microwave spectroscopy data sets consist of vast feature space sizes, much of which contains no information that would be of use in classification; this is due to the fact that only small parts of the microwave spectrum will be impacted by the substances tested for.

5.3.3 Wrapper feature selection

The second stage of the feature selection process applied a wrapper feature selection approach using four machine learning algorithms: Random Forest[68] , Support Vector Machine[59] (SVM), Gradient Boosted Model[71-73] (GBM) and K-Nearest Neighbour (KNN). The wrapper feature selection made use of the caret and mlbench R libraries for implementations of the machine learning algorithms and feature importance extraction.

Each machine learning algorithm was provided with the same training control and random seed to ensure that each algorithm started from the same point providing a level playing field for all algorithms. The algorithms used a normalised data set to perform a wrapper feature selection.

5.4 Alpha Cypermethrin data set case study

5.4.1 Data set introduction

The data set consisted of 50 readings per class with 3 classes giving a total of 150 readings across the entire data set, each reading consisted of 4,000 features. Each class within the data set represented a different concentration of Alpha-Cypermethrin; a common pesticide used in India to prevent the spread of malaria by treating the walls of homes with it. The classes represented the following concentration levels of Alpha-Cypermethrin: 25 mg/m^2, 400 mg/m^2 and 425 mg/m^2.

5.4.2 Data collection

Data was collected using a VNA (Vector Network Analyser) connected to a horn antenna, the horn antenna was placed in a 2 mm airgap between the antenna and the tile treated with a dosage of Alpha-Cypermethrin. Readings were taken and stored using LabVIEW software on a laptop computer which was connected to the VNA, the software captured the data and stored it in csv format for later processing.

Concentrations of Alpha-Cypermethrin were applied to the back of ceramic tiles, in the following concentrations: 25 mg/m^2, 400 mg/m^2 and 425 mg/m^2 which were applied by treating the tile with 22.5 ml of Alpha-Cypermethrin wettable powder and Alpha-Cypermethrin to create the desired concentration.

Once prepared the tiles were left to dry completely and were stored in cool and dry conditions until readings were taken. Readings were taken using a horn antenna, the antenna was connected to a Vector Network Analyser using an SMA cable. Figure 5-5 shows an example of the experimental set-up.

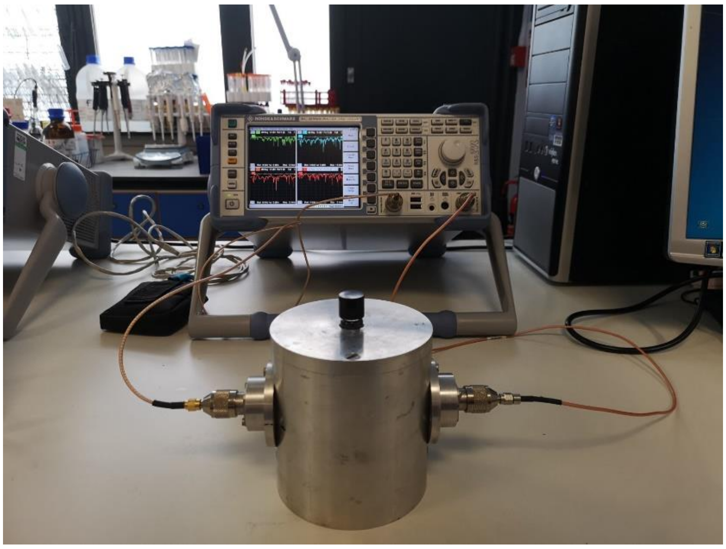

Figure 5-5 Example of the experimental set-up

5.4.3 Results

Table 5-2 shows the top-ranking features as identified through using an information gain filter and wrapper feature selection approaches. The results from the data analysis identified multiple frequencies within the 1.7 GHz to 2 GHz and 4.4 GHz to 4.9 GHz ranges identifying an area for potential further investigation in later works.

Table 5-2 The top 10 ranking frequencies for the detection of Alpha-Cypermethrin concentrations selected using machine learning and feature selection techniques

Top 10 ranking frequencies for the detection of Alpha-Cypermethrin concentrations in (Hz)				
Information Gain (IG)	**Random Forest (RF)**	**Gradient Boosted Model (GBM)**	**Support Vector Machine (SVM)**	**K-Nearest Neighbour (KNN)**
1947736960	1900225024	1947736960	5969992704	5608652288
4279569920	1960240000	4279569920	4477119488	4687171584
4284571136	4895973888	5353588224	1707676928	3438109440
4285821440	1947736960	5997499392	1896474112	4982245376
4312077824	4284571136	1701425408	4917229056	4619655168
4314578432	1952738176	1960240000	4883470848	3734433536
4315828736	1928982272	4289572352	5838709760	4969742336
4319579648	1920230016	1692673152	4412103168	4568392192
4410852864	1703926016	1950237568	3646911744	3641910528
4412103168	4304576000	4358339584	3668167168	5383596032

Figure 5-6 shows the top three ranking features using an information gain filter. The three highest ranking features were 1947736960 Hz, 4279569920 Hz and 4284571136 Hz. The plot shows a clear clustering for each of the three classes. The clusters formed are clearly separated from each other showing a clear separation between the classes using the three identified frequencies.

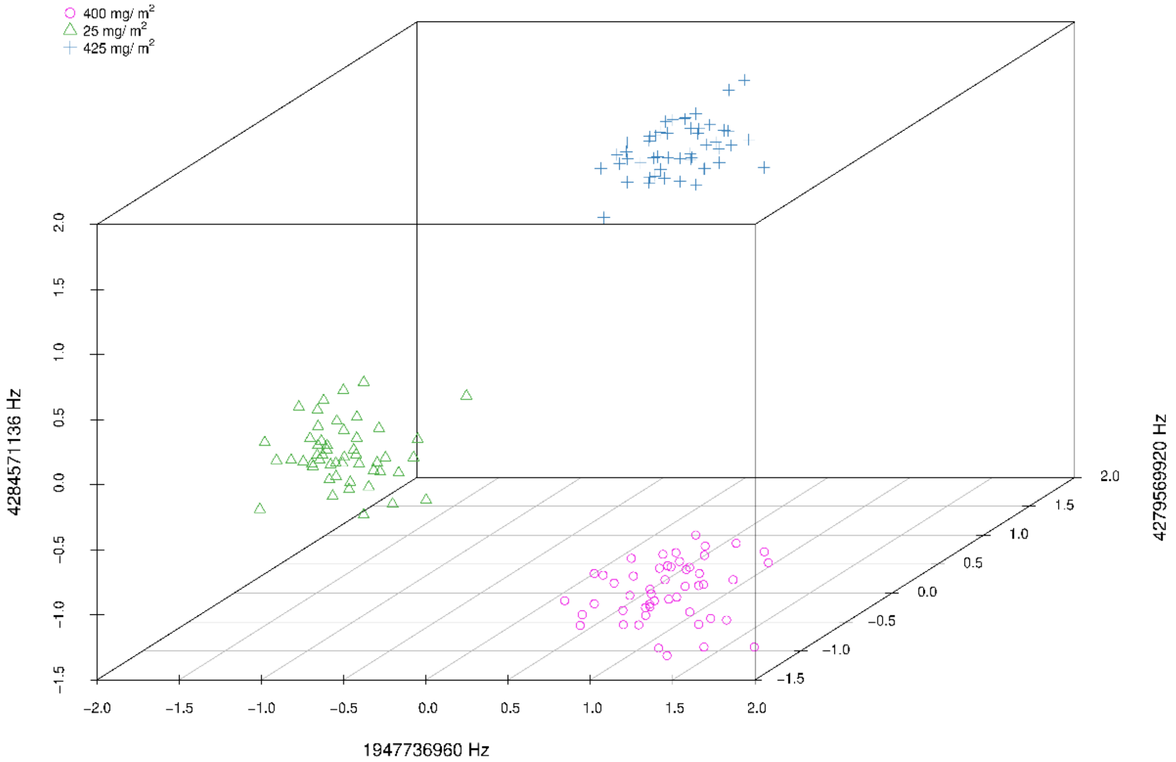

Figure 5-6 A plot of features identified for alpha-cypermethrin using an information gain filter

Figure 5-7 shows the top three ranking features identified using a Support Vector Machine (SVM) with a radial kernel. The features identified were 5969992704 Hz, 4477119488 Hz and 1707676928 Hz. The plot shows two clear clusters formed for the 25 mg/m^2 and 425 mg/m^2 classes. The plot shows that that the 400 mg/ m^2 class formed a more dispersed cluster which groups into two looser clusters.

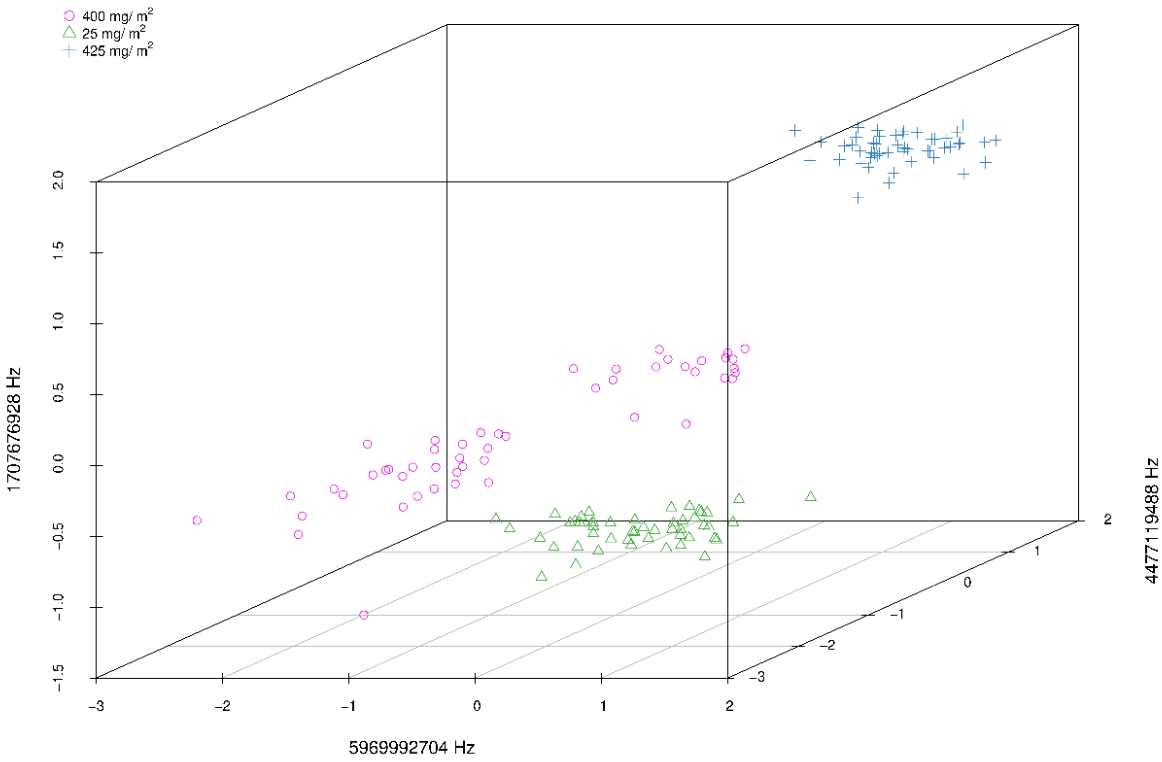

Figure 5-7 A plot of features identified for alpha-cypermethrin using Support Vector Machines

Figure 5-8 shows the top three ranking features identified using Gradient Boosted Models (GBM). The feature identified were 1947736960 Hz, 4279569920 Hz and 5353588224 Hz. The plot shows two tightly formed clusters for the 25 mg/m^2 and 425 mg/m^2 classes. The plot shows that the 400 mg/m^2 class formed a slightly more dispersed cluster than the other two classes. All clusters formed show clear separation from each other although not as separated as the clusters formed using features identified using an information gain filter in Figure 5-8.

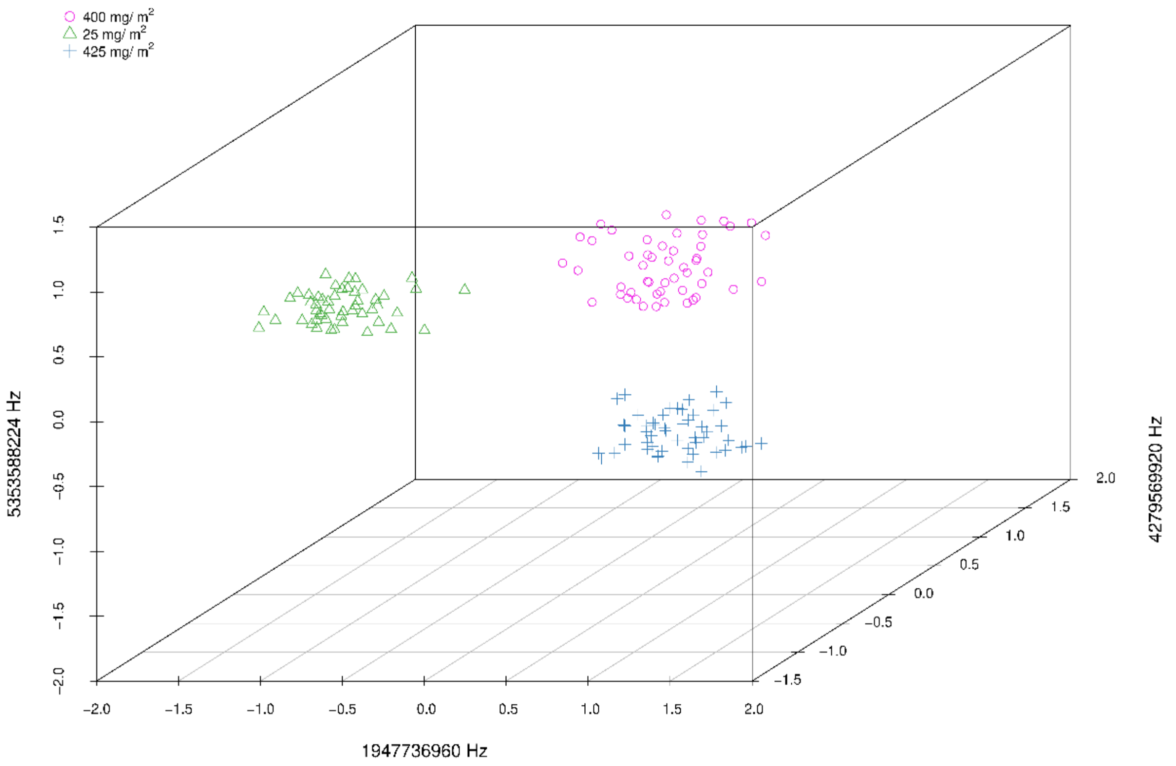

Figure 5-8 A plot of features identified for alpha-cypermethrin using a Gradient Boosted Model

Figure 5-9 shows the top three ranking features identified using Random Forest (RF). The features identified were 1900225024 Hz, 1960240000 Hz and 4895973888 Hz. The plot shows a strongly formed cluster for the 25 mg/m^2 class with a weaker cluster formed for the 425 mg/m^2 class with a more dispersed class formed for the 400 mg/m^2. The plot shows significant separation between the three clusters with a significant separation between the more dispersed 400 mg/m^2 and the 25 mg/m^2 and 425 mg/m^2 classes.

