Report No.380

epa Research

Innovative Water Monitoring

Authors: Kevin Fitzgibbon, William Whelan-Curtin, Chinna Devarapu, Patricia Loren, Colin O'Sullivan and Ian Aherne

www.epa.ie

Rialtas na hÉireann
Government of Ireland

ENVIRONMENTAL PROTECTION AGENCY

The Environmental Protection Agency (EPA) is responsible for protecting and improving the environment as a valuable asset for the people of Ireland. We are committed to protecting people and the environment from the harmful effects of radiation and pollution.

The work of the EPA can be divided into three main areas:

Regulation: *We implement effective regulation and environmental compliance systems to deliver good environmental outcomes and target those who don't comply.*

Knowledge: *We provide high quality, targeted and timely environmental data, information and assessment to inform decision making at all levels.*

Advocacy: *We work with others to advocate for a clean, productive and well protected environment and for sustainable environmental behaviour.*

Our Responsibilities

Licensing

We regulate the following activities so that they do not endanger human health or harm the environment:

- waste facilities *(e.g. landfills, incinerators, waste transfer stations)*;
- large scale industrial activities (*e.g. pharmaceutical, cement manufacturing, power plants)*;
- intensive agriculture *(e.g. pigs, poultry)*;
- the contained use and controlled release of Genetically Modified Organisms *(GMOs)*;
- sources of ionising radiation *(e.g. x-ray and radiotherapy equipment, industrial sources)*;
- large petrol storage facilities;
- waste water discharges;
- dumping at sea activities.

National Environmental Enforcement

- Conducting an annual programme of audits and inspections of EPA licensed facilities.
- Overseeing local authorities' environmental protection responsibilities.
- Supervising the supply of drinking water by public water suppliers.
- Working with local authorities and other agencies to tackle environmental crime by co-ordinating a national enforcement network, targeting offenders and overseeing remediation.
- Enforcing Regulations such as Waste Electrical and Electronic Equipment (WEEE), Restriction of Hazardous Substances (RoHS) and substances that deplete the ozone layer.
- Prosecuting those who flout environmental law and damage the environment.

Water Management

- Monitoring and reporting on the quality of rivers, lakes, transitional and coastal waters of Ireland and groundwaters; measuring water levels and river flows.
- National coordination and oversight of the Water Framework Directive.
- Monitoring and reporting on Bathing Water Quality.

Monitoring, Analysing and Reporting on the Environment

- Monitoring air quality and implementing the EU Clean Air for Europe (CAFÉ) Directive.
- Independent reporting to inform decision making by national and local government *(e.g. periodic reporting on the State of Ireland's Environment and Indicator Reports)*.

Regulating Ireland's Greenhouse Gas Emissions

- Preparing Ireland's greenhouse gas inventories and projections.
- Implementing the Emissions Trading Directive, for over 100 of the largest producers of carbon dioxide in Ireland.

Environmental Research and Development

- Funding environmental research to identify pressures, inform policy and provide solutions in the areas of climate, water and sustainability.

Strategic Environmental Assessment

- Assessing the impact of proposed plans and programmes on the Irish environment *(e.g. major development plans)*.

Radiological Protection

- Monitoring radiation levels, assessing exposure of people in Ireland to ionising radiation.
- Assisting in developing national plans for emergencies arising from nuclear accidents.
- Monitoring developments abroad relating to nuclear installations and radiological safety.
- Providing, or overseeing the provision of, specialist radiation protection services.

Guidance, Accessible Information and Education

- Providing advice and guidance to industry and the public on environmental and radiological protection topics.
- Providing timely and easily accessible environmental information to encourage public participation in environmental decision-making *(e.g. My Local Environment, Radon Maps)*.
- Advising Government on matters relating to radiological safety and emergency response.
- Developing a National Hazardous Waste Management Plan to prevent and manage hazardous waste.

Awareness Raising and Behavioural Change

- Generating greater environmental awareness and influencing positive behavioural change by supporting businesses, communities and householders to become more resource efficient.
- Promoting radon testing in homes and workplaces and encouraging remediation where necessary.

Management and structure of the EPA

The EPA is managed by a full time Board, consisting of a Director General and five Directors. The work is carried out across five Offices:

- Office of Environmental Sustainability
- Office of Environmental Enforcement
- Office of Evidence and Assessment
- Office of Radiation Protection and Environmental Monitoring
- Office of Communications and Corporate Services

The EPA is assisted by an Advisory Committee of twelve members who meet regularly to discuss issues of concern and provide advice to the Board.

EPA RESEARCH PROGRAMME 2021–2030

Innovative Water Monitoring

(2016-W-LS-12)

EPA Research Report

Prepared for the Environmental Protection Agency

by

Munster Technological University and Hydrolight Ltd

Authors:

Kevin Fitzgibbon, William Whelan-Curtin, Chinna Devarapu, Patricia Loren, Colin O'Sullivan and Ian Aherne

ENVIRONMENTAL PROTECTION AGENCY
An Ghníomhaireacht um Chaomhnú Comhshaoil
PO Box 3000, Johnstown Castle, Co. Wexford, Ireland

Telephone: +353 53 916 0600 Fax: +353 53 916 0699
Email: info@epa.ie Website: www.epa.ie

ACKNOWLEDGEMENTS

This report is published as part of the EPA Research Programme 2021–2030. The EPA Research Programme is a Government of Ireland initiative funded by the Department of the Environment, Climate and Communications. It is administered by the Environmental Protection Agency, which has the statutory function of co-ordinating and promoting environmental research.

The authors would like to acknowledge the members of the project steering committee, namely Charlotte Picard (Irish Water), Gaëtane Suzenet (International Impact Partners), Simon O'Toole and Dr John Feehan (EPA), and from EPA Research Lisa Sheils and Aisling O'Connor, as well as Margaret Keegan (ex-EPA) and Noreen Layden (ex-EPA) for their management of the project.

The authors also acknowledge the following additional members of the Project Team for their contributions to the work: Fionán Murray, Dr Eoin Byrne, Madhumidha Murugan and Karl Maxwell.

The project received support and assistance from the members of the EU Interreg SWIM Project Team at University College Dublin, whose contributions are gratefully acknowledged, particularly Professor Wim Meijer and Dr Laura Sala-Comorera.

Further collaborative assistance was provided by the University of Jena, Germany; in particular, the authors would like to thank Professor Ute Neuberger.

Finally, the authors wish to express their thanks to Mr Rod O'Connor for his very helpful contributions to the design and fabrication of the buoy system reported below.

This report is based on research carried out/data from April 2017 to February 2020. More recent data may have become available since the research was completed.

The EPA Research Programme addresses the need for research in Ireland to inform policymakers and other stakeholders on a range of questions in relation to environmental protection. These reports are intended as contributions to the necessary debate on the protection of the environment.

EPA RESEARCH PROGRAMME 2021–2030
Published by the Environmental Protection Agency, Ireland

ISBN: 978-1-80009-001-9 June 2021

Price: Free Online version

Project Partners

Mr Kevin Fitzgibbon
Nimbus Centre
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Tel.: +353 21 433 5095
Email: kevin.fitzgibbon@cit.ie

Dr William Whelan-Curtin
Centre for Advanced Photonics and Process Analysis
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Tel.: +353 21 433 5391
Email: william.whelancurtin@cit.ie

Dr Chinna Devarapu
Centre for Advanced Photonics and Process Analysis
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Email: chinna.devarapu@cit.ie

Dr Patricia Loren
Centre for Advanced Photonics and Process Analysis
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland

Mr Ian Aherne
Centre for Advanced Photonics and Process Analysis
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Email: ian.aherne@cit.ie

Mr Fionán Murray
Hydrolight Ltd
Rubicon Centre
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Tel.: +353 86 606 3606
Email: fin@hydrolight.ie

Mr Colin O'Sullivan
Hydrolight Ltd
Rubicon Centre
Munster Technological University, Cork
(formerly Cork Institute of Technology)
Bishopstown
Cork
Ireland
Email: colin@hydrolight.ie

Contents

List of Figures

List of Tables

Executive Summary

Water resources can be damaged by pollution of many kinds, from a variety of sources. European Union directives have been transposed into Irish national legislation to ensure that the quality of natural waters is protected. Water quality monitoring programmes play a key role, but testing is predominantly laboratory based, costly and inherently intermittent. These factors drive research to develop affordable, reliable sensors that can be deployed in the field, act autonomously, and provide readings in real time. A wide variety of sensing technologies have been created by the global research community. Typically, individual sensors target specific parameters, and some approaches require chemical reagents as part of the method. The concept behind this project was to create a sensing technology that does not require reagents and can detect multiple parameters.

Previous research and the literature indicated that Raman spectroscopy combined with artificial intelligence (AI) methods could be a feasible way to detect certain target analytes in water. The vision of the project was to use these technologies to pursue an innovative, low-cost autonomous system for detection of nutrients (nitrates and phosphates) and pathogens (*Escherichia coli* in particular) in water, and capable of operating in close to real-time. A Lab-on-Chip (LOC) model was envisaged as the ideal project outcome.

The project has verified the hypothesis by developing and testing Watermon, an end-to-end Raman spectroscopy-based detection system that uses suitably trained AI models to detect nitrates and, to some extent, *E. coli*. A user interface allows results to be viewed on mobile phones or personal computers. However, it was not possible during this project to adequately isolate phosphate Raman signatures to allow this approach to succeed.

An iterative approach to system development allowed parallel progress on different aspects of the project, which mitigated the risks and technical challenges of creating the LOC version.

Four hardware versions were developed:

1. Laboratory version: the basis for the project, to enable results verification.
2. Transportable Car-boot version: to conduct testing and rapid detection in the field and gather data for model training.
3. Buoy version: for autonomous operation when deployed in a river or other water body.
4. LOC version: a reduced-scale sensing instrument, compatible with either a Car-boot or Buoy platform; this version was partially achieved.

An AI model for nitrate detection was very effective (>99% accuracy) at detecting nitrates at 30 mg/l or above in river and drinking water. The accuracy of the method may be affected by changing the system variables, i.e. integration time or number of averages used. Accuracy would be expected to increase and/or the detection limit would be expected to reduce, by using larger labelled training datasets.

The *E. coli* labelled training datasets model was able to detect a level of *E. coli* at 250 colony-forming units (cfu) with cross-validation accuracy of 83.4%. At such levels, it was not possible to identify positive *E. coli* samples by traditional analysis of the data, i.e. inspection of a single spectrum, underscoring the superior ability of the AI model, albeit at relatively higher concentrations, i.e. 250 cfu. The success of the Watermon platform is predicated on good-quality calibrated data being used to train the AI model. It could be expected that, with larger quantities of calibrated *E. coli* training data, detection accuracy levels of the *E. coli* model could improve significantly.

We drew the following specific conclusions:

1. AI models can be successfully trained to detect water quality parameters, in close to real time, in conjunction with Raman spectroscopy.
2. The system detected nitrate at 30 mg/l, at an accuracy (confidence) level of 99%.
3. The system can detect *E. coli* at concentrations c. 250 cfu, with 83.4% confidence levels.
4. The full end-to-end Car-boot and Buoy systems operate in close to real time, i.e. sub-5 minutes from the end of collection of the Raman spectra to receipt of the results on a mobile device.

5. The system depends on the availability of good-quality suitable labelled data to train the AI models. Additional data can improve the system accuracy, through iterative re-training of the models.
6. The LOC version of the system has successfully detected *E. coli* at high concentrations.
7. Phosphate detection proved inconclusive in testing – the Raman signature of this compound was not as amenable to detection using the equipment employed in this project.

1 Introduction

1.1 Project Background

Ireland's water resources face continuous challenges in several respects, in common with countries across the European Union (EU) and other regions of the world. Significant problems can arise because of pollution of natural waters. Pollution can be of many types, and from various sources; its effects can include damage to human health and aquatic life, either from direct contact with the water or from ingestion of contaminants.

In developed countries, drinking water is commonly produced by abstracting water from natural watercourses, and treating it to a potable standard. However, some polluting compounds can be difficult and/or expensive to remove during treatment, resulting in those compounds emerging into the public water supply. On the other hand, natural watercourses and coastal areas are also the most common route for disposal of wastewater from human activity, whether the wastewater originates from agricultural, industrial or municipal sources These activities also present risks to the quality of natural waters that are used for recreational purposes, such as bathing.

The EU has adopted directives for Member States that specifically address the issues relating to the quality of natural waters. In particular, the Water Framework Directive (WFD) (EU, 2000) requires Member States to implement a whole-of-system approach to catchment management, including monitoring water quality in a systematic way. In addition, the Bathing Water Directive (EU, 2006) specifies water quality standards for recreational uses, again involving programmes of water quality monitoring. Such monitoring typically takes the form of regular water sampling of the water body in question, followed by laboratory analysis of the samples for parameters of interest.

This approach has some inherent drawbacks. First, it is costly in terms of staff time to collect samples and carry out laboratory testing. Second, by its nature, testing is periodic, not continuous, and so is best suited to monitoring the overall/background level of water quality. However, individual pollution incidents can easily be missed by this approach. These factors provide the motivation to develop water quality sensors that can operate "in the field", in a continuous, autonomous manner, and communicate results via a suitable protocol, chiefly a wireless method such as 3G or long-range radio (LoRa). In this way, a network of affordable field sensors can act as an early-warning/"red flag" detection system, which would trigger action on the part of relevant authorities or operators of the system. Responses could include:

- dispatching personnel in a targeted manner to collect samples for detailed laboratory testing, to verify the initial reading;
- initiating preventative action at a water treatment works, such as turning off abstraction pumps, to avoid drawing unwanted pollutants into the plant; or
- initiating pollution control action.

For these reasons, a significant amount of research has been carried out internationally to develop such sensors, using a variety of approaches. In December 2016, the Environmental Protection Agency (EPA) in Ireland, as part of its competitive Research Call, awarded the Cork Institute of Technology (CIT) – now incorporated into Munster Technological University (MTU) – funding for a project entitled Innovative Water Monitoring, or Watermon. The project was a collaboration between MTU and a private company, Hydrolight Ltd. Two MTU research centres contributed to the project: the Centre for Advanced Photonics and Process Analysis (CAPPA) and the Nimbus Research Centre for Embedded and Cyber-Physical Systems (Nimbus). Previous research and the literature indicated that Raman spectroscopy (RS) can be effective in detecting target analytes in water, when combined with artificial intelligence (AI) methods (Wang *et al*., 2018; Mei and Liu, 2019; Xu and Jackson, 2019). Therefore, the project aimed to create and demonstrate an innovative technique for water quality monitoring using RS and AI, capable of operating in "close to real time".

The project was subdivided into the following elements:

- RS (the main sensing technology), led by CAPPA;

- AI software for the detection of selected target analytes, led by Hydrolight; and
- other necessary hardware and systems to support the sensing technology, e.g. data and power management, communications, led by Nimbus.

The project received material support from the team working on the System for Bathing Water Quality Monitoring (SWIM) Project at University College Dublin (UCD), an EU INTERREG VA Programme project. This support was in the form of access to water samples and laboratory test results arising from SWIM's work in the Dublin area.

The project also collaborated with the University of Jena, Germany, in the form of access to innovative technology to help improve pathogen detection via concentration techniques, at the 'Lab-on-Chip' (LOC) scale.

1.2 Objectives

The ultimate aim of the project was to develop an adaptable integrated water quality monitoring system, based on RS and AI, having the following features: autonomous operation; rapid detection; reasonable detection accuracy; low cost of manufacture; and flexibility of analyte targeted.

The project plan envisaged a LOC system, i.e. a miniaturised integrated sensing platform for detection of water quality parameters, aimed at both drinking water and bathing water applications. In overall terms, the project achieved a significant step towards the realisation of this vision, through the development and demonstration of a prototype RS-based sensing system/platform [to technology readiness level (TRL) 7]; as well as an early-stage version of a LOC sensor using RS (to TRL c 3), for future deployment in the sensing platform, once it has been advanced to a higher TRL.

The main objectives of the project are to develop:

1. an RS device into a transportable unit suitable for taking readings in the field;
2. AI algorithms for RS-based detection, targeting nutrients and *E. coli*;
3. a database with operational data from trials;
4. a fully operational end-to-end detection system, as a LOC and to demonstrate it;
5. suitable monitoring software and demonstrate it.

The objectives have been substantially achieved, or progress has been made towards achieving them, in all cases.

1.3 Methodology

The primary objective of the project was to develop an initial version of the proposed system, i.e. an operational, end-to-end water quality monitoring system, using a combination of RS and AI, to detect selected target analytes: nitrates, phosphates and *E. coli*. The project aimed to combine a novel mix of technologies and processes, namely photonics, microfluidics, RS, glass microchip fabrication, AI software and big data management. An iterative hardware development was used to advance different aspects in parallel, i.e. hardware and collection of AI data. So, the Raman hardware was developed in four successive versions. The AI system places reliance on the availability of labelled calibration data, generated in part through the project team's work and in part through the assistance of the SWIM Project at UCD. The AI system consists of two models, one each for nitrates and *E. coli* detection, as well as the necessary data management and presentation layers.

1.4 Report Structure

The remainder of this report is structured as follows:

- Chapter 2 reports on a literature review carried out primarily to determine the state of the art in RS and AI for water quality parameters.
- Chapter 3 provides an overview of the available hardware relating to RS and the project's hardware development strategy.
- Chapters 4–7 describe the four versions of the hardware system that were developed as part of the project, and the objectives, description, testing and results arising from each version.
- Chapter 8 describes the AI diagnostic models.
- Chapter 9 covers the main project outcomes.
- Chapter 10 presents conclusions and recommendations.

2 Literature Review

2.1 Introduction

The Innovative Water Monitoring (Watermon) project was carried out in the context of national and international water quality concerns and related issues. It aimed to develop an innovative approach to water quality monitoring, harnessing spectroscopic techniques combined with AI for data analysis, in close to real time. Therefore, this literature review is arranged in three main parts. First, the context and background to the project is described. Second, the state of the art in RS, the main sensing technology being developed, and specifically as it applies to water quality monitoring, is investigated. Finally, the state of the art regarding the application of AI methods to this domain is reported.

2.2 Water Quality and the Irish Environment

Water in Ireland is a major and valued natural resource, and access to good-quality potable water offers significant competitive advantages to sectors such as industry, agriculture and tourism. Water is a resource that must be managed effectively, and the improvement of its quality has been noted as a national priority for Ireland (EPA, 2014). Water is under pressure from increasing pollution, rising global populations and changes in demographics, increasing urbanisation, climate change, land use and economic activities (e.g. industry, agriculture). As a resource, water not only plays a part in sustaining life, but it also plays a fundamental role in ecosystem support, economic recovery and development, as well as community and social well-being (EPA, 2013).

As observed in the EPA report *Ireland's Environment – An Assessment 2016*, "while Ireland's waters might be among the best in Europe, we are still a long way from meeting the full legal requirements of the Water Framework Directive, against which water quality is measured". The same report also states "elevated nutrient concentrations (phosphorus and nitrogen) continue to be the most widespread water quality problem in Ireland, arising primarily from human activities such as agriculture and waste water discharges to water from human settlements, including towns, villages and rural houses" (EPA, 2016). These activities, as well as physical impacts on habitats from excess fine sediment and forestry, are also described as "significant pressures impacting water quality" in the EPA report *Water Quality in Ireland 2013–2018* (O'Boyle *et al*., 2019).

Good-quality bathing water is also a highly desirable natural resource for recreational use as well as being an important economic factor for tourism. Public health safety and reducing the risk of illnesses, especially for outbreaks of *E. coli* and intestinal enterococci, is an over-riding concern. In particular, the impacts of pollution from urban run-off, wastewater discharges and agricultural sources – especially after heavy rain – are a continuing threat, particularly in more built-up areas (Webster and Lehane, 2015). The release of nutrients (nitrogen, phosphorus and potassium) and agrochemicals from intensive agriculture and animal waste accelerates the eutrophication of freshwater and coastal marine ecosystems and increases this pollution (Herbert *et al*., 2015). High nutrient concentrations in water can result in adverse human health impacts, such as 'blue baby' syndrome, which can be caused by high nitrate levels in drinking water (WHO, 2016). Inappropriate use of pesticides can pollute water resources with carcinogens and other toxic substances that can adversely affect human health and aquatic life (Raich, 2013). In addition to nutrients and pesticides, the accumulation of salt in soils can also be very harmful to water bodies. High-saline water alters geochemical cycles of other major elements – e.g. carbon, nitrogen, phosphorus, sulfur, silica and iron – thereby impacting ecosystems.

Legislation has been developed and implemented at national level to transpose the EU WFD (EU, 2000), which was adopted in 2000 as a single piece of legislation, covering rivers, lakes, groundwater, and transitional (estuarine) and coastal water. Ireland's national regulations implementing the directive are the European Communities (Water Policy) Regulations 2003 (Government of Ireland, 2003). Article 10(1) of these regulations requires a programme of monitoring of water status, to provide a coherent

and comprehensive overview of water status within each of the river basins in the State in accordance with Articles 7(1) and 8 of the Directive (EPA, 2006). Similarly, the national regulations giving effect to Bathing Water Directive 2 also require monitoring of bathing water quality (Government of Ireland, 2008).

The drivers described above have led to the implementation of water quality monitoring programmes in Ireland and across the EU. However, a rapid alert and response to a pollution incident can be critical, to minimise impacts. Apart from knowing where and when to take samples for analysis, one of the biggest challenges is the time required and cost of investigating whether or not the water is safe for drinking, bathing, or for other uses. This is because the methods usually involve field collection and transportation of samples to a laboratory, and in the case of some tests, such as for pathogens, sample preparation by culturing, followed finally by analysis. These challenges apply for nutrients, pathogens and other compounds, such as priority hazardous substances (PHSs) (EU, 2007; Regan *et al*., 2013). Together, these factors act as drivers in the effort to develop water quality monitoring sensors that can act continuously in real time, detecting the presence and concentration of relevant parameters in a cost-effective manner.

There are numerous examples of previous projects in this overall area of real-time water quality monitoring. These can be focused on particular sensor development, typically targeting single parameters for detection, or on systems integration to demonstrate end-to-end functionality. A variety of detection technologies have been explored through such research. The following project examples, and their associated summary descriptions, illustrate some of the related research activity in the field:

- Aquarius: H2020-ICT-2016-1 – 731465 (2017–2020): “Broadband Tunable QCL Based Sensor for Online and Inline Detection of Contaminants in Water.” The project aims to detect hydrocarbon contaminants in water (oil-in-water contaminations). It “aims to provide improved online and inline sensors in terms of quality and effectiveness, permitting reliable and continuous real-time monitoring on site. The new sensors will be made possible by the use of a new class of external cavity (EC) quantum cascade lasers (QCL) and detectors.” (AQUARIUS, 2020).
- WaterSpy: H2020-ICT-2016-1 – 731778 (2016–2020): “High-sensitivity, portable photonic device for pervasive water quality analysis.” WaterSpy uses “photonics technology suitable for inline, field measurements, operating in the 6- to 10-μm region. The solution is based on the combined use of advanced, tuneable quantum cascade lasers (QCLs) and fibre-coupled, fast and sensitive higher operating temperature photodetectors. Together with these new components, optimised laser driving and detector electronics as well as laser modulation will be developed. Attenuated total reflectance spectroscopy will be used to give rise to the biochemical profile of the surface chemistry of the sample. Targeted analytes will be specific heterotrophic bacterial cells. Several novel techniques are employed in order to increase the SNR [signal-to-noise ratio], including antibodies capable of binding the targeted analytes and a novel pre-concentration method.” (CORDIS, 2020a).
- SmartWater4Europe (2014) Seventh Framework Programme project. This is “an example of a systems integration and technology demonstration project aimed to deploy innovative sensing, information and communication technology (SICT) solutions of European SMEs [small and medium-sized enterprises] on 4 demonstration sites for selected smart water networks (SWNs) applications to improve current practice of water supply management and quality control, under the headings of leak detection, water quality management, energy optimisation and customers’ interaction.” (SW4EU, 2020).
- EU-SWIM: INTERREG VA Programme. This is “a cross-border research programme for developing a system for live bathing water monitoring… The aim is to develop a system that will allow bathers to check the water quality of their chosen bathing spots live … through use of a specially designed app on their smart phones.” (EUSwim, 2020).
- LIFE Ecosens Aquamonitrix – LIFE17 ENV/IE/000237: “Enhanced Portable Sensor for Water Quality Monitoring, moving to genuinely integrated Water Resource Management.” This project “consolidates research outputs from previous R&D [research and development] projects to demonstrate & bring to market a novel water quality monitoring integrated solution, focusing on nutrients in water, to meet the requirements for

frequent water quality monitoring under the Water Framework Directive." (Ecosens, 2020).

- ColiSense Online: "Online and Automated *E. coli* Monitoring for 100% Safe Drinking Water. This is a Horizon 2020 (H2020) SME instrument, i.e. an industry-led project, whose objective is to develop a low-cost, low-threshold *E. coli* detection method for drinking water, using flow cytometry with a laser diode operating at 488 nm." (CORDIS, 2020b; bNovate Technologies, 2020).
- CYTO-WATER: "Integrated and Portable Image Cytometer for Rapid Response to Legionella and *Escherichia coli* in Industrial and Environmental Waters." This project uses a fluorescence method for detecting Legionella and *E. coli*, with quantification requiring *c*. 2 hours; it requires pre-concentration of samples. The project has provided some promising results (CORDIS, 2017).

2.3 Review of Current State of Knowledge – Raman Spectroscopy

Photonics-based sensors have particular practical advantages for water quality monitoring, in that they are non-destructive, can be deployed *in situ*, can operate without the use of consumables, and can provide results in close to real time (unlike chemical or microbiological analysis). Techniques such as fluorescence analysis are popular and powerful, but the spectral signatures are broad and the emission spectra of different substances often overlap. Both RS and mid-infrared (MIR) spectroscopy directly measure the energies of the chemical bonds that make up the substances of interest. However, However, MIR wavelengths can be absorbed during passage through water, which creates a fundamental obstacle to their use for water quality monitoring. RS is therefore the most promising photonics-based sensing option to exploit the above advantages for water quality monitoring.

RS is a potentially powerful detection basis, as it offers the possibility of monitoring water *in situ* and in real time. RS is based on Raman scattering, one of the effects that occurs during the interaction of light with matter.

The interaction of light and matter can result in light being absorbed or scattered (and scattering may be elastic or inelastic). Alternatively, there may be no interaction and light may be transmitted or reflected (Long, 2002). If the scattering is elastic (without energy losses), the wavelength of the emitted light will be the same as that of the incident light, as shown by the energy transition levels in Figure 2.1a. This effect is also known as Rayleigh scattering, and it is responsible for the blue colour of the sky, for example. (In this case, particles present in the atmosphere interact with sunlight, causing scattering of light. However, this light scattering strongly depends on the size of the particles, their separation and the wavelength of the incident light. If the particles are sufficiently separated and their size is small, then they cause strong scattering of short wavelengths in the visible spectra. This selective scattering of visible light causes blue light to scatter significantly more than red light. As a result, blue light reaches our eyes more effectively than the rest of the spectra during clear days, thus making sky appear blue.)

Alternatively, scattering can be inelastic, i.e. involving energy loss or gain, and this is called Raman scattering. However, in this process, only a small fraction of light is scattered, which depends on the chemical structure of the sample.

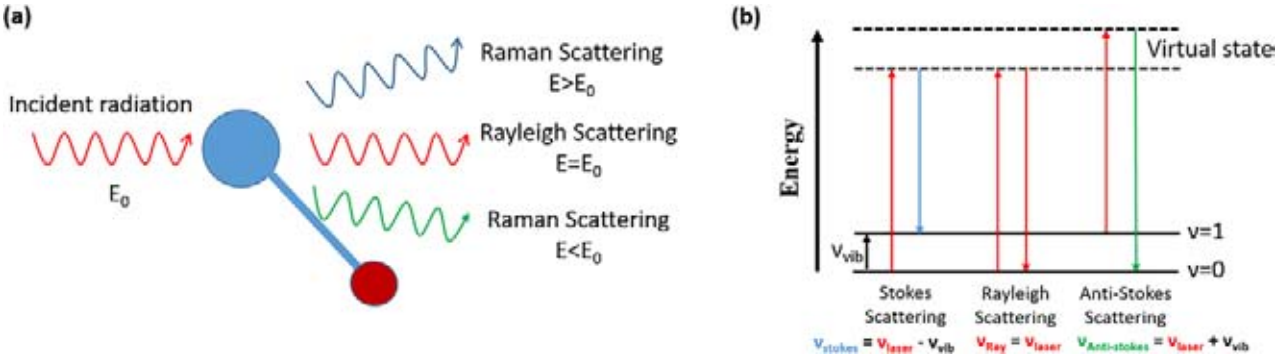

Figure 2.1. Raman scattering scheme. (a) Classical theory and (b) quantum mechanics.

From the point of view of classical physics, the Raman effect is based on the distortion of the electron distribution (electron cloud) of a molecule, caused by the electric field of the incident radiations during the scattering process. An induced electric dipole moment, caused by the oscillating incident radiation, results in periodic deformation of molecules, as a result of which molecules start vibrating with a characteristic frequency. This can be shown as $P = \alpha E$, where P is the dipole moment, α is the polarisability of the molecule (the ease of distortion of the electron cloud of a molecule by an electric field) and E is the strength of the electric field (Long, 2002).

From the point of view of quantum mechanics, when photons interact with a molecule, excitation of the molecule results in an increase in its energy state from the ground energy state to a virtual energy state (Smith and Dent, 2005). This virtual state is not a real energy level; it is transient, since the state is not stable, and when the molecule relaxes a photon is re-radiated.

In Rayleigh scattering the photon released in the molecule relaxation process has the same energy as the incident photon (laser incident radiation), as shown in Figure 2.1b. This forms the majority of the scattered energy.

However, if during the relaxation process from the virtual state the molecule drops back down to a different vibrational state – via processes called Stokes or anti-Stokes scattering – the difference in energy level results in a shift in the wavelength of the scattered light (University of Cambridge, 2020). RS detects this shift.

In RS, excitation energy is typically provided by laser light and a spectrometer is used to detect photons arising from all types of scatter. However, compared with Rayleigh scattering, the Raman signal is much weaker, being responsible for only 0.001% of incident light (Princeton Instruments, undated). For this reason, in order to reduce Rayleigh scattered light, and to obtain useful information from Raman effects, suitable filters must be employed on the light received by the spectrometer.

The energy difference between the incident light (laser) and the transitions allowed in Figure 2.1b can be plotted. This plot is known as Raman shift and is usually expressed in wavenumbers. The transitions provide specific information about the molecules, or fingerprint, allowing the molecule to be identified.

The potential of RS to analyse and monitor contaminants has long been reported (Gowen, 2012; Zhen *et al*., 2016). In addition, because water is a weak Raman scatterer, RS is superior to other vibrational spectroscopies (Zhiyun *et al.*, 2014), allowing the detection of specific target compounds in liquid samples, and also tissue imaging in some applications such as biomedical diagnosis (Keren *et al.*, 2008; Popp *et al*., 2011). RS has become established as an optical rapid detection technique (Das and Agrawal, 2011) that generates a spectroscopic fingerprint for the samples, and provides quantitative and qualitative information that can be used to characterise, discriminate and identify at a cellular level (Harz *et al*., 2014).

