

A Novel Geosmin Detection Method Based on Microwave Spectroscopy

Samuel Peter Ryecroft
Built Environment and Sustainable
Technologies (BEST) Institute
Liverpool John Moores University
Liverpool, United Kingdom
S.P.Ryecroft@2012.ljmu.ac.uk

Andy Shaw
Built Environment and Sustainable
Technologies (BEST) Institute
Liverpool John Moores University
Liverpool, United Kingdom
A.Shaw@ljmu.ac.uk

Paul Fergus
Built Environment and Sustainable
Technologies (BEST) Institute
Liverpool John Moores University
Liverpool, United Kingdom
P.Fergus@ljmu.ac.uk

Patryk Kot
Built Environment and Sustainable
Technologies (BEST) Institute
Liverpool John Moores University
Liverpool, United Kingdom
P.Kot@ljmu.ac.uk

Khalid Hashim
Built Environment and Sustainable
Technologies (BEST) Institute
Liverpool John Moores University
Liverpool, United Kingdom
K.S.Hashim@ljmu.ac.uk

Laura Conway
United Utilities,
Warrington
WA5 3LP, United Kingdom
Laura.Conroy@uuplc.co.uk

Adam Moody
United Utilities,
Warrington
WA5 3LP, United Kingdom
Adam.Moody@uuplc.co.uk

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

Keywords— *Electromagnetics, Geosmin, Microwave spectroscopy, Sensors, Water quality,*

I. INTRODUCTION

Geosmin is a leading cause of odour related complaints from water consumers in the UK. Geosmin is a naturally occurring organic compound and is often introduced into water intake supplies through the breakdown of dead algae by microorganisms[1] in water catchment areas such as reservoirs. Geosmin can be removed from water supplies using a weak acidic environment to break Geosmin down, however this process is only undertaken when Geosmin contamination is detected in water intakes. In 2017 out of 37,186 samples taken in England 100 failed to meet the standards for odour, with 17,000 contacts to companies regarding odour issues[2].

Humans can detect Geosmin in extremely low concentrations between 4 ng/L to 29 ng/L [3], meaning even small quantities of Geosmin contamination can be detected by end consumers. With complaints to water providers often starting when contamination reaches 7 ng/L or 0.007 ppb.

Water odour is an important indicator for many consumers, an unexpected and unexplained odour can cause consumers to become concerned over the quality of the water

supply increasing complaints received by water suppliers. Unexpected odours can also have an impact on the state of mental health of consumers

Legislation in the England and Wales does not specify acceptable quantities of Geosmin in end consumers water, however legislation dose state with regards to odour that water supplied must be: ‘Acceptable to consumers and no abnormal change’[4] this stipulation of the legislation must be complied with at the consumers water tap.

R.L.Bristow et al[5] conducted an extensive review in to the methods currently available for the detection of Geosmin in water. The work covered methods including mass spectrometry, gas chromatography, bromine reaction and enzyme-linked immunosorbent assay. The work examines and compares the various available methods for Geosmin detection including required sample volume and detection limits.

One Geosmin detection method developed by M. Son et al[6] developed a method of real-time monitoring of Geosmin using a bioelectronic nose. The approach used cloned human Olfactory Receptors sensitive to Geosmin to create a bioelectric nose. The resultant sensor could detect Geosmin at low levels of concentration that would make it suitable to Geosmin detection in water. The approach allows for the detection of Geosmin at 10 ng/L, however the manufacturing process requires several steps to produce using biological material. The work dose not identify how the sensor functions over long periods of time, it is possible that the performance of the sensor will decay over time making it unsuitable for prolonged deployment.

The most common method of Geosmin detection is using gas chromatograph mass spectrometry using a variety of extraction methods such as headspace solid-phase microextraction [7] and stir bar sorptive extraction [8].

Work by T. P. Hensarling and S. K. Waage[9] established a bromine-based colour reaction for the detection of Geosmin.

The work used a method of titrating a 4% bromine solution into a Geosmin sample until a faint yellow colour persisted in the sample. Five drops of formic acid were added and a blue colour was observed. The solution was monitored at 15 minutes intervals using a spectrophotometer, the absorbance intensified for 1.5 hours before slowly decreasing. The authors highlighted that the work does not provide a direct detection method for Geosmin, but can be used to indicate the presence of Geosmin.