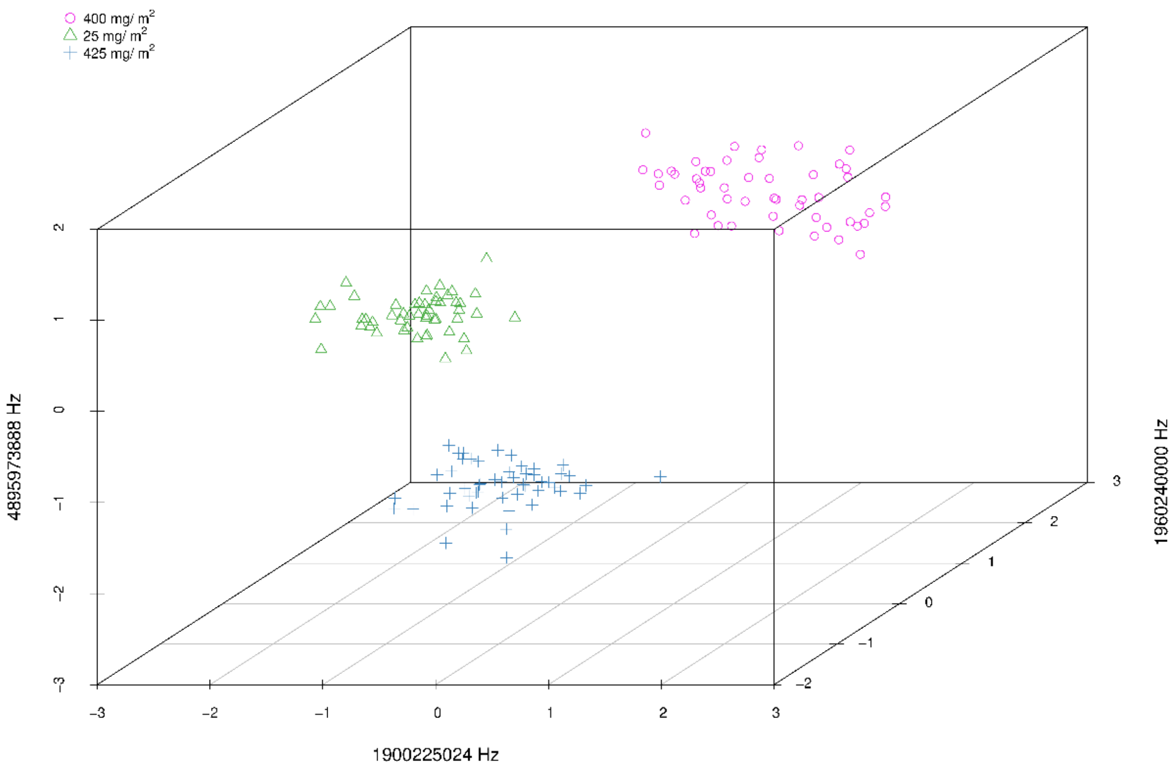

Figure 5-9 A plot of features identified for alpha-cypermethrin using a random forest model

Figure 5-10 shows the top three ranking features identified using K-Nearest Neighbour (KNN). The features identified were 5608652288 Hz, 4687171584 Hz and 3438109440 Hz. The plot shows a dispersed but significantly separated cluster for the 425 mg/m^2 class; the plot shows a strong cluster formed for the 25 mg/m^2. The plot shows that the 400 mg/m^2 class is significantly dispersed interfering with the cluster for the 25 mg/m^2 class.

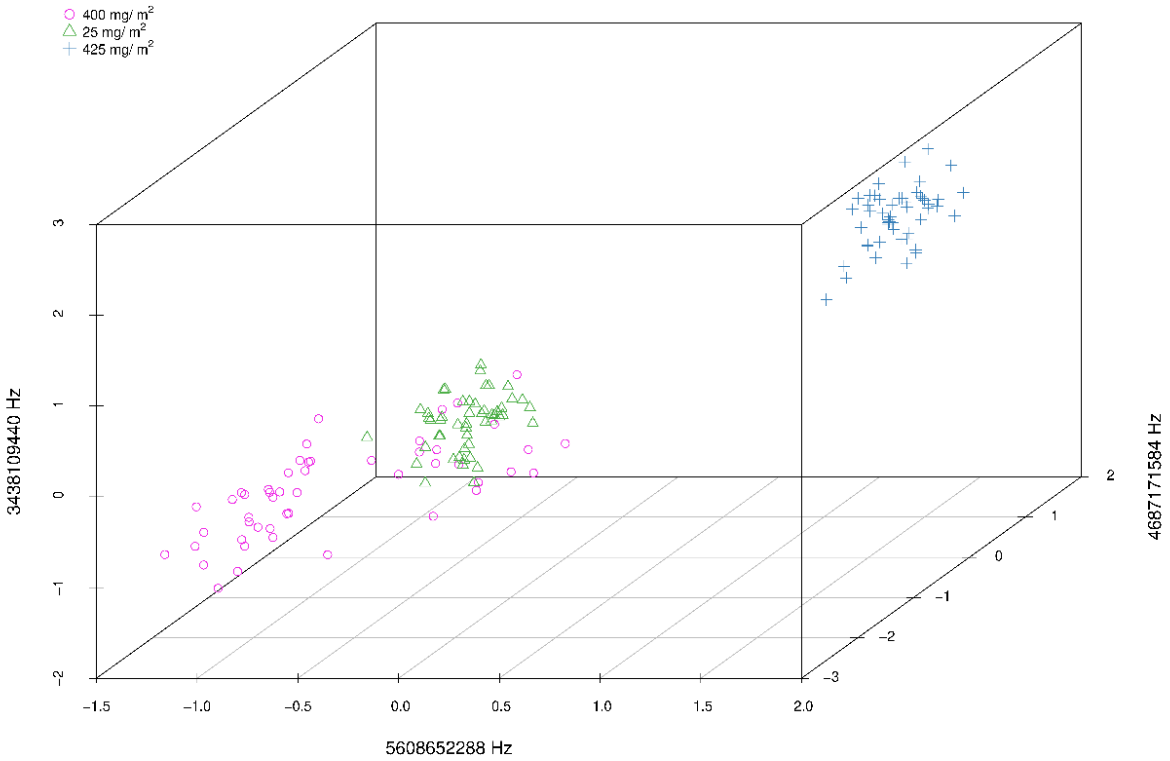

Figure 5-10 A plot of features identified for alpha-cypermethrin using a KNN model

5.5 Geosmin data case study

5.5.1 Data set introduction

The data set used throughout these experimental works examined four Geosmin samples with the following concentrations: 5 ng/l, 10 ng/l, 0.5 mg/l and 1.0 mg/l the Geosmin within the samples was dissolved in methanol which was equalised in all samples. In addition to the Geosmin samples three methanol blank samples were used with the following levels of methanol: of 5 mg/l, 10 mg/l and 20 mg/l which was used to ensure that any variance within the methanol levels between samples did not impact the results by influencing the machine learning algorithms.

5.5.2 Data collection

Data was collected using a VNA (Vector Network Analyser) connected to a microwave cavity with a resonant frequency of 2.4 GHz; samples were contained in a plastic centrifuge tube. A laptop computer was connected to the VNA and a custom LabVIEW application was used to collect data from the VNA sweeping a range of frequencies between 1 GHz and 13 GHz. Figure 5-11 shows an example of the experimental set-up used for data capture.

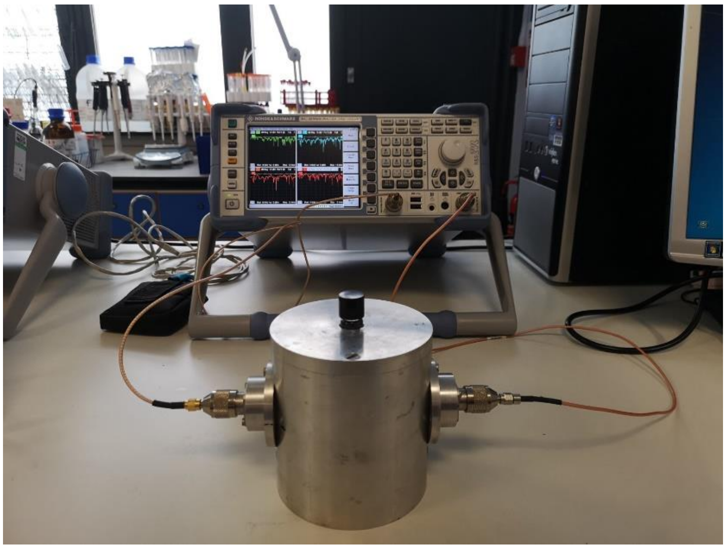

Figure 5-11 A photo of the experimental set-up with a large cavity connected to a ZVL vector network analyser for data capture

Concentrations of Geosmin were created using distilled water, Geosmin and methanol which the Geosmin samples were dissolved in when supplied. Samples were created with the following concentrations: 1 mg/l, 0.5 mg/l, 10 ng/l and 5 ng/l, the quantity of methanol was equalised for all samples so that methanol concentrations did not play a factor. Three methanol blank samples containing deionised water and methanol concentrations of 5 mg/l, 10 mg/l and 20 mg/l were also created. The methanol samples were treated as a single class, desensitising the process to any slight variances in the concentrations of methanol between samples.

The experimental work used a range of concentrations with two low level samples of 5 ng/l and 10 ng/l being included in the results to represent levels at which humans can detect geosmin at. Two high concentration samples of 1 mg/l and 0.5 mg/l were used to identify if the it was possible to detect higher levels of geosmin as well as the extremely low concentrations used.

Geosmin samples were kept in plastic centrifuge tubes and refrigerated when not in use. During use a sample was placed into the microwave cavity and a reading taken, the sample was then removed agitated and replaced into the cavity, the sample was left to settle for 30 seconds before the next reading was taken.

The experimental set-up used a resonant microwave cavity resonating in the 2.4 GHz frequency range.The experiment used the cavity to take S12 readings, altough an S11 reading could also have been captured using the cavity.

5.5.3 Results

The results from the data analysis identified multiple frequencies within the 5.4 GHz to 6.0 GHz and 6.4 GHz to 6.6 GHz frequency ranges; identified frequencies could be used to aid the development of a sensor targeted at the detection of Geosmin contamination. The top 10 ranking frequencies identified through the feature selection process are shown in Table 5-3.

Table 5-3 The top 10 most important features identified using the feature selection process in Geosmin data samples in order of most important to least important.

Top 10 ranking frequencies for the detection of Geosmin concentrations in (Hz)				
Information Gain (IG)	**Random Forest (RF)**	**Gradient Boosted Model (GBM)**	**Support Vector Machine (SVM)**	**K-Nearest Neighbour (KNN)**
8282820608	6461365248	12954988544	7436609024	7322580480
6485371392	6476368896	9294073856	5729182208	5807201792
6476368896	5900225024	4666916864	7364591104	10416354304
5789197312	10374343680	12960990208	5642160640	5618154496
6461365248	10659415040	9270067200	5480119808	7172542976
7151537664	5501125120	12723931136	5846211584	6551387648
6482370560	10974493696	12969992192	7202550784	5963240960
6488372224	5507126784	7862715904	4960990208	7133533184
6470367744	10980495360	12942986240	8351838208	7421605376
3649662464	8276819456	11409602560	6470367744	6491373056

Figure 5-12 shows the top three ranking features using an information gain filter. The three highest ranking features were 8282820608 Hz, 6485371392 Hz and 6476368896 Hz. The Figure shows a strong clustering of 5 ng/l and 0.5 mg/l classes with a clear separation between them and the other classes except for one instance of the 0.5 mg/l class which sits within the methanol blanks classes. The 10 ng/l and 1 mg/l classes show more dispersion between the two classes with instances of them overlapping with each other. The plot shows a clear separation between the Geosmin samples and the Methanol blanks.

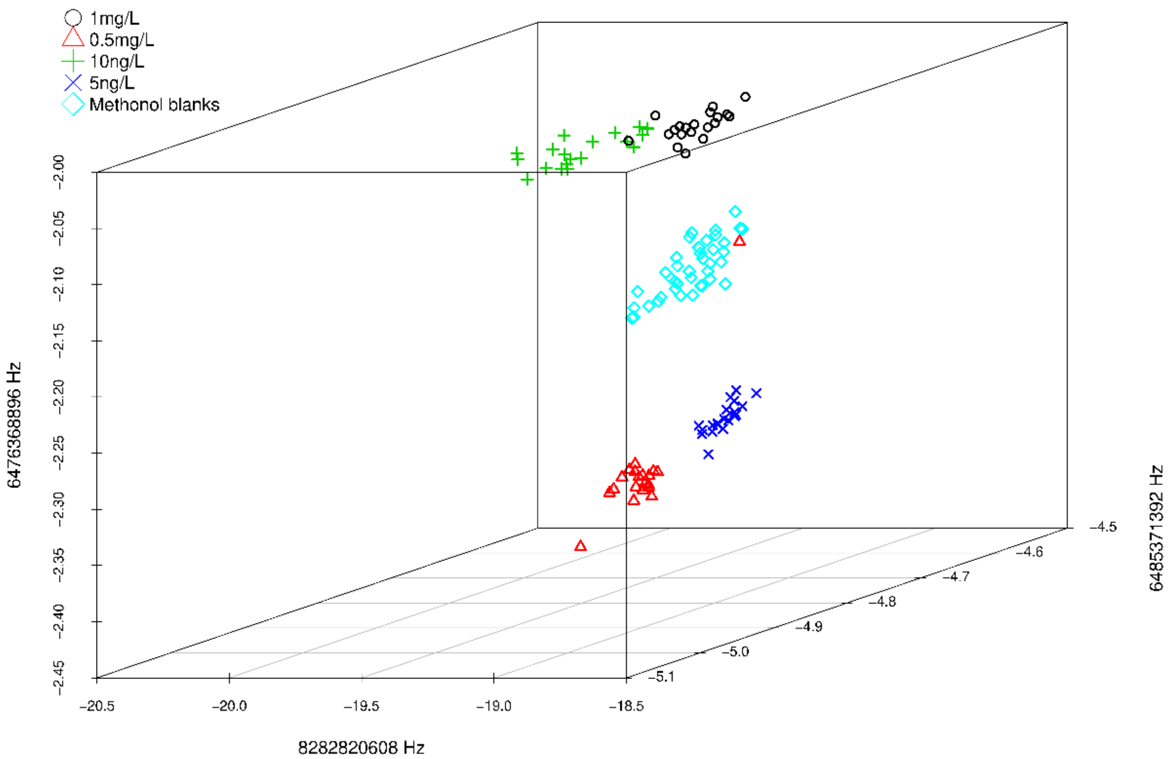

Figure 5-12 A plot of the top three ranking features identified using an information gain filter for detecting Geosmin concentrations in water

Figure 5-13 shows the top three ranking features identified using a Support Vector Machine (SVM) with a radial kernel. The features identified were 7436609024 Hz, 5729182208 Hz and 7364591104 Hz. The figure shows them significantly more dispersed with the Methanol blank class dispersed over a wide range, the figure shows that the 10 ng/l and 1 mg/l classes overlap significantly. The plot also shows a significant dispersion within the 5 ng/l class which disperses into the 0.5 mg/l class. The 0.5 mg/l class also shows significant dispersion.