Zhang *et al*. (2006) found that Raman spectra can be used to measure pesticides on the surfaces of vegetables and fruits *in situ*, and that current Raman instrumentation technology enables the recording of reliable and repeatable spectra in pesticide monitoring applications. It was found that at 1064 nm excitation wavelength, pesticides can be detected (Zhang *et al.*, 2006). A range of contaminants have been detected using spectrographic techniques. These include *Cryptosporidium* and *Giardia, E. coli,* enterococci*,* pesticides, polyaromatic hydrocarbons (PAHs) and heavy metals (Zhiyun *et al*., 2014).

The detection of nutrients such as nitrates using RS was reported by Sadate *et al*. (2011), who used a portable stand-off RS system to detect ammonium nitrate (NH_4NO_3) and sodium nitrate ($NaNO_3$) dissolved in water. They reported a limit detection of 200 mg/l for ammonium nitrates and of 50 mg/l for sodium nitrates. Durickovic and Marchetti (2014) were able to detect a minimum nitrate concentration of 100 mg/l, also using a portable Raman system. However, they used a laser source in the visible range (532 nm wavelength), which can produce high fluorescence signal when used to analyse river water that contains a large amount of organic matter that fluoresces when excited with light in the visible range.

Ben Mabrouk *et al*. (2013) reported the analysis of different sulfates in two states: solid and aqueous solution. They reported on the effect of ionic substitutions on the Raman band positions of different vibrations and showed that the sulfate environment

is barely affected by the nature of the cations when the salt is in the form of aqueous solution; however, when sulfates are found in solid form, the sulfate environment can be affected by the nature of the cations. Kauffmann and Fontana (2015) reported the simultaneous detection by RS of pollutants, such as nitrates, sulfates and chlorides, dissolved in water, and also provided the concentrations for the same compounds using chemometric methods.

The development of substrates to enhance Raman scattering is also explored in the literature; for example, Gajaraj *et al.* (2013) used a gold nanosubstrate (a gold-coated silicon material) to detect nitrate and nitrite in water and wastewater. Profiting from surface-enhanced RS (SERS), they were able to detect nitrate with linear ranges of 1–10,000 mg NO_3^-/l. The authors also pointed out that phosphate appeared to be the major interfering anion among the common anions affecting nitrate measurement. Nevertheless, the percentage error of nitrate measurement in wastewater by the proposed SERS method was comparable to that of ion chromatography.

More recently, Zeng *et al*. (2019) proposed a portable Raman SERS chip based on commonly used filter paper and silver nanoparticles (AgNP). They presented a smartphone-based portable Raman spectrometer equipped with a laser of 785 nm wavelength and operated by a touch-screen application as the human–machine interface. The Raman detection module can be attached to or removed from the smartphone through the smartport interface, i.e. it does not affect the normal use of the smartphone. Moreover, Zeng *et al*. (2019) appear to be the first to report the use of a smartphone as a miniaturised Raman spectral analyser. Although the system is portable, the use of AgNPs may cause some problems owing to silver oxidation affecting the SERS efficiency in biosensing. In addition, the authors applied the sensor only for standard sample dyes such as rhodamine 6G and crystal violet, which provides the proof of concept. There are currently no tests using real-world samples to prove sensor reliability; however, this work demonstrates the improvement and progress towards miniaturisation of instrumentation using RS, enabling the implementation of more compact set-ups, albeit the particular set-up described has significant limitations on its usefulness.

The scientific community has also explored to some extent the use of more compact microfluidic-based systems targeting bacterial detection. McClain *et al*. (2001), for example, demonstrated flow cytometry of *E. coli* on microfluidic devices and found that microchip flow cytometry has the potential to be a cost-effective and portable alternative to conventional flow cytometry. The potential of such a microenvironment offers the prospect of fine manipulation of target analytes using on-chip techniques such as flow cytometry, laser tweezer and hydro-focusing. However, the microfluidic environment also suffers from practical disadvantages, such as potential for channel blockage.

The application of silicon nitride (Si_3N_4) waveguide layers instead of silicon can extend the range of excitation and detection frequencies to the entire visible and near-infrared wavelength range, which is particularly relevant for RS (Dhakal *et al.*, 2014). One important advantage is the fact that the high refractive index contrast in a Si_3N_4 waveguide helps to enhance the electric field strength of a guided mode for a given mode power (Dhakal *et al.*, 2013). This enhancement can be boosted further through either resonant cavity enhancement and/or plasmonic enhancement by means of metallic nanostructures in the vicinity of the waveguides (Peyskens *et al.*, 2013).

Modern LOC devices are hybrids that combine glass, silicon and various polymers such as acrylic, polyester, polycarbonate, resists, thermoplastics or moulds like polydimethylsiloxane (PDMS). Precision, miniaturisation, cost-effectiveness, large-scale production and ability to incorporate electronics make these materials very attractive (Giannitsis, 2011). Ashok *et al*. (2011) reported the implementation of fibre-based microfluidic RS, in which they embedded optical fibre probes into a PDMS-based microfluidic chip. They were able to detect urea at a concentration of 80 mM over an acquisition time of 5 seconds with 200 mW excitation power. The microfluidic chip can be used to replace the fibre-optic Raman probe. The use of soft lithography-based fabrication and standard optical fibres makes this device at least two orders of magnitude cheaper than commercially available fibre-optic Raman probes. In addition, PDMS presents advantages such as optical transparency, cost-effectiveness and ease of moulding, it allows complex fluidic systems to be built, is inert and is non-toxic. However, such systems are usually not rugged,

which may cause flow profile problems due to leakage and/or uneven pressure (Kim *et al*., 2010; Chen *et al.*, 2016). The fabrication efficiency of soft lithography is low, making it labour-intensive (McDonald *et al*., 2001). Moreover, fast adjustment of device features is almost impossible without the fabrication of a new mould.

A technology that has recently emerged as an alternative fabrication method for microfluidics is three-dimensional (3-D) printing, which has shown the potential to address many of the problems associated with PDMS devices (Gross *et al.*, 2014). Compared with the labour-intensive and multiple-step soft lithography process (Xia and Whitesides, 1998), 3-D printing can potentially fabricate a microfluidic device in one step (Kitson *et al.*, 2012). In addition, 3-D printing allows quick adjustment of device features with each print by changing the design in the computer-aided design (CAD) software (Chen *et al*., 2016).

De Coster *et al*. (2015) presented a microfluidic chip in polymethyl methacrylate (PMMA) for optical trapping of particles through microchannel and single-mode fibres separated from the microfluidic channel by thin PMMA walls with a width of 70 µm. The walls prevent contamination of the trapping fibres by the sample fluid flowing in the microchannel (De Coster *et al.*, 2015). Particle trapping can assist in obtaining the Raman signal.

Another study that makes use of microfluidic and optical fibres is described by Dochow *et al*. (2013), who built quartz microfluidic chips, embedded innovative multi-core single-mode fibres integrated with Bragg gratings for detection. The Bragg gratings work as a notch filter for the Rayleigh scattering. They reported an improvement of more than two orders of magnitude compared with previous fibre-based microfluidic Raman detection schemes.

The technique of dielectrophoresis allows the spatial manipulation of particles through the interaction of the sample with a non-uniform electric field (Pething, 2010) and can be used to concentrate bacteria to enhance the Raman signal. The technique has great potential because it allows the spatial manipulation of cells and bacteria without the need for biochemical labels or other bioengineered tags (Qian *et al.*, 2014; Pething *et al.*, 2014). The University of Jena in Germany obtained the most advanced version of a microfluidic chip to capture the Raman spectra of bacteria and demonstrated clinical applicability of the dielectrophoresis chip device by testing *E. coli* susceptibility to the commonly prescribed second-generation fluoroquinolone antibiotic ciprofloxacin (Schröder *et al.*, 2017).

Microstructured waveguides built by femtosecond laser inscriptions also have also proven to be a good alternative to more compact RS (Chen *et al.*, 2018). Such waveguides could be incorporated in integrated Raman laser platforms for biomedical applications. Professor Roberto Osellame's group in Italy is expert in the field; a recent paper described the principles and applications of femtosecond laser 3-D micro- and nanofabrication for LOC applications (Sima *et al.*, 2018). The unique advantage of femtosecond laser processing over conventional methods resides in the capability of sculpturing complex 3-D shapes at micro- and nanoscales in both inorganic and organic transparent materials. Indeed, by employing focused ultrashort pulses with extremely high peak intensities, one can precisely set the interaction region at a localised area of either surfaces or in volume (Sima *et al*., 2018). Although these systems have not been applied so far as Raman platforms, they could be explored in further work.

2.4 Application of Artificial Intelligence to Sensing

Water systems should maintain a good chemical and ecological status to protect natural ecosystems and biodiversity, water supply and, of course, human health. Reviewing the ecological status of water systems in a comprehensive way would greatly benefit from improved water quality monitoring approaches (Voulvoulis *et al*., 2017). To date, many substances such as nitrates and phosphates in water systems are monitored through low-frequency discrete assessment campaigns. To reduce the possibility of eutrophication, nutrient monitoring is crucial. Eutrophication, the overenrichment of nutrients in the water system, is a major water quality issue that can result in fish deaths, unhealthy ecosystems and public health issues. Monitoring for substances such as nitrates, phosphates and *E. coli* would benefit from the introduction of real-time monitoring leveraging well-trained AI data models. The use of always-online water monitoring platforms has been acknowledged as providing great value (Viviano *et al.*, 2014; Valkama and Ruth, 2017), whereas low-frequency testing can

lead to under-reporting of risks (Brack *et al*., 2017), with consequences.

A number of water-testing systems are commercially available, such as ion-selective electrodes (ISEs), but these have been shown to be prone to error, e.g. readings can be temperature dependent and frequent calibration can be required (Thompson, 2005). Another approach is, of course, to use wet chemical analysers or optical sensors, which can show greater detail but have the downside of being expensive (Pellerin *et al*., 2016). Traditional models are used to compute models such as effluent discharge models and watershed models, but, of course, the complex relationships of input variables or data sources and infrequency of the inputs to these models can result in a weak model. An alternative to these more traditional models are machine learning models, which can be deployed online with high availability ready to receive data from source and provide a fast alternative to time-consuming, expensive, infrequent laboratory chemical analysis (Wang *et al*., 2018).

The use of machine learning models to predict water quality has been attempted in many research projects (Schilling *et al*., 2017; Steffy and Shank, 2018). It has been demonstrated that non-linear machine learning models do have advantages over linear regression. Non-linear models have been shown to perform poorly on high-dimensional complex datasets. In this case the RS signal is a non-trivial dataset (Kuhar *et al*., 2018), and machine learning models have been shown to perform well with similarly complex datasets whose key features or relationships can be difficult to identify (Xu and Jackson, 2019). Ensemble-based tree-based machine learning models have been shown to perform particularly well with such datasets, including Random Forests (RFs) and quantile regression forests (Francke *et al*., 2008).

Arising from these factors, continuous high-frequency water monitoring is becoming a critical part of water management (Bowes *et al*., 2015). Sensor technology development continues to provide improvements in data gathering; however, processing such data effectively using traditional testing or statistical models has proved challenging. This is where AI can play a key role in processing high-throughput data samples. AI models have been used successfully to make estimations by means of data-driven models (Casanovas-Massana *et al*., 2015; Zhang *et al*., 2017). In many cases machine learning models such as RFs have been used with success, as have algorithms such as Gaussian process, M5P and Random Tree (RT). In many research projects various model types were trained to identify which algorithm performed best according to the specific input datasets (Bui *et al*., 2020).

Using collected quality sample datasets, advanced AI models permit the training of advanced data models to create accurate generalised cost functions, which in turn permit more accurate predictions for a specific assessment use case, such as the testing for a specific element in a water system (Mei and Liu, 2019). When compared with the predictions of linear modelling, a significant reduction in the root-mean-square error (RMSE) is often found, improving the classification of data samples.

The method of using the Raman spectrum to analyse the structure of water was described by Cross *et al*. (1937) and the Raman signal provided very detailed information. Since then, the use of optofluidic RS for water monitoring applications has continued to be a busy area of research. More recent studies have examined areas such as using high-power lasers for exciting the samples and increasing the collection efficiency of Raman scattered photons, improving the sensitivity of overall RS systems, as well as looking at the opportunities for creating portable RS devices (Persichetti and Bernini, 2016).

Studies combining the use of optical data collected using RS and AI machine learning techniques such as deep learning and ensemble models such as RF to analyse blood samples are being published at an increasing rate (e.g. Lussier *et al*., 2020). The combination of this optical technique with machine learning methods such as RF or neural networks (NNs) has been used for the analysis of complex signals in human serum infected with Dengue virus (Khan *et al*., 2017a), for screening for hepatitis B-infected blood sera (Khan *et al*., 2019), for screening for asthma (Ullah *et al*., 2019) and with principal component analysis (PCA) for screening for nasopharyngeal cancer in human blood sera (Khan *et al*., 2017b). In a very interesting study, a combination of RS with an RF classifier was used to understand the molecular structure of milk from different species (Amjad *et al*., 2018). The study also leveraged dimensionality reduction and analysed the

variations in intensity in the Raman peaks of the milk samples with high precision.

Arising from the above, it was concluded that leveraging proven ensemble AI/machine learning methods would work well with the high dimensionality produced by continuous Raman datasets. It was also concluded that the literature supports the feasibility of applying the method to the domain of water quality testing in the field, targeting specific analytes with good Raman signatures, as well as, possibly, to the detection of bacteria.

2.5 Summary

There are a number of drivers for the development of innovative water quality monitoring platforms operating in real time or close to real time, such as the type reported here. RS offers a potential sensing method to achieve this objective, but in isolation it is not sufficient for in-field analysis. However, AI algorithms offer powerful tools for analysing large datasets, such as those produced using RS. The literature supports the hypothesis that a combination of RS and AI can be used to detect specific target analytes. This project resulted from the recognition of the potential utility of combining these technologies, and as a response to the clear benefits that would accrue from a continuous, real-time monitoring platform that does not require consumables/reagents to operate. In this respect, the project aimed to combine technologies that were not reflected in the existing portfolio of research projects.

3 Hardware Overview

3.1 Raman Spectroscopy Overview

RS is a non-invasive, non-destructive, analytical technique that can provide molecular fingerprints of various substances present in water, without the use of consumable items or chemicals. The RS system is based on detecting the change in frequency of the incident monochromatic light interacting with the sample. The Raman instrument consists of a monochromatic source (laser), focusing objective and a spectrometer, as shown in Figure 3.1a. Bench-top Raman spectroscopes (see Figure 3.1b) are usually bulky and costly, and it is not possible to deploy them near water bodies for continuous monitoring of water quality. The main aim of the project was to seek to reduce the scale of the hardware, in other words to make progress towards a miniaturised version, for affordability and durability.

3.2 Overview of the Hardware Development Strategy

To achieve the project's primary objective, it was recognised that an iterative approach to developing the necessary hardware would be required. This approach envisaged a number of versions of the system to be developed, with insights and learning incorporated in successive stages, leading towards the LOC version. The versions envisaged were:

1. Laboratory version: to act as a basis for the project, establish baselines, enable verification of results, etc.
2. Car-boot version: to provide a means to conduct testing and rapid detection in the field, without deploying an instrument, e.g. onto a river;
3. Buoy version: to develop an instrument for autonomous operation and parameter detection when deployed in a river or other water body;
4. Lab-on-Chip (LOC) version: a reduced-scale sensing instrument capable of installation in either a Car-boot or Buoy platform.

This approach had the advantage that it allowed progress to be made on different aspects of the system in parallel. It also acted as a risk mitigation against the technical challenges of creating the LOC version, which was unlikely to generate reliable data or be robust enough for deployment at the early stages of the project. The idea was that the earlier versions would provide useful guidance for both of the eventual LOC detection systems in terms of data analysis.

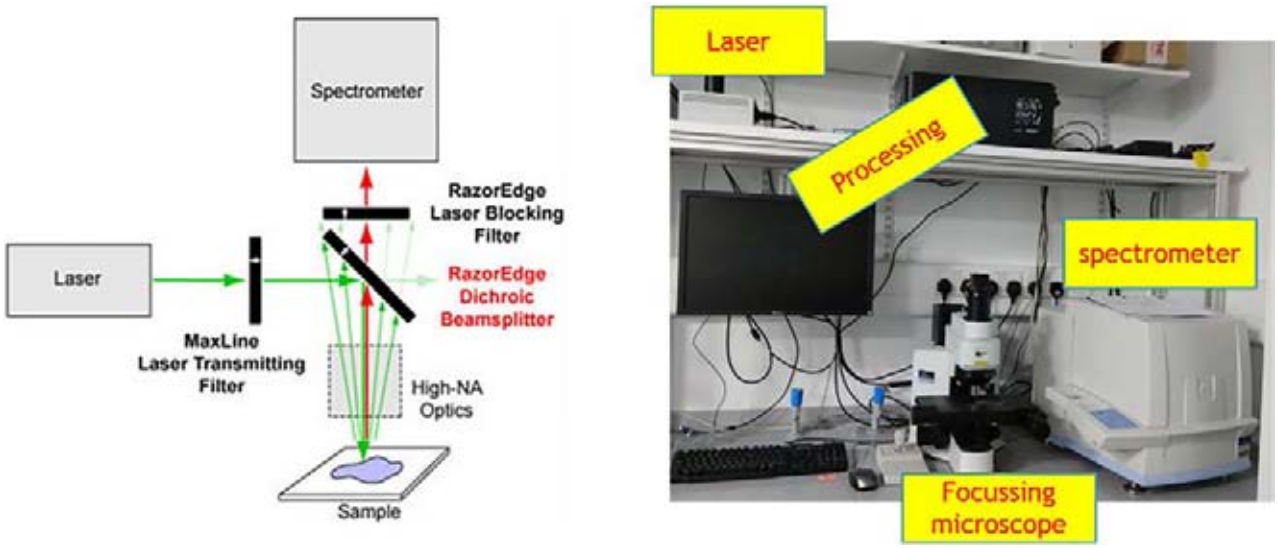

Figure 3.1. Raman spectroscopy system. (a) Schematic and (b) bench-top system.

In general terms, one of MTU's main roles in the project was to create the hardware versions of the system. Within MTU, the CAPPA Research Centre focused on developing the sensing technology, i.e. RS, for the different versions. The Nimbus Research Centre was responsible for the design and integration of the sensing systems for the Car-boot and Buoy versions with the other functions that were necessary. These included power management, data management and data communications, for example, as well as the necessary firmware to operate the overall system. The LOC version was developed by CAPPA to an early stage; it was not deployed in the field as part of the project.

The next four chapters of this report describe the versions of the hardware that were developed through the project and the methods used to test and evaluate their performance in capturing accurate Raman spectra. Chapter 7 also describes the evaluation carried out on the original concepts for the LOC version, and the adaptation that was required, in the light of the insights obtained.

4 Hardware Version 1 – Laboratory-based Version

4.1 Objectives

The main objective of the laboratory (lab) version of the Raman spectrometer is to generate the Raman spectra of the pollutants in the water. The data generated are transferred to the software team to build the machine learning models being used throughout the project. This lab version acts as a precursor to both the portable Car-boot and Buoy versions of the Raman spectroscopes built during the project, and provides guidance to the subsequent systems on the best means of collecting and analysing the data. Table 4.1 summarises the principal elements of the four hardware versions developed as part of the project.

4.2 System Description

The lab version of Raman set-up consists of a fibre probe, benchtop 350-mW laser Master Oscillator Power Amplifier (MOPA) laser from Innovative Photonic Solutions Ltd and a thermoelectrically (TE) cooled WP 785 spectrometer from Wasatch Photonics Ltd. Both laser and spectrometer run on alternating current (AC) power. Fibre-based probes provide optical filtering of the elastically scattered Rayleigh line and high-signal collection in a compact design. The Raman probe employed and the example spectra obtained are as shown in Figure 4.1. The Raman probe is fibre-coupled, and contains lenses, dichroic mirrors and various filters to excite the sample and collect the signal, as shown in the schematic of the Raman probe in Figure 4.2. The collected signal is directed into the spectrometer for analysis.

4.3 Testing

The samples used were isopropyl alcohol (IPA) then freshly prepared solutions of nitrates in water using nitrate salts of various concentrations. The spectra acquired using the spectrometer software comprise a Raman signature of glass vial, nitrate, water peaks and cosmic ray spikes. However, the inherent weakness of the Raman effect, coupled with spectral variability due to spurious signals from sample holders, can produce significant problems for chemometric-based high-throughput assays. To obtain a good SNR it is vital to avoid Raman peaks that are not of interest. After some experimentation, glass vials were replaced with aluminium sample holders, because the light does not penetrate the container glass, thus avoiding spurious background.

4.4 Results

Using this lab version, spectral signatures of individual nutrients in powder form and diluted water were collected. The nutrients characterised were nitrates (NO_3), nitrites (NO_2), phosphates (PO_4, HPO_4, H_2PO_3) and sulfates (SO_4), which give distinct single spectral peaks, unlike bacteria, which give complex, and generally weak, Raman spectra. It was found that the spectral signatures of these nutrients agreed well with the values in the literature. In later experiments, mixtures of the nutrients were examined.

The results obtained with the lab version spectroscope reduced the measurement time of the Raman spectra by 100 times compared with the bench-top spectrometer tested (a PerkinElmer system), as illustrated by the Raman spectra of nitrates in

Table 4.1. Comparison of four different versions of the project's Raman system hardware

Version	Laser	Raman probe	Spectrometer	Power	Purpose
Laboratory	MOPA-350 mW	Fibre probe	Wasatch	(5 V, 1.2A) + (12 V, 0.8A)	*E. coli* detection and machine learning
Car-boot	Ondax-600 mW	Fibre probe	Wasatch	(5 V, 1.2A) + (12 V, 0.8A)	Nitrate and other nutrients
Buoy	Ondax-600 mW	Fibre probe	QE-Pro AC + 5 V, 5A + computer	AC + 5 V, 5A + computer	Nitrate and other nutrients
LOC	TO can lasers	Waveguides	DIY	5 V, 0.25A (estimated)	*E. coli* and nutrients

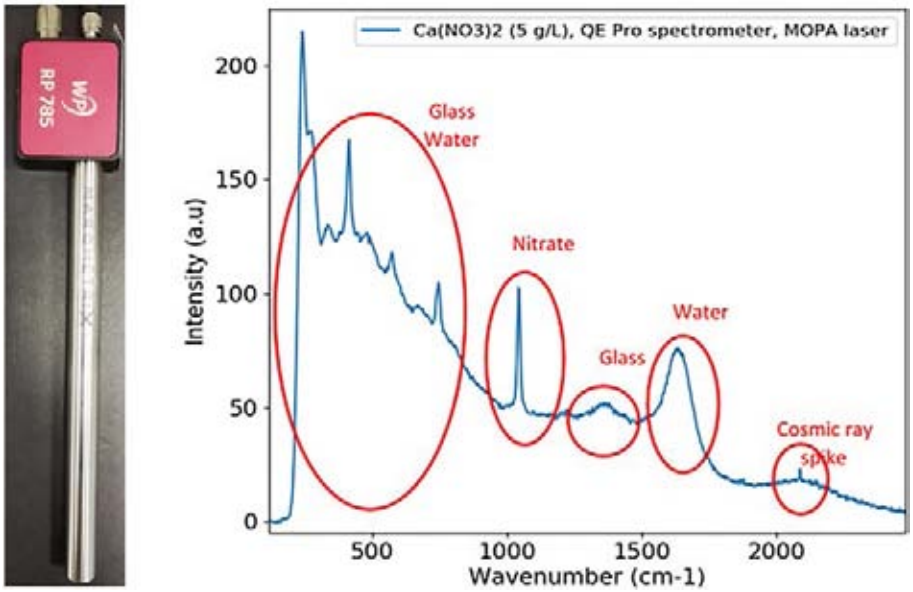

Figure 4.1. Image of the Raman immersion probe and the Raman spectra of a nitrate sample. The Raman photons intensity on the *y*-axis is represented in arbitrary units (a.u.).

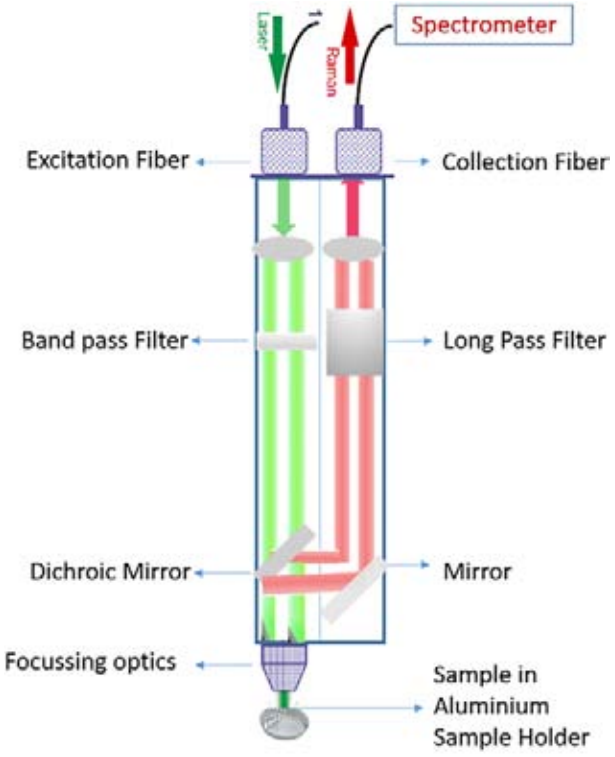

Figure 4.2. Schematic of a Raman probe with an aluminium sample holder.

water in Figure 4.3. This significant improvement in measurement time is possible owing to the choice of the fibre-probe as the focusing objective. This probe allows the impinging light to directly interact with the water, unlike the bench-top spectrometers, in which light can interact only with water passing through sample holders. Owing to the improvements obtained with the lab version of the Raman spectroscope, it was possible to consistently detect nitrates at levels below 50 mg/l (Figure 4.3).

To provide more realistic results that were more representative of real-world field testing, one unfiltered sample of water from the River Lee in Cork City was spiked with nitrates and phosphates, and analysed. The resulting spectrum is shown in Figure 4.4. The associated spectral peaks can clearly be identified against the background signal (mainly fluorescence) from other substances present in the sample.

The result shows that the laboratory system accurately captures and identifies Raman spectrum features, even in the presence of signals from multiple complex background compounds.

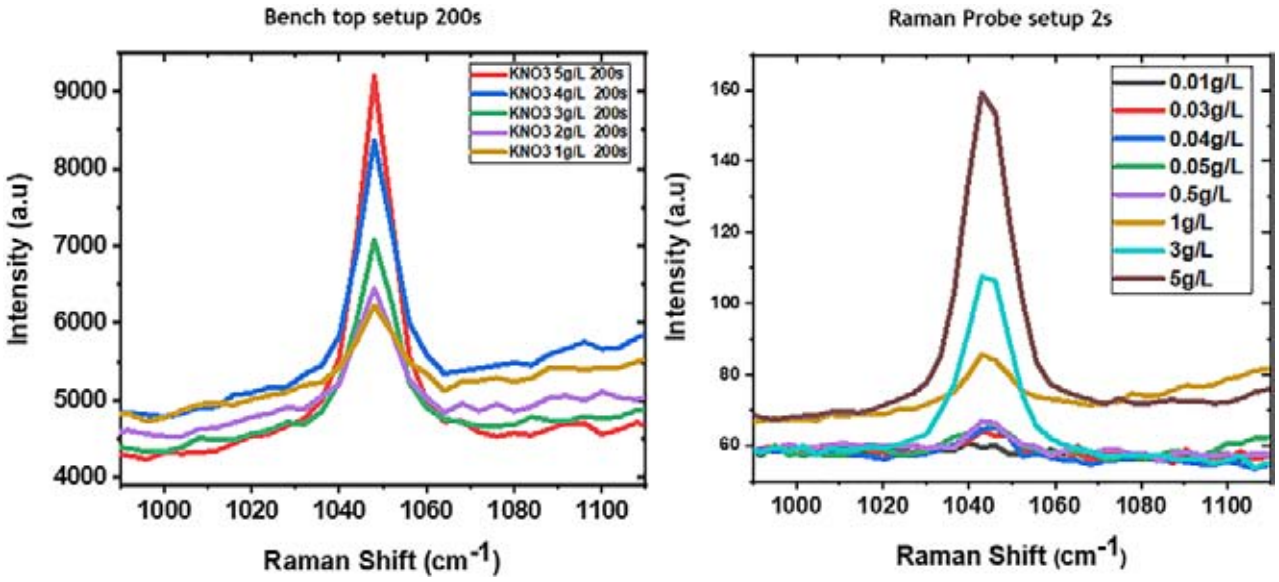

Figure 4.3. Time consumption comparison between the bench-top and fibre-probe system for nitrates.

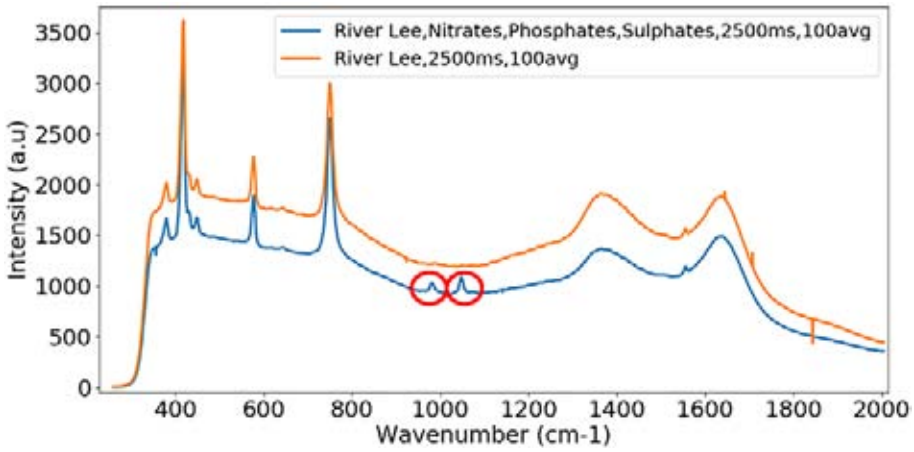

Figure 4.4. River Lee water sample spiked with nitrates and phosphates; Raman spectrum signatures for phosphate (left circle) and nitrate (right circle) are shown.

5 Hardware Version 2 – Car-boot Version

5.1 Objectives

The objective of developing a transportable or Car-boot version of the system was to be able to conduct convenient on-site testing of water contamination. This version was built such that it could be run on 12 V direct current (DC) if required. This meant that the main items of equipment would not be the same as the (large and power-hungry) laboratory-based equipment, e.g. the TE-cooled spectrometer and the AC-powered MOPA laser.

5.2 System Description

Extensive testing of various spectrometers and lasers was undertaken to identify suitably accurate yet power-efficient combinations of instruments for the Car-boot version. The spectrometers tested for the Car-boot version, i.e. the QE-Pro, Maya, Hamamatsu and Wasatch, are shown in Figure 5.1. It was concluded that the Wasatch spectrometer was an appropriate model, with superior performance and moderate power consumption. Similarly, the DC-compatible Ondax laser (see Figure 5.1) was chosen as the light source for the Car-boot version of the Raman spectroscope. The Ondax laser is hosted in a metal chassis for better thermal management and mechanical stability. Furthermore, a 12-V battery with charging circuit and power outlets to power both the laser and spectrometer was designed by the Nimbus team. Before assembling, a suitable housing was designed and 3-D printed for the Ondax laser. The components were fitted in a Peli case to withstand field conditions (Figure 5.2).