Work undertaken by G. S. Braga[10] et al developed an electronic tongue for the detection of Geosmin in water. The approach used conducting polymer sensors to detect Geosmin at concentrations of 25 ng/L with the sensor signal being saturated at 300 ng/L, the lower threshold of 25 ng/L is significantly below that demand by industry.

Work produced by S. Chung[11] et al established the use of enzyme-linked immunosorbent assay (ELISA) for the detection of Geosmin in water. The approach described allows for a detection rate of 1 ug/mL (1000000 ng/L). The low detection threshold makes the use of ELISA for detecting Geosmin in the small quantities for water quality monitoring purposes unsuitable.

Microwave spectroscopy sensors provide a non-destructive method of testing that does not require reagents. Microwave spectroscopy sensors have been developed for a wide range of applications including the detection of zinc in water[12], moisture content of concrete[13] and glyphosate in deionised water[14].

The development of a sensor that can be used in a continuous monitoring approach combined with development in underwater communications could enable the remote monitoring of Geosmin enabling water providers to begin treatment for Geosmin contaminations sooner.

Machine learning, and feature selection techniques have provided powerful techniques to analyse large data sets. Feature selection techniques have previously been used to reduce the vast feature spaces involved in microwave spectroscopy data sets, allowing for areas of potential interest to be quickly identified.

Previous works undertaken by P. Kot et al. [15] have used a two-stage approach to reducing microwave spectroscopy data sets, applying an information gain filter to reduce large feature spaces before applying a wrapper feature selection approach using supervised machine learning algorithms: K Nearest Neighbour (KNN), Gradient Boosted Models (GBM), Support Vector Machine (SVM) and Random Forest (RF). The results demonstrated that Feature Selection Techniques can be used to identify frequencies that could be applied to detect Alpha Cypermethrin in microwave spectroscopy data sets.

Environmental monitoring through the use of wireless sensor networks has been explored with requirements of such as system considered by other works[16]. With developments in the use of radio frequency communications in underwater environments[17] the feasibility of deploying a real-time monitoring system for Geosmin as well as other contaminants has significantly improved if sensor technologies can be developed to allow for continuous monitoring without the need for reagents that are capable of being deployed for long periods of time without the need for maintenance.

II. DATA CAPTURE

Samples were created using Geosmin, Methanol and Distilled water. Samples of Geosmin were made in the following concentrations; 1 mg/l, 0.5 mg/l, 10 ng/l and 5 ng/l. The concentration of methanol remained the same across all samples. Methanol blank samples consisting of water and Methanol were created in the following concentrations; 5mg/l, 10 mg/l and 20 mg/l. The methanol samples provide a reference of varying levels of methanol allowing for verification that the results are based on detecting Geosmin and not Methanol. Samples were created and stored in 15ml plastic centrifuge tubes and were refrigerated when not in use.

Readings were taken using two microwave resonant cavities using and ZVL Vector Network Analyser (VNA) manufactured by Rhode and Schwartz. Readings were captured through lab view software that captured data from the VNA and stored it in CSV format for later analysis.

Readings were taken 20 times for each sample, each time the sample was removed from the cavity and placed in an alternative position to the previously tested position. This was done to ensure that the readings were not influenced by factors such as inconsistencies in the centrifuge tube. Figure 1 shows an example of the experimental setup of a cavity sensor.

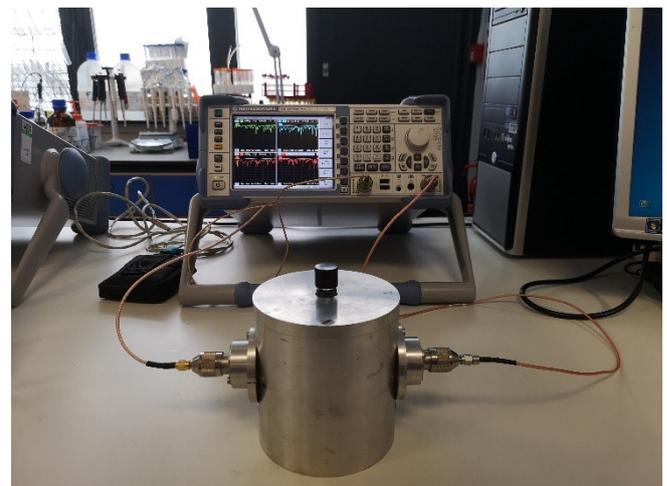


Figure 1 Example of the experimental setup for data capture using a cavity connected to a VNA

III. DATA ANALYSIS

Data was analysed using R statistical language, data was loaded in from CSV format for further manipulation. During the loading processes the data was labelled based on the class that the data represented, with the following classes used: 5 ng/l, 10 ng/l, 0.5 mg/l and 1 mg/l and methanol blanks being represented as one class. The methanol blanks were treated as one class to exclude any variances in methanol concentrations that could be present in the Geosmin samples. Data was first normalised, this was done to ensure that the processing was not influenced by the magnitude of the reading values.