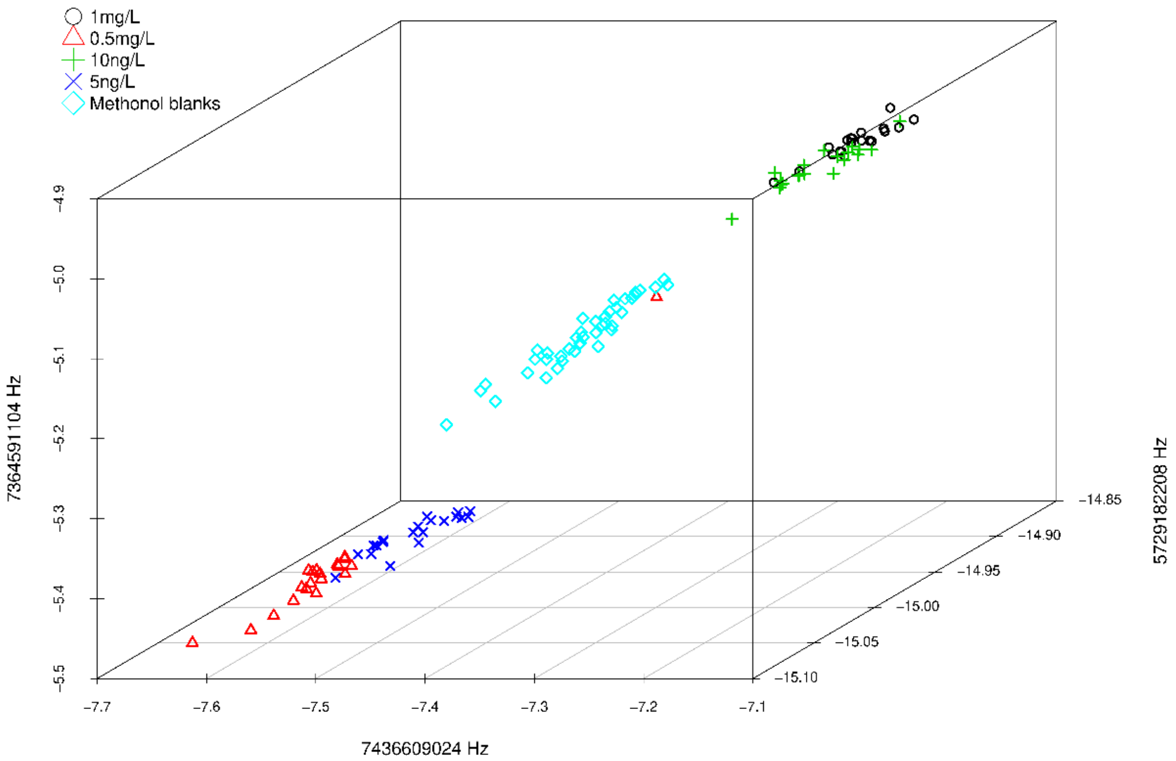

Figure 5-13 - A plot of the top three ranking features identified using wrapper feature selection with a support vector machine with a radial kernel for detecting Geosmin concentrations in water

Figure 5-14 shows the top three ranking features identified using Gradient Boosted Models (GBM). The feature identified were 12954988544 Hz, 9294073856 Hz and 4666916864 Hz. The plot shows significant dispersion within all classes with a significant overlap between the 1 mg/l, 0.5 mg/l, 10 ng/l and methanol blank classes. The plot shows a significantly dispersed but separate cluster for the 5 ng/l class.

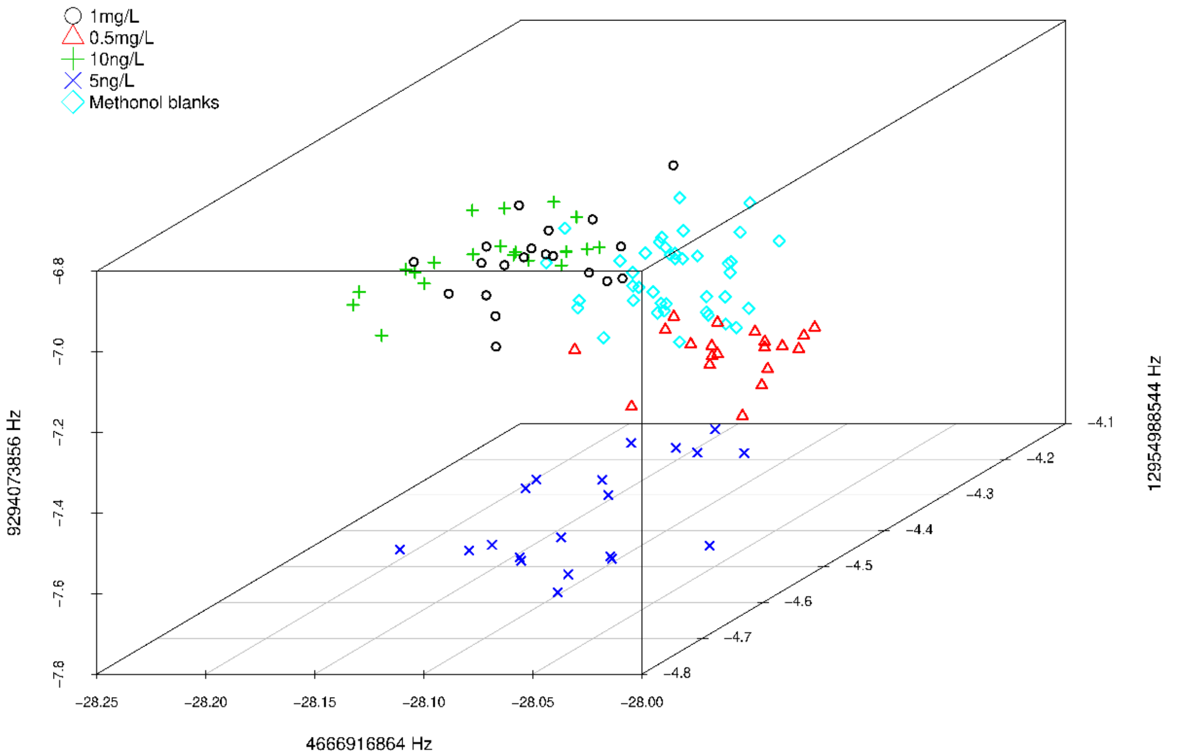

Figure 5-14 A plot of the top 3 features as identified using GBM

Figure 5-15 Shows the top three ranking features identified using Random Forest (RF). The features identified were 6461365248 Hz, 6476368896 Hz and 5900225024 Hz. The plot shows a strong clustering of the 0.5mg/l, 5 ng/l and methanol blank classes. The plot shows dispersion within the 1 mg/l and 10 ng/l classes causing significant overlap between the two classes.

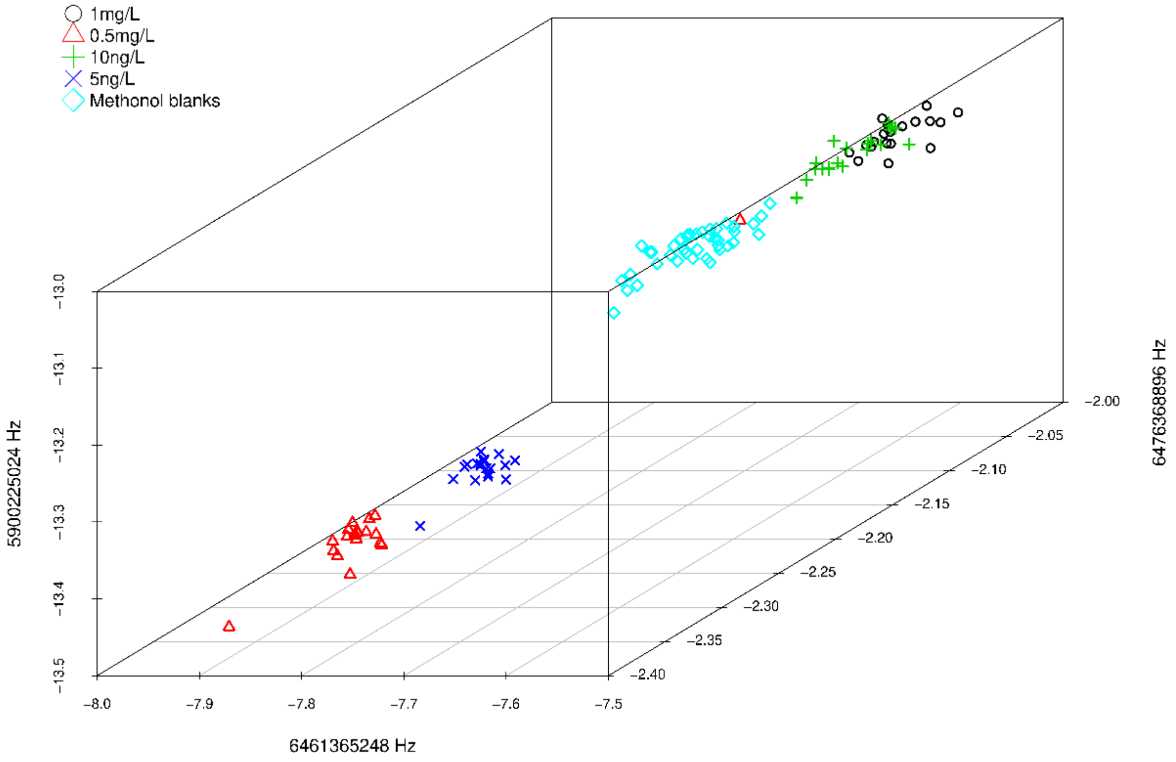

Figure 5-15 Features identified using random forest wrapper feature selection

Figure 5-16 shows the top three ranking features identified using K-Nearest Neighbour (KNN). The features identified were 7322580480 Hz, 5807201792 Hz and 10416354304 Hz. The figure shows that there is significant dispersion within all classes. The methanol blanks clearly cluster but do form an elongated cluster. The 10 ng/l and 1 mg/l classes significantly overlap each other with the classes dispersed amongst one another. The 5 ng/l and 0.5 mg/l classes also significantly overlap, some 5 ng/l readings appear to cluster away from the main cluster of the 5 ng/l class.

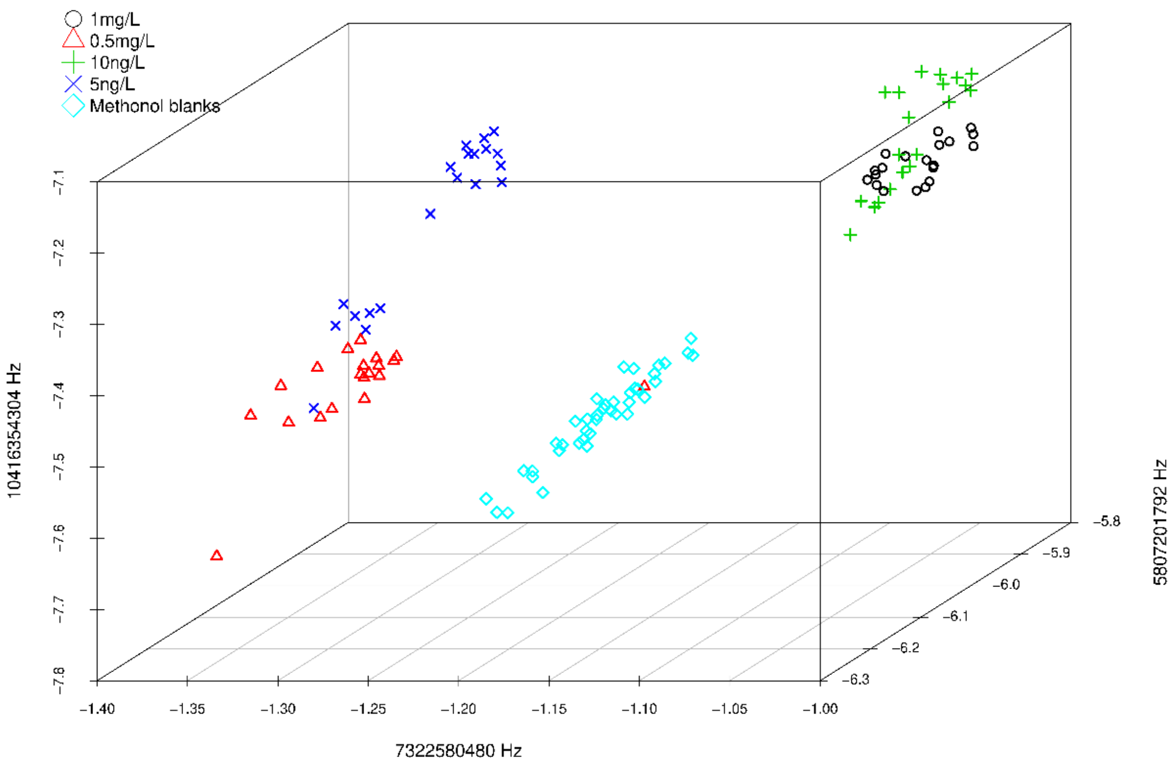

Figure 5-16 A plot of the top 3 features as identified using KNN

5.6 Discussion of results

The results from data sets identified multiple frequencies which could be used to separate defined classes within microwave data sets. Both data sets examined applied the two-step feature selection process to the magnitude data sets outlined in 5.3 Analysis approach.

In both cases results from the information gain filter provided strong results, showing that feature filtering could be used to great effect within these types of data set, as in both cases examined, an information gain filter provided consistently strong clusters. Information gain also provided the most consistent clustering between the two data sets and strengthened the case that feature filtering methods could be used without the wrapper feature selection processes in these types of data set.

In both data sets examined, SVM performed weakly, with at least one class being dispersed, this could be due to the nature of SVM being to take a one versus all approach to multiclass classification leading to a class imbalance. Figure 5-17 shows the decision boundaries of the SVM model trained on the alpha-cypermethrin data set taken from the model that identified the features for Figure 5-7 using the two highest ranking features for SVM, 5969992704 Hz and 4477119488 Hz. The plot shows a clear dispersion within the 400 mg/m^2 class.