5.3 Testing and Results

Figure 5.3 compares the performance of the Lab and Car-boot versions of the system and illustrates the Raman spectra of nitrate-spiked water samples. It is evident that an improvement of five times the Raman signal for the lower concentration of nitrates (50 mg/l) is achieved with the Car-boot version of the Raman spectrometer. This superior performance is possible owing to the combination of the high sensitivity of the Wasatch spectrometer and the large power of the Ondax laser (600 mW).

5.3.1 Collaboration with the EU SWIM project team at UCD

Initial measurements were made using *E. coli* samples provided by the BioExplore group at MTU. These were laboratory samples in which the *E. coli* was diluted in de-ionised water. These samples provided background-free spectra (i.e. no interference from other substances) and so were not realistic tests of the system. It proved challenging to generate sufficient numbers of samples, and it was also difficult to undertake the microbiological analysis to define the number of colony-forming units (cfu)/100 ml required for training the machine learning algorithms. To validate the performance of the system with real-world samples, a collaboration was established

Figure 5.1. Various spectrometers tested for the Car-boot version of the Raman system.

Figure 5.2. Components of the Car-boot system. (a) Wasatch spectrometer, (b) Raman probe, (c) Ondax laser, (d) battery and (e) Peli case.

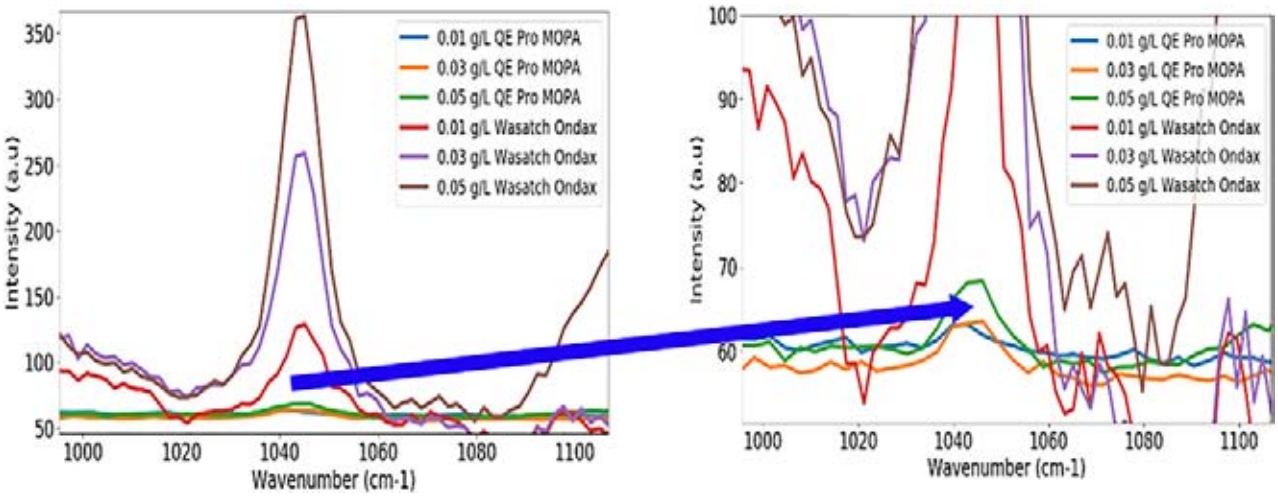

Figure 5.3. Comparison of the performance between the lab and Car-boot versions of the Raman system.

with Professor Wim Meijer's team at UCD (who are participants in the EU SWIM Project (EUSwim, 2020).[1] The EU SWIM Project team, supervised by Dr Laura Sala-Comorera, supplied four batches of water samples (15–20 individual samples in each case) that were collected from rivers and beaches in the Dublin area. The UCD team carried out microbiological analysis on each of the samples and provided the concentration of pathogens, e.g. *E. coli*, for each. In parallel, the samples were sent for chemical analysis by a commercial service provider [the Environmental Research Institute laboratory (ERI)] to determine the nitrate, sulfate and phosphate concentrations. Using the Car-boot Raman system (Figure 5.4), the CAPPA team measured the Raman spectrum of each sample. The chemical and microbiological analysis provided the initial training data for the machine learning model (see below).

Batches 1 and 2 were used for system development and to develop the standard operating procedures (SOPs) for the analysis process. For example, samples must be refrigerated and tested within 2 days of arrival to avoid degradation and minimise bacterial growth. The results from batches 3 and 4 are shown in Figure 5.5. In particular, Figure 5.5a

1 https://swimproject.eu/ (accessed 3 December 2020).

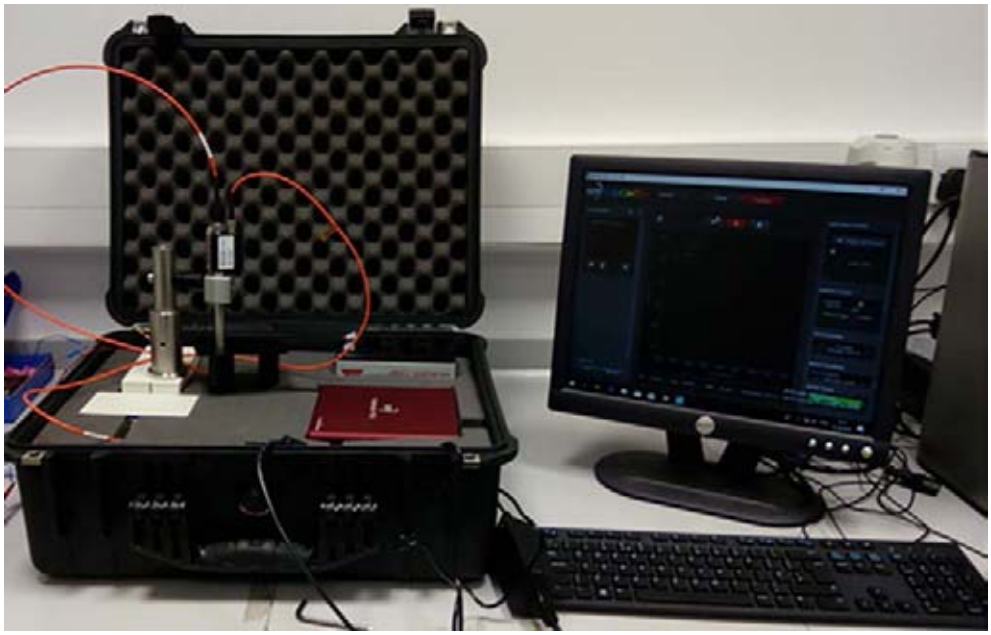

Figure 5.4. Car-boot version of the Raman system.

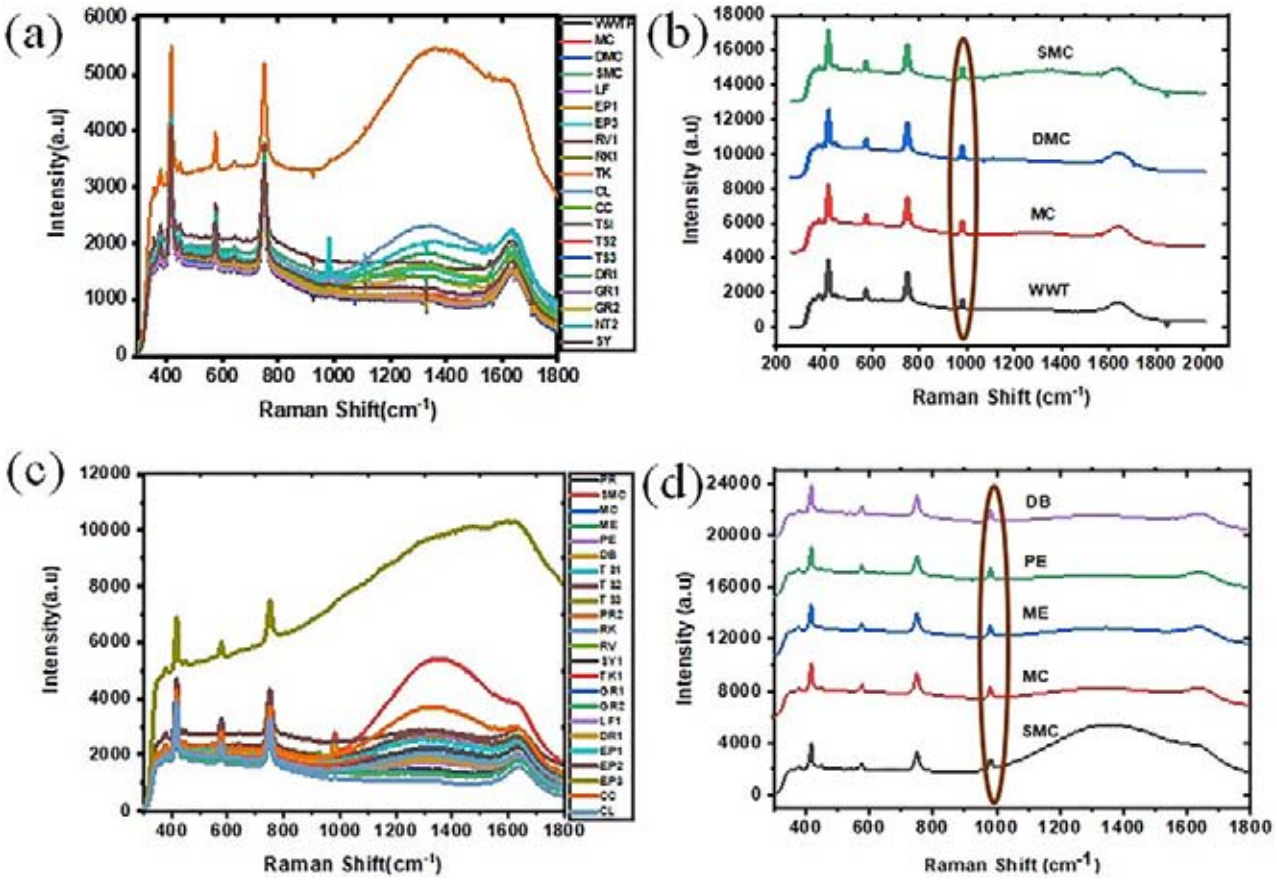

Figure 5.5. Raman results from the EU SWIM Project samples. (a) Water sampled from different rivers and beaches in the Dublin area (batch 3). (b) Selected samples from batch 3, in which sulfates were detected. (c) Water sampled from different rivers in the Dublin area (batch 4). (d) Selected samples from batch 4 in which sulfates were detected.

and b represents the Raman spectra from batch 3. The Raman system showed that the nitrate levels in all of these samples from the EU SWIM Project were below 50 mg/l, in agreement with the chemical analysis from ERI (Table 5.1). Although an encouraging result for the water quality of Dublin, this result was not a sufficient validation of the system. To compare the Car-boot Raman system results against the benchmark laboratory chemical analysis results for a compound that was present in meaningful levels, sulfate concentrations were then considered. Five samples from batch 3 present Raman peaks at around 981 cm^{-1}, which corresponds to the presence of sulfates (see Figure 5.5b), and these results from the Raman system were confirmed by the chemical analysis from ERI, as shown in Table 5.1.

Similarly, Raman spectra of samples from batch 4 are shown in Figure 5.5c. The Raman system predicted four samples with high concentrations of sulfates, as shown in Figure 5.5d. These predictions were validated by the results from ERI, as shown in Table 5.2.

Table 5.1. Concentration of sulfates and nitrates in the EU SWIM Project samples from batch 3

Sample	ASU code	Sulfate (mg/l)	Nitrate (mg/l)
CC	3434-P1	51.68	1.530
CL	3434-P2	32.08	1.036
DMC	3434-P3	2622.31	0.069
DR1	3434-P4	78.92	0.835
EP1	3434-P5	101.88	2.810
EP3	3434-P6	61.76	2.183
GR1	3434-P7	49.22	2.369
GR2	3434-P8	48.30	2.418
LF	3434-P9	49.28	3.842
MC	3434-P10	2535.42	0.083
NT2	3434-P11	2166.77	0.130
RK1	3434-P12	86.81	2.618
RV1	3434-P13	38.47	0.831
SMC	3434-P14	2297.67	0.304
SY	3434-P15	85.19	2.594
TK	3434-P16	77.38	5.673
TS1	3434-P17	60.89	3.167
TS2	3434-P18	54.52	2.979
TS3	3434-P19	54.82	3.624
WWTP	3434-P20	1830.80	0.996

ASU, Aquatic Service Unit of ERI.

Table 5.2. Concentration of sulfates, nitrates and phosphates in the EU SWIM Project samples from batch 4

Sample	ASU code	Sulfate (mg/l)	Nitrate (mg/l)	Phosphate (mg/l)
CC	3500-P1	60.95	1.442	0.109
CL	3500-P2	21.43	0.292	0.011
DB	3500-P3	2632.58	0.070	0.005
DR	3500-P4	2632.63	0.064	0.007
EP1	3500-P5	51.27	0.962	0.056
EP2	3500-P6	58.29	2.009	0.072
EP3	3500-P7	30.15	<0.010	0.149
GR1	3500-P8	51.98	2.232	0.118
GR2	3500-P9	49.41	2.086	0.130
LF1	3500-P10	48.06	2.136	0.025
MC	3500-P11	2444.70	0.062	0.018
ME	3500-P12	2502.65	0.056	0.005
PE	3500-P13	2525.06	0.061	0.020
PR	3500-P14	23.22	1.264	0.039
PR2	3500-P15	57.46	11.949	0.015
RK	3500-P16	69.13	2.971	0.071
RV	3500-P17	88.75	0.674	0.096
SMC	3500-P18	2391.92	0.061	0.045
SY1	3500-P19	73.13	1.746	0.081
TK1	3500-P20	85.22	1.641	0.029
TS1	3500-P21	33.34	2.376	0.093
TS2	3500-P22	26.28	1.236	0.058
TS3	3500-P23	23.97	0.629	0.187
Na_2SO_4	3500-P24	1007.86	NA	NA

ASU, Aquatic Service Unit of ERI. NA, not applicable as this sample contains only Na_2SO_4 sent for calibration and it does not contain any nitrates or phosphates.

It should be noted that while performing the Raman analysis for these samples from the EU SWIM Project, no processing of the water samples took place (e.g. filtering, screening). Therefore, various components in water created a background noise to the Raman measurement. Nonetheless, the system performed well and showed agreement with the results of the chemical analysis, and successfully detected the absence/presence of the targeted contaminants. The Raman spectra obtained from the EU SWIM Project samples were used to initial train and test the AI algorithms, following an 80/20 random split between training data and test data. However, it became apparent that the results were sensitive to system variables, i.e. integration time and the number of averages used. Therefore, although the EU SWIM Project data were very valuable, additional datasets

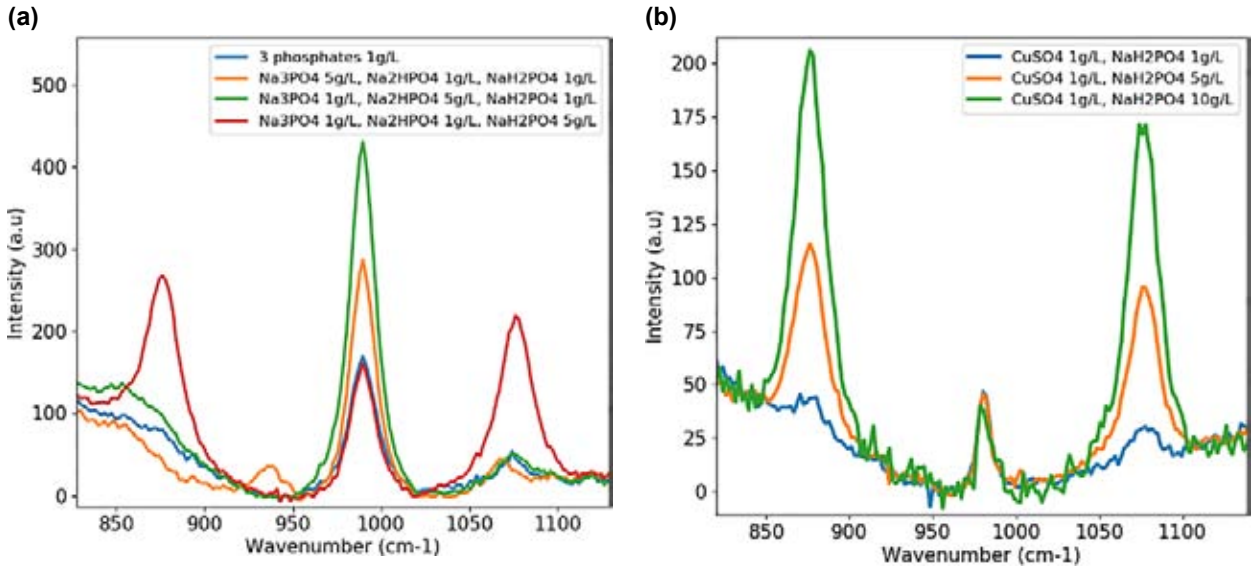

Figure 5.6. Raman spectra of mixtures of (a) phosphates and (b) sulfates and phosphates.

were prepared by the project team, particularly for nitrates, and subsequently used for model training.

5.3.2 Sample preparation for data analysis by machine learning

CAPPA provided the Hydrolight machine learning team with datasets for different salts dissolved in water, first to train the AI algorithm to identify nitrates and phosphates in water, and then to test the algorithm. The following salts were provided:

- nitrates of varying concentrations from 10 mg/l to 25 g/l, where the limit for nitrates is 50 mg/l as per the European drinking water quality directive (EU, 1998) and S.I. No. 122 of 2014 (Government of Ireland, 2014)
- sulfates from 25 g/l to 0.5 g/l;
- phosphates from 25 g/l to 0.8 g/l;
- mixture of sulfates and phosphates from 10 g/l to 1 g/l;
- mixture of nitrates and phosphates from 12.5 g/l to 1 g/l;
- datasets with water and no salts.

Stock solutions were prepared from these salts, to give known concentrations.

Sample Raman spectra for mixtures of phosphates and mixtures of sulfates and phosphates can be seen in Figure 5.6. In summary, a set of SOPs were developed that allowed the hardware and samples to be used consistently, to ensure that high-fidelity data were generated. Using these techniques, consistent Raman spectra were produced by the Car-boot version, appropriately labelled with the results from chemical and biological analysis; these were passed to the Hydrolight team to train and test the AI algorithms.

6 Hardware Version 3 – Autonomous Buoy Version

6.1 Objectives

The main objective of the Buoy version was to create a full end-to-end operational model incorporating all of the functionality of an autonomous water quality monitoring platform, using Raman spectroscopy and AI for parameter detection. This would allow the full end-to-end system to be demonstrated in the field. It was also envisaged that this version would form the basis for the operational LOC version, in terms of core functionality and design.

6.2 System Description

6.2.1 Overview

The primary technique used for water analysis in this project is RS, which generally involves using expensive laboratory-grade equipment, specifically a spectrometer combined with a Raman probe, a suitable laser, filters, etc. Ordinarily, such equipment is located in a laboratory. However, for this project a system was required that could operate autonomously, harvest data from the spectrometer and send the data wirelessly to the cloud for analysis, while managing power consumption and ensuring integrity of the data. Figure 6.1 shows a block diagram of the overall system needed to support the operation of the spectrometer.

The spectrometer used was a QE Pro from Ocean Optics, using a 5-V, 5-A supply; an Ondax Laser Module was used. Data from the spectrometer are normally acquired using Ocean Vie software, on a lab-based PC. As this was an embedded application, a microcontroller was needed to acquire the data from the spectrometer using serial communication. To switch the spectrometer on and off, appropriate power supply hardware was used to supply the necessary voltage and current. On-board data storage required an external memory, and wireless transmission of data was also needed. Table 6.1 lists the main components used.

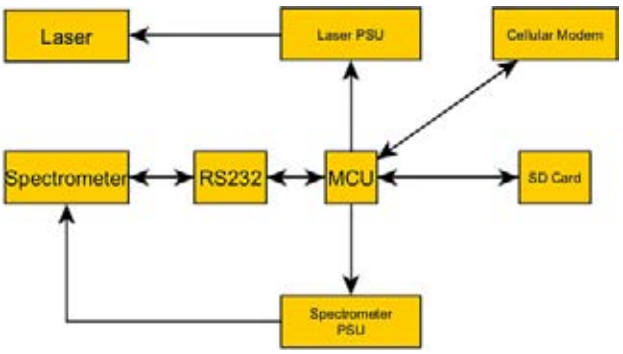

Figure 6.1. Simplified block diagram of system. MCU, main control unit; PSU, power supply unit; RS232, recommended standard 232 data cable.

6.2.2 Printed circuit board design

A schematic of the system was drafted in Altium Designer and committed to printed circuit board (PCB) and manufactured by Eurocircuits (Figure 6.2). Assembly was done by hand in Nimbus (MTU).

6.2.3 Other aspects

Cooling was achieved using four 12-V DC fans that blow air through pipes connected to the enclosure,

Table 6.1. Schedule of main components in the Buoy version

Component	Description
Spectrometer	Ocean Optics QE Pro scientific-grade with Hamamatsu S7031-1006S Detector and RS-232 interface
Laser	Ondax Raman Butterfly laser module
Printed circuit board	Custom-designed by the project team, using a modular “plug-in” approach
Main control unit	ST Microelectronics Nucleo F401RE development board
Main control unit firmware	Developed with ST/Atollic True Studio Eclipse, coded in embedded C
RS232 converter	Mikroe MAX232 board with a MAX232N drive/receiver IC
On-board data storage	Sparkfun SD card module using SDIO protocol
Spectrometer/ laser power	Murata UEI15 series DC/DC converter
Wireless communications	Mikroe GSM4 click-board using UBLOX SARA-G350 2.5G GSM module

GSM, global system for mobile communication; SD, secure digital.

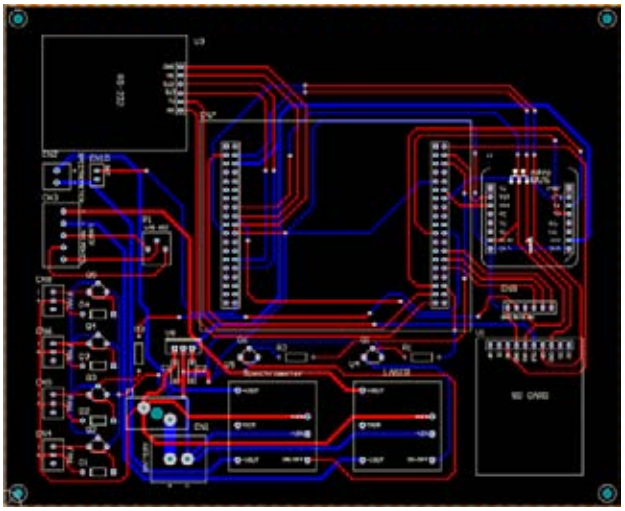

Figure 6.2. Two-dimensional image of PCB layout.

and mounted under the buoy hull, acting as heat exchangers for the circulating air. The fans are switched using BC547 bipolar transistors controlled by the main control unit (MCU). Various mounting brackets were designed and fabricated using an Ultimaker 3-D printer. All components were housed in a sealed enclosure (Figure 6.3) and tested for end-to-end operation. During testing, spectral data were sent to a post-testing server website for verification. The data consisted of 4176 bytes of spectral data; every 4 bytes represented one pixel of the spectrum.

6.2.4 Buoy integration

Once the end-to-end transfer of data was completed and the transfer communication system was debugged, the enclosure and system were installed in the buoy housing, developed by boat-building company Customworks.

Figure 6.4 shows the completed buoy with Perspex deck, and hinged lid and separate housing for the Raman probe.

Figure 6.5 shows the completely wired-up system installed in the buoy; on the left side of the picture the two 12-V batteries can be seen.

6.3 Testing

6.3.1 Testing strategy

Testing the system was carried out in a structured way to prove its functionality. The sequence of testing followed the order of the steps required for data collection, processing, transmission, verification and analysis; and included testing in both laboratory and field environments. The main stages were:

- harvesting data from the spectrometer;
- processing the data locally;

Figure 6.3. The top view of the system in the enclosure.

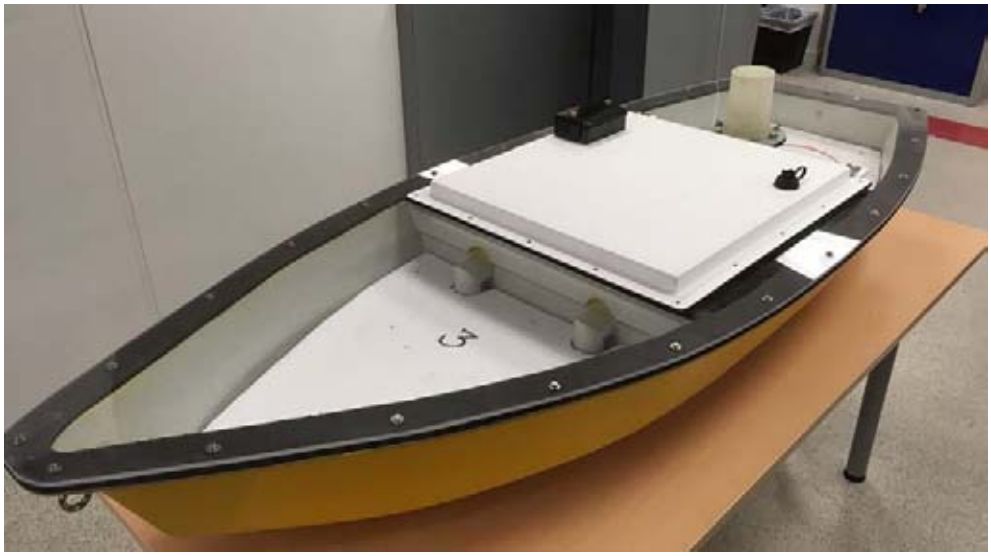

Figure 6.4. Buoy-side view.

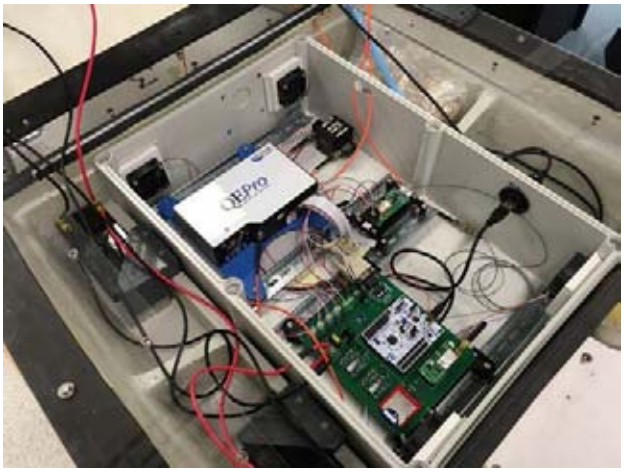

Figure 6.5. Enclosure mounted in the buoy.

- sending the data to the generic cloud server;
- data format verification;
- Rinodrive[2] server connectivity;
- calibration testing;
- laboratory end-to-end testing;
- field deployment;
- end-to-end functionality testing.

6.3.2 Testing – the initial stages

- Harvesting spectrometer data:
  - extracting data using RS232 communication, via a custom firmware driver; and
  - ensuring that data are successfully captured.
- Processing the data locally:
  - customising firmware driver for data storage on an SD card;
  - ensuring integrity of the stored data; and
  - locally converting the data into the correct format for sending to the cloud.
- Sending data to the cloud (generic server):
  - creating a firmware driver for the cellular modem; and
  - during functional testing, sending the spectral data to a post-testing server website.
- Data format verification:
  - ensuring that the correct number of data were received, at the server end, and in the correct order.
- Rinodrive connectivity:
  - creating a communication bridge between the buoy and the Rinolab server; and
  - sending randomised data to the Rinodrive platform to test connectivity.

6.3.3 Raman spectroscopy calibration testing (original equipment manufacturer software versus the Watermon system software)

Controlled calibration tests were conducted in the lab using a control solution of IPA with a defined Raman spectrum signature. In this way, the characteristics of the data sent to the cloud could be checked for

2 Rinodrive is the proprietary name of Hydrolight's data management toolkit, which was used for the project.

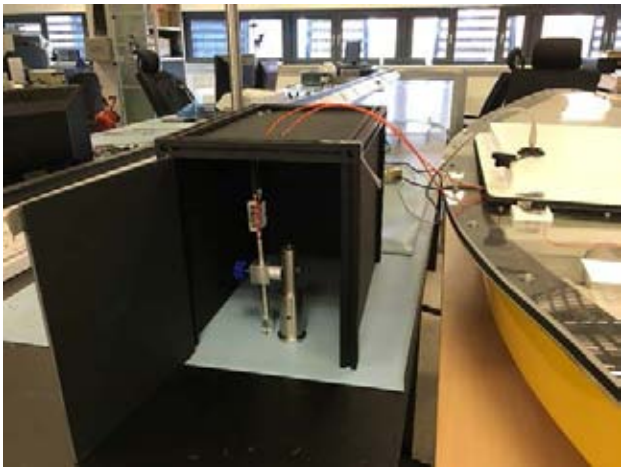

Figure 6.6. Raman probe immersed in IPA solution.

correct reading of the Raman probe, data capture, transmission and display. Figure 6.6 shows the probe removed from the buoy sitting in a vial of IPA control sample.

A side-by-side test using a control IPA sample was conducted using the spectrometer controlled using either a laptop and the software (OceanView) provided by the original equipment manufacturer (OEM) or the buoy system. The results are shown in Figures 6.7 and 6.8 and the graphed data were confirmed by CAPPA team members to display the correct Raman signature of IPA. This confirmed that the RS system in the buoy was reading correctly.

6.3.4 Laboratory Raman spectroscopy functionality testing (using nitrates as a case study)

A range of laboratory-prepared samples with defined levels of nitrates were tested, using different spectrometer integration times and read-averaging numbers. These were compared with the spectrograph obtained using the manufacturer's OceanView software. The set of data generated for this aspect of the project comprised 72 spectra. See section 6.4 for a discussion of the results.

6.3.5 Field deployment end-to-end functionality testing

The buoy was deployed for field trials of the full end-to-end system, including using AI to detect target analytes. Chapter 9 describes the findings.