The initial feature space consisted of 4000 features, to reduce the feature space to a more manageable level where wrapper selection methods could be easily applied an information gain filter to the feature space. The information gain filter was used to rank features in order of mutual importance of each of the features within the data set. The 500 highest ranking features were carried forward to the next stage

of the feature selection processes, the 500 limit was established after empirical evaluation of other thresholds, it was found that the 500 threshold gave a strong balance of quick computational performance while maintaining a without narrowing the features available to the wrapper feature selection approach.

To reduce the feature space further the 500 features carried over from the information gain filtering process the features were used in a wrapper feature selection approach using four machine learning algorithms: K-Nearest Neighbour (KNN), Support Vector Machines (SVM) and Random Forest (RF). The wrapper feature selection process split the data with 70% being used for training and the remaining 30% being used as an evaluation set. The models were evaluated to identify the features that most influenced the predictions of the machine learning algorithms used.

IV. RESULTS AND DISCUSSION

The most important frequencies identified using the feature selection process outlined can be observed in Table 1. Table 1 contains the top 10 features from the KNN, SVM and RF algorithms applied as well as the top 10 ranking features of the information gain filter used provided from most important to least important as identified through the respective algorithm.

Table 1 shows that many of the frequencies identified focused around the 6.4 GHz to 6.5 GHz range across multiple machine learning algorithms with the information gain filter and Random Forest models all identifying multiple frequencies within this range, the K-Nearest Neighbour and Support Vector Machines also identified frequencies in this range though these were ranked lower than other frequencies.

Table 1

IG	RF	KNN	SVM
8282820608	6461365248	7322580480	7436609024
6485371392	6476368896	5807201792	5729182208
6476368896	5900225024	10416354304	7364591104
5789197312	10374343680	5618154496	5642160640
6461365248	10659415040	7172542976	5480119808
7151537664	5501125120	6551387648	5846211584
6482370560	10974493696	5963240960	7202550784
6488372224	5507126784	7133533184	4960990208
6470367744	10980495360	7421605376	8351838208
3649662464	8276819456	6491373056	6470367744

Features from each of the used wrapper feature selection methods and the information gain filter in order of most to least importance

Figure 2 shows a plot of the most important features as identified by the information gain filter. The plot shows a clear and significant grouping of the five classes except for one 0.5 mg/L Geosmin sample that is closely located to the Geosmin sample. The plot shows that each class within the plot is clearly separated from other classes with the 5 ng/L and 0.5 mg/L classes being well grouped and separated from each other. The 10 ng/L and 1 mg/L classes show slightly more dispersion with the 10 ng/L readings being dispersed into the 1 mg/L class.

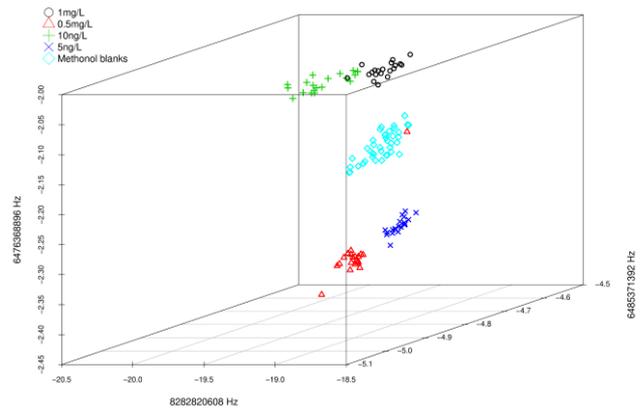


Figure 2 Top three features identified through information gain filter

Figure 3 shows a plot of the most important features as identified using the RF wrapper feature selection. The features identified through the RF wrapper feature selection process shows a clear separation between the Geosmin concentrations and the methanol blanks except for one 0.5 mg/L, which clusters with the methanol blanks. Unlike the features identified through the information gain filter the Geosmin samples were more dispersed and loosely clustered.