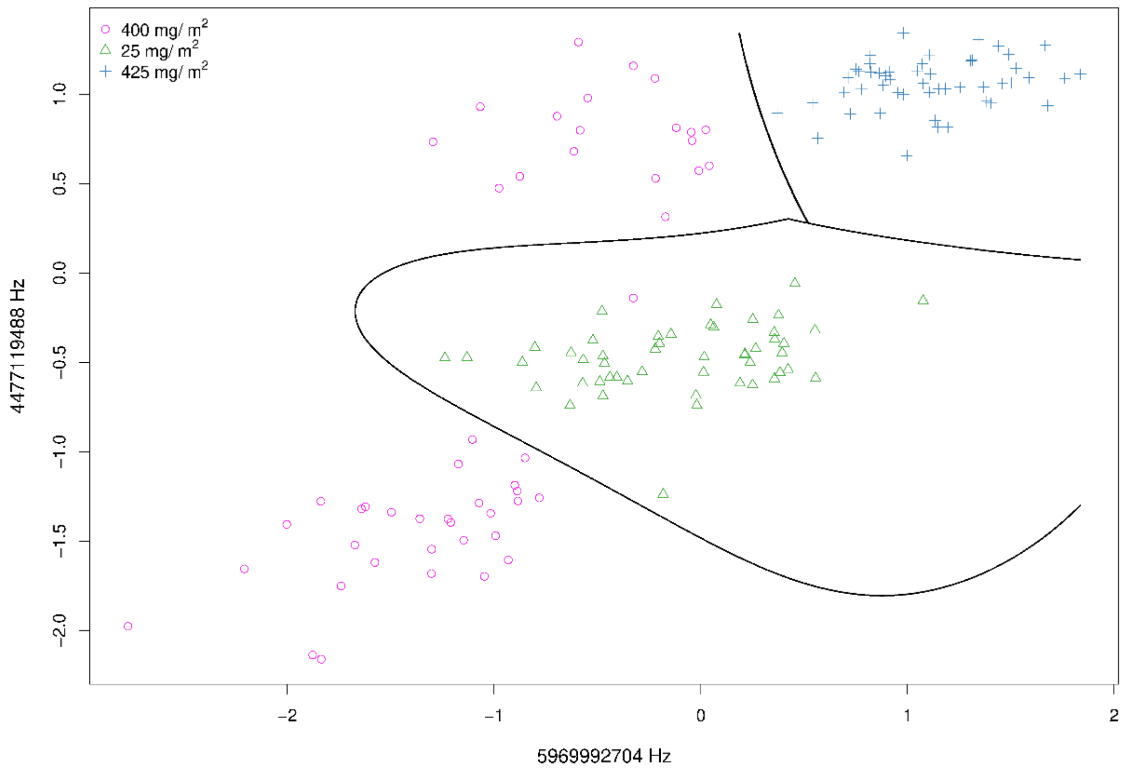

Figure 5-17 Decision boundary plot of alpha-cypermethrin data set

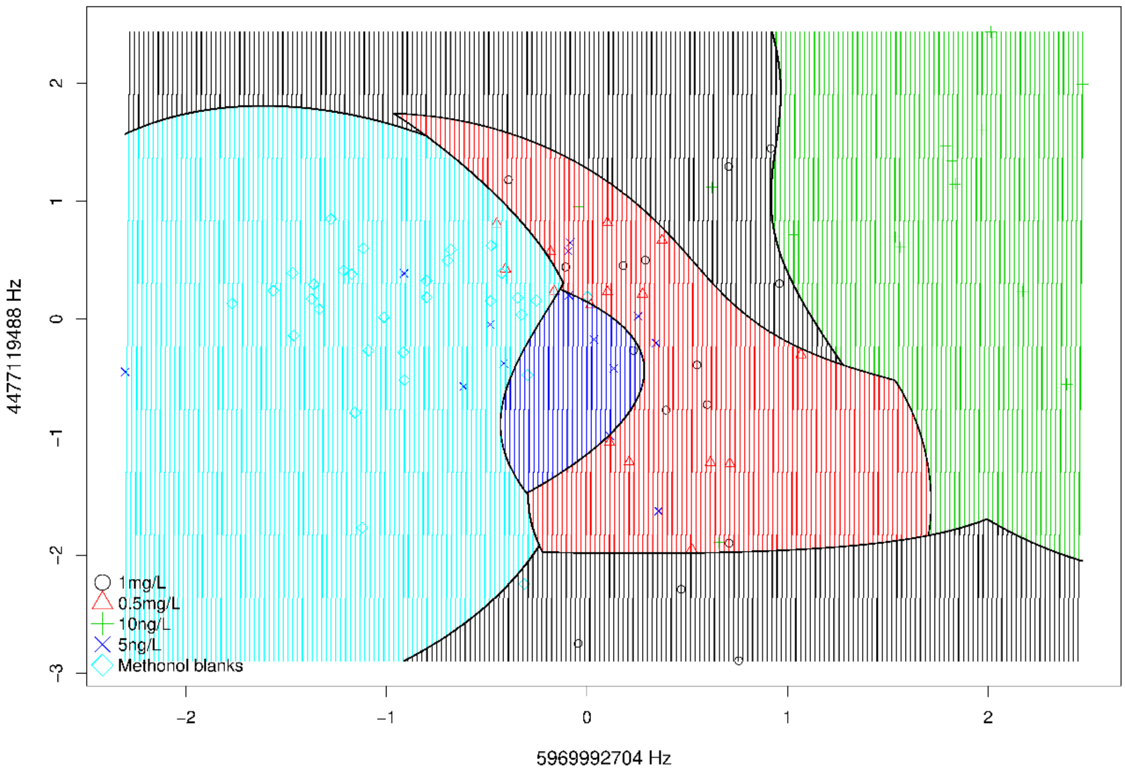

Figure 5-18 A plot of decision boundaries of the SVM model trained using the Geosmin data set

Figure 5-18 shows a plot of the decision boundaries in the Geosmin data set using the two highest ranked features for the SVM model that identified the features in Figure 5-13, 7436609024 Hz and 5729182208 Hz. The plot shows a large boundary region for methanol classification while concentrations of 5 ng/l have considerably smaller bounded region. The decision boundary plot also shows that many readings for the 1 mg/l class fall into the 0.5 mg/l boundary region within the 5729182208 Hz frequency, the plot suggests that the SVM model produced good boundaries for Methanol and 10 ng/l with large regions being allocated to the classes within the decision boundary plot. The 5 ng/l class performed the worst with only three points within the decision boundaries for the class. The 5 ng/l class also has the smallest decision bounded region.

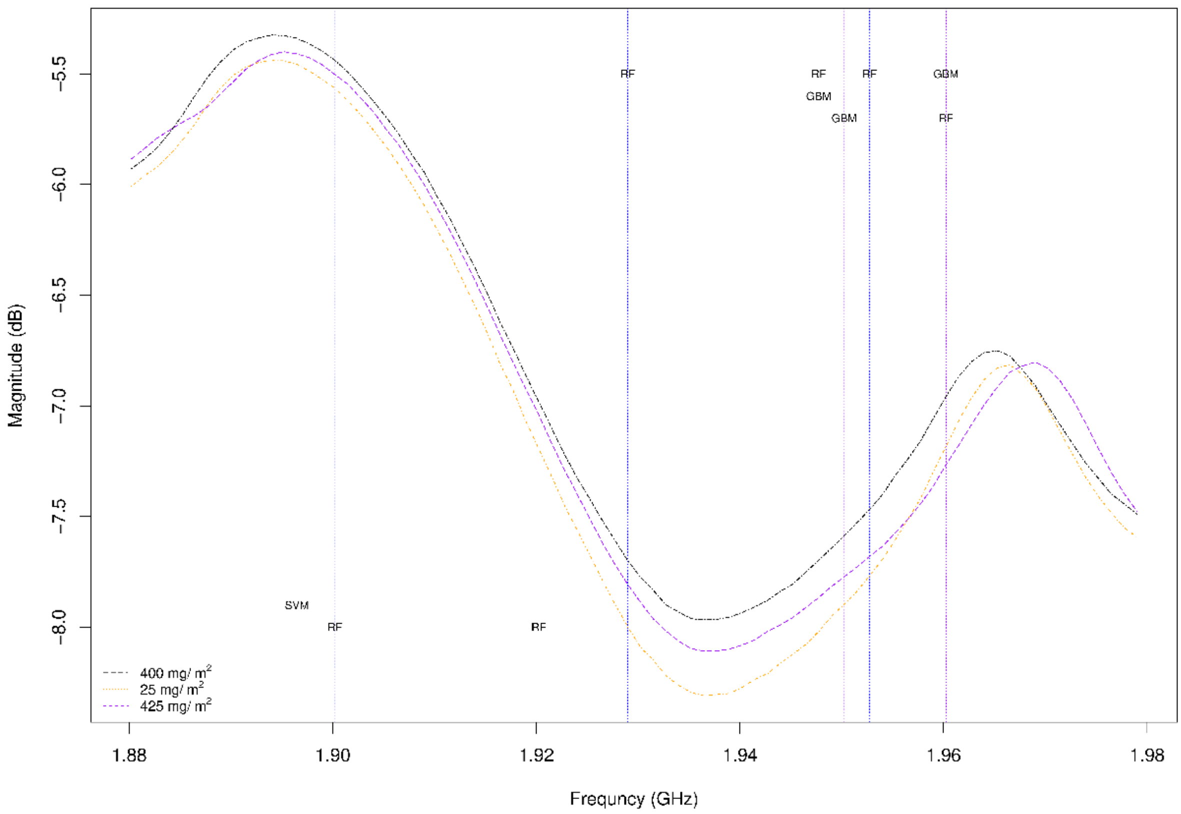

Figure 5-19 Alpha cypermethrin data set mean from 1.88 GHz to 1.98 GHz

Figure 5-19 shows the mean for each of the three classes within the alpha cypermethrin data set examined. The plot covers the frequency range of between 1.88 GHz and 1.98 GHz, this region contained multiple frequencies identified using the outlined process. The plot shows features identified, the plot shows clear separation between the three classes displayed within the plot, with 8 frequencies being identified, in some instances by multiple machine learning algorithms with a total of 10 identifications within the 1.88 GHz to 1.98 GHz frequency range.

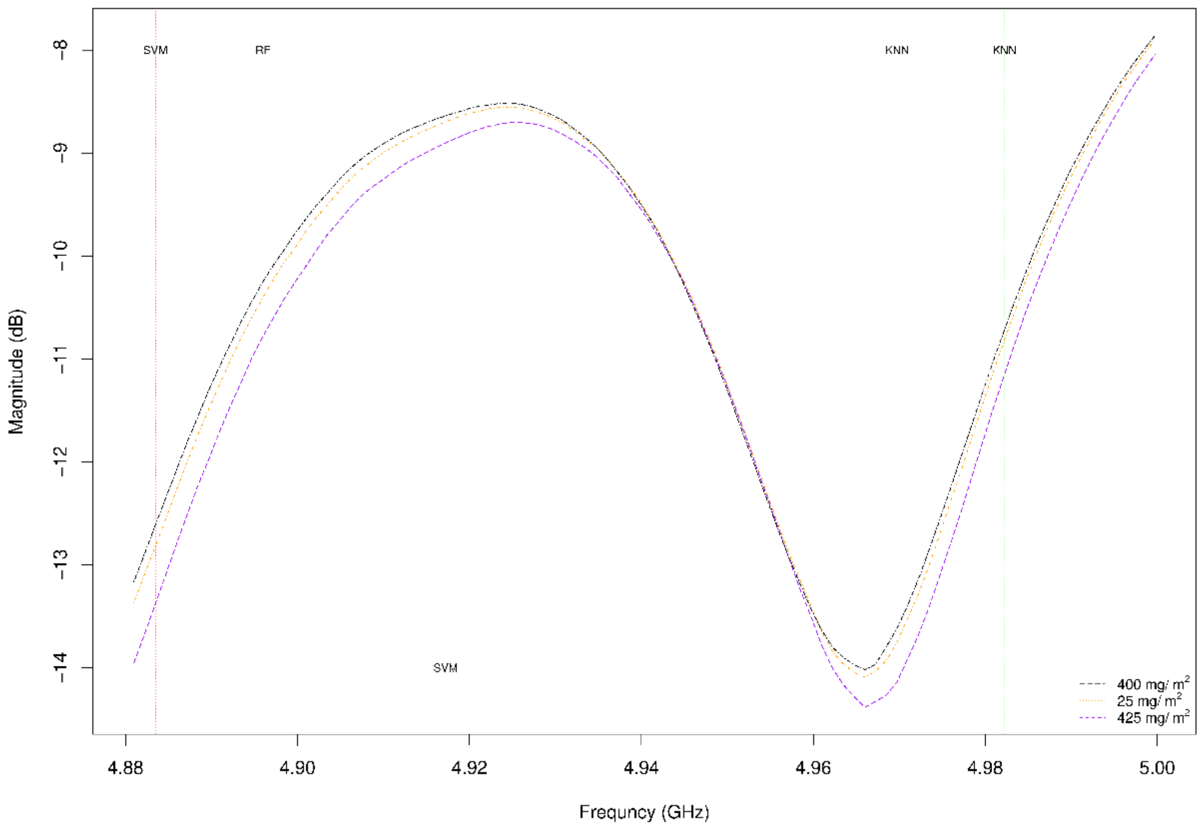

Figure 5-20 Alpha cypermethrin data set mean from 4.88 GHz to 5.00 GHz

Figure 5-20 shows the mean for each of the three classes within the alpha cypermethrin data set examined. The plot shows the frequency range of 4.88 GHz to 5.00 GHz. A total of 5 frequencies were identified within this region. All the points identified within the region show a clear separation between the mean values of the classes.

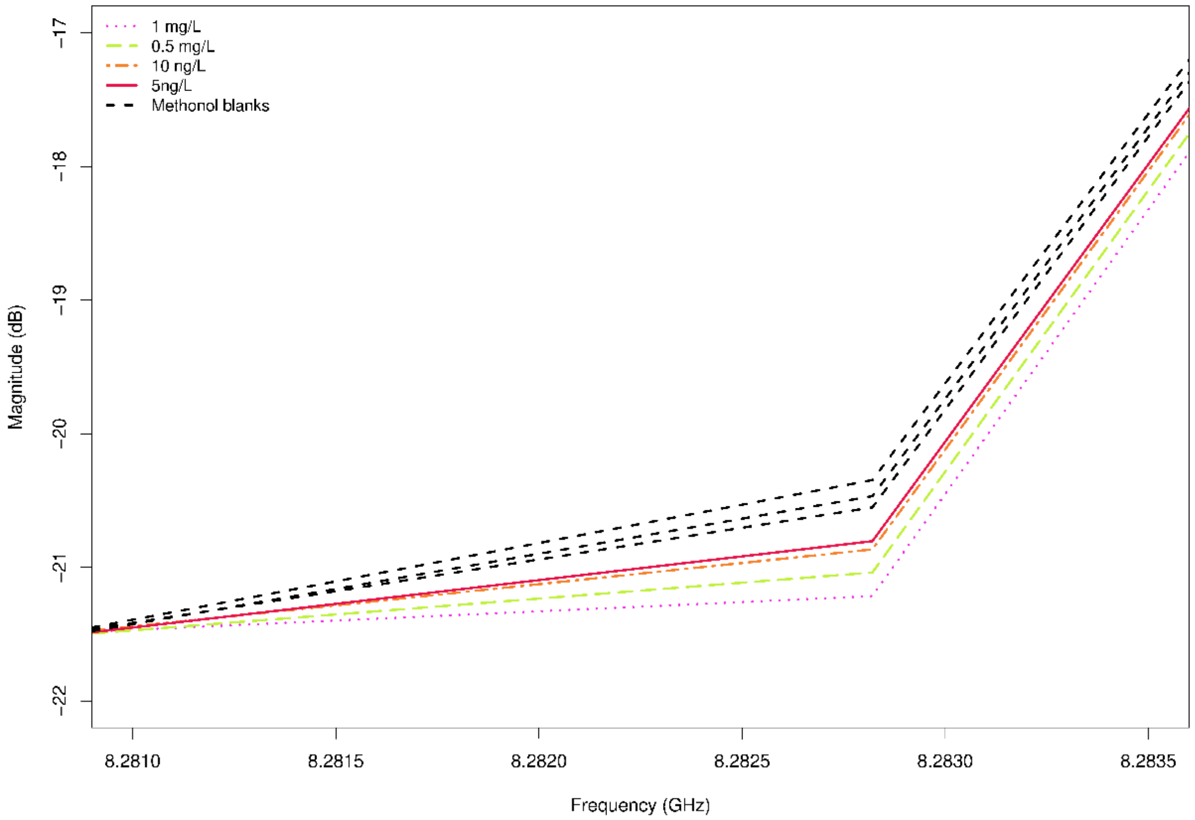

Figure 5-21 A plot of means for Geosmin data from 8.281 GHz to 8.2835 GHz

Figure 5-21 shows the mean for each of the 5 classes within the Geosmin data set examined. The plot shows the frequency range of 8.2810 GHz to 8.2835 GHz, this plot includes the frequency of 8282820608 HZ which was identified as the most important feature of an information gain filter.