6.4 Results of the Raman Spectroscopy System Testing

6.4.1 Water sample testing

Figures 6.9 and 6.10 show the results of a water sample containing a high concentration of nitrates (5g/l) obtained using the OEM software and the buoy system, respectively. This high concentration was used

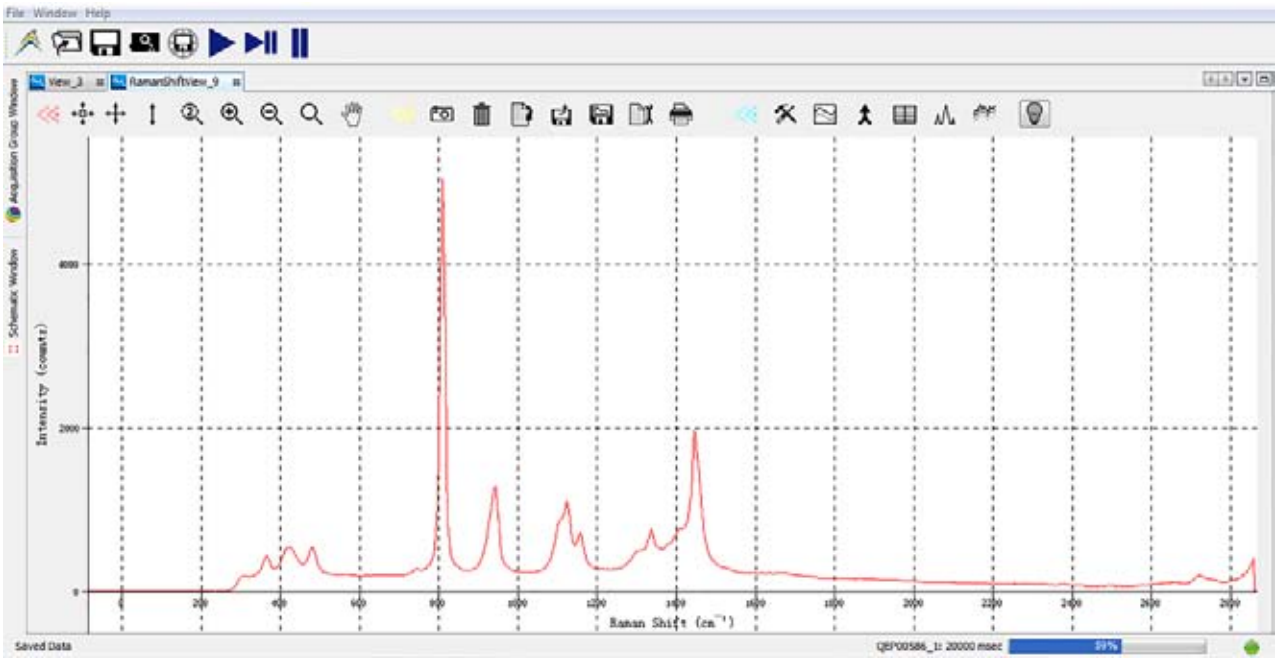

Figure 6.7. OEM software IPA spectrograph.

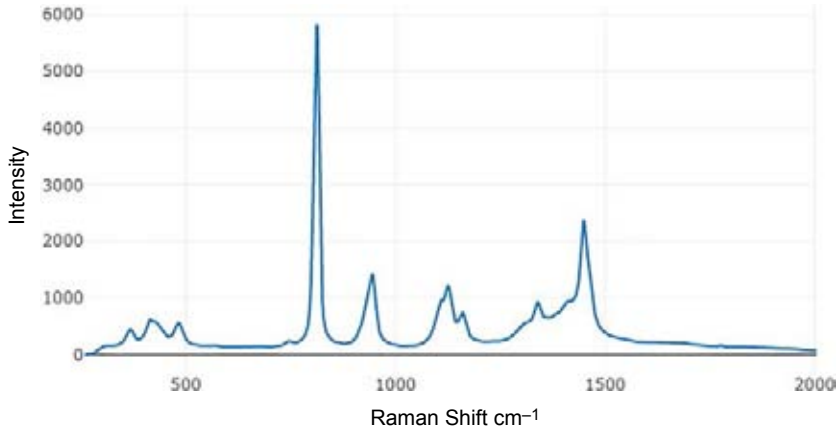

Figure 6.8. Buoy system IPA spectrograph; result graphed on the Rinodrive platform.

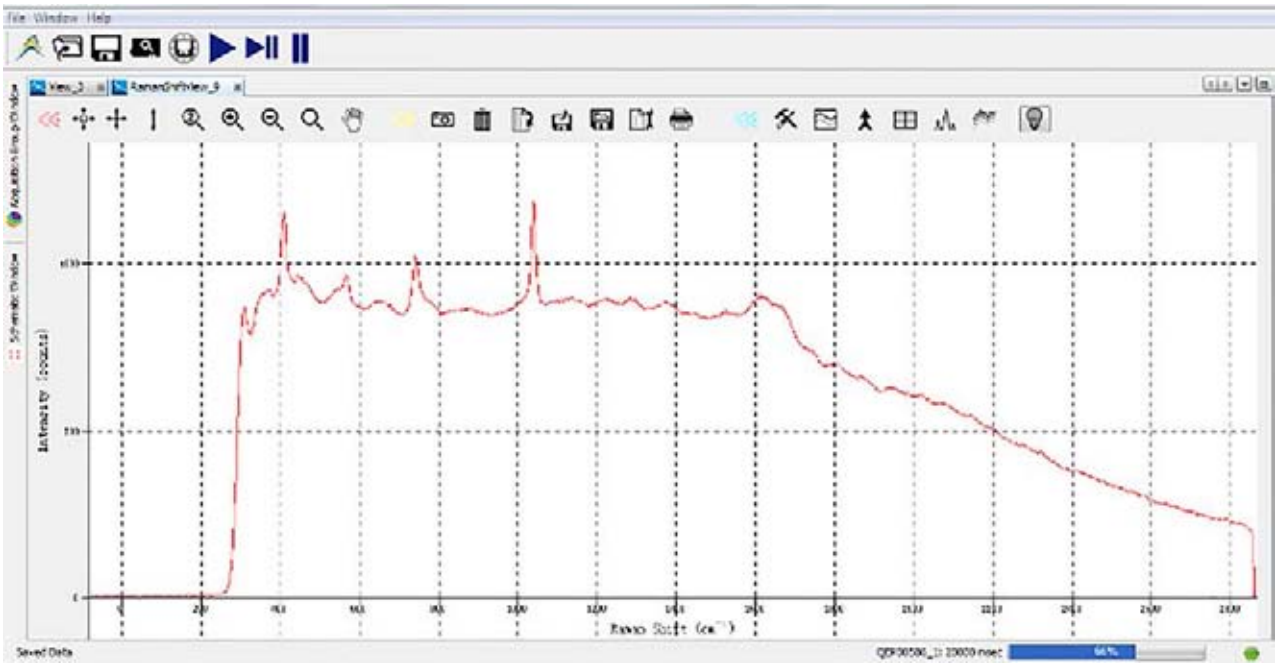

Figure 6.9. Sample nitrate concentration (5 g/l) determined using the OEM software.

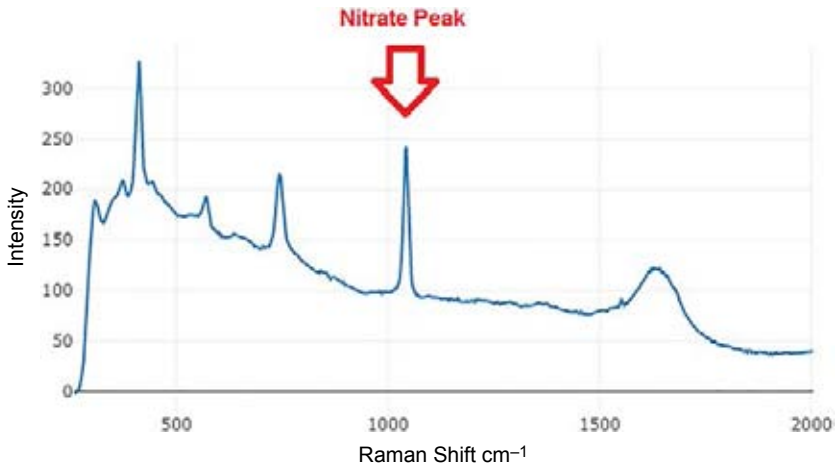

Figure 6.10. Sample nitrate concentration (5 g/l) determined using the buoy system.

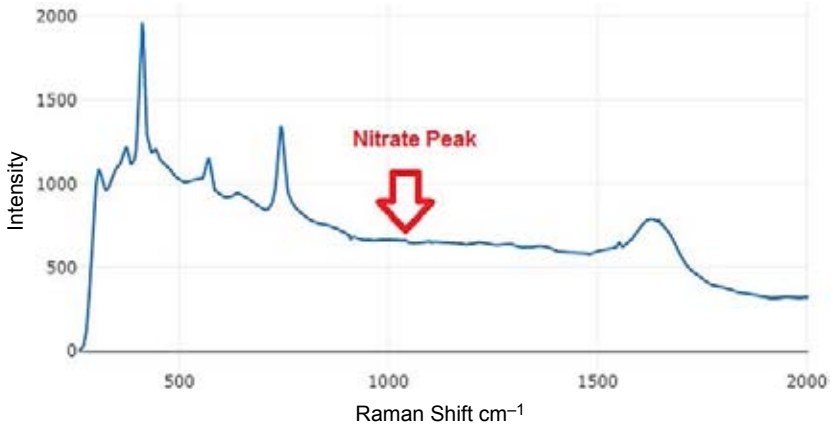

Figure 6.11. Spectrum for samples containing 40 mg/l nitrate.

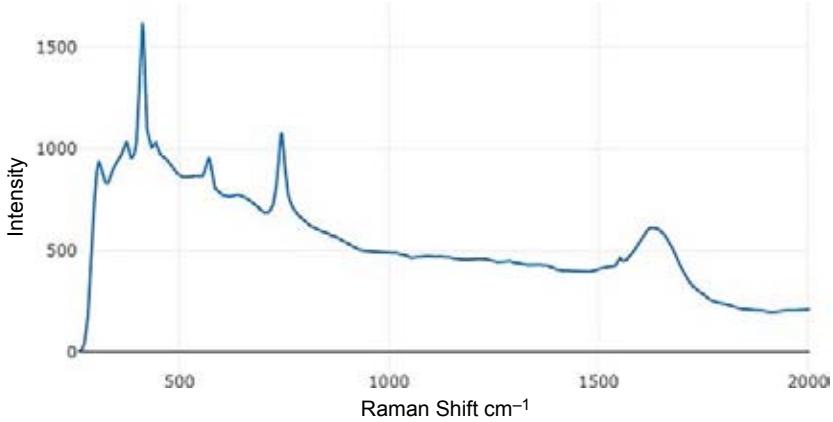

Figure 6.12. Raman profile of water.

to clearly illustrate the nitrate peak at *c*. 1043 cm^{-1}. The buoy system accurately detected both the profile of the water sample and the spike of nitrate contaminant. All test results shown below were conducted using a spectrometer integration time of 10 seconds and 50 averages.

By contrast, Figure 6.11 shows a barely visible nitrate signature for 40 mg/l nitrates at 1043 cm^{-1}. At this concentration, the spectrum peak is almost imperceptible on a graphical plot; traditional analysis of a single spectrum such as this, in isolation, may be inadequate for detection.

Results from testing potable water using the same spectrometer configuration are shown in Figure 6.12, to illustrate the lack of obvious differences in Raman spectra between the samples when concentrations of nitrates are low (Figure 6.11 vs. Figure 6.12).

6.5 Conclusion

The Buoy version was comprehensively tested for end-to-end functionality, in terms of capture of Raman spectra, data management and transmission to the cloud-based platform. The system is fully operational and functioning correctly; it reproduces calibrated (OEM) spectra correctly. Chapter 8 describes the data processing workflow from when the data arrive in the cloud. The graphs of the spectra for low nitrate concentrations illustrate the reasons for using AI to facilitate detection at such concentrations – see Chapter 8 for further details.

7 Hardware Version 4 – Lab-on-Chip Version

7.1 Objectives

The objective of the LOC version is to decrease size and cost, and improve the large-scale manufacturing potential of the Watermon for water quality monitoring. Under the present project, the MTU–CAPPA team has advanced the LOC version in a succession of steps.

Early-stage work examined specific details of the envisaged LOC. Through that work, it became clear that the optofluidics element presented challenges regarding aspects such as biofouling/clogging and the “optical tweezers” (for bacteria concentration). Therefore, the proposed channel configuration was altered to avoid fluid confinement. For the tweezing, a dielectrophoresis trapping technique from collaborators at the University of Jena was sourced. This modified approach was presented to the project steering committee for discussion and feedback, after which development of the LOC version proceeded. In parallel (as described above), the Car-boot and Buoy versions also advanced, to provide meaningful data for the project in the short term. These versions were developed to a higher TRL and had the ability to provide results to the AI portion of the team, in the required timeframe.

7.2 System Development and Testing

7.2.1 Lab-on-Chip schematic

In high-level terms, any LOC version must have:

1. a laser light source;
2. a method to position/concentrate the target sample;
3. a method to conduct the laser light onto the sample; and
4. a spectrometer to capture the Raman spectra induced within the sample.

The most challenging aspects for the project were items (2) and (3), and it was in these areas that most attention was concentrated during the system development. Work was also carried out to develop a new spectrometer, item 4, as described in section 7.3.

The lasers used in the process were available “off the shelf” in all cases. A schematic of the initial LOC design is shown in Figure 7.1.

A number of other factors affect the performance of LOC-type Raman systems, as demonstrated by the project team during the project. These include the material used as the chip substrate, the nature and orientation of light-conducting fibres, the use of surface enhancements, the ability to concentrate and accurately target a suitable sample with laser light, the level of signal interference/noise generated by material defects in the system, etc. The following sections describe the iterative process of development and testing that was undertaken for the LOC version, and illustrate the progress made.

7.2.2 Alternatives to the optofluidics glass chip concept

Alternatives were explored for the material of the chip, the layout and fabrication method for the laser light-channelling system, and the sample handling method within the LOC. Alternative candidate materials included polymer resin, SU8 (a commonly used

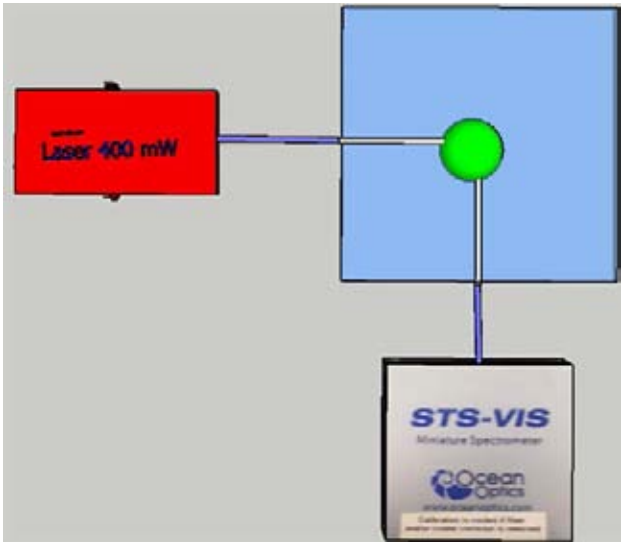

Figure 7.1. Schematic of the LOC design with grooves to hold the fibres.

epoxy-based negative photoresist) and aluminium. The alternative sample handling pursued was surface placement on the chip. The physical arrangement of the waveguides was reviewed. Finally, it was recognised that considerations such as anti-fouling would depend on the outcome of these investigations. Given the change in approach, the open configuration (avoiding microchannels for fluids) greatly mitigates many issues. The open channel configuration offers advantages with respect to microfluidic channels when it comes to biofouling. Mechanical cleaning is a viable solution in the open configuration (and with the probe), whereas in the channel very complicated techniques would have been required. It should be noted that in the open (and probe) configuration there is no need to bring samples into the unit.

7.2.3 Open channel fibre-grooves

A chip with open channels to hold optical fibres was fabricated using a high-quality stereolithography-based 3-D printer using polymer resin. The CAD design and the 3-D printed device of the LOC are shown in Figure 7.2a and b, respectively. This LOC platform consists of grooves that hold excitation and collection fibres, positioned orthogonal to each other, connected to the LOC laser and spectrometer. IPA droplets were placed in the sample slot and covered with a calcium fluoride (CaF_2) slide to avoid evaporation. The IPA Raman spectrum from using this LOC is shown in Figure 7.2c. This corresponds well with the results from the full-scale laboratory equipment.

7.2.4 Closed channel fibre grooves

To improve the performance, in an improved design closed channels were used to hold the fibres more securely, minimising the use of adhesives, as shown in Figure 7.3a and b. This closed-channel-based system gives better Raman signals (Figure 7.3c) than the open-channel LOC.

Figure 7.2. LOC Raman system with grooves to hold the fibres. (a) CAD design, (b) 3-D printed chip with a stereolithography apparatus (SLA) 3-D printer using resin and (c) Raman spectra of IPA using the resin LOC.

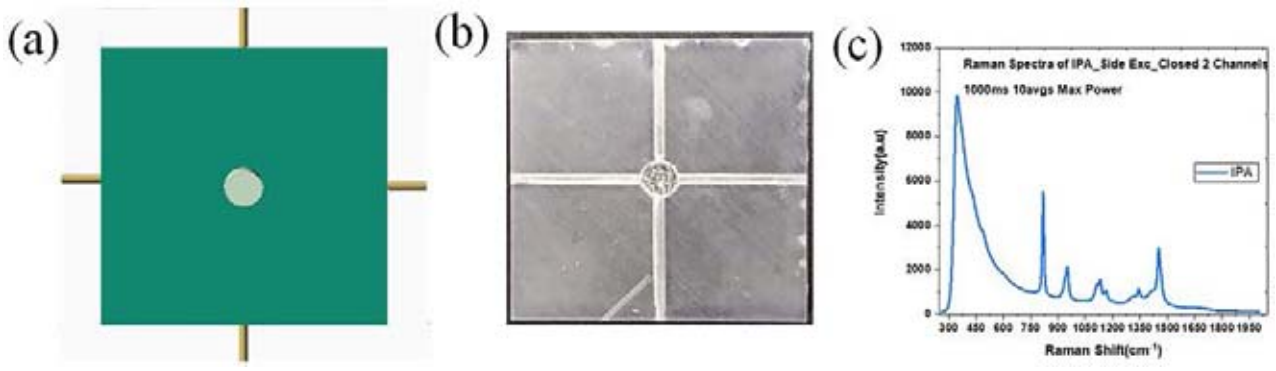

Figure 7.3. LOC Raman system with closed fibre holders. (a) CAD layout, (b) 3-D printed LOC using resin and (c) Raman spectra of IPA using the device.

7.2.5 Closed-channel optimisation

To further improve the performance of the closed-channel Raman LOC, different arrangements of the light-conducting fibres close to the sample chamber were tested, using IPA as a test sample. The results are presented in Figure 7.4, which demonstrates that, of the three arrangements tested, the best option is that shown in Figure 7.4b, because it provides the sharpest spectrum; in the configuration in Figure 7.4a, diffraction reduces the amount of Raman scattered light collected, whereas in the configuration in Figure 7.4c, the collection fibre may collect some of the pump light, increasing the background signal.

7.2.6 SU8-based designs

The 3-D printed LOCs described in section 7.2.5 provided insights for improving the LOC design and rapidly testing ideas. However, to bring this technology to market, the MTU–CAPPA team has been employing designs more suited to mass manufacturing, namely photolithography-based techniques. For this purpose, a set of SU8-based LOC designs were fabricated and tested at MTU and Tyndall National Institute, Cork, using their semiconductor device manufacturing tools, which are capable of low-volume mass production (for example, Figure 7.5). These devices continue to be evaluated at CAPPA in ongoing work.

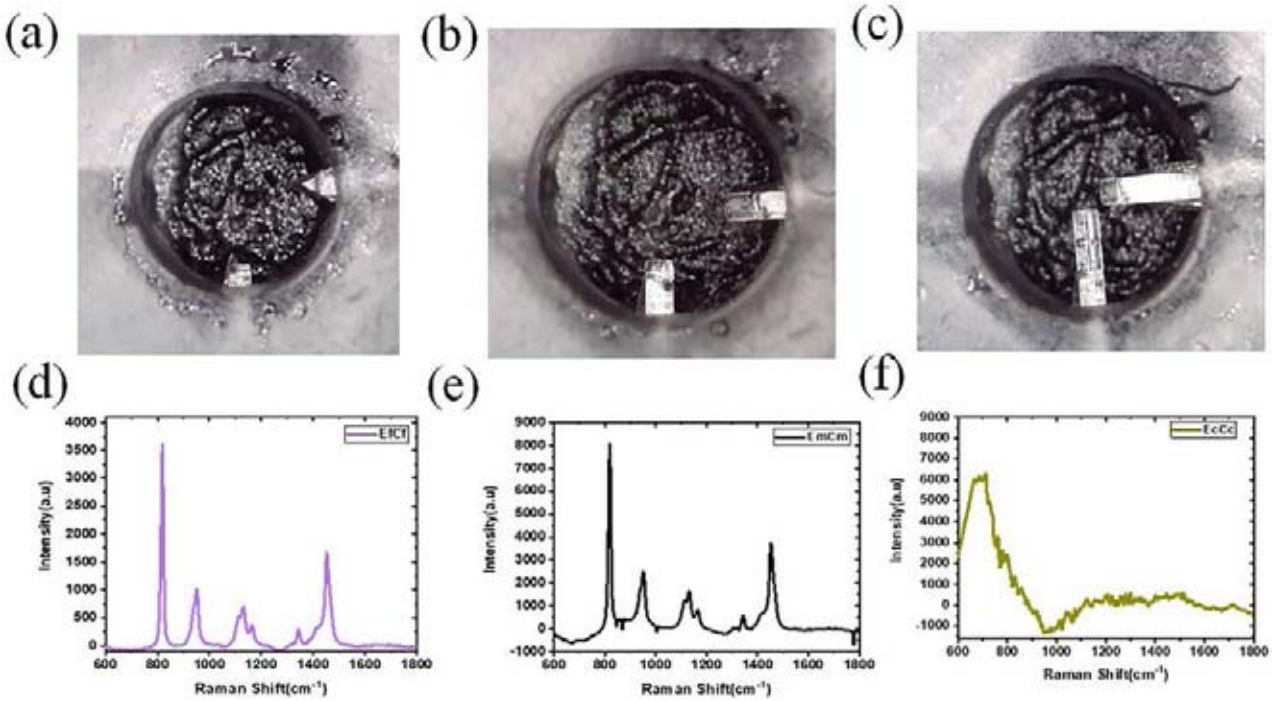

Figure 7.4. Optimisation of fibre positions in the 3-D printed LOC platform. (a–c) The LOC of Raman systems with various configurations of fibres. (d–f) The corresponding baseline-corrected Raman spectra of IPA at 1000 ms, 5 read-averages, 415 mW, maximum laser power.

Figure 7.5. SU8-based LOC device. (a) Layout, (b) fabricated device and (c) characterisation of the device.

7.2.7 *Surface-enhanced Raman scattering-improved systems*

SERS effects can improve the performance of RS. To explore this option, a groove-based LOC design was 3-D printed on aluminium substrate (Figure 7.6a). The experimental results demonstrated an increase in spectrum intensity but at the same time an undesirable increase in fluorescence, as shown in Figure 7.6b. To overcome this problem, MTU–CAPPA is also evaluating the computer numerical control (CNC) machining of the LOC system (Figure 7.6c) as a further alternative.

7.2.8 *Dielectrophoresis*

To further improve the detection limits of the LOC version, particularly for *E. coli*, the MTU–CAPPA team established a collaboration with the University of Jena in Germany, which has developed technology based on dielectrophoresis (DEP), which enables *E. coli* to be concentrated in a small region of a chip-based detection system, thus enhancing the Raman signals (Cheng *et al.*, 2007). The system is illustrated in Figure 7.7. DEP allows the spatial manipulation of particles through the interaction of the sample with a non-uniform electric field. The technique offered an interesting avenue of investigation for the present project, since it allows the spatial manipulation of cells and bacteria, without the need for biochemical labels or other bioengineered tags. Figure 7.7 shows *E. coli* under an electric field, with an applied voltage of 4 V peak to peak. Visualisation is achieved through a 40× objective microscope. Figure 7.8 shows the Raman spectra obtained from the *E. coli* strain Nissle using the DEP with the Horiba Xplora Raman system at CAPPA.

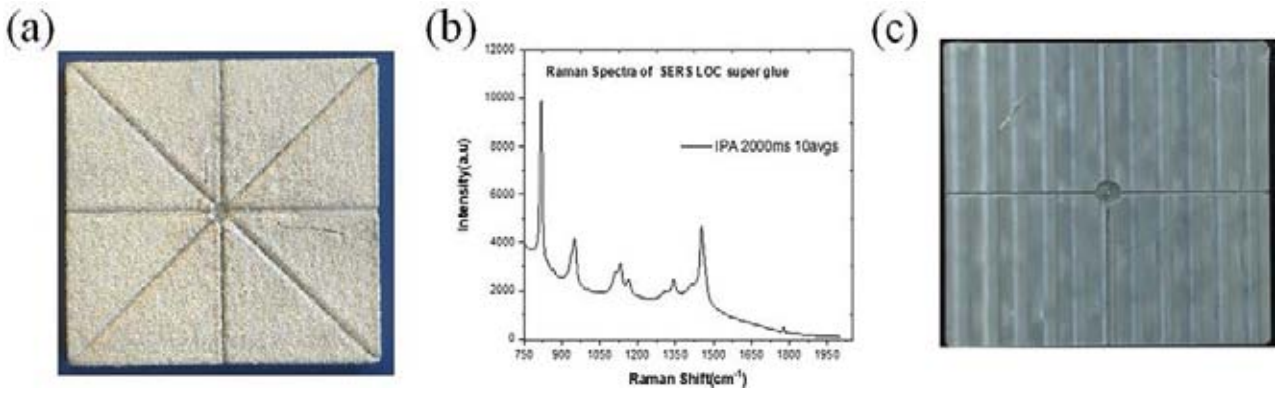

Figure 7.6. Alternative fabrication methods tested. (a) 3-D metal printed LOC platform on aluminium, (b) Raman spectra of IPA using an aluminium LOC and (c) CNC-based LOC device.

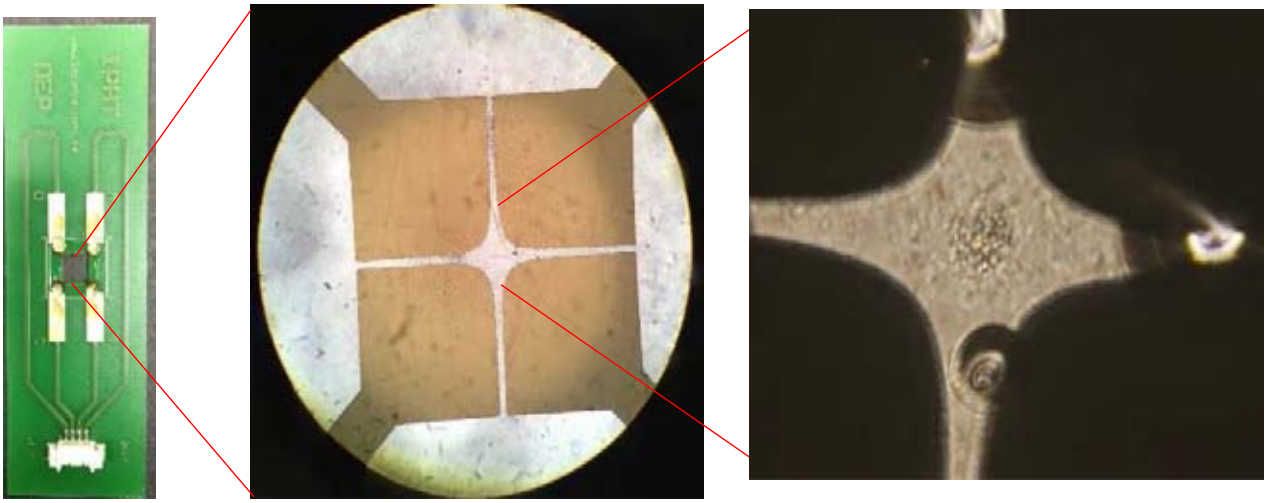

Figure 7.7. Dielectrophoresis Jena University chip, measuring *E. coli*.

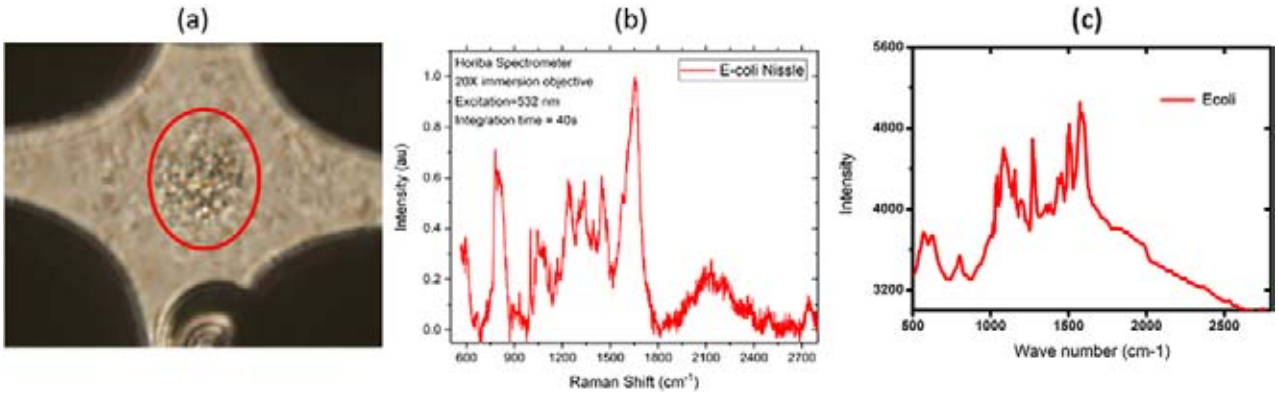

Figure 7.8. Results of testing DEP with *E. coli* Nissle. (a) DEP with concentration of *E. coli* Nissle in the middle of the chip, (b) Raman spectrum of *E. coli* and (c) Raman spectrum of *E. coli* using spectral data from Cheng *et al*. (2007).

The DEP chips from Jena were tested with *E. coli* Nissle and CAPPA's benchtop microscopy Raman system (Horiba Xplora), with an immersion objective 20×, and excitation wavelength of 532 nm. The concentration of *E. coli* was 2.4×10^8 cells/ml. The resulting Raman spectrum (Figure 7.8b) displays potential for DEP to be used as a basis for a fully operational LOC *E. coli* detection system, having some similarities to *E. coli* Raman spectra found in the literature (Figure 7.8c) (Cheng *et al*., 2007). It proved difficult to combine the DEP chip with Raman probe-based systems owing to probe alignment issues. However, the DEP electrodes are fully compatible with the processes required to fabricate the SU8 waveguides. A more promising avenue of development appears to be to integrate photo-lithographically realised polymer (SU8) waveguides on top of DEP chips. These polymer waveguides can efficiently collect the Raman signal from concentrated bacteria, and it is planned that the system envisaged will form part of further work for the project team (Cheng *et al.*, 2007).

7.3 Customised Spectrometer for the Lab-on-Chip Raman System

One challenge to produce a low-cost, compact RS system is the high cost of the spectrometer. Hence, the project team took early-TRL steps to develop a low-cost spectrometer. This involves hardware development and software models, collection optics, light power sources, etc. A compact opto-mechanical module with customised wavenumber region and acceptable resolution and SNR was assembled.