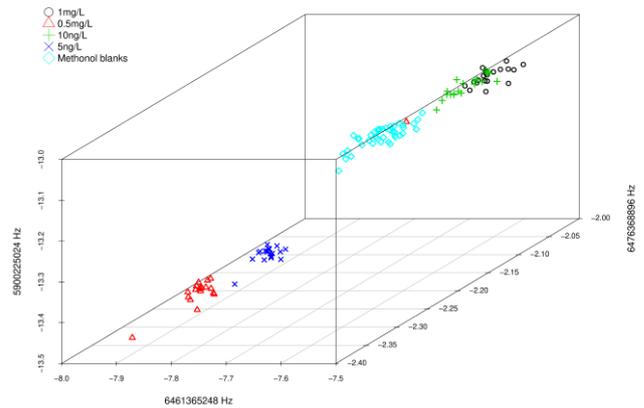


Figure 3 Top three features identified through Random Forest feature selection

Figure 4 shows a plot of the top three ranking features of the KNN wrapper feature selection. The plot shows a more dispersed methanol control group though all methanol samples remain clustered together. The Geosmin classes show separation away from the methanol control group though the Geosmin classes are dispersed and in the case of the 1 mg/L and 10 ng/L overlap significantly. Some samples from the 5 ng/L class are significantly separated from the others within the class, instead overlapping more with the 0.5 mg/L class rather than the 5 ng/L class.

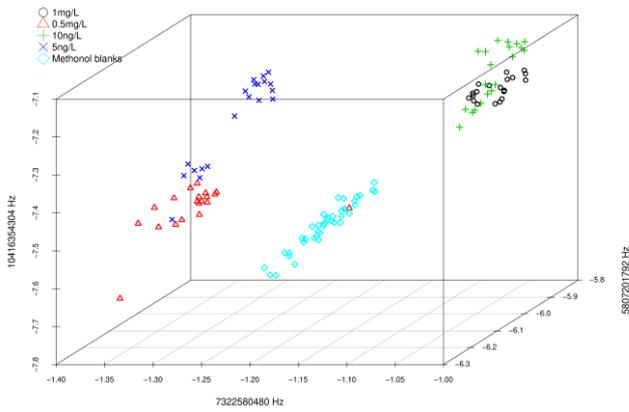


Figure 4 Top three features identified through KNN feature selection

Figure 5 shows a plot of the most important features identified through SVM wrapper feature selection. The plot shows a clear distinction between the methanol control samples and the Geosmin classes. The separation between Geosmin classes is significantly less defined with the 1 mg/L and 5 ng/L samples overlapping significantly. The 0.5 mg/L and 10 ng/L samples were also more loosely clustered with overlap between the two dispersed clusters.

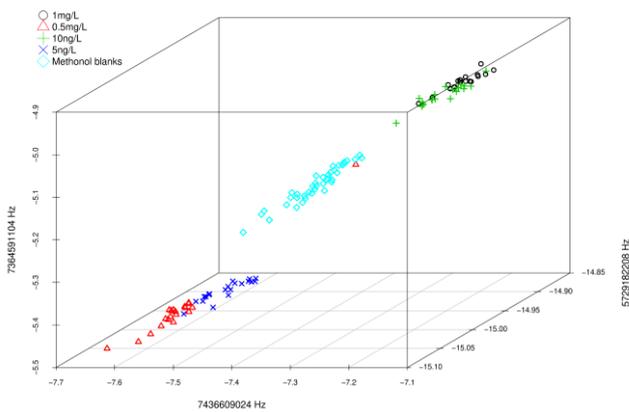


Figure 5 Features identified through SVM feature selection

Figure 6 shows a plot of frequency and magnitude readings surrounding the 8282820608 Hz frequency identified as the top ranking feature through using an information gain filter. The plot shows a clear difference between the mean average of each of the methanol blanks and each of the Geosmin sample means. This clear difference could be used to distinguish between blank samples and samples contaminated with Geosmin.