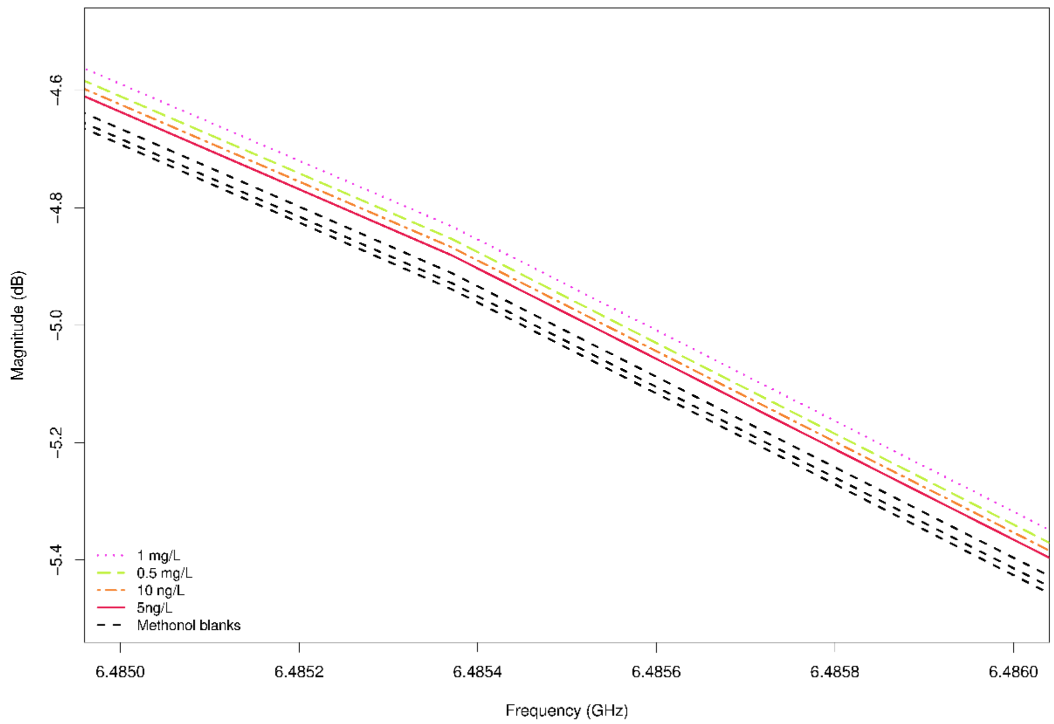

Figure 5-22 A plot of the 6.485 GHz to 6.4860 GHz frequency range in the Geosmin data set

Figure 5-22 shows the mean for each of the 5 classes within the Geosmin data set examined. The plot shows the frequency range of 6.485 GHz to 6.486, this plot includes the frequency of 6485371392 HZ which was identified as the second most important feature for an information gain filter. The plot shows a grouping of the three methanol blank classes used with some separation, the plot shows that the four Geosmin classes are separated from one another and the methanol control samples.

5.7 Development of a microwave sensor board

After the identification of frequency ranges that could be used to detect a substance the next stage is to develop a sensor targeted at the appropriate frequency ranges. The sensor can then be used to take readings within the range and can be carried forward to collect more data in a range of environments during field trials and studies where the use of a VNA would not be practical or feasible.

Such a sensor board requires a number of key components including the following: RF source, RF coupler, RF power detector, Microcontroller, Digital to analogue converter, analogue to digital converter, these are just the key components and other supporting components would also be required to produce a complete circuit.

5.7.1 Microwave Source

The radio frequency source is a vital component for a microwave sensor, it is responsible for producing the microwave that is applied to a test substance. There are several ways in which the source wave can be produced, with one such way being a Voltage Controlled Oscillator (VCO). A VCO produces a microwave output at a frequency based upon a tuning voltage, varying the value of the tuning voltage enables the sensor to alter the output frequency that is produced. A VCO can be particularly useful for covering a range of frequencies which are closely located together, as a VCO is limited to a specified frequency range.

Another possible source of a microwave is through a crystal oscillator source, the source crystal can produce an exact resonant frequency. The output frequency can then be either directly applied to a test substance or processed through a range of microwave circuits to produce an alternative frequency or an amplified frequency.

There are a range of circuits that can be used including frequency mixers which enable two frequencies to be mixed together to produce two new frequencies based on the sum of the frequencies and on the difference of the frequencies. Mixers are often used with a frequency filter, either a low pass filter or a high pass filter, the use of a filter enables the isolation of a single frequency output from the mixer.

Another common type of microwave circuit used for frequency manipulation is a frequency multiplier. A frequency multiplier produces a harmonic of the input frequency enabling a higher frequency to be produced from a low frequency input.

A combination of microwave circuits combined with a resonant crystal oscillator can be used to produce a specific desired frequency. However, in many instances it may be desirable to cover a range of frequencies close together, particularly during the development stages of the sensor. Once final frequencies are identified it would be possible to design a specific microwave circuit to output the desired frequency.

5.7.2 Microwave Coupler

A coupler is another important component within microwave circuits but is only used where reflected measurements (S_{11}) need to be taken, such as when using an antenna. The coupler enables the input from the microwave source to be isolated from the incoming reflected power allowing the reflected power to be transmitted to the power detector.

In cases where a pass-through measurement(S_{12}) measurement is desired a coupler is not used. In place of a coupler two connections are made, one with the output power connected to the microwave source and the second to the power and phase detector.

5.7.3 Microwave Power Detector

A radio frequency power detector is vital in microwave circuits where a magnitude reading needs to be taken. The power detector takes a microwave input and normally outputs a voltage based on the power of the microwave that was detected. The voltage output corresponds to a reading in dB based upon the manufacture's documentation. Power detectors can often cover a wide range of powers and frequencies but it is important to ensure that the appropriate detector is used based upon experimental data using the VNA, to ensure that the detector is capable of working with the required frequency and the magnitude range that was observed.

It is possible to amplify a signal that is input into a power detector to ensure that it is within the detection range of the power detector, although this comes at a cost of distorting the incoming microwave and adding noise to the overall microwave circuit.

5.7.4 Microwave Phase Detector

A microwave phase detector is like a power detector, though rather than detect the power of an incoming microwave it measures the change in phase from the outgoing wave. A phase detector requires two microwave inputs, the first is a reference wave for which to measure the phase against, this is the outgoing microwave produced by the Microwave source, the second is the incoming microwave that is to be measured. Depending on the type of phase detector used the detector will output the change in phase that is detected, this can then be interpreted and read.

5.7.5 Digital to Analogue Converter

A Digital to Analogue Converter (DAC) is a common Integrated Circuit (IC) that can normally be connected to a microcontroller. The DAC can be used by a microcontroller to take a digital input which is communicated to the IC through a protocol such as I2C or SPI. The DAC then outputs the set output voltage based upon the one set by the micro controller. DACs form an important part of many circuits, allowing for a variable voltage to be output.

Where a VCO is used within a circuit there must be a controllable constant voltage source to enable a microcontroller to select the required frequency. An important factor in an analogue to digital converter is the resolution that is supported, this is normally measured in bits and limits the granularity of voltage that can be controlled by the microcontroller. For example, a 1 Bit DAC would only be able to represent two levels, compared to a 10 bit DAC which could represent 1,024 values.

5.7.6 Analogue to Digital Converter

An analogue to Digital Converter (ADC) provides the opposite functionality to a DAC, providing a method to take an analogue voltage and convert it into a binary value that a microcontroller can process.

In cases where a power detector is used within a microwave circuit an DAC must be connected to the reading output in order to allow the microcontroller to read the output voltage of the power detector.

5.7.7 Microcontroller

A microcontroller plays an important role within any device, it is responsible for coordinating other peripherals such as the ADC and DAC. The microcontroller implements and manages the logic of the sensor, setting the appropriate frequency required using the DAC when required and taking a frequency. Once the frequency is set the microcontroller is responsible for recording the magnitude and phase readings.

Depending upon the purpose of the sensor the microcontroller can either use the captured readings to produce a classification or alternatively the readings can be captured and stored or communicated across a communication platform for storage or processing.

5.7.8 Construction of a microwave sensor

Figure 5-23 shows an example block diagram of how the key components of a microwave sensor are connected using a VCO microwave source to capture magnitude readings. The diagram highlights the key components and connections of components for a microwave sensor and ignores additional components such as capacitors, resistors and voltage regulators as well as any mechanism for data storage or communication.

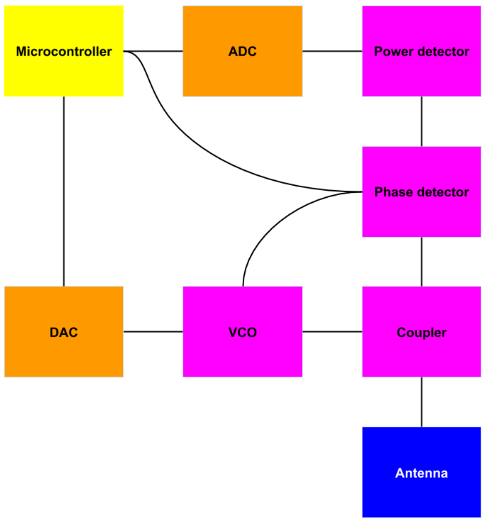

Figure 5-23 A block diagram of a microwave circuit

The block diagram given uses a VCO to enable a range of frequencies to be produced by the sensor, allowing for it to sweep the entire range, or select specific frequencies through the manipulation of a tuning voltage.

In many instances this initial type of circuit would be used to collect additional data after an initial frequency range has been identified using data collected using a VNA. In these cases, an application needs to be capable of collecting data at intervals and storing it for later processing. Figure 5-24 shows the flow of an application that can be used to achieve this through the repeated manipulation of the tuning voltage and the taking and storage of readings alongside the frequency at which the reading was taken.

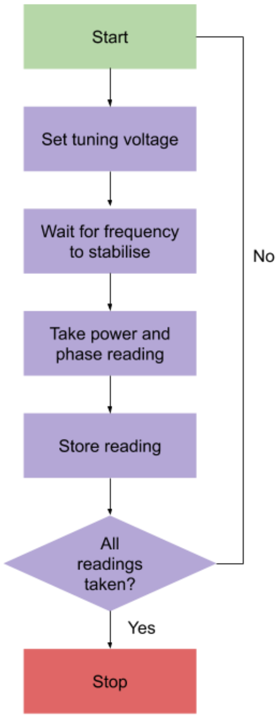

Figure 5-24 A flow chart for a simple microwave sensor for data capture

Based upon the block diagram an initial prototype system was developed using a development board combined with microwave components designed for experimental work and research produced by mini-circuits. Figure 5-25 A shows the internals of the sensor module including two VCOs, a coupler and power detector with the two VCOs connected to a switch to enable the sensor to change between VCOs.

Figure 5-25 B Shows the interface of the sensor, the interface allows the user to input data to allow for a label data set to be produced with labelling including the positioning of the sensor on the wall and the material of the wall with data being saved to an SD card connected to the development board.

A **B**

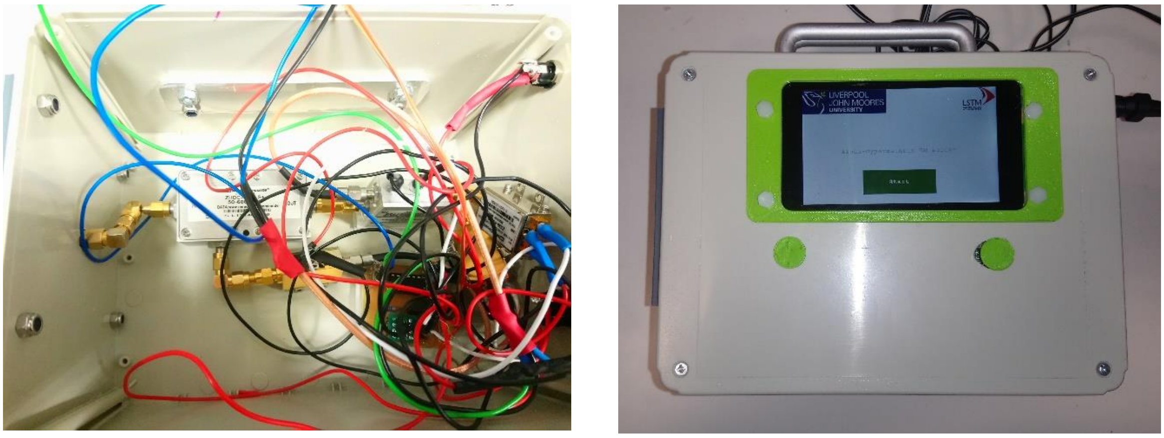

Figure 5-25 A) The internals of the prototype sensor using mini-circuits B) touch screen interface of the sensor

The VCOs selected for the initial prototype were selected based on the experimental work and subsequent analysis undertaken for the work presented in 5.4 which examined the detection of Alpha-Cypermethrin using an S_{11} measurement. Based upon the findings two frequency ranges were identified 1.8 GHz to 2.0 GHz and 4.2 GHz to 4.5 GHz with multiple features identified that could be used to detect Alpha-Cypermethrin.

To enable the two VCOs to be used by the sensor a microwave switch was used between the VCOs and the coupler. The switch enables a single VCO to be used as the microwave source for the circuit by allowing a single VCO to be used as the input to the microwave coupler, selected by the microcontroller using a high or low voltage to select the appropriate VCO.

Subsequent work identified multiple frequencies within the 2.6 GHz to 3.1 GHz frequency that appeared to be sensitive to the presence of Alpha Cypermethrin. Based upon these findings the next stage was to develop a more targeted sensor board that could be used for further data collection in the field with the ability to produce several prototype systems that were easy to use and portable.

The previous generation prototype was large and, when using the integrated 12V lead acid battery, weighed a significant amount. The first-generation prototype was also hand made using wire connections, during field trials and transportation this wiring could come loose with no way for the user to identify that the prototype was no longer functional it could lead to erroneous data being collected. To extend the use of the microwave sensor in field trials it was necessary to develop several boards that were consistent, meaning a single standard board would be advantageous.

The issues highlighted by the previous generation of prototype led to the identification of three key requirements for the next generation of prototype, firstly the prototype needed to be lightweight, using a power source that was easily accessible in the field such as a 9V battery. The second was that the next generation needed to be durable and withstand drops and bumps during use and transportation. Finally, there needed to be consistency between the prototypes developed, enabling a number to be produced and distributed for field trials and further data collection.

Based upon the need to ensure consistency among the sensor prototypes it was decided to use a single PCB. Using a universal PCB provides several advantages including the ability to produce several sensors with the same layout and connections, reducing inconsistencies produced by wiring and soldering in the previous generation of prototype. Another advantage of using a single PCB is that it is more resistance to impacts during usage and transportation as components are secured to a board.

The power source for the board was another key consideration, due to the use in the field it was necessary to ensure that the power source was easily accessible and changeable when required. The power source also needed to be able to provide the required voltage for all components, with the largest voltage requirement coming from the VCO as other components such as DACs and microcontrollers commonly function at 3.3 V or 5 V.

Based upon the frequencies selected, an appropriate VCO was selected that required 10 V meaning a minimum input voltage of over 10 V was required to ensure a voltage regulator could be used to provide a stable 10 V power rail. In addition to the 10 V power rail a 5 V power rail and a 3.3 V power rail were also included in the design to provide power to other components, the use of an intermediary 5 V power regulator also reduced the amount of voltage that the 3.3 V regulator would have to handle compared to if it was directly connected to the 10 V rail.

Based upon the prototype the next stage of development involved the development of a specific PCB sensor board. The creation of a specific PCB enables a significant reduction in the size of the components and therefore the overall size of the sensor. The developed sensor board was designed using Auto Desk Eagle software and used a single VCO removing the need for a switch.

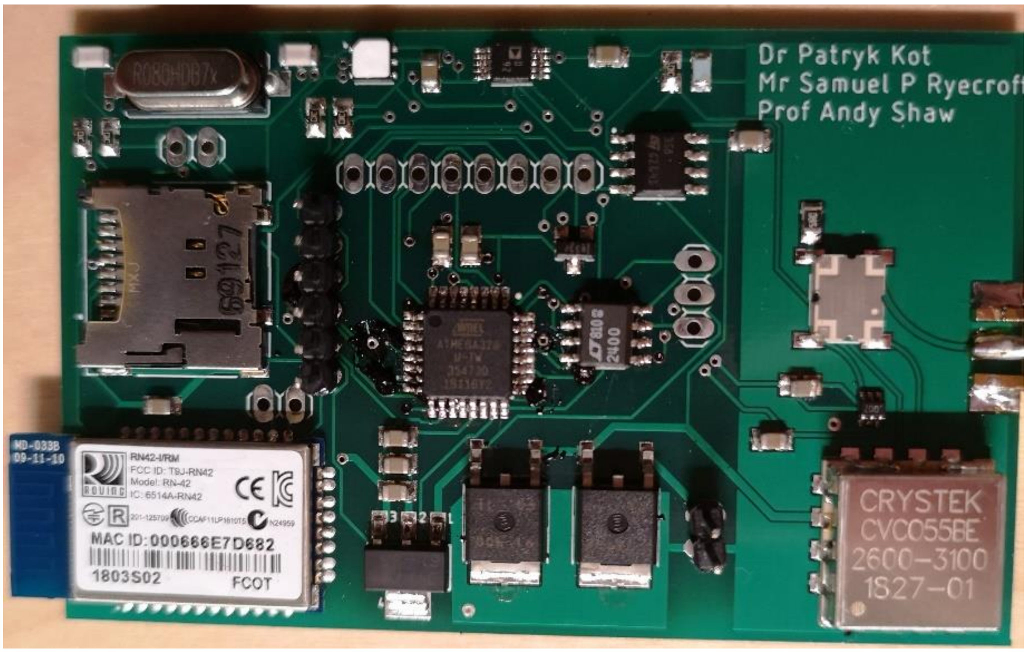

Figure 5-26 Developed sensor PCB

Figure 5-26 shows the developed sensor board developed to take S_{11} magnitude readings, using a single VCO to produce microwave frequencies within the 2.6 GHz to 3.1 GHz range. The board uses a power detector and microwave coupler to enable magnitude readings to be taken, with the power detector connected to a Linear Technology ADC to mesure the voltage produced by the power detector.