Figure 7.9 shows the assembly of the diffraction grating, the image sensor consisting of a PCB soldered in a 3-D housing. This custom spectrometer was calibrated using three light-emitting diodes (LEDs), 405 nm, 650 nm and white LED, allowing each element of the charge-coupled device (CCD) array to be assigned an appropriate wavelength. The spectra

Figure 7.9. Diffraction grating and spectroscope assembled in 3-D printed casing for the custom spectrometer.

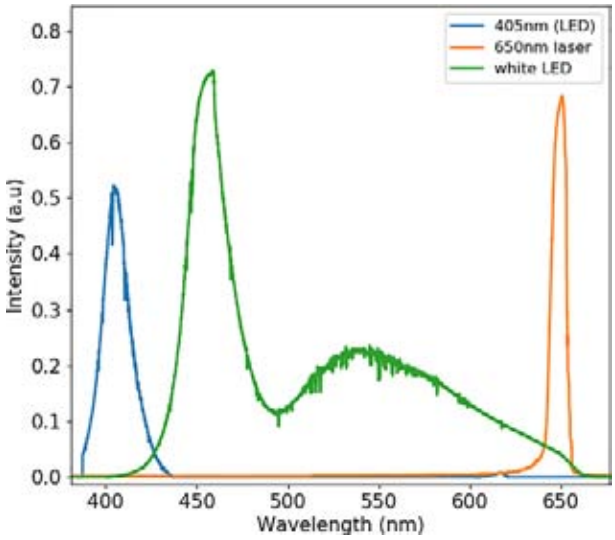

Figure 7.10. Calibration of the custom spectrometer.

from these LEDs is shown in Figure 7.10. The CAPPA team continues to develop this miniaturised spectrometer, with the goal of combining it with the LOC version Raman and the DEP chips to realise a high-performance low-cost sensing platform.

7.4 Summary

A series of different compact RS systems were developed during the project. By virtue of its robustness and maturity, the Car-boot version was used for deployment on the buoy and to generate the data required for machine learning analysis. As the machine learning would require re-training if the hardware was altered, the Car-boot version was frozen once a satisfactory level of performance was reached, so that consistent datasets were generated. Considerable know-how was developed on how to use the Raman system, e.g. on the use of aluminium sample holders and appropriate analysis techniques, such as background subtraction. Such analysis techniques are key to successful RS, and this process took longer than expected. Once development of the Car-boot version was complete, efforts shifted to the LOC versions, which successfully

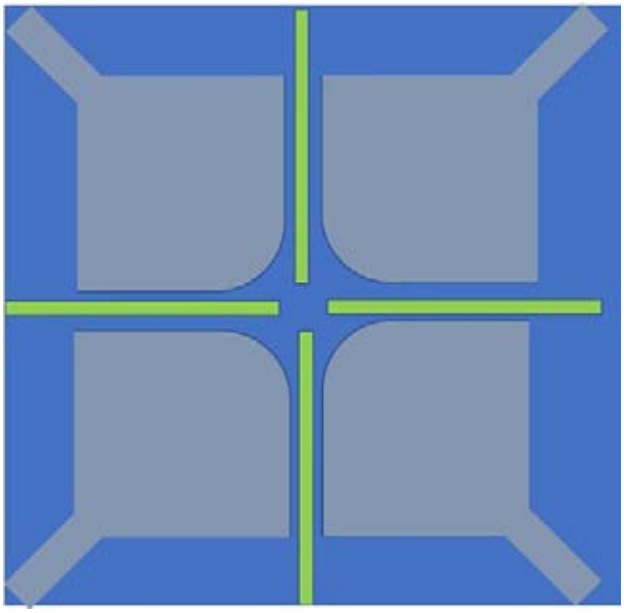

Figure 7.11. Plan view of the proposed LOC/DEP combination. The SU8 waveguides are shown in green and the electrodes in grey. The waveguides can be fabricated after electrode deposition.

demonstrated nutrient detection. *E. coli* concentration was successfully demonstrated using DEP. The strategy selected was chosen for the compatibility of the SU8 waveguides and DEP, although the project ended before they could be combined into a full LOC version. Figure 7.11 shows a schematic for this system, which represents the ultimate form that the LOC Raman system will take, and which will continue to be developed.

It proved very difficult to reliably detect *E. coli* using traditional, deterministic analysis,[3] i.e. by analysis of peaks in the spectra, including using techniques such as background subtraction and others. This is because of the weakness of *E. coli* as a Raman scatterer and the complexity of the spectrum. In other words, *E. coli* produces a large number of weak peaks, whereas nitrate produces a single peak. Machine learning techniques developed by the project team proved more successful than deterministic analysis, as discussed in the next chapter.

3 “Deterministic” analysis refers to the traditional analysis, which, for example, calculates the area under a peak, the peak maximum and other numerical analysis techniques. These techniques analyse only a single spectrum at a time.

8 Artificial Intelligence Software and Platform

8.1 Artificial Intelligence/Machine Learning Approach

8.1.1 Overview

The state-of-the-art review indicated that proven ensemble class machine learning methods would work well with the high dimensionality produced by continuous Raman datasets. The focus for this project was the use of RF and boosting (XGBoost) classifiers (Zhang and Ma, 2012).

8.1.2 Technology stack key components

The technology stack has a number of key components (Figure 8.1):

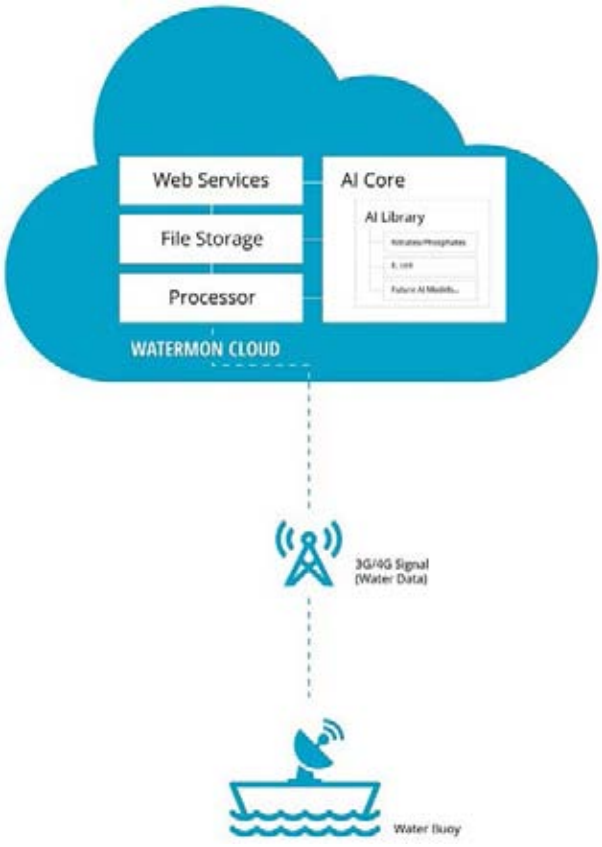

Figure 8.1. Component diagram for the AI Watermon cloud.

- cloud infrastructure for the high availability and low cost of AI computations;
- the latest micro-frameworks for Python; and
- well-established machine learning libraries in Scikit-learn.

Web services were used as integration points between the buoy hardware and the AI Core web platform. Plotly graphs were used for signal visualisation and Bootstrap was used to create a standardised mobile-friendly front-end dashboard. For data storage, the secure Rinodrive platform was used with MySQL Database.

8.1.3 Details of components

Cloud availability

RS data are stored securely on cloud storage selected for high availability, where they can be wrangled and processed using the required processing power ready for input to the AI model library for classification (Kandel *et al*., 2011).

Artificial intelligence core platform

This is a web-based services application for data handling, error logging and hosting the trained AI models for parameter detection, a lean, fast micro-framework type of application. This is the functional heart of the AI platform.

Signal interpolation function

The received data file is processed through an interpolation function developed in line with the Raman spectroscope technical guidelines.

Artificial intelligence library

This contains the Watermon pre-trained AI models, which process incoming Raman data sequentially and store the results in the database. The library is extendable with additional models.

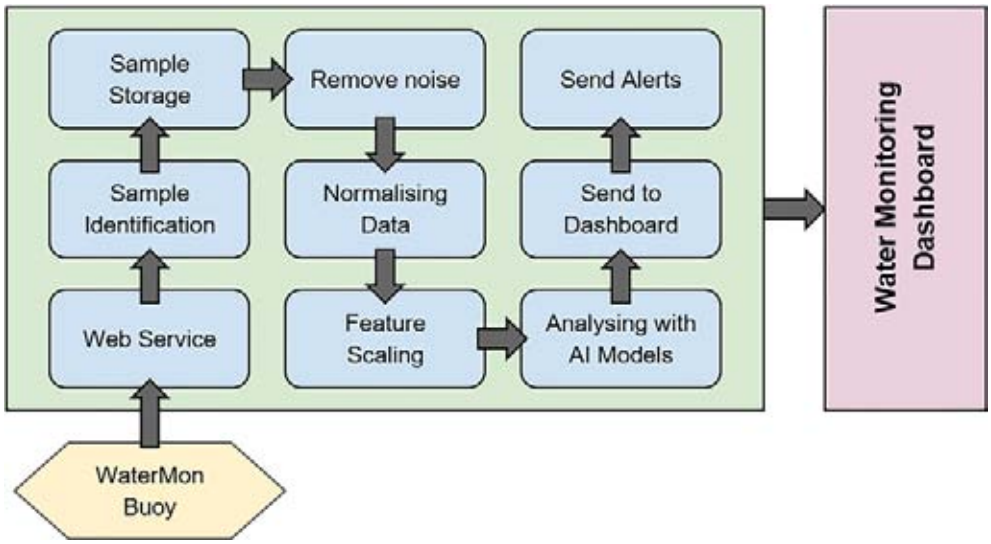

Figure 8.2. The data processing workflow.

8.2 Data Processing Workflow

The workflow for processing data from the autonomous buoy platform is shown in Figure 8.2.

8.2.1 Watermon buoy

The buoy sends data from RS analysis over a cellular connection to the cloud-based water monitoring platform. The data are sent as a dark.txt (background noise) file and a water.txt file containing the actual water sample.

8.2.2 Web service

The platform receives the data over an HTTP message. This file is processed through a web service on the cloud platform.

8.2.3 Sample file identification

On arrival at the web service the file type is checked to determine whether it is the file for a dark sample or an actual water sample. If it is file for the dark sample, it is stored in the database for pairing with its matching water sample file. When the water sample file is received, the dark.txt file is used in the background noise reduction stage.

8.2.4 Sample file storage

Both dark and water sample files are retained in the application database for use during analysis. The database has two distinct database tables for storing of the samples, as shown in Figure 8.3.

8.2.5 Removing background noise

Each data point of the dark sample (laser off) is subtracted from the equivalent water sample (laser on) (Figure 8.4). The rationale for this procedure is described in the OceanView Guide (OceanView, 2013). In summary, a sample-specific background spectrum is captured and used to eliminate undesirable effects, including ambient light, other

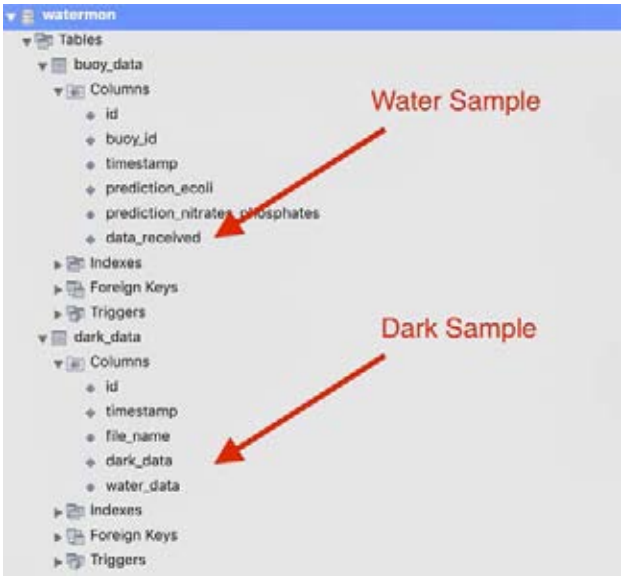

Figure 8.3. Watermon database table schemas for dark and water samples.

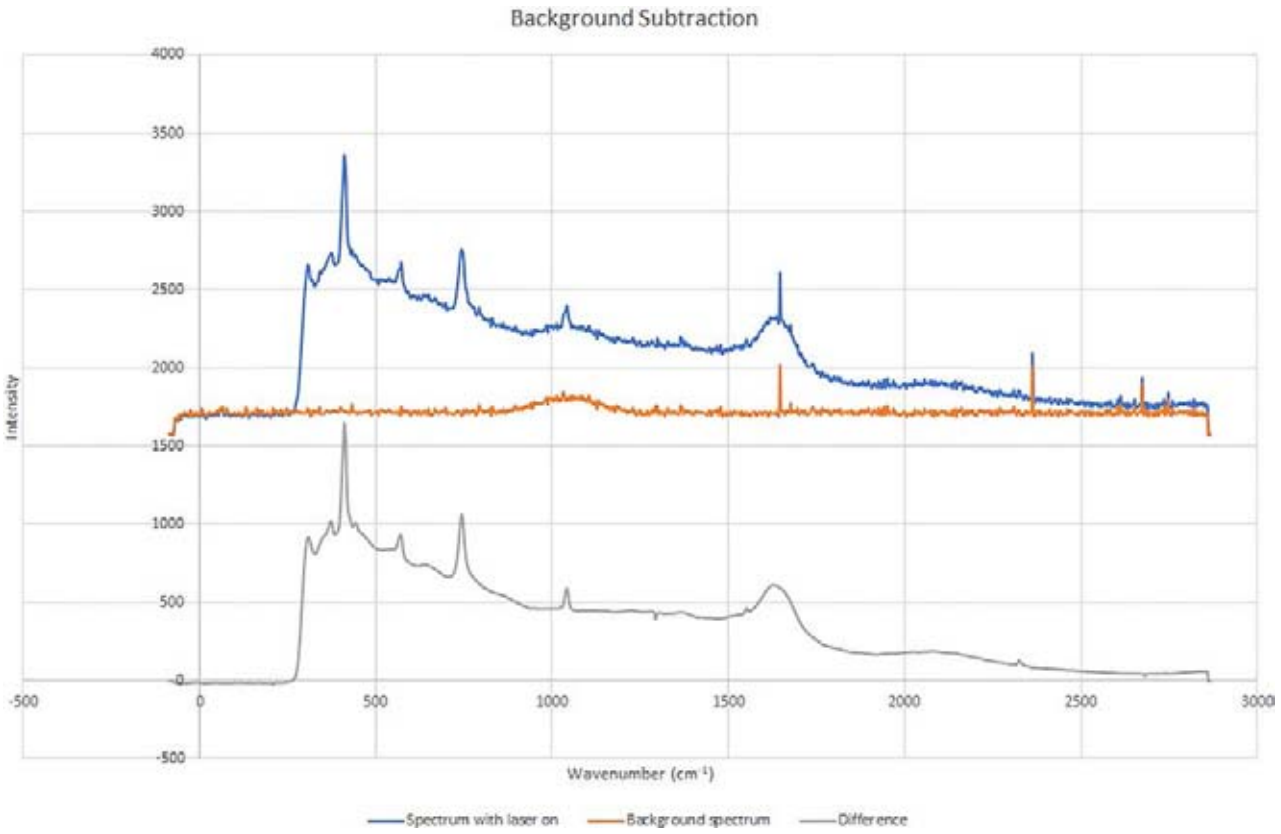

Figure 8.4. Background spectrum subtraction.

light sources, thermal noise, etc., all of which can compromise the usefulness of the data.

8.2.6 Normalising data

Because the range of intensity values can vary significantly between samples, each sample must be normalised. This is the case for both the samples used to train the model and the samples sent from the buoy for testing. The normalisation used was min-max feature scaling. Each normalised data point is calculated as follows:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8.1)$$

where X is the original data point and X_{max} and X_{min} are the highest and lowest intensity values, respectively, which occur below 500 cm^{-1}. The nitrate and *E. coli* datasets used to train the models are shown in Figures 8.5 and 8.6.

8.2.7 Feature scaling

From the Scikit-learn description of the StandardScaler function: “Standardisation of a dataset is a common requirement for many machine learning estimators…” (Pedregosa *et al.,* 2011; Scikit-learn, 2020).

For each feature of a dataset (each wavenumber in this case), a standard score, Z, is calculated:

$$Z = \frac{X - \mu}{\sigma} \quad (8.2)$$

where X is the value of each datapoint for that feature, while μ and σ are the mean and standard deviation of all data points in the dataset for that feature. These standard scores are used as final values of the dataset used to train the AI model. The standard scores for each feature in a signal coming from the buoy must be calculated using the same μ and σ from the training dataset. Scikit-learn’s StandardScaler function simplifies this process.

8.2.8 Analysing with artifical intelligence models

At the training and testing stage, in addition to the normalised and feature-scaled RS data, each sample used to create the dataset must also have an associated label that gives the value of the

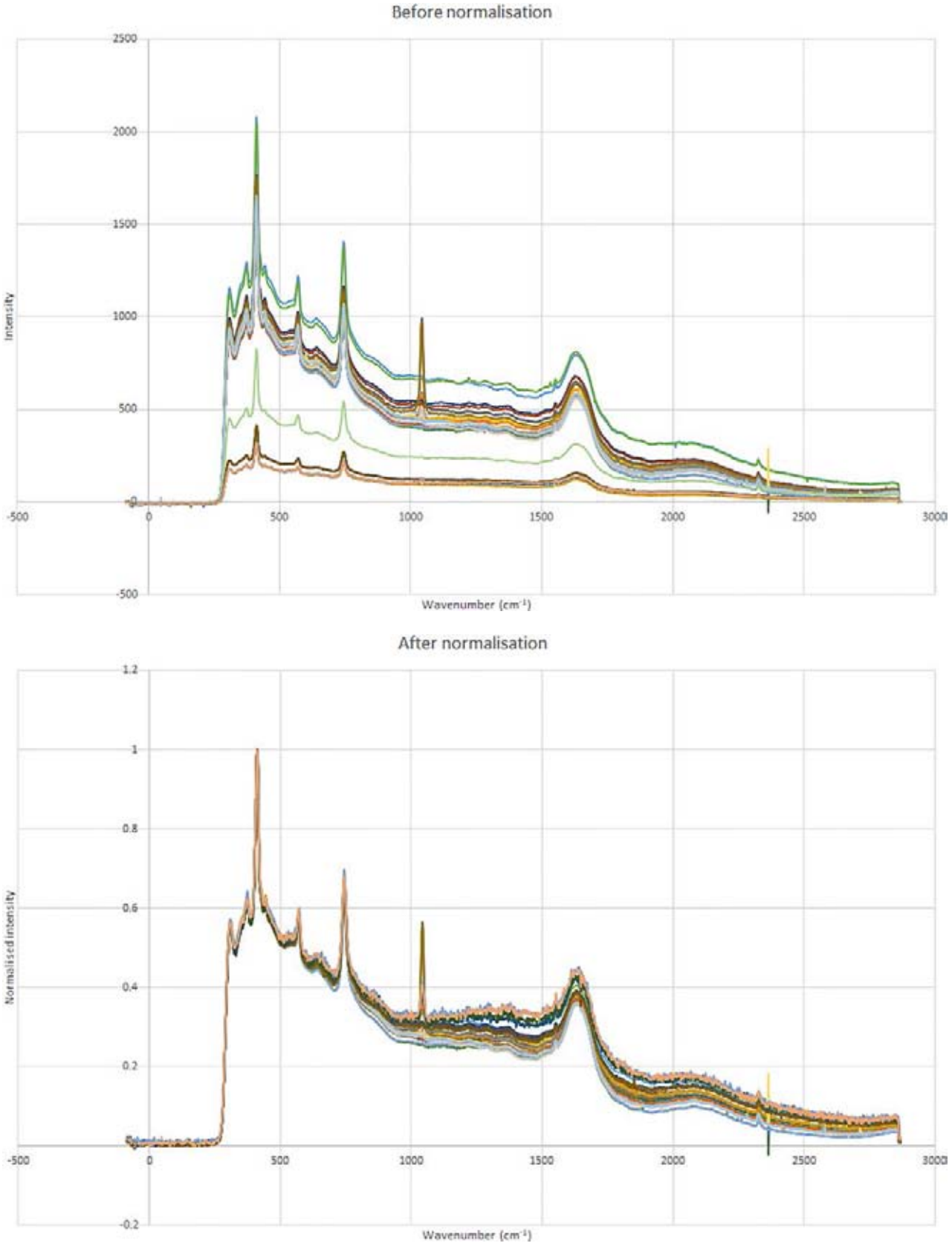

Figure 8.5. Nitrate training dataset before and after the normalisation stage.

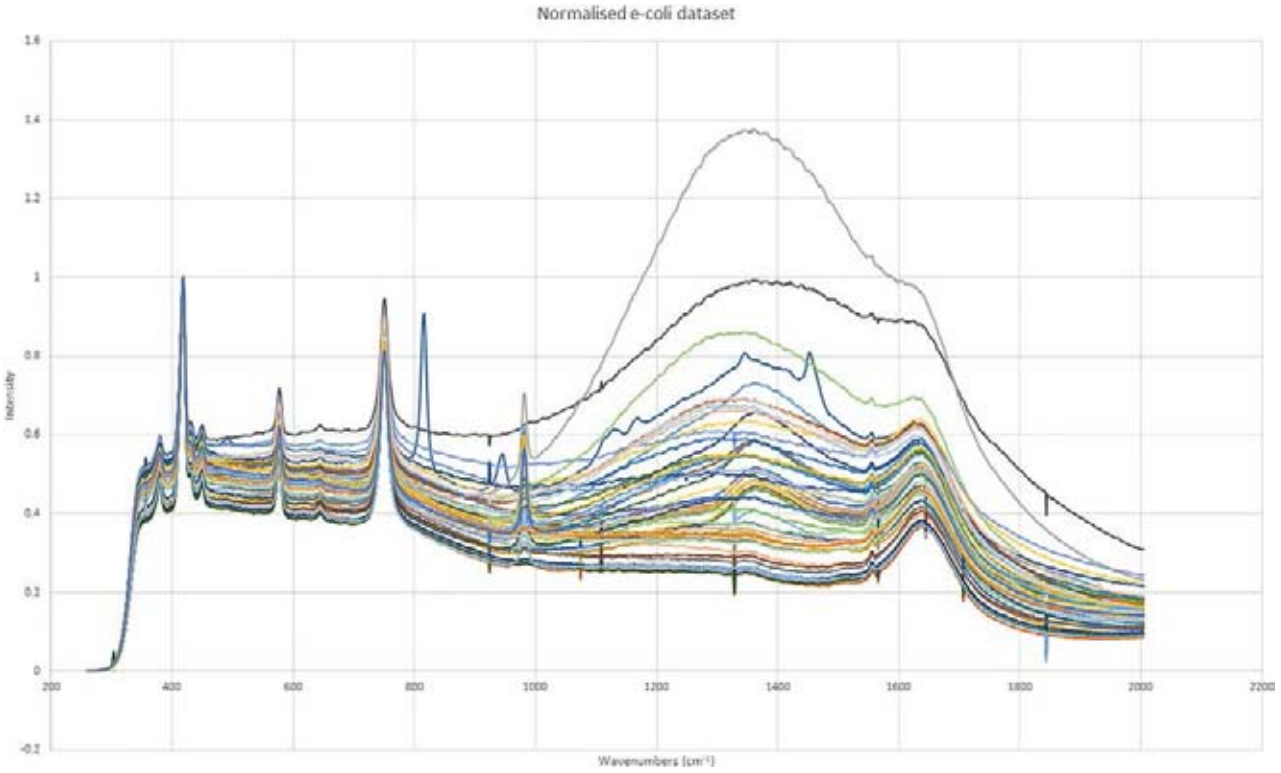

Figure 8.6. *E. coli* training dataset after normalisation stage.

parameter that the new AI model is to detect, e.g. nitrate concentration. With a completed AI model, i.e. one that has been trained and tested successfully, the normalised and feature-scaled RS data are passed through the trained model. The result is a yes/no flag that shows the presence of the target analyte above its detection threshold. A single Raman spectrum can be passed through multiple AI models, each trained for a specific parameter, to check for multiple parameters.

8.2.9 Watermon dashboard and alerts

Finally, the results of the analysis are transmitted to the Watermon Dashboard, a prototype web-based water-monitoring dashboard that uses a mobile-friendly framework. This dashboard provides an interface for the Watermon team to review real-time test results or to send alerts. Incoming data, i.e. Raman spectra, are graphed in real time to allow them to be visualised; and stored in data array format for further processing. Alerts can be sent via email or SMS text, on detection of the target parameters. The total time taken for the overall data flow process varies, mainly depending on the settings for the spectroscopy; longer integration time and higher number of averages lead to a longer time being taken to capture the RS spectra. The trade-off is that higher sensitivity is achieved. Once the spectra have been captured, the other data-processing steps generally all take place within less than 5 minutes.

8.3 Artifical Intelligence Model Development

8.3.1 Overview

The workflow for AI development is illustrated in Figure 8.7; model validation is a key step. The models are developed through the tuning of the algorithms' hyperparameter settings, and model training and testing. Raman spectrum data were uploaded to Rinodrive by the CAPPA team, as one file per spectrum. These Raman spectrum data files were 'wrangled' into a training dataset, randomly split 80% for training and 20% to test model accuracy.

8.3.2 Artificial intelligence model basis – Scikit-learn library

A range of open source AI model libraries are available for use by researchers. In this project, the

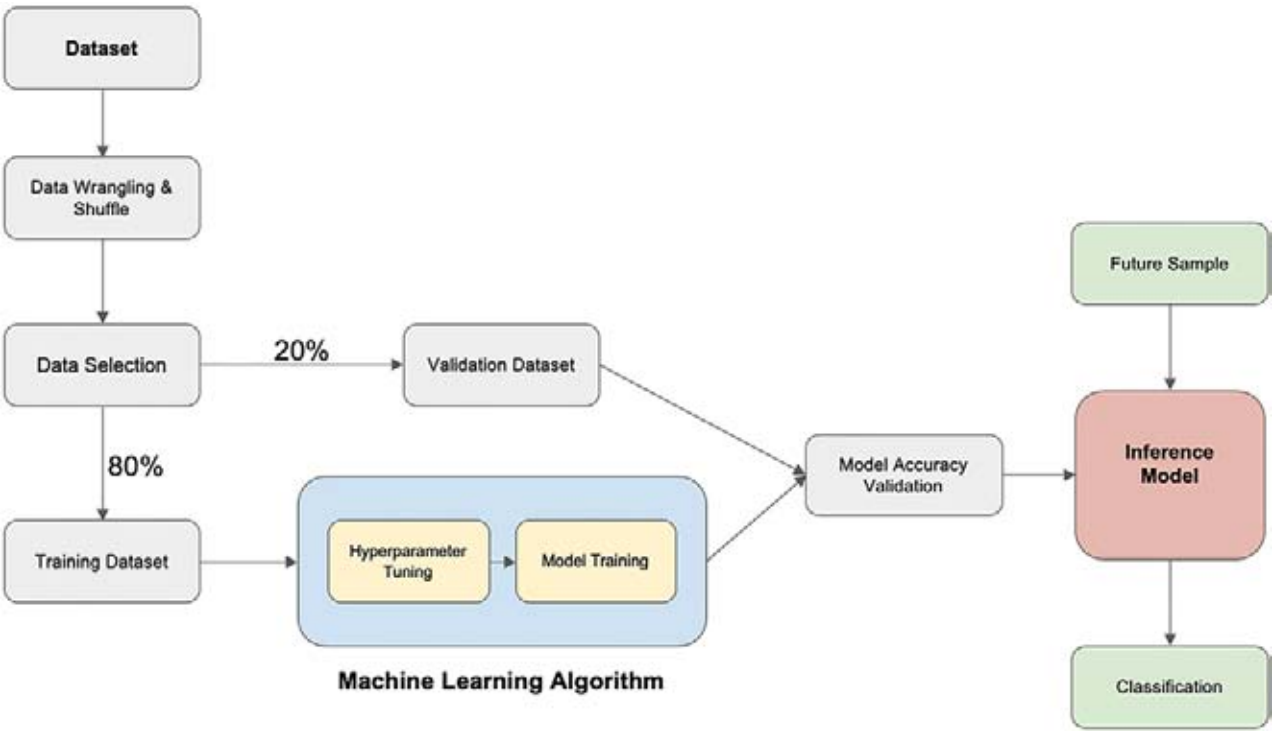

Figure 8.7. AI model training workflow diagram.

basis for the AI models was the Scikit-learn suite, an open source software machine learning library for the Python programming language. This library has been used in many research studies (Abraham *et al*., 2014; Richter and Khoshgoftaar, 2019; Kumar *et al*., 2020). The data structure from RS is suited to two classes of model, namely ensemble models and NNs. Models from both of these classes were applied to the datasets and gave results of similar accuracy; however, NN models were slower to run (and so more expensive in computing time) than ensemble models. Therefore, the project team focused on using ensemble models.

8.3.3 Hyperparameter tuning

For the models used in this project, hyperparameters are passed as arguments to the constructor of the estimator classes. It is possible, and recommended, to search the hyperparameter space for the best cross-validation score. Any parameter provided when constructing an estimator may be optimised in this manner; specifically, to find the names and current values for all parameters for a given estimator. In this case GridSearchCV was the method to identify optimised hyperparameters.

8.4 Final Artificial Intelligence Model – Nitrates

8.4.1 Nitrates Random Forest model

For nitrates, an RF classifier provided the highest accuracy scores. The specific model used from the Scikit-learn library was RandomForestClassifier (Lateef, 2019). RF is a learning method that operates by constructing a large number of decision trees that operate as an ensemble. Each branch of the tree represents a possible decision, occurrence or reaction. Each individual tree generates a class prediction and the class with the most votes becomes the AI model's prediction (see Figure 8.8).