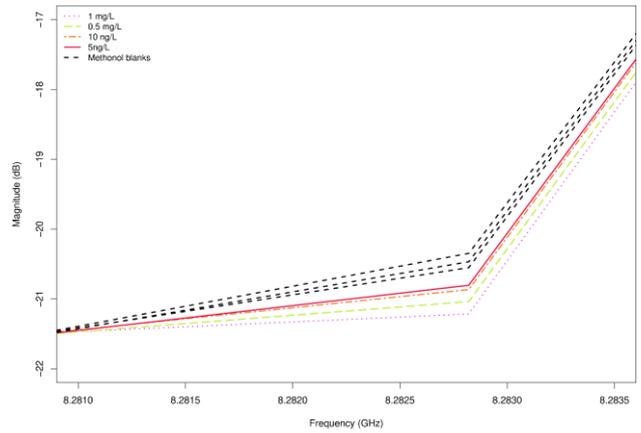


Figure 6 A plot of frequency against magnitude in the surrounding region of the 8.2 GHz frequency

Figure 7 shows a plot of frequency and magnitude readings surrounding the 6485371392 Hz frequency identified as the second most important feature using an information gain filter. The plot again shows a clear separation between the Methanol blank means and the Geosmin means, the plot shows a separation between each of the Geosmin samples while the Methanol samples display less separation between one and another.

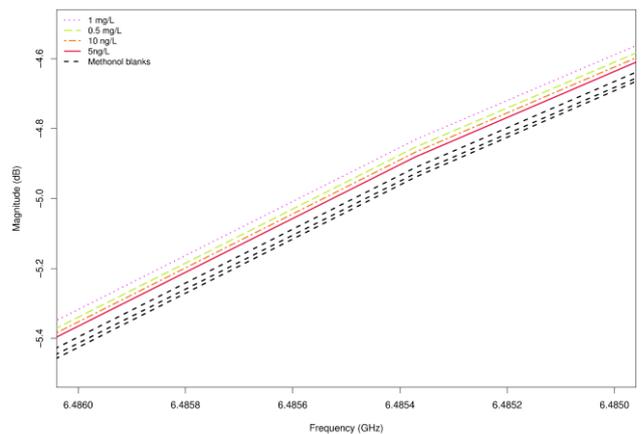


Figure 7A plot of frequency against magnitude in the surrounding region of the 6.48 GHz frequency

The results show that the use of microwave spectroscopy enables the detection of Geosmin at levels that are detectable by humans making it an ideal approach for water industries. The presented approach allows for a continual monitoring of Geosmin throughout a prolonged deployment period unlike some of the existing methods such as mass spectrometry and ELISA that require lab-based testing and equipment to detect and quantify Geosmin levels.

V. CONCLUSION

The results demonstrated the feasibility of using microwave spectroscopy to detect Geosmin contamination. The results show that the following frequencies could be used by a sensor to detect Geosmin contamination.

The results identified that frequencies in the range of 6.4 GHz to 6.5 GHz could be used to identify Geosmin

contamination in water. These frequencies showed Geosmin samples of each class clustered together and clearly separated from the control methanol samples. This frequency range not only indicates the ability to detect Geosmin, but could be applied to quantify the amount of Geosmin present in the water.

Other frequency ranges such as 8.2 GHz to 8.4 GHz could also be applied for the detection of Geosmin as well as the measurement of the quantities of Geosmin presented in water.

The results indicated that Information Gain identified more strongly clustered classes showing, agreeing with previous results obtained in similar data sets by P. Kot et al [15], which showed that information gain filters and Random Forest algorithms provided strong clustering of classes in similar data sets.

The results showed that Random Forest achieved more weekly formed clusters even though the random forest wrapper feature selection identified two frequencies in a similar range to the more closely formed Information Gain results.

The results found that other frequencies could be used to detect Geosmin such as those identified through SVM. The SVM results identified one feature in the 6.4 GHz frequency range although this was ranked the lowest of the 10 features in order of feature importance. The SVM feature selection identified multiple features in the 5.4 GHz to 5.9 GHz and 7.2 GHz to 7.5 GHz frequency ranges.

The results shown could be used to develop a dedicated sensor targeted at detecting Geosmin contamination. Such as sensor could be combined with other work in underwater communications to develop a sensor network that could be deployed to detect Geosmin contamination over large water catchment areas such as reservoirs.

VI. FUTURE WORK

The results presented support the use of a microwave spectroscopy for the detection of Geosmin in water. Further work needs to evaluate the feasibility of using microwave spectroscopy in water containing other contaminants.

Additional work needs to be undertaken to validate results in a larger sample group using additional samples contaminated with additional substances including contaminants commonly found in raw water and other likely deployment environments to further validate these initial findings in more complex substances.

Once further work has been undertaken to validate these initial findings a suitable sensor prototype can be developed to integrate with an underwater wireless communication system that is capable of relaying readings from sensors deployed over a large area such as a reservoir.