The frequency produced by the VCO is controlled using a tuning voltage provided using a DAC. Due to limitations of the DAC only being able to produce a voltage of up to 3.3V and the range of voltages required for tuning the VCO being 10V, the board uses an amplification circuit to multiply the tuning voltage. The use of this amplification reduces the granularity of tuning but allows the use of the full range of the VCO.

Due to the intended application of the sensor board it needs to support a wide range of input voltages and provide the required voltages for the components on the board. Due to this the board uses three voltage regulators to provide 10V, 5V and 3.3V voltage rails while allowing input voltages of up to 24V, enabling power to be provided from a range of sources including two 9V batteries in series or a 12V lead acid battery.

The sensor board provides two methods for data transmission and storage, the first is through an SD card adapter which connects directly to the microcontroller using SPI, enabling the sensor to output data files to the SD card. The board also includes a Bluetooth module connected to the microcontroller using UART, which can be used to relay data to a connected device.

The sensor board also includes a temperature and humidity sensor to capture additional data. In future work these additional parameters could be used to better understand the readings captured and how the changes in humidity and temperature impact on the readings captured by the board.

5.8 Conclusion

Results from both data sets examined support that machine learning can be used to identify potential frequencies within microwave data sets that could be used to highlight areas that merit further investigation as well as frequencies that could be used to develop a custom microwave sensor board based on frequencies identified.

The work presented focuses on using the approach to magnitude readings however it is possible that the approach could also be applied to phase features in place of magnitude features, it is also possible that the approach could be used with a combination of magnitude and phase features.

In both cases examined, the information gain filter provided strong results with clear clusters being formed. Wrapper feature selection gave mixed results, in some instances ensemble methods such as Random Forest and Gradient Boosted Models performed well, while in others they did not.

Support Vector Machine feature selection did not perform well in both instances, the clusters formed were dispersed, and in both instances, classes overlapped. The poor performance of SVM may in part be attributed to the one versus all approach that SVM takes when used in non-binary classification.

Random Forest appears to have worked well within both data sets, with some overlap being observed within the Geosmin data set. Similarly, this was observed in the results produced through the gradient boosted machine approach, with strong clear clustering shown in the case of alpha-cypermethrin but considerably more dispersed clustering in the case of Geosmin.

The consistent performance of an information gain filter combined with its low computational cost makes it a strong candidate for being regularly used in microwave spectroscopy data sets for identification of features that could be used to identify frequencies that could be used in classification, with both data sets showing a clear separation of features identified through information gain on its own.

Using the frequencies identified by the feature selection processes it is possible to carry forward this information to create a targeted microwave sensor that can be used for fieldwork and further data collection. Though the use of a controller combined with an appropriate microwave source and accompanying circuitry it is possible to develop a sensor that can generate a microwave frequency within a desired frequency range and capture and store microwave readings for further processing and analysis.

The development of a targeted microwave sensor board offers several advantages over capturing data through the use of a VNA. The first advantage is the associated risk with using a VNA within a field environment with the cost of a VNA being significantly higher than that of the developed microwave sensor board, during fieldwork there may be times where the equipment is at risk of damage where it is preferable to risk the less costly microwave board. As well as the risk there is also the advantage of being able to produce multiple sensor boards enabling a larger amount of data to be captured.

The microwave sensor board also offers the advantage of being smaller, more lightweight, and portable than a VNA. As well as being portable the developed microwave board is also simpler to use, allowing those without training in the usage of a VNA to quickly and easily take readings with the developed sensor board.

The developed sensor board does pose some issues such as it being less sensitive then a VNA. Therefore, it is important that the selected frequencies show magnitude readings within the rage of the selected power detector. The developed microwave board also poses some issues with calibration and consistency; while the developed microwave board improves significantly with consistency compared to the previous generation, there is still likely to be some inconsistency between the assembled microwave boards due to the assembly process.

While the developed microwave sensor board focuses on the capture of S_{11} readings, using a coupler to pass the reflected power on to the power detector it would be possible to develop a microwave sensor board for capture of S_{12} readings. The developed board would only require modest changes with the removal of the microwave coupler and connecting the microwave source to a suitable connector with a second connector connected directly to the power detector.

The developed microwave board also offers the opportunity to be coupled with a wireless communication platform due to its small size and flexibility. Connecting the developed board to a communication platform could allow for data to be continually captured and processed enabling remote monitoring and data capture. The data collected could then be used for further processing to build a larger data set or ultimately identify substances within the monitored environment.

6 Conclusions and future work

6.1 Introduction

This thesis presents the importance of water quality, how it can impact on people's daily lives and how contamination can pose serious issues to human health as well as having serious environmental impacts. Current approaches to monitoring are limited by the need to manually capture samples before transporting them back for analysis. Not only does this create additional toil for water companies, but it limits the amount of data companies can collected due to the physical constraints of manual sample capture and introduces a delay between contamination of water intake sites and the detection of the contaminant.

This thesis focuses on the need for the ability to carry out wider coverage of water catchment sites through the capture of more data across a water catchment site. This raises two fundamental issues, for such a system we need to be able to reliably transmit this large quantity of data to a place where it can be further processed, the second issue is the need for a cost-effective sensor capable of detecting contaminants without the need for manual intervention or the production of waste product that would add additional maintenance requirements.

To achieve a communication mechanism that could relay data collected from sensors over a distributed site such as a reservoir we have examined wireless sensor networks, and the considerations that must be taken. We have seen that while conventional wireless sensor networks rely on radio frequency communications underwater wireless sensor networks currently use optical communications or acoustic communications, each of which suffers its own respective problems. We have seen that work has been undertaken with success to use Radio Frequency communications in an underwater environment.

We have seen that the current state-of-the-art sensing technology for two key contaminants for water industry: pesticides and Geosmin, are limited, with the current state of the art unable to meet the demands required for deployment within an underwater wireless sensor network, or unable to detect contaminants demanded by industry. We have seen the potential that microwave sensors have to offer, and their applications in detecting other contaminants within water. We have also examined the use of machine learning techniques to examine and process data, and their previous applications within microwave sensor data sets.

In attempting to achieve the development of an underwater wireless sensor network for the detection of contaminants in water catchment sites, we need to address two major challenges that prevent this. Firstly, exploring the use of radio frequency communications within an underwater wireless sensor network. Starting with lab-based experimentations exploring the role of conductivity on signal attenuation, before moving to two field trials, one in the Leeds Liverpool canal which explored the impact that baud rates play on achievable communication distances. Further work built upon the first set of field trials, undertaken in a real-world water catchment site, in a significant body of water, showing improved communication distances and demonstrating multi-hop routing with signals crossing the water-air boundary and the air-water boundary.

The second challenge was developing a sensor that can detect Geosmin in pesticides at levels which may be of use to industry and importantly in a manner that could work with a long-term deployment as part of an underwater wireless sensor network. This thesis presents a feature selection approach to microwave spectroscopy data, that has been applied to microwave spectroscopy data from two types of microwave sensor with two contaminants. The results show that in both cases the approach was able to identify microwave frequencies within the data sets that could be used to detect the respective contaminants.

6.2 Contribution to novelty

This thesis has presented the use of underwater wireless communications using radio frequency to achieve communication distances of 17 metres to form a multi-hop wireless sensor network to relay data in real time from sensors to a sink node, examining the requirements of an underwater sensor network in S Ryecroft et al[136] and presenting real world experimental results in S Ryecroft et al[137] .

The results from the field trials undertaken at the Leeds Liverpool canal and Hurlstone revisor show the ability to communicate in an underwater environments at a data rate of 1.2 Kbps with a separation of up to 17 meters which could be used to form a underwater wireless sensor network in an underwater environment. Results from the Leeds Liverpool canal also show that a faster data rate of 25 Kbps could be used at shorter distances of 5 meters which could be used where quicker data rates could be useful at shorter distances such as wireless data transfer between a base station and a UAV.

This thesis has also presented the use of machine learning to identify frequencies that could be used in the development of targeted microwave sensors to detect contaminants including Geosmin and alpha cypermethrin as presented in S Ryecroft et al[138] and P Kot et al[139] with a patent application in progress in part covering the data processing approach.

The results of the application of machine learning to microwave data sets show that the large data sets produced by microwave spectroscopy can be reduced significantly and frequencies sensitive to substances can be quickly identified. Using the machine learning feature selection process to identify frequencies it is then possible to develop a targeted sensor that can emit the identified frequencies.

6.3 Further work

The implementation of an underwater multi-hop wireless sensor network demonstrates a contribution to knowledge, however the research work presented raises further research challenges and questions including the impact that temperature and depth play in communication distances achieved. Similarly, the contributions made to use of machine learning within microwave sensors and the development of microwave sensors for the detection of Geosmin and alpha cypermethrin raise further challenges such as accounting for the build-up of biofilm on the sensor and temperature compensation to readings.

6.3.1 Accounting for deployment environment factors in underwater RF communications

The work presented has shown that radio frequency communications can be used in an underwater environment. Previous works have indicated that signal propagation happens by crossing the water-air boundary meaning that deployment depth will impact on the communication distances achieved, however this is an important factor that will need to be examined to develop the concept of an underwater wireless sensor network for the detection of contaminants in water. Similar other environmental factors such as temperature are likely to also influence the maximum effective communication range due to thermal noise.

6.3.2 Large scale deployment

The work put forward supporting evidence for the use of radio frequency communications in an underwater environment establishing data rates and distances that can be achieved. The work has shown a small-scale application of a multi-hop wireless sensor network using controlled flooding to route network traffic. The work also presents simulations of the likely achievable network lifetimes for given scenarios with a real-world proven hardware platform. Further work now needs to be undertaken into the long-term performance of an underwater wireless sensor network using the presented platform, examining the lifetime of the network against the simulated results.

6.3.3 Biofilm build up

The work presented has examined the use of microwave sensor in lab-controlled environments, however in real-world environments and prolonged deployment periods there is likely to be a build-up of biofilm over time. One of the benefits of microwave sensors over other sensor technologies is the potential for the sensor to penetrate and pass through these biofilm build-ups; however more research needs to be undertaken into how this biofilm will impact on the readings taken. It is possible that the work presented using machine learning techniques could be developed and expanded to account for and cancel out biofilm as it builds over time enabling longer term deployment of microwave sensors in these types of application.

6.3.4 Accounting for temperature

The data collected during experimental works was taken during the same lab session, meaning that the temperature of samples was uniform for all readings. However, it is a well-known fact that temperature can be a considerable factor when working with microwave sensors. It is possible that sensors could be calibrated through the use of a known reference sample deployed alongside the sensor such as a methanol blank that was used in the case of the experimental work with Geosmin. Taking readings from this reference sample deployed alongside the sample enabled readings to be taken at any temperature.

7 Appendix

7.1 References

[1] W. H. Organization. "Drinking-water." https://www.who.int/news-room/fact-sheets/detail/drinking-water (accessed 21/02/2020.

[2] R. Damania, S. Desbureaux, A.-S. Rodella, J. Russ, and E. Zaveri, *Quality Unknown*. Washington, DC: World Bank, 2019, p. 114.

[3] H. Stan and T. Heberer, "Pharmaceuticals in the aquatic environment," *Analusis*, vol. 7, no. 25, pp. M20-M23, 1997.

[4] K. Kümmerer, "Drugs in the environment: emission of drugs, diagnostic aids and disinfectants into wastewater by hospitals in relation to other sources – a review," *Chemosphere*, vol. 45, no. 6, pp. 957-969, 2001/11/01/ 2001, doi: https://doi.org/10.1016/S0045-6535(01)00144-8.

[5] J. Fick, H. Söderström, R. H. Lindberg, C. Phan, M. Tysklind, and D. J. Larsson, "Contamination of surface, ground, and drinking water from pharmaceutical production," *Environmental Toxicology and Chemistry*, vol. 28, no. 12, pp. 2522-2527, 2009.

[6] T. Heberer, "Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: a review of recent research data," *Toxicology Letters*, vol. 131, no. 1, pp. 5-17, 2002/05/10/ 2002, doi: https://doi.org/10.1016/S0378-4274(02)00041-3.

[7] D. W. Kolpin *et al.*, "Pharmaceuticals, Hormones, and Other Organic Wastewater Contaminants in U.S. Streams, 1999−2000: A National Reconnaissance," *Environmental Science & Technology*, vol. 36, no. 6, pp. 1202-1211, 2002/03/01 2002, doi: 10.1021/es011055j.

[8] R. H. Lindberg, P. Wennberg, M. I. Johansson, M. Tysklind, and B. A. V. Andersson, "Screening of Human Antibiotic Substances and Determination of Weekly Mass Flows in Five Sewage Treatment Plants in Sweden," *Environmental Science & Technology*, vol. 39, no. 10, pp. 3421-3429, 2005/05/01 2005, doi: 10.1021/es048143z.

[9] T. Steger-Hartmann, K. Kümmerer, and A. Hartmann, "Biological degradation of cyclophosphamide and its occurrence in sewage water," *Ecotoxicology and environmental safety*, vol. 36, no. 2, pp. 174-179, 1997.

[10] D. Adriano, "Trace elements in the terrestrial environment.− Springer-Verlag," *New York*, vol. 533, 1986.

[11] M. C. Jung, "Heavy metal contamination of soils and waters in and around the Imcheon Au–Ag mine, Korea," *Applied Geochemistry*, vol. 16, no. 11, pp. 1369-1375, 2001/08/01/ 2001, doi: https://doi.org/10.1016/S0883-2927(01)00040-3.

[12] I. Thornton, "Geochemical aspects of heavy metal pollution and agriculture in England and Wales," *Reference Book, Ministry of Agriculture, Fisheries and Food*, no. 326, pp. 105-125, 1980.

[13] G. Merrington and B. Alloway, "The transfer and fate of Cd, Cu, Pb and Zn from two historic metalliferous mine sites in the UK," *Applied Geochemistry*, vol. 9, no. 6, pp. 677-687, 1994.

[14] R. Khaleel, K. Reddy, and M. Overcash, "Transport of potential pollutants in runoff water from land areas receiving animal wastes: a review," *Water research*, vol. 14, no. 5, pp. 421-436, 1980.

[15] S. Z. Cohen, S. M. Creeger, R. F. Carsel, and C. G. Enfield, "Potential Pesticide Contamination of Groundwater from Agricultural Uses," in *Treatment and Disposal of Pesticide Wastes*, vol. 259, (ACS Symposium Series, no. 259): American Chemical Society, 1984, ch. 18, pp. 297-325.

[16] A. N. Sharpley and P. J. Withers, "The environmentally-sound management of agricultural phosphorus," *Fertilizer research*, vol. 39, no. 2, pp. 133-146, 1994.