8.4.2 Preprocessing data

To train and test the nitrate model, the water spectra data were shuffled using the shuffle method in Scikit-learn (sklearn.utils.shuffle) and then split using the train–test–split method. The nitrates detection AI model is trained with 36 samples from the identical hardware deployed in the buoy. It became apparent during the project that calibrating/training the model based on the same hardware and system variables – integration time and number of reading averages – that

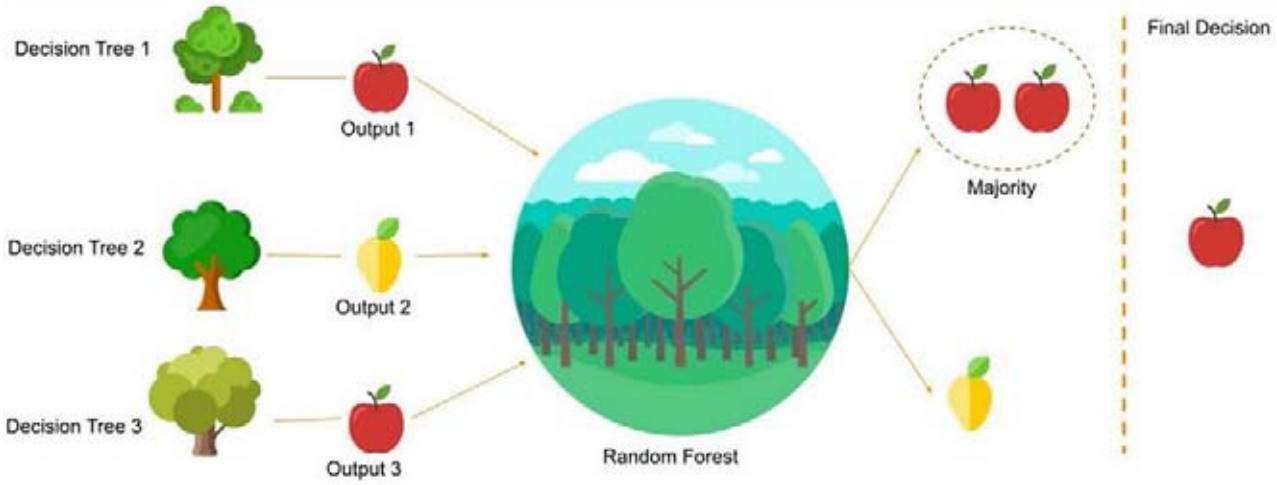

Figure 8.8. Graphical illustration of the Random Forest classifier method.

will be deployed in the field is key to getting good results from the AI models. The training set contains 12 water samples and 24 nitrate samples of varying concentrations. The accuracy score calculated through the Scikit-learn accuracy function is 99.8%.

8.4.3 Setting parameters

The GridSearchCV process was used to establish the optimised hyperparameters:

```
RandomForestClassifier(max_depth = 6,
max_features = 'auto',
n_estimators=n_e,
criterion='gini',
random_state=2,
class_weight='balanced')
```

The n_estimators parameter is important in the RF classifier; it represents the number of "trees in the forest". In this case 'n_estimators' was set to 130 decision trees, this being the number of trees required to arrive at the optimal prediction score (see Figure 8.9).

8.4.4 Feature importance

During training, an AI model typically isolates key features of interest and assigns relative importance values to them. In the case of RS, this translates into a

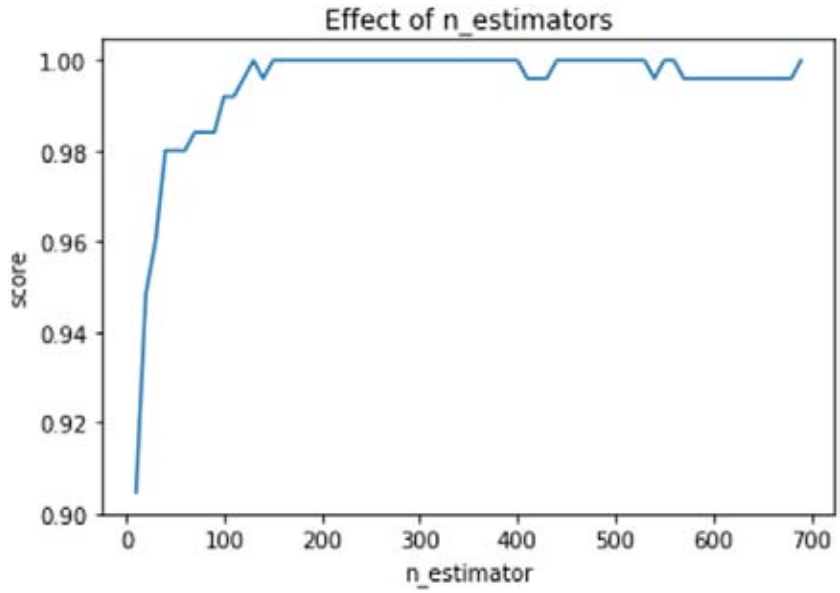

Figure 8.9. Graph of results optimising the n_estimator parameters.

shortlist of key wavenumbers that the model finds are significant in determining the presence or absence of the parameter in question.

8.4.5 Nitrate concentration 100 mg/litre

Figure 8.10 shows a nitrate spectrum from a concentration level of 100 mg/l using the buoy equipment. Overlaid on the graph are dots at the feature points that the model identified; the size of the marker shows how influential the feature is in deciding whether or not nitrate is present. Feature 1 is the primary feature, and features 2 and 7 are also located in the same area of the spectrum. This corresponds well with the known Raman wavenumbers for nitrate, *c*. 1037–1043 cm^{-1}. When we zoom in on this area of the spectrum, the expected spike is readily apparent (see Figure 8.11).

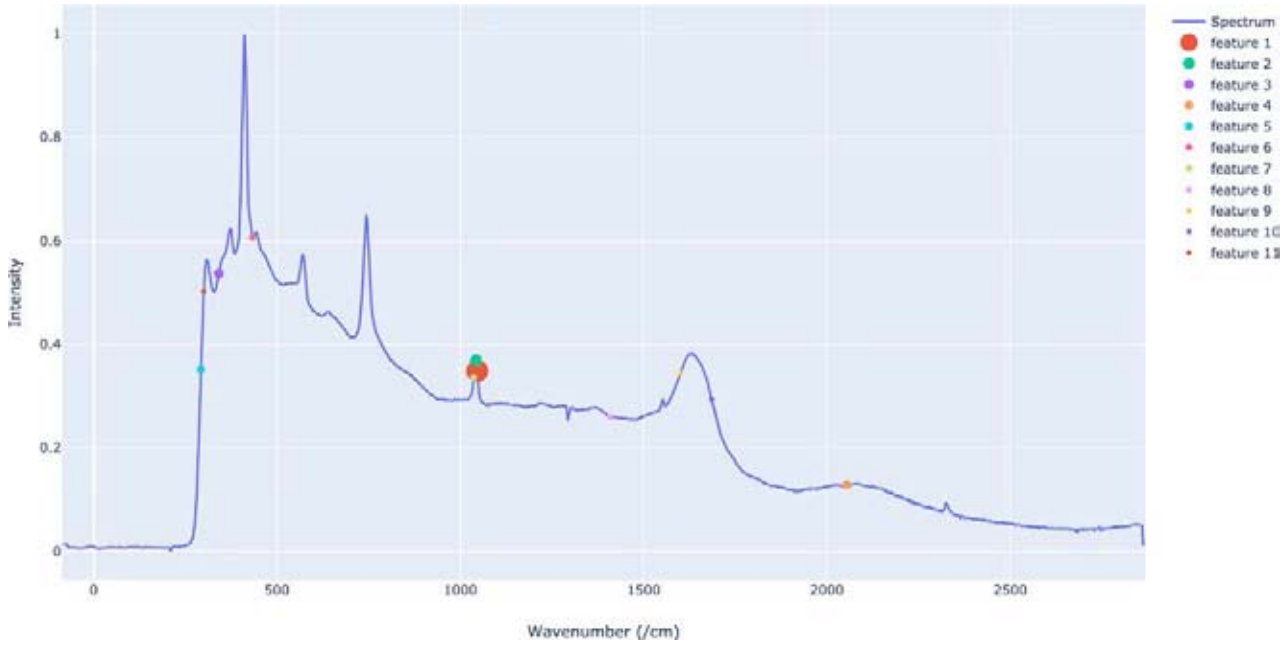

Figure 8.10. RS nitrate spectrum with AI features overlaid (concentration 100 mg/l).

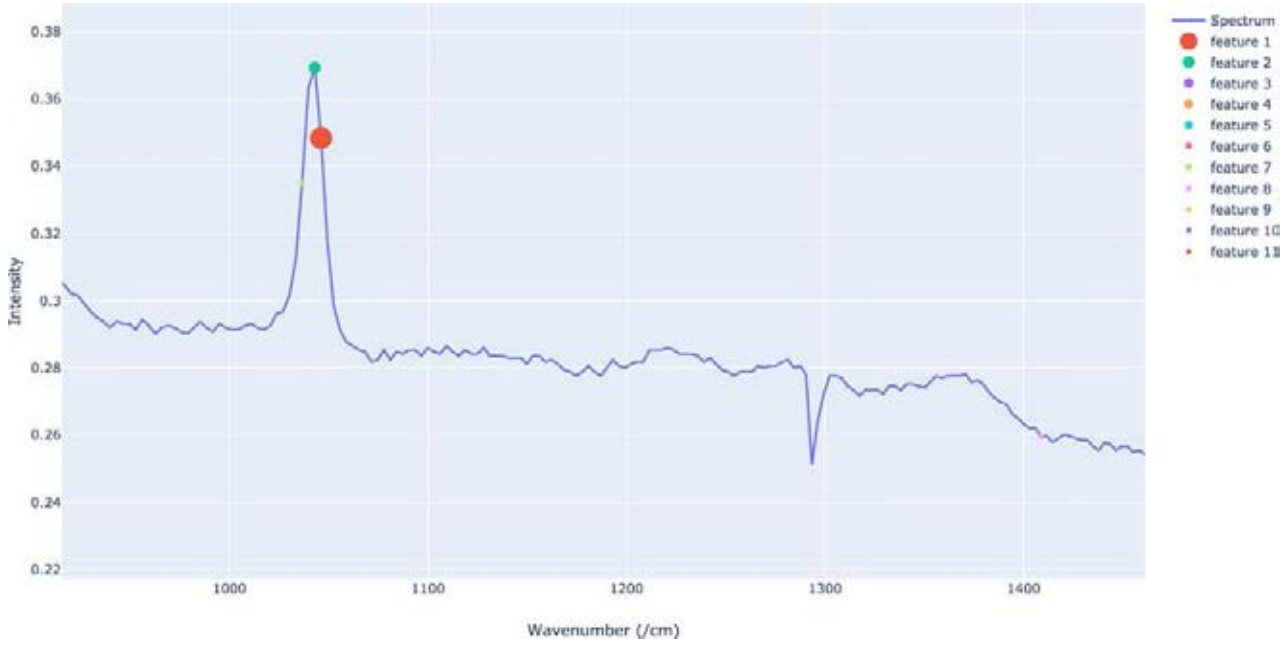

Figure 8.11. Magnified portion of RS nitrate spectrum from Figure 8.10 showing the primary feature, feature 1, of the AI model, *c*. wavenumber 1040 cm^{-1} (concentration 100 mg/l).

8.4.6 Nitrate concentration 80 mg/l

Figure 8.12 shows the RS spectrum from a nitrate concentration of 80 mg/l; the difference from water without nitrates is less distinct on visual inspection – see Figure 8.13. The AI model identifies samples of this concentration as a nitrate. A distinct small peak at wavenumber 1040 cm^{-1} on the spectrum can still be seen, made clearer by zooming in (see Figure 8.14).

8.4.7 Nitrate concentration 30 mg/l

At levels of 30 mg/l and below the nitrates are not identifiable visually, even when the spectrum is examined more closely (see Figures 8.15 and 8.16). However, the generalised function created within the AI model uses the full set of features together to analyse the spectrum, and flags a positive nitrate result.

8.4.8 Validating models

It is common practice when performing a (supervised) machine learning experiment to retain some of the available data (20% in this case) as a test dataset, X_test, Y_test, to evaluate the algorithm's accuracy. Figure 8.17 shows a flow chart of a typical cross-validation workflow in model training.

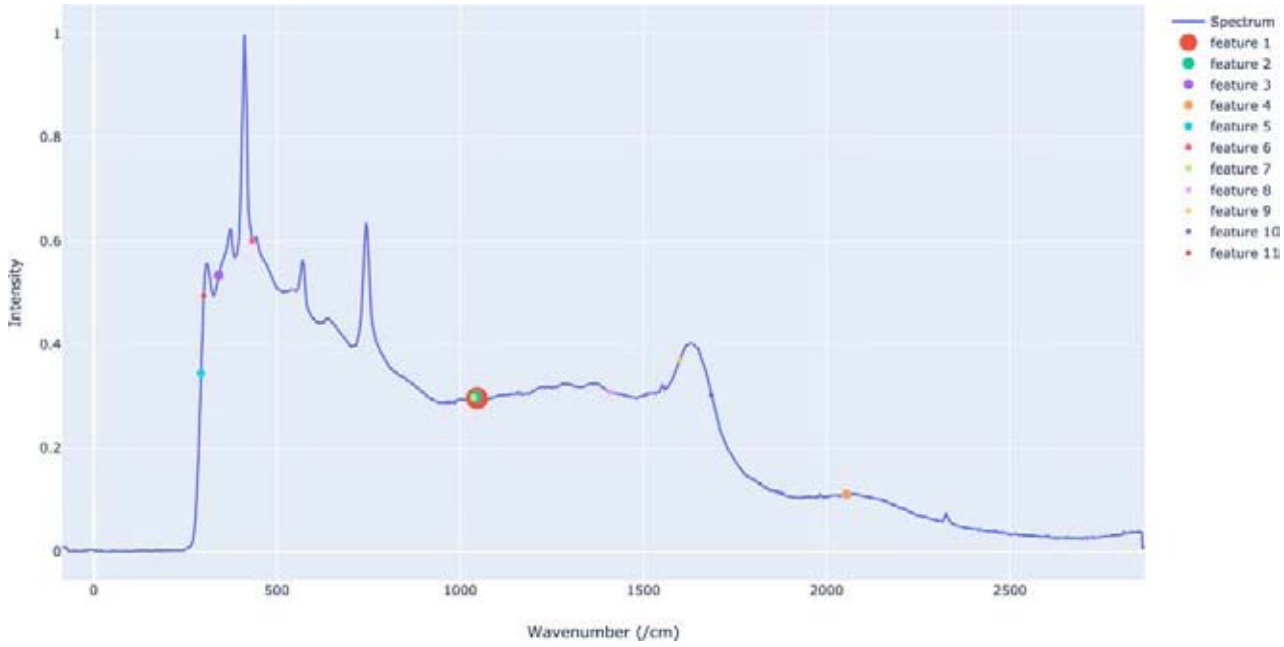

Figure 8.12. Full RS nitrate spectrum (concentration 80 mg/l).

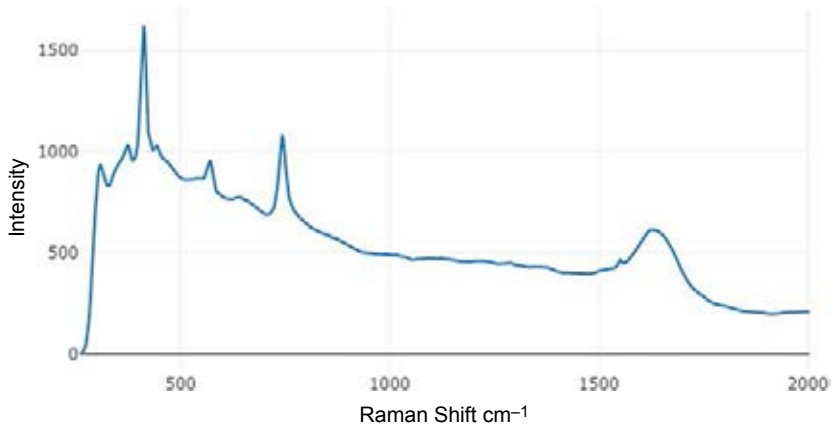

Figure 8.13. Raman profile of water.

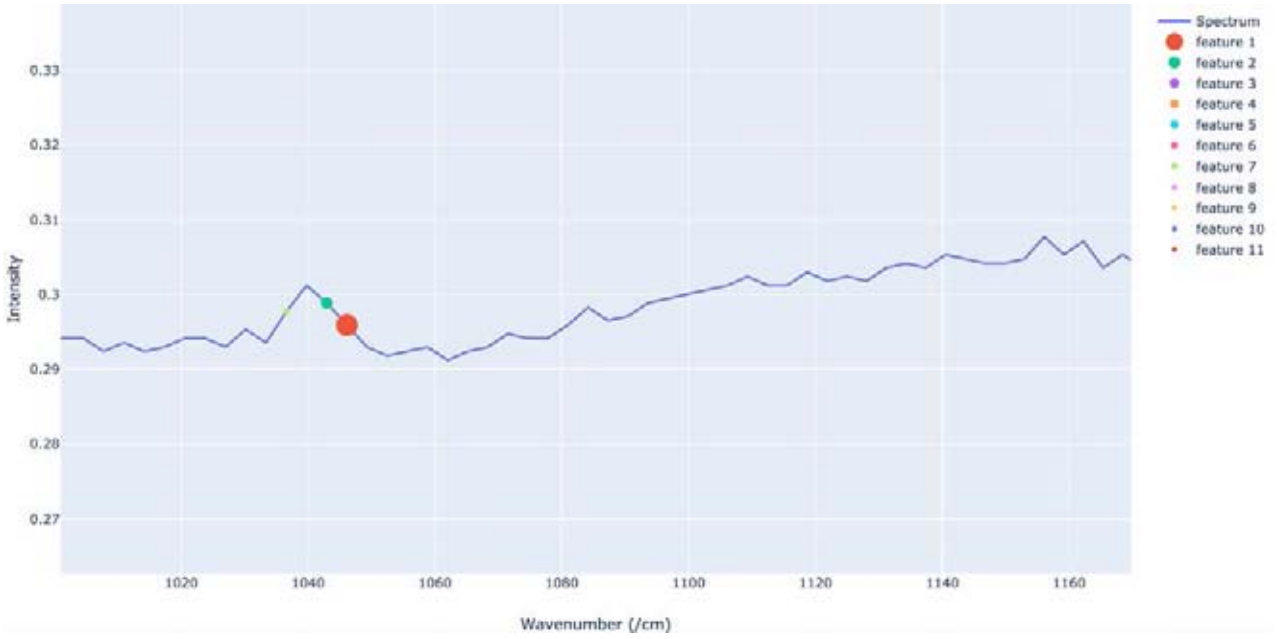

Figure 8.14. Magnified portion of the RS nitrate spectrum from Figure 8.12 zooming in on the wavenumber 1040 cm^{-1} area (concentration 80 mg/l).

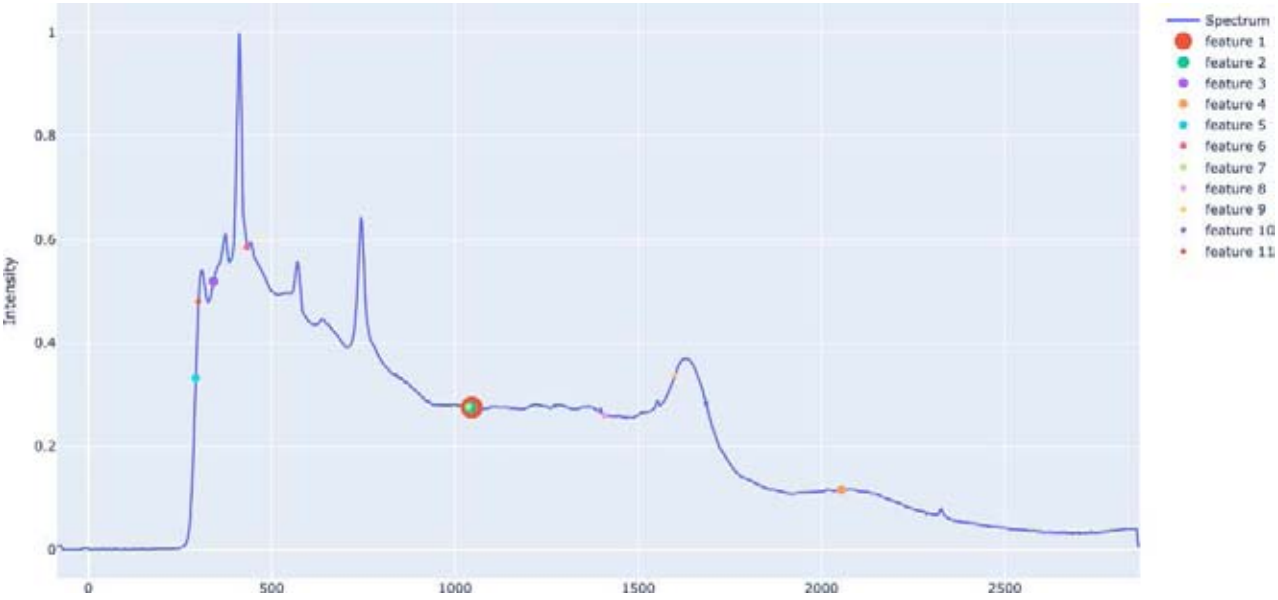

Figure 8.15. Full RS nitrate spectrum (concentration 30 mg/l).

The cross_val_score function was used to calculate the cross-validation scores. This indicated 99.8% accuracy of nitrate detection using the model.

8.4.9 Model deployment

Finally, the trained model was saved (pickled) to file for use on the AI Core web platform, and, therefore, became available for the analysis of incoming real-time data, from either laboratory or field RS readings.

8.5 Final Artificial Intelligence Model – *Escherichia coli* XGBoost Artificial Intelligence Model

E. coli was a more difficult parameter to detect using the equipment and there are health risks associated with testing *E. coli* samples, which imposed limitations on the type of testing that could be carried out. The AI model for *E. coli* is based on a boost classifier that performed best in testing, specifically the

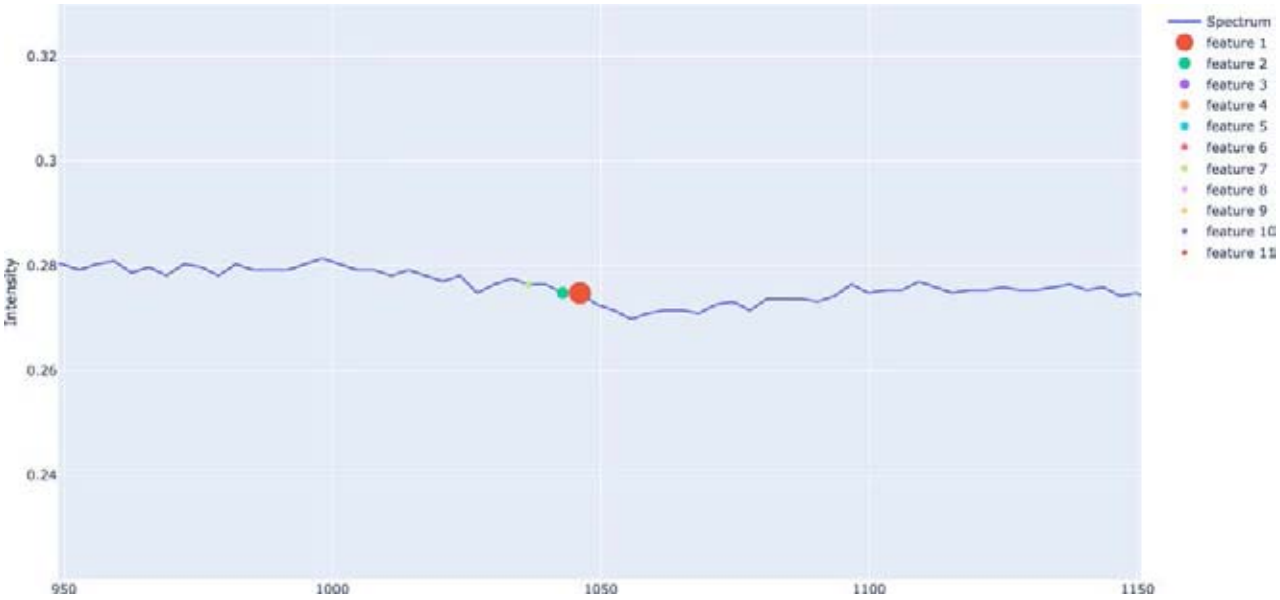

Figure 8.16. Magnified portion of the RS nitrate spectrum from Figure 8.15 zooming in on the wavenumber 1040 cm^{-1} area (concentration 30 mg/l).

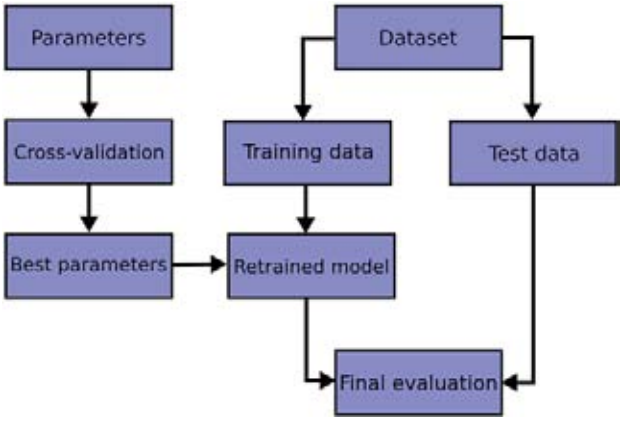

Figure 8.17. Flow chart of typical AI model cross-validation workflow.

XGBClassifier from XGBoost in the Scikit-learn library. In terms of bacteria detection, the project focused on investigating *E. coli*. This aspect proved challenging for practical reasons – safety of laboratory personnel, sample preparation and handling, limitations on storage capacity, etc. – and, therefore, other pathogens such as enterococci could not be investigated.

8.5.1 Boosting models

Boosting is an ensemble learning technique that uses machine learning algorithms to convert weak learners to strong learners, to improve the model. Ensembling is a process that reduces error in AI models. A number of boosting methods, and different ensembling methods, are available. XGBoost is an implementation of a gradient-boosting decision tree algorithm that uses parallel ensembling. It performed best in testing and was selected for this project. Details of XGBoost are available online (Lateef, 2019; Morde, 2019).

8.5.2 XGBoost parameters

For the *E. coli* AI XGBClassifier model, the relevant parameters were set as shown in Figure 8.18.

8.5.3 Model training and feature importance

The model was developed using a training dataset consisting of Raman spectra taken from 61 samples from the different batches supplied by Professor Wim Meijer's team under the EU SWIM Project at UCD as described in section 5.3.1. Biological analyses were carried out at UCD to determine *E. coli* concentrations. The training dataset was made up of 48 samples of the 61. The validation dataset consisted of the remaining 13 samples. The model was assessed for a range of sensitivity levels, to assess its effectiveness at different *E. coli* concentrations. For each training level, e.g. 250 cfu, the RS spectrum of each sample was labelled as positive or negative, based on the cfu count being above or below that level. The model was then trained and tested on the labelled data.