ACKNOWLEDGMENT

The authors are grateful for providing a funding for this research to United Utilities.

REFERENCES

- [1] N. N. Gerber, "Geosmin, from microorganisms, is trans-1, 10-dimethyl-trans-9-decalol," *Tetrahedron Letters*, vol. 9, no. 25, pp. 2971-2974, 1968/01/01/1968.
- [2] C. I. o. D. Water, "Drinking water 2017," Drinking Water Inspectorate, www.dwi.gov.uk2017, Available: http://www.dwi.gov.uk/about/annual-report/2017/Summary_CIR_2017_England.pdf, Accessed on: 13/05/2019.
- [3] D. M. C. Rashash, A. M. Dietrich, and R. C. Hoehn, "FPA of selected odorous compounds," *Journal - American Water Works Association*, vol. 89, no. 4, pp. 131-141, 1997.
- [4] HM Government, *Water, England and Wales. The Water Supply (Water Quality) Regulations 2016* (no. 614). England and Wales: HM Government, 2016, pp. 11-15, 36-42.
- [5] R. L. Bristow, I. S. Young, A. Pemberton, J. Williams, and S. Maher, "An extensive review of the extraction techniques and detection methods for the taste and odour compound geosmin (trans-1, 10-dimethyl-trans-9-decalol) in water," *TrAC Trends in Analytical Chemistry*, vol. 110, pp. 233-248, 2019/01/01/2019.
- [6] M. Son *et al.*, "Real-time monitoring of geosmin and 2-methylisoborneol, representative odor compounds in water pollution using bioelectronic nose with human-like performance," *Biosensors and Bioelectronics*, vol. 74, pp. 199-206, 2015/12/15/2015.
- [7] K. Saito, K. Okamura, and H. Kataoka, "Determination of musty odorants, 2-methylisoborneol and geosmin, in environmental water by headspace solid-phase microextraction and gas chromatography-mass spectrometry," *Journal of Chromatography A*, vol. 1186, no. 1, pp. 434-437, 2008/04/04/2008.
- [8] S. Nakamura, N. Nakamura, and S. Ito, "Determination of 2-methylisoborneol and geosmin in water by gas chromatography-mass spectrometry using stir bar sorptive extraction," *Journal of Separation Science*, vol. 24, no. 8, pp. 674-677, 2001.
- [9] T. P. Hensarling and S. K. Waage, "A bromine-based color reaction for the detection of geosmin," *Journal of Agricultural and Food Chemistry*, vol. 38, no. 5, pp. 1236-1237, 1990/05/01/1990.
- [10] G. S. Braga, L. G. Paterno, and F. J. Fonseca, "Use of an Electronic Tongue to Detect Geosmin in Distilled Water," *AIP Conference Proceedings*, vol. 1362, no. 1, pp. 287-288, 2011/09/06/2011.
- [11] S.-Y. Chung, P. B. Johnsen, and P. Klesius, *Development of an ELISA using polyclonal antibodies specific for 2-methylisoborneol*. 1990.
- [12] I. Frau, O. Korostynska, P. Byrne, and A. Mason, "Continuous monitoring of Zn in water with bismuth oxide thick-film using microwave and electric techniques," in *2017 Eleventh International Conference on Sensing Technology (ICST)*, 2017, pp. 1-6.
- [13] K. H. Teng *et al.*, "Embedded Smart Antenna for Non-Destructive Testing and Evaluation (NDT&E) of Moisture Content and Deterioration in Concrete," *Sensors*, vol. 19, no. 3, p. 547, 2019.
- [14] S. Cashman, O. Korostynska, A. Shaw, P. Lisboa, and A. Mason, *Detection of glyphosate in deionised water using machine learning techniques with microwave spectroscopy*. 2017, pp. 253-256.
- [15] P. Kot *et al.*, "Identification of Optimal Frequencies to Determine Alpha-Cypermethrin Using Machine Learning Feature Selection Techniques," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, 2018, pp. 1-7.
- [16] S. P. Rycroft *et al.*, "Requirements of an underwater sensor-networking platform for environmental monitoring," in *11th International Conference on Developments in eSystems Engineering (DeSE)*, Cambridge, UK, 2018, pp. 95-99.
- [17] S. Rycroft *et al.*, "A First Implementation of Underwater Communications in Raw Water Using the 433 MHz Frequency Combined with a Bowtie Antenna," *Sensors*, vol. 19, no. 8, p. 1813, 2019.