[17] F. Edition, "Guidelines for drinking-water quality," *WHO chronicle*, vol. 38, no. 4, pp. 104-8, 2011.

[18] *Water, England and Wales. The Water Supply (Water Quality) Regulations 2016*.

[19] H. Govenment, "Water Act," 1989. [Online]. Available: http://www.legislation.gov.uk/ukpga/1989/15/contents.

[20] (1973). *Water Act*. [Online] Available: https://www.legislation.gov.uk/ukpga/1973/37/contents

[21] L. Jen-Hao *et al.*, "Developed urban air quality monitoring system based on wireless sensor networks," in *2011 Fifth International Conference on Sensing Technology*, 28 Nov.-1 Dec. 2011 2011, pp. 549-554, doi: 10.1109/ICSensT.2011.6137040.

[22] M. Ryu, J. Yun, T. Miao, I. Ahn, S. Choi, and J. Kim, "Design and implementation of a connected farm for smart farming system," in *2015 IEEE SENSORS*, 1-4 Nov. 2015 2015, pp. 1-4, doi: 10.1109/ICSENS.2015.7370624.

[23] P. P. F. Dheena, G. S. Raj, G. Dutt, and S. V. Jinny, "IOT based smart street light management system," in *2017 IEEE International Conference on Circuits and Systems (ICCS)*, 20-21 Dec. 2017 2017, pp. 368-371, doi: 10.1109/ICCS1.2017.8326023.

[24] D. Ibrahim, B. Froberg, A. Wolf, and D. E. Rusyniak, "Heavy Metal Poisoning: Clinical Presentations and Pathophysiology," *Clinics in Laboratory Medicine*, vol. 26, no. 1, pp. 67-97, 2006/03/01/ 2006, doi: https://doi.org/10.1016/j.cll.2006.02.003.

[25] P. Nicolopoulou-Stamati, S. Maipas, C. Kotampasi, P. Stamatis, and L. Hens, "Chemical pesticides and human health: the urgent need for a new concept in agriculture," *Frontiers in public health*, vol. 4, p. 148, 2016.

[26] M. E. E. Alahi, L. Xie, S. Mukhopadhyay, and L. Burkitt, "A temperature compensated smart nitrate-sensor for agricultural industry," *IEEE Transactions on industrial electronics*, vol. 64, no. 9, pp. 7333-7341, 2017.

[27] F. Gao, L. Tang, L. Dai, and L. Wang, "A fluorescence ratiometric nano-pH sensor based on dual-fluorophore-doped silica nanoparticles," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 67, no. 2, pp. 517-521, 2007.

[28] J. Cleary, C. Slater, C. McGraw, and D. Diamond, "An autonomous microfluidic sensor for phosphate: On-site analysis of treated wastewater," *IEEE Sensors Journal*, vol. 8, no. 5, pp. 508-515, 2008.

[29] B. O'Flynn *et al.*, "SmartCoast: a wireless sensor network for water quality monitoring," in *32nd IEEE Conference on Local Computer Networks (LCN 2007)*, 2007: Ieee, pp. 815-816.

[30] (2001). *Factors causing off-taste in waters, and methods and practices for the removal of off-taste and its causes* [Online] Available: http://dwi.defra.gov.uk/research/completed-research/reports/0820.pdf

[31] S. Fried, B. Mackie, and E. Nothwehr, "Nitrate and phosphate levels positively affect the growth of algae species found in Perry Pond," *Tillers,* vol. 4, 01/01 2003.

[32] N. N. Gerber, "Geosmin, from microorganisms, is trans-1, 10-dimethyl-trans-9-decalol," *Tetrahedron Letters,* vol. 9, no. 25, pp. 2971-2974, 1968/01/01/ 1968, doi: https://doi.org/10.1016/S0040-4039(00)89625-2.

[33] D. Rashash, A. Dietrich, and R. Hoehn, "FPA of Selected Odorous Compounds," *Journal / American Water Works Association,* vol. 89, pp. 131-141, 04/01 1997, doi: 10.1002/j.1551-8833.1997.tb08213.x.

[34] M. J. McGuire, "Off-flavor as the consumer's measure of drinking water safety," *Water Science and Technology,* vol. 31, no. 11, pp. 1-8, 1995/01/01/ 1995, doi: https://doi.org/10.1016/0273-1223(95)00448-V.

[35] "Drinking water 2018," 2019. [Online]. Available: http://www.dwi.gov.uk/about/annual-report/2018/CIR-2018-England.pdf

[36] R. McCallum, P. Pendleton, R. Schumann, and M.-U. Trinh, "Determination of geosmin and 2-methylisoborneol in water using solid-phase microextraction and gas chromatography-chemical ionisation/electron impact ionisation-ion-trap mass spectrometry," *Analyst,* 10.1039/A803621G vol. 123, no. 10, pp. 2155-2160, 1998, doi: 10.1039/A803621G.

[37] K. Saito, K. Okamura, and H. Kataoka, "Determination of musty odorants, 2-methylisoborneol and geosmin, in environmental water by headspace solid-phase microextraction and gas chromatography–mass spectrometry," *Journal of Chromatography A,* vol. 1186, no. 1, pp. 434-437, 2008/04/04/ 2008, doi: https://doi.org/10.1016/j.chroma.2007.12.078.

[38] G. De Souza Braga, L. Paterno, and F. Fonseca, "Use of an Electronic Tongue to Detect Geosmin in Distilled Water," vol. 1362, 11/23 2011, doi: 10.1063/1.3651654.

[39] M. Son *et al.*, "Real-time monitoring of geosmin and 2-methylisoborneol, representative odor compounds in water pollution using bioelectronic nose with human-like performance," *Biosensors and Bioelectronics,* vol. 74, 12/15 2015, doi: 10.1016/j.bios.2015.06.053.

[40] T. P. Hensarling and S. K. Waage, "A bromine-based color reaction for the detection of geosmin," *Journal of Agricultural and Food Chemistry,* vol. 38, no. 5, pp. 1236-1237, 1990/05/01 1990, doi: 10.1021/jf00095a017.

[41] S. Y. Chung, J. R. Vercellotti, P. B. Johnsen, and P. H. Klesius, "Development of an enzyme-linked immunosorbent assay for geosmin," *Journal of Agricultural and Food Chemistry,* vol. 39, no. 4, pp. 764-769, 1991/04/01 1991, doi: 10.1021/jf00004a030.

[42] S W Kang, M. Lim, and N. A. Lee, "Gold seeded magnetic nanoparticles as a multifunctional platform for a competitive Enzyme Linked Immunosorbent Assay (ELISA) to detect geosmin in water," 2014. [Online]. Available: https://elsevier.conference-services.net/viewsecurePDF.asp?conferenceID=3603&abstractID=780959.

[43] K. Sato and S. Tanaka, "DETERMINATION OF ORGANIC COMPOUNDS IN WATER USING INHIBITORY EFFECT ON THE LUMINOL-HYDROGEN PEROXIDE-PEROXODISULFATE CHEMILUMINESCENCE REACTION," *Analytical Sciences*, vol. 7, no. Supple, pp. 1167-1170, 1991, doi: 10.2116/analsci.7.Supple_1167.

[44] D. R. Mimmo and L. C. McEwen, *Pesticides* (Handbook of Ecotoxicology). Blackwell Scientific Publications, Oxford, 1994.

[45] W. H. O. I. A. F. R. O. CANCER, *Some Thyrotropic Agents* (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans no. 79). IARCPress

[46] W. H. Organization, *Public health impact of pesticides used in agriculture*. World Health Organization, 1990.

[47] B. Alewu and C. Nosiri, "Pesticides and human health," *Pesticides in the Modern World–Effects of Pesticides Exposure. InTech*, pp. 231-50, 2011.

[48] S. Berijani, Y. Assadi, M. Anbia, M.-R. Milani Hosseini, and E. Aghaee, "Dispersive liquid–liquid microextraction combined with gas chromatography-flame photometric detection: Very simple, rapid and sensitive method for the determination of organophosphorus pesticides in water," *Journal of Chromatography A*, vol. 1123, no. 1, pp. 1-9, 2006/08/04/ 2006, doi: https://doi.org/10.1016/j.chroma.2006.05.010.

[49] S. Suwansa-ard, P. Kanatharana, P. Asawatreratanakul, C. Limsakul, B. Wongkittisuksa, and P. Thavarungkul, "Semi disposable reactor biosensors for detecting carbamate pesticides in water," *Biosensors and Bioelectronics*, vol. 21, no. 3, pp. 445-454, 2005/09/15/ 2005, doi: https://doi.org/10.1016/j.bios.2004.11.005.

[50] A. A. Boyd-Boland, S. Magdic, and J. B. Pawliszyn, "Simultaneous determination of 60 pesticides in water using solid-phase microextraction and gas chromatography–mass spectrometry," *Analyst*, 10.1039/AN9962100929 vol. 121, no. 7, pp. 929-937, 1996, doi: 10.1039/AN9962100929.

[51] M. Huang, D. Shen, L. M. Chow, and M. Yang, "Correlations of the impedance parameters and conductivity and permittivity of liquid and gel phases in a series piezoelectric quartz crystal sensor," *Sensors and Actuators B: Chemical*, vol. 72, no. 1, pp. 21-27, 2001/01/05/ 2001, doi: https://doi.org/10.1016/S0925-4005(00)00627-4.

[52] T. E. o. E. Britannica. "Dielectric constant." Encyclopædia Britannica, inc. https://www.britannica.com/science/dielectric-constant (accessed 26/06/2020, 2020).

[53] T. E. o. E. Britannica. "Dielectric loss." Encyclopædia Britannica, inc. https://www.britannica.com/science/dielectric-loss (accessed 26/06/2020, 2020).

[54] K. H. Teng, I. Idowu, P. Kot, A. Shaw, and M. Muradov, "Non-destructive Electromagnetic Wave Sensor for Hazardous Biological Materials," in *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, 7-10 Oct. 2019 2019, pp. 651-655, doi: 10.1109/DeSE.2019.00122.

[55] K. Zhang, R. K. Amineh, Z. Dong, and D. Nadler, "Microwave Sensing of Water Quality," *IEEE Access,* vol. 7, pp. 69481-69493, 2019, doi: 10.1109/ACCESS.2019.2918996.

[56] R. Mirzavand, M. M. Honari, B. Laribi, B. Khorshidi, M. Sadrzadeh, and P. Mousavi, "An Unpowered Sensor Node for Real-Time Water Quality Assessment (Humic Acid Detection)," *Electronics,* vol. 7, no. 10, 2018, doi: 10.3390/electronics7100231.

[57] I. Frau, O. Korostynska, A. Mason, and P. Byrne, "Comparison of electromagnetic wave sensors with optical and low-frequency spectroscopy methods for real-time monitoring of lead concentrations in mine water," *Mine Water and the Environment,* vol. 37, no. 3, pp. 617-624, 2018.

[58] A. Azmi *et al.*, "Performance of Coating Materials on Planar Electromagnetic Sensing Array to Detect Water Contamination," *IEEE Sensors Journal,* vol. 17, no. 16, pp. 5244-5251, 2017, doi: 10.1109/JSEN.2017.2720701.

[59] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[60] Z. Xinfeng, W. Shengchang, and Z. Yan, "Application of support vector machine and least squares vector machine to freight volume forecast," in *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, 24-26 June 2011 2011, pp. 104-107, doi: 10.1109/RSETE.2011.5964227.

[61] S. Choudhury, S. Bandyopadhyay, S. Mukhopadhyay, and S. Mukherjee, "Vector quantization and multi class support vector machines based fingerprint classification," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, 26-27 Aug. 2016 2016, vol. 2, pp. 1-4, doi: 10.1109/INVENTIVE.2016.7824828.

[62] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *Journal of machine learning research,* vol. 2, no. Dec, pp. 265-292, 2001.

[63] Y. Lee, Y. Lin, and G. Wahba, "Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data," *Journal of the American Statistical Association,* vol. 99, no. 465, pp. 67-81, 2004.

[64] J. Weston and C. Watkins, "Support vector machines for multi-class pattern recognition," in *Esann*, 1999, vol. 99, pp. 219-224.

[65] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of machine learning research,* vol. 5, no. Jan, pp. 101-141, 2004.

[66] V. Bijalwan, V. Kumar, P. Kumari, and J. Pascual, "KNN based machine learning approach for text and document mining," *International Journal of Database Theory and Application,* vol. 7, no. 1, pp. 61-70, 2014.

[67] K. Alkhatib, H. Najadat, I. Hmeidi, and M. K. A. Shatnawi, "Stock price prediction using k-nearest neighbor (kNN) algorithm," *International Journal of Business, Humanities and Technology,* vol. 3, no. 3, pp. 32-44, 2013.

[68] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.

[69] M. S. Alam and S. T. Vuong, "Random forest classification for detecting android malware," in *2013 IEEE international conference on green computing and communications and IEEE Internet of Things and IEEE cyber, physical and social computing*, 2013: IEEE, pp. 663-669.

[70] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forest classification of multisource remote sensing and geographic data," in *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, 2004, vol. 2: IEEE, pp. 1049-1052.

[71] L. Breiman, "Arcing the edge," Technical Report 486, Statistics Department, University of California at …, 1997.

[72] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189-1232, 2001.

[73] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean, "Boosting algorithms as gradient descent," in *Advances in neural information processing systems*, 2000, pp. 512-518.

[74] I. Semanjski and S. Gautama, "Smart city mobility application—Gradient boosting trees for mobility prediction and analysis based on crowdsourced data," *Sensors*, vol. 15, no. 7, pp. 15974-15987, 2015.

[75] M. Zięba, S. K. Tomczak, and J. M. Tomczak, "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction," *Expert Systems with Applications*, vol. 58, pp. 93-101, 2016/10/01/ 2016, doi: https://doi.org/10.1016/j.eswa.2016.04.001.

[76] S. Cashman, O. Korostynska, A. Shaw, P. Lisboa, and L. Conroy, "Detecting the presence and concentration of nitrate in water using microwave spectroscopy," *IEEE Sensors Journal*, vol. 17, no. 13, pp. 4092-4099, 2017.

[77] L. Harrsion, M. Ravan, D. Tandel, K. Zhang, T. Patel, and R. K Amineh, "Material Identification Using a Microwave Sensor Array and Machine Learning," *Electronics*, vol. 9, no. 2, p. 288, 2020.

[78] M. Soltani and M. Omid, "Detection of poultry egg freshness by dielectric spectroscopy and machine learning techniques," *LWT - Food Science and Technology*, vol. 62, no. 2, pp. 1034-1042, 2015/07/01/ 2015, doi: https://doi.org/10.1016/j.lwt.2015.02.019.

[79] R. K. Kodali, V. S. K. A, S. Bhandari, and L. Boppana, "Energy efficient m- level LEACH protocol," in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 10-13 Aug. 2015 2015, pp. 973-979, doi: 10.1109/ICACCI.2015.7275737.

[80] BlueComm. "BlueComm 200 series." https://www.sonardyne.com/product/bluecomm-200-wireless-underwater-video/ (accessed 26/12/2019, 2019).

[81] F. Hanson and S. Radic, "High bandwidth underwater optical communication," *Appl. Opt.*, vol. 47, no. 2, pp. 277-283, 2008/01/10 2008, doi: 10.1364/AO.47.000277.

[82] J. Snow *et al.*, *Underwater propagation of high-data-rate laser communications pulses* (San Diego '92). SPIE, 1992.

[83] D. Anguita, D. Brizzolara, G. Parodi, and Q. Hu, "Optical wireless underwater communication for AUV: Preliminary simulation and experimental results," in *OCEANS 2011 IEEE - Spain*, 6-9 June 2011 2011, pp. 1-5, doi: 10.1109/Oceans-Spain.2011.6003598.