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0,
              learning_rate=0.1, max_delta_step=0, max_depth=3,
              min_child_weight=1, missing=None, n_estimators=100, n_jobs=1,
              nthread=None, objective='binary:logistic', random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
              silent=None, subsample=1, verbosity=1)
```

Figure 8.18. XGBoostClassifier model parameters.

As for the nitrate model, during the training process the AI algorithm for *E. coli* extracted the most significant features of the spectrum for *E. coli* detection. The feature importance graph in Figure 8.19 shows the important Raman spectrum features in classifying samples as identified by the boosting method used.

8.5.4 Model results and deployment

This accuracy score was calculated using the Scikit-learn accuracy_score method. The accuracy score of the XGBClassifer model for *E. coli* was calculated as 83.4%. This was based on the available *E. coli* training dataset as described above, and for a detection limit of 250 cfu. It is expected that the accuracy of the method will improve as further datasets are used to train and validate the model. Finally, the trained model was saved (pickled) for use on the AI Core platform and, therefore, it became available for analysis of incoming real-time data, from either laboratory or field RS readings.

8.6 Calibrating Artificial Intelligence Models for Deployment – Equipment Types and Settings

The above models were trained with data from the specific RS units, as set out in Chapter 6. To ensure

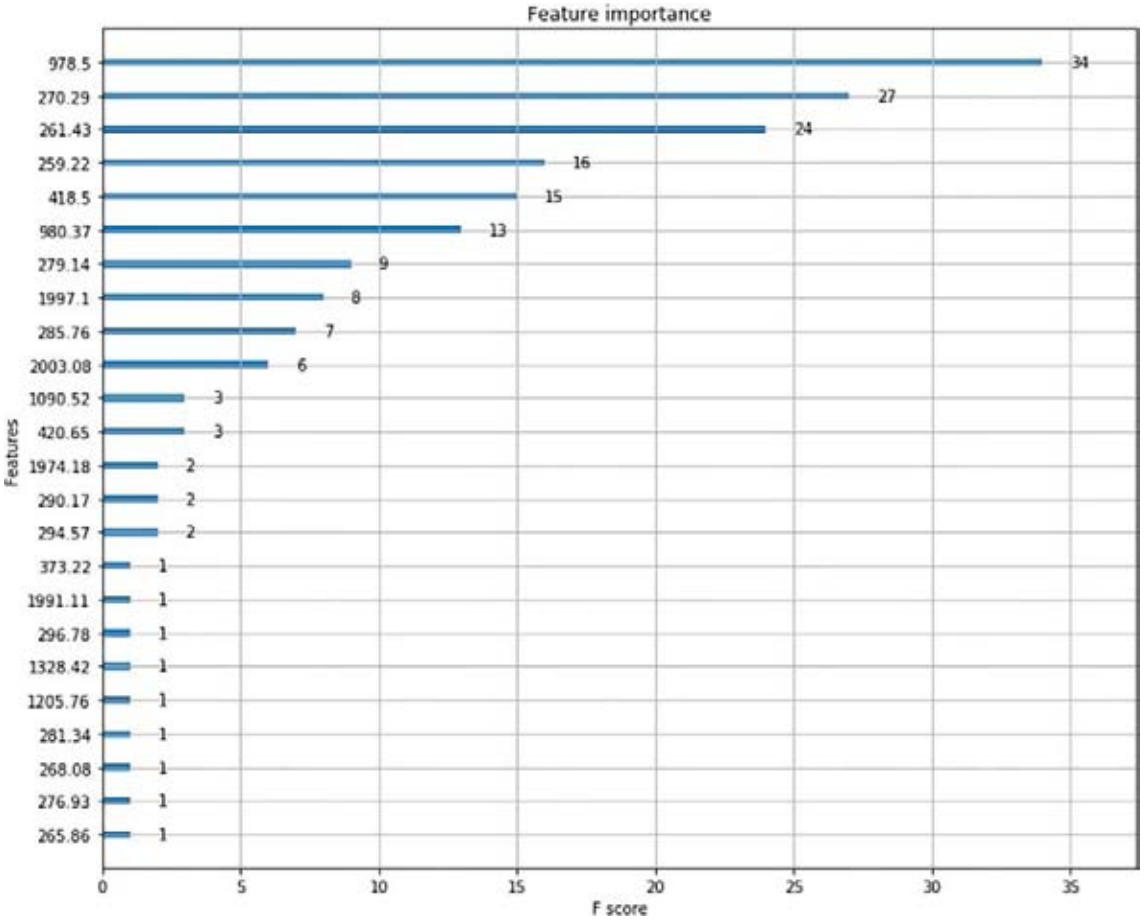

Figure 8.19. *Escherichia coli* model feature importance results.

the accuracy of the validated models, it became clear that the training dataset for model training should be data collected from the same RS equipment with the identical settings, to reduce false positives or negatives caused by the equipment or setting variations. Any changes in equipment types or settings may require a recalibration of the AI models, by training the models with an updated dataset from the new equipment choices.

8.7 Data Storage Platform – Rinolab.com

All data for the project are stored on a dedicated secure data server, in this instance from Rinolab, and not on a mass-market consumer-grade general data platform. This allows science teams to collaborate securely and safely across different geographical locations. Team members are notified when new datasets are added to the data repositories or when any activity takes place on the files. A complete audit trail or activity feed is visible to team members, ensuring data integrity across the project. Water sample data were uploaded to the platform for inclusion in the machine learning models as required.

See Figure 8.20 for a view of the Rinolab secure data server file structure, accessible at https://app.rinolab.com/cit, subject to user authorisation.

8.8 User Interface

The water monitoring dashboard is delivered to users through a modern web application. This dashboard lists samples in chronological order for ease of use. For a more detailed analysis, the user can access raw sample data, or obtain a graphical representation of the Raman signal. The user can also interrogate areas of the signal using a dynamic graph function. The interface was built using a Flash web application framework, which permits tight integration with the AI models as well as integration with the database abstraction layers. Front-end templating was used to allow for consistent presentation on mobile and other devices. Bootstrap styling was used to present well on all browser types. The interface allows module additions, in terms of dashboard types, as required. Figure 8.21 shows the presentation of data analysis on the web interface. The dashboard can be accessed at http://watermon.rinodrive.com/, subject to user authorisation.

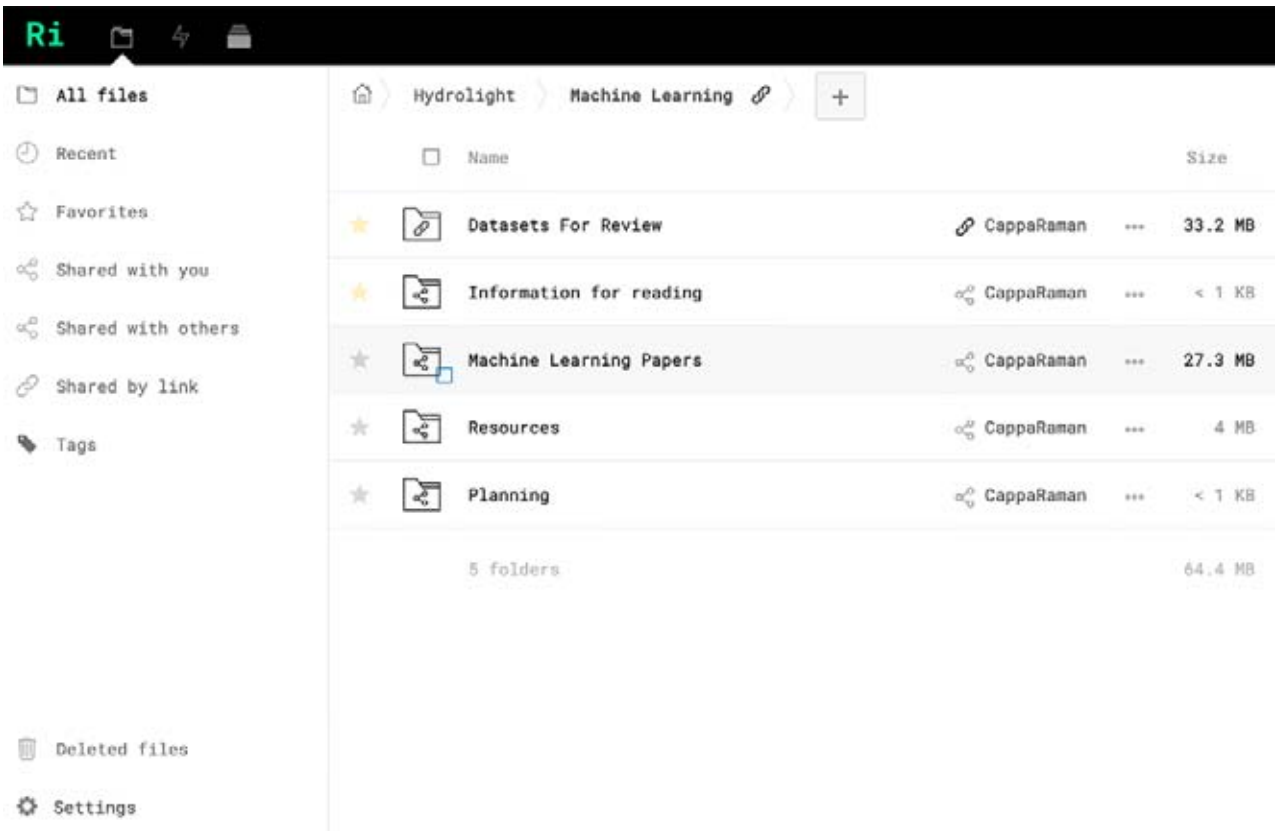

Figure 8.20. Screenshot of the Rinolab secure data server.

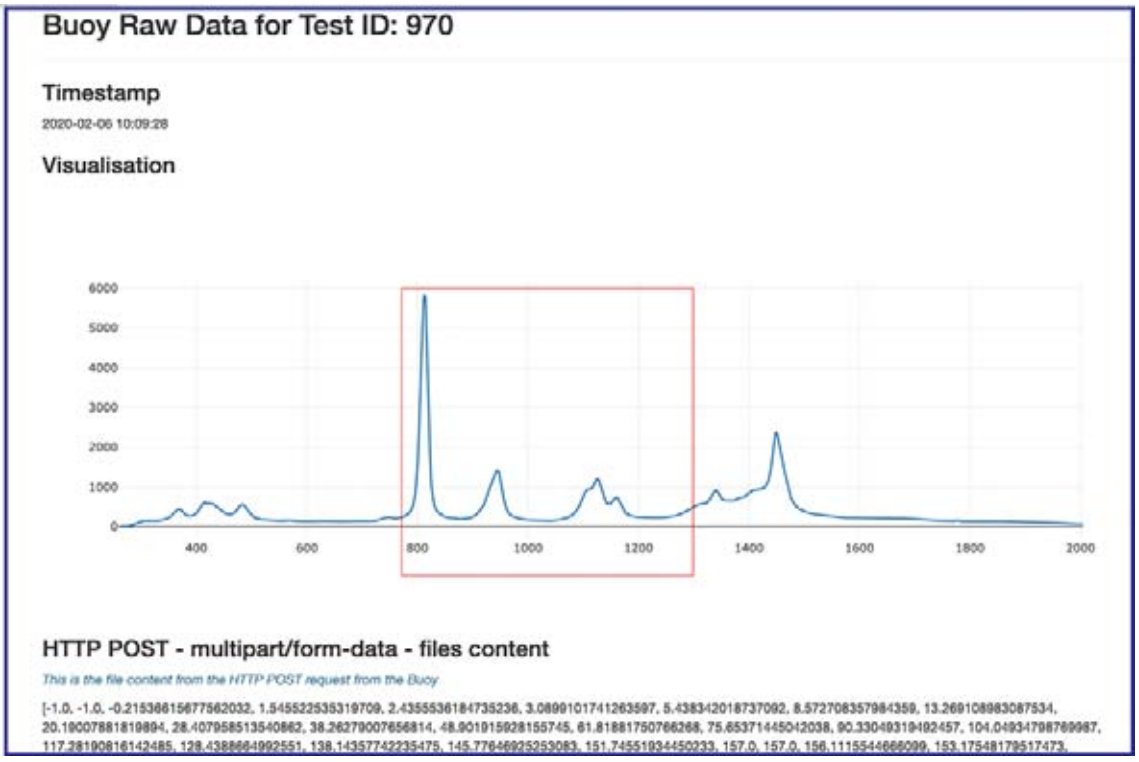

Figure 8.21. Raw data screen on the Watermon web application (the area of interest is highlighted).

8.9 Artificial Intelligence/Machine Learning Software and Platform Summary

The literature indicated that RS, combined with machine learning methods and cross-validation, can be accurate in detecting target analytes in water. On that basis, the project team has successfully developed and tested AI/machine learning tools that use suitably trained AI models to detect nitrates and *E. coli*.

The data management aspect has been implemented by using the proprietary platform, Rinolab. This platform is specifically designed to securely capture and transmit laboratory data in encrypted form and to store them securely, and to maintain complete time-stamped and auditable records of all interactions with the data by authorised users.

A presentation layer/user interface has been created that allows authorised users to access, interrogate, download and visualise the data results from live monitoring of water bodies – all in close to real time, i.e. with time delays measured as low as single-figure numbers of minutes.

As a general principle, larger quantities of training data will always create a more robust AI model, resulting in higher prediction accuracy. For example, arising from this project, *E. coli* detection accuracy would improve by using larger quantities of calibrated *E. coli* training data.

It was not possible to develop the phosphate AI model; further work may enable this aspect to be achieved, subject to the availability of sufficient calibrated data.

9 Discussion of the Results

9.1 Outcomes

The overall aim of the project was to create a fully operational end-to-end water quality monitoring platform targeting nitrates and *E. coli* using RS and AI and operating in close to real time. This objective was achieved, and the fully operational system, complete with 'close-to-real-time' parameter detection, was deployed for demonstration in two locations in County Cork. The system performed as expected: Raman spectrum data were collected, transmitted to the cloud-based database and analysed using the AI models developed during the project, and the results transmitted back for viewing on a mobile phone. During the field trials, no nitrate or *E. coli* contamination was detected in the natural waters tested (unlike the spiked river samples tested in the laboratory, when contamination was detected; see Chapters 4 and 5). Figure 9.1 shows the completed Watermon Buoy version ready for deployment, with trailer.

The Watermon buoy described in Chapter 6 was deployed in Cork Harbour at Crosshaven, County Cork (Figure 9.2), and on the Owenabue River at Ballinhassig, County Cork (Figure 9.3), for live testing. Data were collected using the probe installed in the buoy, which protruded to make contact with the water flowing below the hull of the buoy. No physical samples were taken into the buoy; this will form part of future work, as the LOC version progresses (it was not necessary for the buoy version developed as part of this project). Tests were conducted over a 2-hour and 4-hour period, respectively, and the interval between Raman cycles varied between 10 and 60 minutes. A full testing cycle can take up to 18–20 minutes because of long integration processing times for the spectrometer's imaging sensor; it is expected that this time could be shortened with further experience of the system. Data were harvested from the Raman spectrometer during the test periods and transmitted to the Rinolab secure data storage platform via the

Figure 9.1. Completed Watermon Buoy version, ready for deployment.

Figure 9.2. Live demonstration at Crosshaven, County Cork.

Figure 9.3. Live demonstration at Owenabue River, Ballinhassig, County Cork.

cellular network in the area. The AI models were applied to the data as they arrived at the platform, and results were displayed immediately on the web interface described in Chapter 8. The web interface is viewable on electronic devices that have internet connectivity – PC, tablet or mobile phone. The time taken from the conclusion of the Raman data collection cycle in the field to successful access of the results on a mobile phone while in the field was less than 15 minutes. The primary factor influencing the time taken was the required communication time over the network at these locations.

The purpose of the Watermon platform was to identify the presence of nitrates or *E. coli* in the water body in which it was deployed. With the samples tested, the Buoy version and associated AI model successfully detected nitrates at concentrations as low as 30 mg/l to 99% accuracy. An example of a river water test result is shown in Figure 9.4; the AI analysis result for this sample was 'negative', which indicates that the level of nitrates was below 30 mg/l, i.e. below the detection limit of the system, insofar as it has been shown to date. The system can detect *E. coli* at levels *c*. 250 cfu to 83% accuracy, lower than for nitrates, which reflects the fact that it is more difficult to detect this bacterium than nutrients. *E. coli* was not present at this level at the time of the demonstrations.

The project has also developed a proof-of-concept LOC version of the Watermon system that is capable of detecting nitrates using a probe system that is 100× cheaper than that deployed in the Car-boot and Buoy versions. The LOC version was developed to a lower TRL than the Buoy version, and was demonstrated in the laboratory. The innovative DEP technology for *E. coli* detection, in collaboration with the University

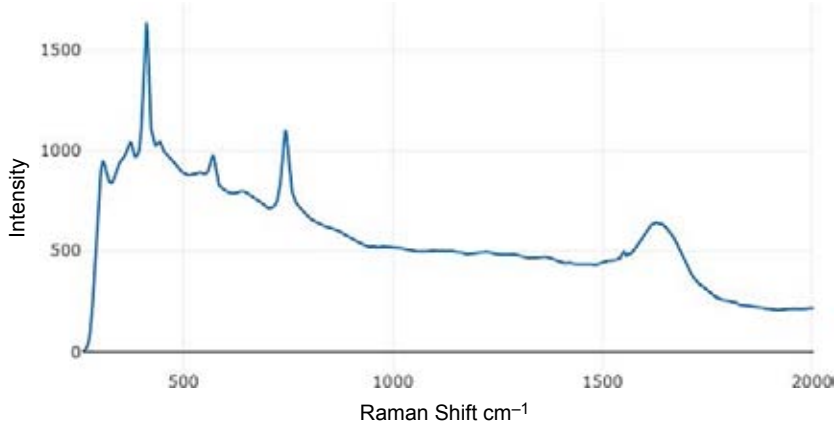

Figure 9.4. Example of Raman spectrum results from Owenabue River (nitrates not detected).

of Jena, Germany, was also demonstrated and will be combined with the LOC version in follow-up work.

This follow-up work would entail further development and testing of the new LOC version, and combining it with the AI model, as well as building a substantial water definition database for various regions to train local models. This would allow the AI models to be trained for region-specific water differences. As this approach to testing is rolled out we would expect future work to involve collating and training these latest models. For example, the current model could be retrained for use in peatland areas (which have a high humic content and extra Raman background) using an estimated 200 laboratory-calibrated samples.

9.1.1 Outcome on the nitrates model

The AI model for nitrate detection has been shown to be very effective at detecting nitrates at 30 mg/l or above in river and drinking water. The test scores indicate > 99% accuracy at this detection limit. The accuracy of the method may be affected by reducing the system variables used, i.e. integration time or number of averages. Accuracy would be expected to increase and/or the detection limit would be expected to reduce when using larger labelled training datasets.

9.1.2 Outcome on the Escherichia coli *model*

In cross-validation, the AI model was able to detect a level of *E. coli* above the threshold of 250 cfu with a high level of accuracy (83.4%). Further development would be required to target levels relevant to drinking water, for example.

At such levels, it was not possible to identify positive *E. coli* samples by deterministic inspection of the spectrum, or by analysis of a single spectrum in isolation. This emphasises the superior ability of the AI model, albeit at relatively higher concentrations, i.e. 250 cfu. The model is able to identify *E. coli* using the set of important features identified in model generation.

The *E. coli* model was trained on data from samples that contained some level of *E. coli*. The team proposes to develop, in the future, a larger training set comprising samples containing *E. coli* and also samples in which no *E. coli* is present. This process is very likely to improve the accuracy and lower the concentration at which such accuracy is obtained. Based on the project team's experience, and as can be inferred from the literature (Beleites *et al*., 2013), it is reasonable to expect a further 5–10% improvement in prediction accuracy with an additional 100–200 samples.

9.2 Learnings

9.2.1 Learnings relative to the project vision

The vision of the project was to make progress towards a low-cost autonomous system using RS and AI, and capable of operating in close to real time, to detect contaminants and pathogens in water. The project has succeeded in creating an end-to-end operational system using this approach for detecting *E. coli* and nitrates and has demonstrated the autonomous probe-based system prototype in the operational environment, representing a TRL of 7. Given that the initial system TRL was 2 (and

the individual components at TRL 3 or 4), this is a remarkable achievement. It is worth noting that this has been achieved despite delays and a significant loss of know-how from the original project team before the commencement of the project.

Considerable effort was devoted to the detailed analysis of various options for the spectroscopy equipment to be used in the Car-boot and Buoy versions. Factors investigated included not only the functional requirements for high-quality, accurate Raman spectra to be collected but also issues related to power management, such as the appropriate trade-offs between power use versus spectrum quality and cooling needs within the systems, and the trade-offs between all of the above and equipment cost.

The analysis clearly showed that, to be a low-cost system, the Raman sensing equipment would have to be significantly miniaturised, ideally down to LOC scale. The fully operational versions that have been created through this project – the Car-boot and the Buoy versions – both contain equipment that is relatively expensive and somewhat bulky. They also draw more power than would ideally be the case – resulting in a need for cooling in the Buoy version, for example. The nature and cost of the present versions are obstacles both to deploying greater numbers and to the long-term unsupervised deployment of the Buoy version. However, a LOC-scale unit, with associated reduced housing, would be cheaper to build and also less conspicuous, thereby reducing the risk of unwanted interference or theft. The project has made significant progress towards such a LOC version, as described in Chapter 7.

9.2.2 Factors affecting the success of this method

A number of factors affect the potential success of the RS/AI method investigated in this project. There are factors relating to the physical equipment used that cannot be adjusted by the operator, such as the power requirements of equipment (e.g. the laser), the sensitivity of the spectrometer, the nature of light filtering, etc. There are also operator-selected variables. RS determines a spectrum by integrating measurements over a time period, which is a variable that can be set. RS uses an averaging method to arrive at final spectrum results, for a given reading cycle, and the number of averages is also a variable that can be adjusted. All of these affect the detail and noisiness of the RS spectrum received, and therefore affect the ability of the model to detect relevant features during training and the ability to detect the target parameter with the completed AI model. The best results were obtained by using a single set-up of the equipment and keeping the variables unchanged, to create the data training set for the nitrates model. In addition, longer integration times and a higher number of averages improved the sensitivity of the method, i.e. lowered the detection threshold for the presence of nitrates.

Several key aspects of the project have been proven fully and can be reapplied to further iterations of the overall sensing platform, such as for the expected LOC version that the project team envisage being developed to TRL 6/7. These include all the elements related to data handling and communication within the platform/buoy, transmission to the cloud, the method of analysing the data via AI algorithms and communicating results that can be viewed via PC or mobile phone, all in close to real time, i.e. under 10 minutes from time of data capture.

Key areas for further work include:

- Further calibration of the AI model for *E. coli* detection, by creating an expanded training dataset.
- Further development of the LOC version, including addressing sample handling and preparation within the context of the autonomous platform.
- Developing additional AI models to detect more parameters of interest, for both bacteria and other non-pathogen parameters. In that regard, the learning from Watermon will significantly help when addressing the challenge of working with and detecting live bacteria.

9.3 Commercial Outcomes

The success of the project in applying AI to RS data has resulted in the industry partner, Hydrolight, and sister companies pursuing further commercial opportunities. In particular, the knowledge and skills developed through the project provided the technical depth to allow the company principals to form Labskin Ltd, which has been acquired by Integumen plc, a company listed on the London Stock Exchange. In particular, the project enabled Hydrolight to broaden and deepen its capacity in applying AI algorithms

to high-volume spectroscope data, for anomaly and feature detection. This enhanced capacity forms the basis for Labskin Ltd's innovative approach to monitoring the microbiome balance levels of incubating skin samples, and automates tasks as a result, cost-effectively accelerating the research, test and development of skin products verified on laboratory-grown human skin without the need for animal testing.

Integumen plc has acquired a significant interest in Modern Water, a US water technology company. Together, these companies have raised over €8 million in investment funding since the start of the project, and are pursuing further research grants valued at *c*. €4 million to the companies, stemming from the technology ideas that were pursued through Watermon (see also section 10.2).

10 Conclusions and Recommendations

10.1 Conclusions

The literature indicated that RS, combined with machine learning methods and cross-validation, can be accurate in detecting target analytes in water. Together with the growing capabilities of cloud-based AI software platforms, using this approach it was posited that an improved technique for real-time water quality monitoring could be achieved. The system was envisaged as using a single set of equipment to capture Raman spectra, followed by application of different AI models to the data, one for each target parameter.

The project has verified this hypothesis by developing and testing Watermon, an end-to-end RS-based detection system that uses suitably trained AI models to detect nitrates and *E. coli*.

The AI model for nitrate detection has been shown to be very effective at detecting nitrates at 30 mg/l or above in river and drinking water. The test scores indicate > 99% accuracy at this detection limit. The accuracy of the method may be affected by reducing the system variables, i.e. integration time or number of averages used. It is expected that accuracy could be increased, and/or the detection limit reduced, by using larger labelled training datasets.

In cross-validation, the AI model was able to detect a level of *E. coli* above the threshold of 250 cfu with a high level of confidence (83.4%). At such levels, it was not possible to identify positive *E. coli* samples by visual inspection of the spectrum, or analysis of a single spectrum in isolation. This emphasises the superior ability of the AI model, albeit at relatively higher concentrations, i.e. 250 cfu. The model is able to identify the *E. coli* using the set of important features identified in model generation.

The success of the Watermon platform is predicated on good-quality calibrated data being used to train the AI model. As a general principle, larger quantities of training data will always create a more robust AI model, resulting in higher prediction accuracy. As further work based on this project is undertaken, it could be expected that, with larger quantities of calibrated *E. coli* training data, detection accuracy levels of the *E. coli* model could improve significantly.

10.1.1 Specific conclusions:

1. AI models can be successfully trained and deployed to detect water quality parameters, in close to real time, when used in conjunction with sensing techniques such as RS, which produce large spectrum datasets.
2. RS has been combined with trained AI models to create a system able to detect nitrates at levels as low as 30 mg/l, at an accuracy (confidence) level of 99%.
3. The system can detect *E. coli* at concentrations *c*. 250 cfu, with 83.4% confidence levels.
4. The full end-to-end Car-boot and Buoy systems operate in close to real time, i.e. less than 5 minutes from the end of collection of Raman spectra to the receipt of results on a mobile device.
5. The system depends on the availability of suitable labelled data to train the AI models. These data would ideally be collected using sensing devices similar to those used in the field equipment. The accuracy of detection depends in part on the quantity and quality of training data available. Additional data can improve the system accuracy, through iterative retraining of the models.
6. The LOC version of the system has successfully detected *E. coli* at high concentrations.
7. Phosphate identification proved inconclusive in testing – the Raman signature of this compound was not as amenable to detection as nitrates using the equipment employed in this project.

10.2 Project Outputs

The project resulted in a number of specific outputs.

10.2.1 Physical detection devices

Four versions of an innovative water quality-sensing platform were developed and demonstrated as part of the project in an iterative manner, i.e. further refinements were incorporated at each successive

stage. The versions and the stages of development achieved are:

1. laboratory-scale version: fully operational in a laboratory environment;
2. portable Car-boot version: to TRL 7;
3. autonomous Buoy version: to TRL 7; and
4. LOC version: to TRL *c*. 3.

The iterative approach allowed new learnings to be incorporated during the process.

10.2.2 Detection artifical intelligence software

AI models were created and trained for the detection of two water quality parameters: nitrates and *E. coli*. These have been deployed on a cloud-based platform that interacts with incoming spectroscopy data and issues results of analysis in real time.

10.2.3 Communication and dissemination

The results of Watermon have been disseminated at two conferences, ENVIRON 2019 and the Seventh European Workshop on Optical Fibre Sensors 2019. Two journal publications are planned, one of which is at an advanced stage. Some further experiments are required, which were delayed by the COVID-19 crisis. We expect to complete these works in the near future.

Dr Chinna Devarapu of the Watermon project team demonstrated aspects of the project at the Irish Photonic Integration Centre's (IPIC) Culture Night at Tyndall Research Institute in 2019, attended by around 250 delegates (Figure 10.1).

Figure 10.1. Dr Chinna Devarapu at the IPIC culture night, 2019.

10.2.4 Labskin/Integumen follow-on projects

As a result of the project, Labskin/Integumen has engaged the CAPPA Research Centre at MTU to conduct a direct-funded feasibility study investigating detecting Covid-19 virus in wastewater using photonics-based detection techniques and AI. The company intends to build on this work in a major further research project involving international partners (Murray, 2020).

A collaboration between Integumen and Modern Water is also being developed following the conclusion of the project; it includes the Nimbus Research Centre at MTU and aims to build an integrated network of river water quality-sensing devices in a US city's environs. Integumen has opened discussions with MTU regarding licensing of the intellectual property arising from the Watermon project for this purpose.

10.3 Recommendations

10.3.1 Technology development – sensing system using Raman spectroscopy and artifical intelligence

- The project has shown the potential for further work to develop the LOC version from its current proof-of-concept stage into a full end-to-end, high-TRL prototype. This technology should be developed further; at the time of writing, the project team continues to pursue this opportunity.
- DEP has shown potential to act as a concentrating technique for *E. coli* detection for RS detection, and should be examined further. This avenue of development will be explored by the project team.
- The system has shown the potential to detect additional parameters. A feasibility study to

determine the most appropriate target analytes, in terms of Raman detection with AI, would identify the most fruitful avenues for investigation in that regard, and should be pursued. Further work is required to develop an AI model for phosphates.

- Further work should be undertaken to identify more precisely the level of sensitivity of the system to integration time and number of averages used, for nitrate and *E. coli* detection in the first instance.
- The detection accuracy of the system as developed to date would be improved by further training of the relevant AI model with additional labelled *E. coli* samples. The SOPs developed allow this to be done efficiently. Equally, the AI models have been developed to such a degree that future training of the models no longer requires expert computer scientists, and could be carried out by most researchers with a moderate level of data-handling knowledge.

10.3.2 Policy

- The project aimed to create an additional innovative sensing system for real-time detection of water quality parameters, to support the existing water quality monitoring policies, programmes and requirements under the Water Framework Directive, the Bathing Water Quality legislation, etc. It has demonstrated the ability to detect two such parameters, nitrates and *E. coli*, using the system. This result provides further support for the concept of using networks of catchment monitoring stations to act as real-time red flag warning systems for pollution events.
- In terms of policy development, the project team considers that the outcomes of the project support a policy of developing such sensing networks in river catchments and bathing areas, using suitable ranges of sensors. To enable such networks, consideration should be given to the most suitable communications infrastructure and protocols to adopt, in terms of cost, data transmission requirements and communications reliability. Creating comprehensive multi-modal communication gateway networks could be considered, to enable such whole-of-catchment monitoring.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, A. and Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics* 8: 14. https://doi.org/10.3389/fninf.2014.00014

Amjad, A., Ullah, R., Khanb, S., Bilal, M. and Khan, A., 2018. Raman spectroscopy based analysis of milk using random forest classification. *Vibrational Spectroscopy* 99: 124–129. https://doi.org/10.1016/j.vibspec.2018.09.003

AQUARIUS, 2020. Broadband tunable QCL based sensor for online and inline detection of contaminants in water. Available online: https://aquarius-project.eu/ (accessed 22 September 2020).

Ashok, P., Singh, G., Rendall, H., Krauss, T. and Dholakiaa, K., 2011. Waveguide confined Raman spectroscopy for microfluidic interrogation. *Lab on a Chip* 11: 1262–1270. https://doi.org/10.1039/C0LC00462F

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C. and Popp, J., 2013. Sample size planning for classification models. *Analytica Chimica Acta* 760: 25–33. https://doi.org/10.1016/j.aca.2012.11.007

Ben Mabrouk, K., Kauffmann, T.H., Arouic, H. and Fontana, M.D., 2013. Raman study of cation effect on sulfate vibration modes in solid state and in aqueous solutions. *Journal of Raman Spectroscopy* 44(11): 1603–1608. https://doi.org/10.1002/jrs.4374

bNovate Technologies, 2020. *BactoSense TCC. Automated Flow Cytometer for Online Monitoring of Microbial Cell Number in Drinking Water*. Available online: https://c5672a6d-5551-4a49-ac62-e780d1d7956c.filesusr.com/ugd/05f6a1_cf47bc23bf6244bda9e63030173e6b75.pdf (accessed 10 February 2021).

Bowes, M.J., Jarvie, H.P., Halliday, S.J., Skeffington, R.A., Wade, A.J., Loewenthal, M., Gozzard, E., Newman, J.R. and Palmer-Felgate, E.J., 2015. Characterising phosphorus and nitrate inputs to a rural river using high-frequency concentration–flow relationships. *Science of the Total Environment* 511: 608–620. https://doi.org/10.1016/j.scitotenv.2014.12.086

Brack, W., Dulio, V., Ågerstrand, M., Allan, I. *et al*., 2017. Towards the review of the European Union Water Framework: recommendations for more efficient assessment and management of chemical contamination in European surface water resources. *Science of the Total Environment* 576: 720–737. https://doi.org/10.1016/j.scitotenv.2016.10.104

Bui, D.T., Khosravi, K., Karimi, M., Busico, G., Khozani, Z.S., Nguyen, H., Mastrocicco, M., Tedesco, D., Cuoco, E. and Kazakis, N., 2020. Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm. *Science of the Total Environment* 715: 136836. https://doi.org/10.1016/j.scitotenv.2020.136836

Casanovas-Massana, A., Gómez-Doñate, M., Sánchez, D., Belanche-Muñoz, L.A., Muniesa, M. and Blanch, B.A., 2015. Predicting fecal sources in waters with diverse pollution loads using general and molecular host-specific indicators and applying machine learning methods. *Journal of Environmental Management* 151: 317–325. https://doi.org/10.1016/j.jenvman.2015.01.002

Chen, C., Mehl, B.T., Munshi, A.S., Townsend, A., Spence, D. and Martin, R.S., 2016. 3D-printed microfluidic devices: fabrication, advantages and limitations – a mini review. *Analytical Methods* 8: 6005–6012. https://doi.org/10.1039/C6AY01671E

Chen, G.Y., Piantedosi, F., Otten, D., Kang, YQ., Kang YQ., Zhang, WQ., Zhou, X., Monro, TM. and Lancaster, DG., 2018. Femtosecond-laser-written microstructured waveguides in BK7 glass. *Scientific Reports* 8: 10377. https://doi.org/10.1038/s41598-018-28631-3

Cheng, I.-F., Chang, H.C., Hou, D. and Chang, H.S., 2007. An integrated dielectrophoretic chip for continuous bioparticle filtering, focusing, sorting, trapping, and detecting. *Biomicrofluidics* 1: 021503. https://doi.org/10.1063/1.2723669

CORDIS, 2017. Integrated and portable image cytometer for rapid response to *Legionella* and *Escherichia coli* in industrial and environmental waters. Available online: https://cordis.europa.eu/project/id/642356 (accessed 23 September 2020).

CORDIS, 2020a. High sensitivity, portable photonic device for pervasive water quality analysis. Available online: https://cordis.europa.eu/project/id/731778 (accessed 22 September 2020).

CORDIS, 2020b. Online and automated *E. coli* monitoring for 100% safe drinking. Available online: https://cordis.europa.eu/project/id/859114 (accessed 23 September 2020).

Cross, P.C., Burnham, J. and Leighton, P.A., 1937. The Raman spectrum and the structure of water. *Journal of the American Chemical Society* 59(6): 1134–1147. https://doi.org/10.1021/ja01285a052

Das, R.S. and Agrawal, Y.K., 2011. Raman spectroscopy: recent advancements, techniques and applications. *Vibrational Spectroscopy* 57: 163–176. https://doi.org/10.1016/j.vibspec.2011.08.003

De Coster, D., Ottevaere, H., Vervaeke, M., Van Erps, J., Callewaert, M., Wuytens, P., Simpson, S.H., Hanna, S., De Malsche, W. and Thienpont, H. 2015. Mass-manufacturable polymer microfluidic device for dual fiber optical trapping. *Optics Express* 23(24): 30991. https://doi.org/10.1364/oe.23.030991

Dhakal, A., Peyskens, F., Subramanian, A.Z., Thomas, N.L. and Baets, R., 2013. Enhancement of light absorption, scattering and emission in high index contrast waveguides. In Ewing, K. and Ferreira, M., (eds), *Advanced Photonics*, OSA Technical Digest (online), Optical Society of America, paper ST2B.5.

Dhakal, A., Wuytens, P., Peyskens, F., Subramanian, A., Thomas, N. and Baets, R., 2014. Silicon-nitride waveguides for on-chip Raman spectroscopy. *Proceedings SPIE 9141 Optical Sensing and Detection III*, 91411C. https://doi.org/10.1117/12.2057509

Dochow, S., Becker, M., and Spittel, R., *et al*., 2013. Raman-on-chip device and detection fibres with fibre Bragg grating for analysis of solutions and particles. *Lab on a Chip* 13(6): 1109–1113. https://doi.org/10.1039/c2lc41169e

Durickovic, I. and Marchetti, M., 2014. Raman spectroscopy as polyvalent alternative for water pollution detection. *IET Science, Measurement and Technology* 8(3): 122–128. https://doi.org/10.1049/iet-smt.2013.0143

Ecosens, 2020. *Ecosens Aquamonitrix, quality water sensor*. Available online: https://www.ecosensaquamonitrix.eu/ (accessed 22 September 2020).

EPA (Environmental Protection Agency), 2006. *Water Framework Directive Monitoring Programme*. EPA, Johnstown Castle, Ireland.

EPA (Environmental Protection Agency), 2013 *Environmental Protection through Research Identifying Pressures – Informing Policy – Developing Solutions Diagram 3*. EPA, Johnstown Castle, Ireland. Available online: https://www.epa.ie/pubs/reports/research/abouteparesearch/EPA_STRIVE_PETR.pdf (accessed 22 February 2021).

EPA (Environmental Protection Agency), 2014. *EPA Research Strategy 2014–2020*. Available online: https://www.epa.ie/pubs/reports/research/eparesearchstrategy2014-2020/EPA%20Research%20Strategy%202014-2020.pdf (accessed 3 December 2020).

EPA (Environmental Protection Agency), 2016. *Ireland's Environment – An Assessment 2016*. EPA, Johnstown Castle, Ireland.

EU (European Union),1998. Directive 98/83/EC of the European Parliament and of the Council of 3 November 1998 on the quality of water intended for human consumption. OJ L 330, 5.12.1998, p. 32–54; amended by Directive (EU) 2015/1787 of 6 October 2015.

EU (European Union), 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. OJ L 327, 22.12.2000, p. 1–73.

EU (European Union), 2006. Directive 2006/7/EC of the European Parliament and of the Council of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC. OJ L 064, 4.3.2006, p. 37–51.

EU (European Union), 2007. Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 on environmental quality standards in the field of water policy, amending and subsequently repealing Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council. OJ L 348, 24.12.2008, p. 84–97.

EUSwim, 2020. EU SwimProject: Beach Water Quality | Bathing Water Monitoring. Available online: https://swimproject.eu/ (accessed 3 December 2020).

Francke, T., López-Tarazón, J.A. and Schröder, B., 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. *Hydrological Processes* 22(25): 4892–4904. https://doi.org/10.1002/hyp.7110

Gajaraj, S., Fan, C., Lin, M. and Hu, Z., 2013. Quantitative detection of nitrate in water and wastewater by surface-enhanced Raman spectroscopy. *Environmental Monitoring and Assessment* 185(7): 5673–5681. https://doi.org/10.1007/s10661-012-2975-4

Giannitsis, A.T., 2011. Microfabrication of biomedical lab-on-chip devices. A review. *Estonian Journal of Engineering* 17: 109–139. https://doi.org/10.3176/eng.2011.2.03

Government of Ireland, 2003. *European Communities (Water Policy) Regulations 2003*. Available online: http://www.irishstatutebook.ie/eli/2003/si/722/made/en/print (accessed 3 December 2020).

Government of Ireland, 2008. *Bathing Water Quality Regulations S.I. No. 79 of 2008*. The Stationery Office, Dublin. Available online: http://www.irishstatutebook.ie/eli/2008/si/79/made/en/pdf (accessed 25 May 2021).

Government of Ireland, 2014. *European Union (Drinking Water) Regulations S.I. No. 122 of 2014*. The Stationery Office, Dublin. Available online: http://www.irishstatutebook.ie/eli/2014/si/122/made/en/pdf (accessed 25 May 2021).

Gowen, A., 2012. Vibrational spectroscopy for analysis of water for human use and in aquatic ecosystems. *Critical Reviews in Environmental Science and Technology* 42: 2546–2573.

Gross, B.C., Erkal, J.L., Lockwood, S.Y., Chen, C. and Spence, D.M., 2014. Evaluation of 3D printing and its potential impact on biotechnology and the chemical sciences. *Analytical Chemistry* 86(7): 3240–3253. https://doi.org/10.1021/ac403397r

Harz, M., Rösch, P. and Popp, J., 2014. *Raman Investigation of Microorganisms on a Single Cell*. Horiba Scientific. Available online: www.horiba.com/scientific (accessed 23 September 2020).

Herbert, E.R., Boon, P., Burgin, A.J., *et al*., 2015. A global perspective on wetland salinization: ecological consequences of a growing threat to freshwater wetlands. *Ecosphere* 6(10): 206. https://doi.org/10.1890/ES14-00534.1

Kandel, S., Heer, J., Plaisant, C., *et al*., 2011. Research directions in data wrangling: visualizations and transformations for usable and credible data. *Information Visualization* 10(4): 271–288. https://doi.org/10.1177/1473871611415994

Kauffmann, T.H. and Fontana, M.D., 2015. *Simultaneous Quantification of Ionic Solutions by Raman Spectrometry and Chemometric Analysis*. Available online: https://hal-centralesupelec.archives-ouvertes.fr/hal-01227871 (accessed 23 September 2020).

Keren, S., Zavaleta, C., Cheng, Z., de la Zerda, A., Gheysens, O. and Gambhir, S.S., 2008. Noninvasive molecular imaging of small living subjects using Raman spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* 105(15): 5844–5849. https://doi.org/10.1073/pnas.0710575105

Khan, S., Ullah, R., Khan, A., Sohail, A., Wahab, N., Bilal, M. and Ahmed, M., 2017a. Random forest-based evaluation of Raman spectroscopy for Dengue fever analysis. *Applied Spectroscopy* 71(9): 2111–2117. https://doi.org/10.1177/0003702817695571

Khan, S., Ullah, R., Javaid, S., Shahzad, S., Ali, H., Bilal, M., Saleem, M. and Ahmed, M., 2017b. Raman spectroscopy combined with principal component analysis for screening nasopharyngeal cancer in human blood sera. *Applied Spectroscopy* 71(11): 2497–2503. https://doi.org/10.1177/0003702817723928

Khan, S., Ullah, R., Ashraf, R., Khan, A., Khan, S. and Ahmad, I., 2019. Optical screening of hepatitis-B infected blood sera using optical technique and neural network classifier. *Photodiagnosis and Photodynamic Therapy* 27: 375–379. https://doi.org/10.1016/j.pdpdt.2019.07.001

Kim, K., Park, S.W. and Yang, S.S., 2010. The optimization of PDMS-PMMA bonding process using silane primer. *Biochip Journal* 4(2): 148–154. https://doi.org/10.1007/s13206-010-4210-0

Kitson, P.J., Rosnes, M.H., Sans, V., Dragonea, V. and Cronin, L., 2012. Configurable 3D-printed millifluidic and microfluidic "lab on a chip" reactionware devices', *Lab on a Chip* 12(18): 3267. https://doi.org/10.1039/C2LC40761B.

Kuhar, N., Kuhar, N., Sil, S., Verma, T. and Umapathy, S., 2018. Challenges in application of Raman spectroscopy to biology and materials. *RSC Advances* 8: 25888–25908. https://doi.org/10.1039/c8ra04491k

Kumar, M., Ellis, C.T., Lu, Q., Zhang, H., Capota, M, Wilke T.L., Ramadge, P.J., Turk-Browne, N. and Norman, K.A. 2020. BrainIAK tutorials: user-friendly learning materials for advanced fMRI analysis. *PLOS Computational Biology* 16(1): e1007549. https://doi.org/10.1371/journal.pcbi.1007549

Lateef, Z., 2019. *A Beginners Guide To Boosting Machine Learning Algorithms.* Available online: https://www.edureka.co/blog/boosting-machine-learning/ (accessed 23 September 2020).

Long, D.A., 2002. *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules*. John Wiley & Sons, Inc., Hoboken, NJ, p. 597.

Lussier, F., Thibault, V., Charron, B., Wallace, G.Q., Masson, J.F., 2020. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *Trends in Analytical Chemistry* 124: 115796. https://doi.org/10.1016/j.trac.2019.115796

McClain, M.A., Culbertson, C.T., Jacobson, S.C. and Ramsey, J.M., 2001. Flow cytometry of *Escherichia coli* on microfluidic devices. *Analytical Chemistry* 73(21): 5334–5338. https://doi.org/10.1021/ac010504v

McDonald, J.C., Metallo, S.J. and Whitesides, G.M., 2001. Fabrication of a configurable, single-use microfluidic device. *Analytical Chemistry* 73(23): 5645–5650. https://doi.org/10.1021/ac010631r

Mei, R. and Liu, W.T., 2019. Quantifying the contribution of microbial immigration in engineered water systems. *Microbiome* 7: 144. https://doi.org/10.1186/s40168-019-0760-0

Morde, V., 2019. *XGBoost Algorithm: Long May She Reign!* Available online: https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d (accessed 23 September 2020).

Murray, F., 2020. *The Mission to Detect Covid in Real Time in Wastewater.* Available online: https://www.labskin.co.uk/the-mission-to-detect-covid-in-real-time-in-wastewater/ (accessed 23 September 2020).

O'Boyle, S., Trodd, W. and Bradley, C., 2019. *Water Quality in Ireland 2013-2018.* Environmental Protection Agency, Johnstown Castle, Ireland. Available online: https://www.epa.ie/pubs/reports/water/waterqua/Water%20Quality%20in%20Ireland%202013-2018%20(web).pdf (accessed 3 December 2020).

OceanView, 2013. *Quick Start Guide: Welcome to OceanView*. Available online: http://photos.labwrench.com/equipmentManuals/14863-5822.pdf (accessed 23 September 2020).

Pedregosa, F., Varoquaux, G., Gramfort, A., *et al*., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12: 2825–2830. Available online https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (accessed 5 February 2021).

Pellerin, B.A., Stauffer, B.A., Young, D.A., Sullivan, D.J., Bricker, S.B., Walbridge, M.R., Clyde Jr., G.A. and Shaw, D.M., 2016. Emerging tools for continuous nutrient monitoring networks: sensors advancing science and water resources protection. *Journal of the American Water Resources Association* 52(4): 993–1008. https://doi.org/10.1111/1752-1688.12386

Persichetti, G. and Bernini, R., 2016. Water monitoring by optofluidic Raman spectroscopy for in situ applications. *Talanta* 155: 145–152. https://doi.org/10.1016/j.talanta.2016.03.102

Pething, R., 2010. Dielectrophoresis: status of the theory, technology, and applications. *Biomicrofluidics* 4(2): 022811. https://doi.org/10.1063/1.3456626

Peyskens, F., Subramanian, A.Z., Dhakal, A., Le Thomas, N. and Baets, R., 2013. Enhancement of Raman scattering efficiency by a metallic nano-antenna on top of a high index contrast waveguide. *2013 Conference on Lasers and Electro-Optics (CLEO 2013)*, 9–24 June, San Jose, CA. Available online: https://doi.org/10.1364/cleo_si.2013.cm2f.5 (accessed 22 February 2021).

Popp, J., Krafft, C. and Mayerhöfer, T., 2011. Modern Raman spectroscopy for biomedical applications. *Optik & Photonik* 6(4): 24–28. https://doi.org/10.1002/opph.201190383

Princeton Instruments, undated. *Raman Spectroscopy Basics*. Available online: https://pdf4pro.com/cdn/raman-spectroscopy-basics-portland-state-university-5048f3.pdf (accessed 10 February 2021).

Qian, C., Huang, H., Chen, L., Li, X., Ge, Z., Chen, T., Yang, Z. and Sun, L., 2014. Dielectrophoresis for bioparticle manipulation. *International Journal of Molecular Sciences* 15(10): 18281–18309. https://doi.org/10.3390/ijms151018281

Raich, J., 2013. *Review of Sensors to Monitor Water Quality*. Publications Office of the European Union, Luxembourg.

Regan, F., Jones, L. and Chapman, J., 2013. *Monitoring of Priority Substances in Waste Water Effluents*. Environmental Protection Agency, Johnstown Castle, Ireland.

Richter, A.N. and Khoshgoftaar, T.M., 2019. Efficient learning from big data for cancer risk modeling: a case study with melanoma. *Computers in Biology and Medicine* 110: 29–39. https://doi.org/10.1016/j.compbiomed.2019.04.039

Sadate, S., Kassu, A., Farley, C. W., Sharma, A., Hardisty, J. and Lifson, M.T.K., 2011. Standoff Raman measurement of nitrates in water. *Proc. SPIE 8156, Remote Sensing and Modeling of Ecosystems for Sustainability VIII*. https://doi.org/10.1117/12.893571

Schilling, K.E., Kim, S.W. and Jones, C.S., 2017. Use of water quality surrogates to estimate total phosphorus concentrations in Iowa rivers. *Journal of Hydrology: Regional Studies* 12: 111–121. https://doi.org/10.1016/j.ejrh.2017.04.006

Schröder, U.C., Kirchhoff, J., Hübner, U., 2017. On-chip spectroscopic assessment of microbial susceptibility to antibiotics within 3.5 hours. *Journal of Biophotonics* 10(11): 1547–1557. https://doi.org/10.1002/jbio.201600316

Scikit-learn, 2020. *sklearn.preprocessing.StandardScaler*. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed 23 September 2020).

Sima, F., Sugioka, K., Vázquez, R.M., Osellame, R., Kelemen, L. and Ormos, P., 2018. Three-dimensional femtosecond laser processing for lab-on-a-chip applications. *Nanophotonics* 7(3): 613–634. https://doi.org/10.1515/nanoph-2017-0097

Smith, E. and Dent, G., 2005. *Modern Raman Spectroscopy - A Practical Approach*. John Wiley & Sons, Chichester, UK. https://doi.org/10.1002/0470011831

Steffy, L.Y. and Shank, M.K., 2018. Considerations for using turbidity as a surrogate for suspended sediment in small, ungaged streams: time-series selection, streamflow estimation, and regional transferability. *River Research and Applications* 34(10): 1304–1314. https://doi.org/10.1002/rra.3373

SW4EU, 2020. *Smart Water For Europe*. Available online: https://sw4eu.com/ (accessed 22 September 2020).

Thompson, R.Q., 2005. *Encyclopedia of Analytical Science*, 2nd edn (Worsfold, P., Tonshend, A., Poole, C.), *Journal of Chemical Education* 82(9): 1313. https://doi.org/10.1021/ed082p1313.2

Ullah, R., Khan, S., Farman, F., Bilal, M., Krafft, C. and Shahzad, S., 2019. Demonstrating the application of Raman spectroscopy together with chemometric technique for screening of asthma disease. *Biomedical Optics Express* 10(2): 600. https://doi.org/10.1364/boe.10.000600

University of Cambridge, 2020. *Raman Scattering – DoITPoMS, TLP Library, Raman Spectroscopy*. Available online: https://www.doitpoms.ac.uk/tlplib/raman/raman_scattering.php (accessed 23 September 2020).

Valkama, P. and Ruth, O., 2017. Impact of calculation method, sampling frequency and Hysteresis on suspended solids and total phosphorus load estimations in cold climate. *Hydrology Research* 48(6): 1594–1610. https://doi.org/10.2166/nh.2017.199

Viviano, G., Salerno, F., Manfredi, E.C., Polesello, S., Valsecchi, S. and Tartari, G., 2014. Surrogate measures for providing high frequency estimates of total phosphorus concentrations in urban watersheds. *Water Research* 64: 265–277. https://doi.org/10.1016/j.watres.2014.07.009

Voulvoulis, N., Arpon, K.D. and Giakoumis, T., 2017. The EU Water Framework Directive: from great expectations to problems with implementation. *Science of the Total Environment* 575: 358–366. https://doi.org/10.1016/j.scitotenv.2016.09.228

Wang, K., Wen, X., Hou, D., Tu, D., Zhu, N., Huang, P., Zhang, G. and Zhang, H., 2018. Application of least-squares support vector machines for quantitative evaluation of known contaminant in water distribution system using online water quality parameters. *Sensors* 18(4): 938. https://doi.org/10.3390/s18040938

Webster, P. and Lehane, M., 2015. *Report on Bathing Water Quality for 2015*. Environmental Protection Agency, Johnstown Castle, Ireland. Available online: http://www.epa.ie/pubs/reports/water/bathing/BW_Report_2015.pdf (accessed 22 September 2020).

WHO (World Heath Organization), 2016. *Preventing Disease Through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks*. Available online: https://www.who.int/quantifying_ehimpacts/publications/preventing-disease/en/ (accessed 1 December 2020).

Xia, Y. and Whitesides, G.M., 1998. Soft lithography. *Angewandte Chemie International Edition* 37: 550–575.

Xu, C. and Jackson, S.A., 2019. Machine learning and complex biological data. *Genome Biology* 20: 76. https://doi.org/10.1186/s13059-019-1689-0

Zeng, F., Zeng, F., Mou, T., Zhang, C., Huang, X., Wang, B., Ma, X. and Guo, J., 2019. Paper-based SERS analysis with smartphones as Raman spectral analyzers. *Analyst* 144(1): 137–142. https://doi.org/10.1039/c8an01901k

Zhang, C. and Ma, Y., 2012. *Ensemble Machine Learning*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-9326-7

Zhang, M., Su, Q., Lu, Y., Zhao, M. and Niu, B., 2017. Application of machine learning approaches for protein–protein interactions prediction. *Medicinal Chemistry* 13(6): 506–514. https://doi.org/10.2174/1573406413666170522150940

Zhang, P.X., Zhou, X.F., Cheng, A.Y.S. and Fang, Y., 2006. Raman spectra from pesticides on the surface of fruits. *Journal of Physics: Conference Series* 28(1): 7–11. https://doi.org/10.1088/1742-6596/28/1/002

Zhen, L., Wang, J. and Li, D., 2016. Applications of Raman spectroscopy in detection of water quality. *Applied Spectroscopy Reviews* 51(4): 313–337. https://doi.org/10.1080/05704928.2015.1131711

Zhiyun L., Deen, M.J, Kumar, S. and Selvaganapathy, P.R., 2014. Raman spectroscopy for in-line water quality monitoring — instrumentation and potential. *Sensors* 14: 17275-17303. https://doi.org/10.3390/s140917275

Abbreviations

3-D	Three-dimensional
AC	Alternating current
AgNP	Silver nano-particles
AI	Artificial intelligence
CAD	Computer-aided design
CAPPA	Centre for Advanced Photonics and Process Analysis
cfu	Colony-forming unit
CIT	Cork Institute of Technology
CNC	Computer numerical control
DC	Direct current
DEP	Dielectrophoresis
EC	European Commission
EPA	Environmental Protection Agency
ERI	Environmental Research Institute laboratory
EU	European Union
GSM	Global System for mobile communications
IPA	Isopropyl alcohol
LED	Light-emitting diode
LOC	Lab-on-Chip
MCU	Main control unit
MIR	Mid-infrared
MTU	Munster Technological University
NN	Neural network
OEM	Original equipment manufacturer
PCB	Printed circuit board
PDMS	Polydimethylsiloxane
PHS	Priority hazardous substance
PMMA	Polymethyl methacrylate
QCL	Quantum cascade laser
RF	Random Forest
RS	Raman spectroscopy
RT	Random Tree
SERS	Surface-enhanced Raman spectroscopy
SD card	Secure digital card
SMEs	Small and medium-sized enterprises
SNR	Signal-to-noise ratio
SOP	Standard operating procedure
SU8	A commonly used epoxy-based negative photoresist
SWIM	System for bathing water quality monitoring EU Interreg project
TRL	Technology readiness level
UCD	University College Dublin
WFD	Water Framework Directive
XGB	eXtreme Gradient Boosting

AN GHNÍOMHAIREACHT UM CHAOMHNÚ COMHSHAOIL
Tá an Ghníomhaireacht um Chaomhnú Comhshaoil (GCC) freagrach as an gcomhshaol a chaomhnú agus a fheabhsú mar shócmhainn luachmhar do mhuintir na hÉireann. Táimid tiomanta do dhaoine agus don chomhshaol a chosaint ó éifeachtaí díobhálacha na radaíochta agus an truaillithe.

Is féidir obair na Gníomhaireachta a roinnt ina trí phríomhréimse:

Rialú: *Déanaimid córais éifeachtacha rialaithe agus comhlíonta comhshaoil a chur i bhfeidhm chun torthaí maithe comhshaoil a sholáthar agus chun díriú orthu siúd nach gcloíonn leis na córais sin.*

Eolas: *Soláthraímid sonraí, faisnéis agus measúnú comhshaoil atá ar ardchaighdeán, spriocdhírithe agus tráthúil chun bonn eolais a chur faoin gcinnteoireacht ar gach leibhéal.*

Tacaíocht: *Bímid ag saothrú i gcomhar le grúpaí eile chun tacú le comhshaol atá glan, táirgiúil agus cosanta go maith, agus le hiompar a chuirfidh le comhshaol inbhuanaithe.*

Ár bhFreagrachtaí

Ceadúnú
Déanaimid na gníomhaíochtaí seo a leanas a rialú ionas nach ndéanann siad dochar do shláinte an phobail ná don chomhshaol:
- saoráidí dramhaíola *(m.sh. láithreáin líonta talún, loisceoirí, stáisiúin aistrithe dramhaíola)*;
- gníomhaíochtaí tionsclaíocha ar scála mór *(m.sh. déantúsaíocht cógaisíochta, déantúsaíocht stroighne, stáisiúin chumhachta)*;
- an diantalmhaíocht *(m.sh. muca, éanlaith)*;
- úsáid shrianta agus scaoileadh rialaithe Orgánach Géinmhodhnaithe *(OGM)*;
- foinsí radaíochta ianúcháin *(m.sh. trealamh x-gha agus radaiteiripe, foinsí tionsclaíocha)*;
- áiseanna móra stórála peitril;
- scardadh dramhuisce;
- gníomhaíochtaí dumpála ar farraige.

Forfheidhmiú Náisiúnta i leith Cúrsaí Comhshaoil
- Clár náisiúnta iniúchtaí agus cigireachtaí a dhéanamh gach bliain ar shaoráidí a bhfuil ceadúnas ón nGníomhaireacht acu.
- Maoirseacht a dhéanamh ar fhreagrachtaí cosanta comhshaoil na n-údarás áitiúil.
- Caighdeán an uisce óil, arna sholáthar ag soláthraithe uisce phoiblí, a mhaoirsiú.
- Obair le húdaráis áitiúla agus le gníomhaireachtaí eile chun dul i ngleic le coireanna comhshaoil trí chomhordú a dhéanamh ar líonra forfheidhmiúcháin náisiúnta, trí dhíriú ar chiontóirí, agus trí mhaoirsiú a dhéanamh ar leasúchán.
- Cur i bhfeidhm rialachán ar nós na Rialachán um Dhramhthrealamh Leictreach agus Leictreonach (DTLL), um Shrian ar Shubstaintí Guaiseacha agus na Rialachán um rialú ar shubstaintí a ídíonn an ciseal ózóin.
- An dlí a chur orthu siúd a bhriseann dlí an chomhshaoil agus a dhéanann dochar don chomhshaol.

Bainistíocht Uisce
- Monatóireacht agus tuairisciú a dhéanamh ar cháilíocht aibhneacha, lochanna, uiscí idirchriosacha agus cósta na hÉireann, agus screamhuiscí; leibhéil uisce agus sruthanna aibhneacha a thomhas.
- Comhordú náisiúnta agus maoirsiú a dhéanamh ar an gCreat-Treoir Uisce.
- Monatóireacht agus tuairisciú a dhéanamh ar Cháilíocht an Uisce Snámha.

Monatóireacht, Anailís agus Tuairisciú ar an gComhshaol
- Monatóireacht a dhéanamh ar cháilíocht an aeir agus Treoir an AE maidir le hAer Glan don Eoraip (CAFÉ) a chur chun feidhme.
- Tuairisciú neamhspleách le cabhrú le cinnteoireacht an rialtais náisiúnta agus na n-údarás áitiúil *(m.sh. tuairisciú tréimhsiúil ar staid Chomhshaol na hÉireann agus Tuarascálacha ar Tháscairí)*.

Rialú Astaíochtaí na nGás Ceaptha Teasa in Éirinn
- Fardail agus réamh-mheastacháin na hÉireann maidir le gáis cheaptha teasa a ullmhú.
- An Treoir maidir le Trádáil Astaíochtaí a chur chun feidhme i gcomhair breis agus 100 de na táirgeoirí dé-ocsaíde carbóin is mó in Éirinn.

Taighde agus Forbairt Comhshaoil
- Taighde comhshaoil a chistiú chun brúnna a shainaithint, bonn eolais a chur faoi bheartais, agus réitigh a sholáthar i réimsí na haeráide, an uisce agus na hinbhuanaitheachta.

Measúnacht Straitéiseach Timpeallachta
- Measúnacht a dhéanamh ar thionchar pleananna agus clár beartaithe ar an gcomhshaol in Éirinn *(m.sh. mórphleananna forbartha)*.

Cosaint Raideolaíoch
- Monatóireacht a dhéanamh ar leibhéil radaíochta, measúnacht a dhéanamh ar nochtadh mhuintir na hÉireann don radaíocht ianúcháin.
- Cabhrú le pleananna náisiúnta a fhorbairt le haghaidh éigeandálaí ag eascairt as taismí núicléacha.
- Monatóireacht a dhéanamh ar fhorbairtí thar lear a bhaineann le saoráidí núicléacha agus leis an tsábháilteacht raideolaíochta.
- Sainseirbhísí cosanta ar an radaíocht a sholáthar, nó maoirsiú a dhéanamh ar sholáthar na seirbhísí sin.

Treoir, Faisnéis Inrochtana agus Oideachas
- Comhairle agus treoir a chur ar fáil d'earnáil na tionsclaíochta agus don phobal maidir le hábhair a bhaineann le caomhnú an chomhshaoil agus leis an gcosaint raideolaíoch.
- Faisnéis thráthúil ar an gcomhshaol ar a bhfuil fáil éasca a chur ar fáil chun rannpháirtíocht an phobail a spreagadh sa chinnteoireacht i ndáil leis an gcomhshaol *(m.sh. Timpeall an Tí, léarscáileanna radóin)*.
- Comhairle a chur ar fáil don Rialtas maidir le hábhair a bhaineann leis an tsábháilteacht raideolaíoch agus le cúrsaí práinnfhreagartha.
- Plean Náisiúnta Bainistíochta Dramhaíola Guaisí a fhorbairt chun dramhaíl ghuaiseach a chosc agus a bhainistiú.

Múscailt Feasachta agus Athrú Iompraíochta
- Feasacht chomhshaoil níos fearr a ghiniúint agus dul i bhfeidhm ar athrú iompraíochta dearfach trí thacú le gnóthais, le pobail agus le teaghlaigh a bheith níos éifeachtúla ar acmhainní.
- Tástáil le haghaidh radóin a chur chun cinn i dtithe agus in ionaid oibre, agus gníomhartha leasúcháin a spreagadh nuair is gá.

Bainistíocht agus struchtúr na Gníomhaireachta um Chaomhnú Comhshaoil
Tá an ghníomhaíocht á bainistiú ag Bord lánaimseartha, ar a bhfuil Ard-Stiúrthóir agus cúigear Stiúrthóirí. Déantar an obair ar fud cúig cinn d'Oifigí:
- An Oifig um Inmharthanacht Comhshaoil
- An Oifig Forfheidhmithe i leith cúrsaí Comhshaoil
- An Oifig um Fianaise is Measúnú
- Oifig um Chosaint Radaíochta agus Monatóireachta Comhshaoil
- An Oifig Cumarsáide agus Seirbhísí Corparáideacha

Tá Coiste Comhairleach ag an nGníomhaireacht le cabhrú léi. Tá dáréag comhaltaí air agus tagann siad le chéile go rialta le plé a dhéanamh ar ábhair imní agus le comhairle a chur ar an mBord.

EPA Research Report 380

Innovative Water Monitoring

Authors: Kevin Fitzgibbon, William Whelan-Curtin, Chinna Devarapu, Patricia Loren, Colin O'Sullivan and Ian Aherne

Identifying Pressures

The European Union (EU) Water Framework Directive, transposed into national legislation, has led to the implementation of water quality monitoring programmes in Ireland and across the EU. In Ireland, high nutrient concentrations in watercourses and periodic pathogen contamination of bathing waters continue to be causes of concern. Apart from knowing where and when to take samples for analysis, one of the biggest challenges for authorities is the time required and cost of investigating whether or not the water quality is safe for drinking, bathing or other uses. The methods usually involve field collection and transportation of samples to a laboratory, and in the case of some tests, such as for pathogens, sample preparation by culturing, followed finally by analysis. These testing challenges apply for nutrients, pathogens and other hazardous compounds. Together, these factors act as drivers in the effort to develop low-cost, low-maintenance, reusable water quality monitoring sensors that can act continuously in real time, detecting the presence and concentration of relevant parameters in a cost-effective manner.

Informing Policy

The project aimed to create an additional innovative sensing system for real-time detection of water quality parameters, to support the existing water quality monitoring policies, programmes and requirements under the Water Framework Directive, the Bathing Water Quality legislation, etc. It has demonstrated the ability to detect two such parameters, nitrates and *Escherichia coli* (*E. coli*), using the system. This result provides further support for the concept of using networks of catchment monitoring stations, to act as real-time "red flag" warning systems for pollution events. The outcomes of the project support a policy of developing dispersed autonomous sensing networks in river catchments and bathing areas, using suitable ranges of sensors. To enable such networks, consideration should be given to the most suitable communications infrastructure and protocols to adopt, in terms of cost, data transmission requirements and communications reliability. Creating comprehensive multi-modal communication gateway networks could be considered, to enable such whole-of-catchment monitoring.

Developing Solutions

Previous research indicated that Raman spectroscopy (RS) combined with artificial intelligence (AI) methods could be a feasible way to detect certain target analytes in water. The project aimed to use these technologies to develop an innovative, low-cost autonomous system for detection of water-borne nutrients (nitrates and phosphates) and pathogens (specifically *E. coli*), and which is capable of operating in close to real time. A Lab-on-Chip model was envisaged as the ideal project outcome. An iterative approach allowed parallel progress on different aspects of the system, which mitigated the technical challenges of creating the Lab-on-Chip version.

The project has verified the hypothesis by developing and demonstrating Watermon, an end-to-end RS-based detection system that uses AI models to rapidly detect nitrates and *E. coli*. The AI model for nitrates was very effective (c. 99% accuracy) at detecting nitrates at 30 mg/l or above in river and drinking water. The *E. coli* AI model was able to detect the pathogen at 250 colony-forming units with an accuracy of c. 83%. At such levels, it was not possible to identify positive *E. coli* samples by traditional analysis of the data, i.e. inspection of a single spectrum, underscoring the superior ability of the AI model, albeit at relatively higher concentrations.

EPA Research: McCumiskey House, Richiew, Clonskeagh, Dublin 14.

Phone: 01 268 0100
Twitter: @EPAResearchNews
Email: research@epa.ie