[84] J. Wang, C. Lu, S. Li, and Z. Xu, "100 m/500 Mbps underwater optical wireless communication using an NRZ-OOK modulated 520 nm laser diode," *Opt. Express*, vol. 27, no. 9, pp. 12171-12181, 2019/04/29 2019, doi: 10.1364/OE.27.012171.

[85] C. D. Hodgman and S. Lind, "Handbook of chemistry and physics," *The Journal of Physical Chemistry,* vol. 53, no. 7, pp. 1139-1139, 1949.

[86] P. W. Zitzewitz and P. T. R. Committee, "Physics: Principles and problems," *The Physics Teacher,* vol. 37, no. 5, pp. 292-293, 1999.

[87] Teledynemarine. "ACOUSTIC MODEMS." http://www.teledynemarine.com/acoustic-modems (accessed.

[88] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater acoustic networks," *IEEE journal of oceanic engineering,* vol. 25, no. 1, pp. 72-83, 2000.

[89] M. Stojanovic, "Acoustic (underwater) communications," *Wiley Encyclopedia of Telecommunications,* 2003.

[90] H. Slabbekoorn, N. Bouton, I. van Opzeeland, A. Coers, C. ten Cate, and A. N. Popper, "A noisy spring: the impact of globally rising underwater sound levels on fish," *Trends in Ecology & Evolution,* vol. 25, no. 7, pp. 419-427, 2010/07/01/ 2010, doi: https://doi.org/10.1016/j.tree.2010.04.005.

[91] X. Tu, A. Song, and X. Xu, "Prefix-Free Frequency Domain Equalization for Underwater Acoustic Single Carrier Transmissions," *IEEE Access,* vol. 6, pp. 2578-2588, 2018, doi: 10.1109/ACCESS.2017.2784388.

[92] S. Zheng, X. Cao, F. Tong, G. Zhang, and Y. Dong, "Performance evaluation of acoustic network for underwater autonomous vehicle in confined spaces," in *2018 IEEE 8th International Conference on Underwater System Technology: Theory and Applications (USYS),* 1-3 Dec. 2018 2018, pp. 1-4, doi: 10.1109/USYS.2018.8778974.

[93] A. K. Morozov and J. C. Preisig, "Underwater acoustic communications with multi-carrier modulation," in *OCEANS 2006*, 2006: IEEE, pp. 1-6.

[94] A. Abdi and H. Guo, "A new compact multichannel receiver for underwater wireless communication networks," *IEEE Transactions on Wireless Communications,* vol. 8, no. 7, pp. 3326-3329, 2009, doi: 10.1109/TWC.2009.070566.

[95] J. Lloret, S. Sendra, M. Ardid, and J. J. P. C. Rodrigues, "Underwater Wireless Sensor Communications in the 2.4 GHz ISM Frequency Band," *Sensors*, vol. 12, no. 4, pp. 4237-4264, 2012. [Online]. Available: https://www.mdpi.com/1424-8220/12/4/4237.

[96] A. A. Abdou *et al.*, "A matched Bow-tie antenna at 433MHz for use in underwater wireless sensor networks," *Journal of Physics Conference Series,* vol. 450, p. 2048, 06/01 2013, doi: 10.1088/1742-6596/450/1/012048.

[97] A. Shaw, A. i. Al-Shamma'a, S. R. Wylie, and D. Toal, "Experimental Investigations of Electromagnetic Wave Propagation in Seawater," in *2006 European Microwave Conference*, 10-15 Sept. 2006 2006, pp. 572-575, doi: 10.1109/EUMC.2006.281456.

[98] E. M. Rubino, D. Centelles, J. Sales, J. V. Marti, R. Marin, and P. J. Sanz, "Wireless Image Compression and Transmission for Underwater Robotic Applications?," *IFAC-PapersOnLine*, vol. 48, no. 2, pp. 288-293, 2015/01/01/ 2015, doi: https://doi.org/10.1016/j.ifacol.2015.06.047.

[99] A. A. Abdou, A. Shaw, A. Mason, A. Al-Shamma'a, J. Cullen, and S. Wylie, "Electromagnetic (EM) wave propagation for the development of an underwater Wireless Sensor Network (WSN)," in *SENSORS, 2011 IEEE*, 28-31 Oct. 2011 2011, pp. 1571-1574, doi: 10.1109/ICSENS.2011.6127319.

[100] B. Kelley and K. Naishadham, "RF multicarrier signaling and antenna systems for low SNR broadband underwater communications," in *2013 IEEE 13th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, 21-23 Jan. 2013 2013, pp. 240-242, doi: 10.1109/SiRF.2013.6489492.

[101] B. F. Bush, V. K. Tripp, and K. Naishadham, "Practical modeling of radio wave propagation in shallow seawater," in *Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation*, 8-14 July 2012 2012, pp. 1-2, doi: 10.1109/APS.2012.6348625.

[102] W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design* New York, Chichester, Weinheim, Brisbane, Toronto, Singapore: John Wiley & Sons, inc, 1998.

[103] O. Aboderin, L. M. Pessoa, and H. M. Salgado, "Wideband dipole antennas with parasitic elements for underwater communications," in *OCEANS 2017 - Aberdeen*, 19-22 June 2017 2017, pp. 1-6, doi: 10.1109/OCEANSE.2017.8084865.

[104] S. I. Inácio *et al.*, "Dipole antenna for underwater radio communications," in *2016 IEEE Third Underwater Communications and Networking Conference (UComms)*, 30 Aug.-1 Sept. 2016 2016, pp. 1-5, doi: 10.1109/UComms.2016.7583457.

[105] Z. M. Loni, H. G. Espinosa, and D. V. Thiel, "Floating Monopole Antenna on a Tethered Subsurface Sensor at 433 MHz for Ocean Monitoring Applications," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 4, pp. 818-825, 2017, doi: 10.1109/JOE.2016.2639111.

[106] E. A. Karagianni, "Electromagnetic waves under sea: bow-tie antennas design for wi-fi underwater communications," *Progress In Electromagnetics Research*, vol. 41, pp. 189-198, 2015.

[107] "IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pp. 1-3534, 2016, doi: 10.1109/IEEESTD.2016.7786995.

[108] A. A. Abdou, "AN INVESTIGATION OF SHORT RANGE ELECTROMAGNETIC WAVE COMMUNICATION FOR UNDERWATER ENVIRONMENTAL MONITORING UTILISING A SENSOR NETWORK PLATFORM," A. Shaw, A. Mason, and A. I. Al-Shamma'a, Eds., ed, 2014.

[109] K. N. Alvertos, E. Karagianni, K. D. Vardakis, T. K. Mpountas, and D. Kaklamani, *Bow-tie antenna for underwater Wireless Sensor Networks*. 2017, pp. 323-326.

[110] S. I. Inácio *et al.*, "Antenna design for underwater radio communications," in *OCEANS 2016-Shanghai*, 2016: IEEE, pp. 1-6.

[111] O. Aboderin, L. M. Pessoa, and H. M. Salgado, "Analysis of loop antenna with ground plane for underwater communications," in *OCEANS 2017 - Aberdeen*, 19-22 June 2017 2017, pp. 1-6, doi: 10.1109/OCEANSE.2017.8084864.

[112] A. Massaccesi and P. Pirinoli, "Loop antennas for underwater communications," in *12th European Conference on Antennas and Propagation (EuCAP 2018)*, 9-13 April 2018 2018, pp. 1-4, doi: 10.1049/cp.2018.1147.

[113] A. A. Abdou, A. Shaw, A. Mason, A. Al-Shamma, S. Wylie, and J. D. Cullen, *Wireless Sensor Network for underwater communication*. 2012, pp. 1-6.

[114] F. Tombe, "Maxwell's Original Equations," 12/24 2011.

[115] J. C. Maxwell, "VIII. A dynamical theory of the electromagnetic field," *Philosophical transactions of the Royal Society of London*, no. 155, pp. 459-512, 1865.

[116] I. S. Bogie, "Conduction and magnetic signalling in the sea a background review," *Radio and Electronic Engineer*, vol. 42, no. 10, pp. 447-452, 1972, doi: 10.1049/ree.1972.0076.

[117] A. Glisson, "Advanced engineering electromagnetics, by constantine a. balanis [book review]," *IEEE Antennas and Propagation Society Newsletter*, vol. 31, no. 6, pp. 24-26, 1989.

[118] S. Jiang and S. Georgakopoulos, "Electromagnetic wave propagation into fresh water," *Journal of Electromagnetic Analysis and Applications*, vol. 2011, 2011.

[119] I. I. Smolyaninov, Q. Balzano, C. C. Davis, and D. Young, "Surface Wave Based Underwater Radio Communication," *IEEE Antennas and Wireless Propagation Letters*, vol. 17, no. 12, pp. 2503-2507, 2018, doi: 10.1109/LAWP.2018.2880008.

[120] A. Palmeiro, M. Martín, I. Crowther, and M. Rhodes, "Underwater radio frequency communications," in *OCEANS 2011 IEEE - Spain*, 6-9 June 2011 2011, pp. 1-8, doi: 10.1109/Oceans-Spain.2011.6003580.

[121] G. Goertzel, "An algorithm for the evaluation of finite trigonometric series," *The American Mathematical Monthly*, vol. 65, no. 1, pp. 34-35, 1958.

[122] M. H. Weik, "Nyquist criterion," in *Computer Science and Communications Dictionary*. Boston, MA: Springer US, 2001, pp. 1127-1127.

[123] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10-21, 1949, doi: 10.1109/JRPROC.1949.232969.

[124] T. Qiang, W. Bingwen, and D. Zhicheng, "MS-Leach: A Routing Protocol Combining Multi-hop Transmissions and Single-hop Transmissions," in *2009 Pacific-Asia Conference on Circuits, Communications and Systems*, 16-17 May 2009 2009, pp. 107-110, doi: 10.1109/PACCS.2009.13.

[125] M. Tripathi, M. S. Gaur, V. Laxmi, and R. B. Battula, "Energy efficient LEACH-C protocol for Wireless Sensor Network," in *Third International Conference on Computational Intelligence and Information Technology (CIIT 2013)*, 18-19 Oct. 2013 2013, pp. 402-405, doi: 10.1049/cp.2013.2620.

[126] A. Manjeshwar and D. P. Agrawal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," in *Proceedings 15th International Parallel and Distributed Processing Symposium. IPDPS 2001*, 23-27 April 2000 2001, pp. 2009-2015, doi: 10.1109/IPDPS.2001.925197.

[127] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," in *Proceedings, IEEE Aerospace Conference*, 9-16 March 2002 2002, vol. 3, pp. 3-3, doi: 10.1109/AERO.2002.1035242.

[128] J. F. Kurose and K. W. Ross, M. Hirsch, Ed. *Computer Networking a Top-Down Approach*, 6th edition ed. Pearson, 2013.

[129] P. Jain and A. Chaudhary, "The comparison between leach protocol and pegasis protocol based on lifetime of wireless sensor networks," *International Journal of Computer Science and Mobile Computing*, vol. 6, no. 12, pp. 15-19, 2017.

[130] J. Lloret, S. Sendra, M. Ardid, and J. J. P. C. Rodrigues, "Underwater Wireless Sensor Communications in the 2.4 GHz ISM Frequency Band," *Sensors*, vol. 12, no. 4, 2012, doi: 10.3390/s120404237.

[131] *R: A Language and Environment for Statistical Computing*. (2013). Vienna, Austria. [Online]. Available: http://www.R-project.org/

[132] *reshape2*. (2017). [Online]. Available: https://github.com/hadley/reshape

[133] P. Romanski, L. Kotthoff, and M. L. Kotthoff, "Package 'FSelector'," *URL http://cran/r-project. org/web/packages/FSelector/index. html*, 2013.

[134] F. Leisch and E. Dimitriadou, "Machine learning benchmark problems," *R Package, mlbench*, 2010.

[135] M. Kuhn, "Caret: classification and regression training," *Astrophysics Source Code Library*, p. ascl: 1505.003, 2015.

[136] S. P. Ryecroft *et al.*, "Requirements of an underwater sensor-networking platform for environmental monitoring," in *2018 11th International Conference on Developments in eSystems Engineering (DeSE)*, 2018: IEEE, pp. 95-99.

[137] S. Ryecroft *et al.*, "A First Implementation of Underwater Communications in Raw Water Using the 433 MHz Frequency Combined with a Bowtie Antenna," *Sensors*, vol. 19, no. 8, p. 1813, 2019.

[138] S. P. Ryecroft *et al.*, "A novel gesomin detection method based on microwave spectroscopy," in *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, 2019: IEEE, pp. 429-433.

[139] P. Kot *et al.*, "Identification of Optimal Frequencies to Determine Alpha-Cypermethrin using Machine Learning Feature Selection Techniques," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018: IEEE, pp. 1-7.

7.2 Publications

7.2.1 Identification of Optimal Frequencies to Determine Alpha-Cypermethrin using Machine Learning Feature Selection Techniques

Machine learning feature space reduction techniques allow for vast feature spaces to be reduced with little loss or even significant improvements in the reliability of predictions of models. Microwave spectroscopy with feature spaces of over 8000 are not uncommon when considering magnitude and phase. Applying Machine learning techniques to this type of feature space allows for a quicker reduction and helps to identify the most suitable predictive features. The control of insect vectors that transmit diseases including malaria, visceral leishmaniasis and zika rely on the use of insecticide. These diseases affect millions, malaria alone accounted for 214 million new cases resulting in 438, 000 deaths in 2015. One method used in controlling the vectors is through indoor residual spraying, applying insecticide to the wall surface inside houses. Alpha-cypermethrin is one of the insecticides that is currently sprayed in several countries for vector control. Quality assurance and monitoring of the control activities is challenging relying on the use of laboratory-reared insects. This was improved with a chemical based Insecticide Quantification Kit, but these assays have been challenging to operationalise. An electromagnetic sensor is being developed to investigate the potential to detect alpha-cypermethrin. Preliminary experiments were carried out to differentiate tiles sprayed with Technical Grade alpha-cypermethrin, wettable powder containing 5% alpha-cypermethrin and wettable powder with over dose of alpha-cypermethrin using a horn antenna at a frequency range between 1 GHz to 6 GHz. The experimental results indicated the potential use of electromagnetic waves to determine alpha-cypermethrin in a non-destructive manner.

7.2.2 A First Implementation of Underwater Communications in Raw Water Using the 433 MHz Frequency Combined with a Bowtie Antenna

In 2016, there were 317 serious water pollution incidents in the UK, with 78,000 locations where businesses discharge controlled quantities of pollutants into rivers; therefore, continuous monitoring is vital. Since 1998, the environment agency has taken over 50 million water samples for water quality monitoring. The Internet of Things has grown phenomenally in recent years, reaching all aspects of our lives, many of these connected devices use wireless sensor networks to relay data to internet-connected nodes, where data can be processed, analyzed and consumed. However, Underwater wireless communications rely mainly on alternative communication methods such as optical and acoustic, with radio frequencies being an under-exploited method. This research presents real world results conducted in the Leeds and Liverpool Canal for the novel use of the 433 MHz radio frequency combined with a bowtie antenna in underwater communications in raw water, achieving distances of 7 m at 1.2 kbps and 5 metres at 25 kbps.

7.2.3 A Novel Gesomin Detection Method Based on Microwave Spectroscopy

Geosmin contamination in water is a leading cause of odour related complaints to water companies in UK, tainting water with an earthy smell that is detectable by humans in quantities as low as 4 nanograms per litre. Current Geosmin detection methods depend on lab-based equipment, requiring samples to be collected and transported before Geosmin can be tested. This research presents a novel method for the detection of Geosmin in water using Microwave spectroscopy capable of detecting differentiating between four levels of Geosmin contamination: 5 ng/L, 10 ng/L, 0.5 mg/L and 1 mg/L as well as control samples. Frequencies within the 5.4 GHz to 5.9, 6.4 GHz to 6.5 GHz and 7.2 GHz to 7.5 GHz ranges showed significant separation between the sample classes.