Detecting the presence and concentration of nitrate in water using microwave spectroscopy

Sean Cashman; Olga Korostynska; Andy Shaw; Paulo Lisboa; Laura Conroy[1]

***Abstract*— Nitrate is a common pollutant in surface waters which water companies must monitor for regulatory and safety reasons. The presence of nitrate in deionised water is detected and concentration estimated from microwave spectroscopy measurements in the range 9kHz-6GHz. Experimental results were obtained for 19 solutions (18 salt solutions in deionised water and 1 deionised water), each measured 10 times with 4001 points (total N=190). The resulting data was randomly assigned into equal parts training and test data (N=95 each). Both regression (for the estimation of nitrate concentration) and classification (for detecting the presence of nitrate) methods were considered, with a rigorous feature selection procedure used to identify two frequencies for each of the classification and regression problems.**

For detection classification models were applied with nitrate levels binned using 30mg/l as the threshold. A logistic regression model achieved AUROC of 0.9875 on test data and a multi-layer perceptron achieved AUROC of 0.9871. In each case the positive predictive value of the model could be optimised at 100% with sensitivity of 90% and specificity of 100%. For the concentration estimates, a linear regression model was able to explain 42% of the variance in the training data and 45% of the variance in the test data and an MLP model delivered similar performance, explaining 43% of variance in the training data and 47% of variance in the test data. A sensor based on this model would be appropriate for detecting the presence of nitrate above a given threshold but poor at estimating concentration.

***Index Terms*—Microwave sensors, water pollution, machine learning, feature selection**

I. Introduction

NITRATE is an inorganic anion of nitrogen, anthropogenic sources of reactive nitrogen (particularly from agriculture) have increased considerably since the 1950s and exceed that produced by natural sources, fundamentally altering the nitrogen cycle [1]. While the use of reactive nitrogen as fertiliser plays an important role in food production, careful monitoring and removal of these contaminants from the environment is vital. Alongside other nitrogen ions, nitrate pollution contributes to eutrophication of surface waters [1, 2], posing a threat to biodiversity in aquatic environments, and high concentrations of nitrate in drinking water supplies can cause acute toxic effects in human beings, notably Methemoglobinemia in infants (also known as "blue baby syndrome") [1]. Ingested nitrate may contribute to some cancers and reproductive harms [3]; a 2010 IARC monograph classified ingested nitrate under certain conditions "probably carcinogenic to humans" (IARC class 2A) [4] and a 2015 study of 1,087 patients in Spain found that long term exposure to very high levels of nitrate in drinking water was associated with increased risk of bladder cancer [5]. This paper aims to serve as a proof of concept for applying rigorous feature selection methods to the detection of nitrates using an electromagnetic wave sensor, testing both the ability of microwave spectroscopy to detect the presence and concentration of nitrate ions in simple solutions and the ability of information theoretic feature selection methods to identify the most useful microwave frequencies for this purpose.

Previous work by our group has investigated the use of electromagnetic wave sensors to measure the presence of substances in solution, including small ions such as nitrate, sulphate [6] and organic pollutants such as the herbicide glyphosate [7]. These previous investigations have involved identifying peaks in spectra where some visible difference exists on average in measurements carried out on solutions of varying substance and/or concentration, this paper extends this by incorporating feature selection to identify the best possible frequencies for discerning concentration of a selected substance.

In the UK nitrates in surface water are regulated by the Drinking Water Inspectorate, the relevant legislation is the EU Nitrates Directive. While the Nitrates Directive specifies a limit of 50 mg/l for surface water, DEFRA have historically tracked the percentage of river length with greater than 30 mg/l as an indicator, finding that 29% of river length in England was over this limit as of 2009 [8], for this reason this 30 mg/l level has been used in his paper as a level at which concern might be raised. I.e. 30 mg/l represents a point at which nitrate levels are sufficiently high that the body responsible for monitoring a body of water may wish to investigate further and consider appropriate action to prevent nitrates from reaching levels dangerous to either human populations or the environment).

Currently detection of nitrates and other ions in drinking water management uses ion sensitive electrodes, this is carried out "off-line" and intermittently, an on-line network of sensor systems capable of monitoring either the current concentration of a contaminant or whether that concentration has crossed some critical threshold in surface water bodies in real time would allow for better information and faster response to problems as and when they occur [9]. Online sensors do currently exist on the market, though these are currently not used by many water management companies.

Existing online monitoring systems for nitrates include "wet chemistry" systems that require a supply of reagents to run, ion sensitive electrode sensors and optical/UV sensors such as the Sea-Bird Coastal SUNA V2, which uses UV spectroscopy to quantify nitrate levels in the range 0.03-277.90 mg/l NO_3^- [10]. Optical sensors have a high degree of precision in monitoring exact nitrate levels, but can perform poorly in bodies of water with high turbidity as visible and UV light cannot pass as easily through these media. These systems rely on telemetry stations, which are large, expensive pieces of equipment. Demand exists

[1] This work was supported by funding from United Utilities Plc.

within industry for a cheaper online monitoring system which works within a wide range of environments and does not require an additional supply of reagents. This study serves as a proof of concept for such a sensor grounded in rigorous statistical methodology.

The high dimensionality of spectral measurements runs the risk of models overfitting the data unless the dimension of the regression space is minimised [11]. Feature selection is the elimination of those features in a dataset which are either irrelevant or redundant. In addition, features that are strongly correlated with each other make it difficult to identify a consistent minimal subset even using linear statistical models. For electromagnetic spectroscopy, feature selection is essential to the identification of a minimum set of frequencies at which S measurements can be used to make predictions about a variable of interest [12]. In this paper we describe the application of feature selection to microwave spectra and show the predictive value that can be obtained for making a smart measurement with these sensors. Identifying the appropriate minimum subset of frequencies can then inform the development of a smaller, more specific sensor in future applications.

II. Materials and Methods

The experimental data comprises 19 solutions (18 of a salt in deionised water and one of pure deionised water) prepared in 15ml plastic centrifuge tubes in the formulations listed in table 1. Each sample was placed in a cylindrical aluminium cavity with internal radius 17mm and depth 110mm attached to a Rohde & Schwarz ZVL vector network analyser (VNA) and electromagnetic (microwave range) spectroscopy readings in the range 9kHz to 6GHz with 4001 points were carried out, similar to the experimental set up in [6]. For each solution, 10 measurements were taken, each of which is treated as an independent sample for the purposes of this paper due to the variation in magnitude across multiple readings on the same solution.

TABLE I
Concentration of Ions in Salt Solutions

Solution	Na^+ (mg/l)	K^+ (mg/l)	NO_3^- (mg/l)	Cl^- (mg/l)
Deionised water	0.0000	0.0000	0.0000	0.0000
20 mg/l $NaNO_3$	5.4099	0.0000	14.5901	0.0000
40 mg/l $NaNO_3$	10.8197	0.0000	29.1803	0.0000
60 mg/l $NaNO_3$	16.2296	0.0000	43.7704	0.0000
80 mg/l $NaNO_3$	21.6395	0.0000	58.3605	0.0000
100 mg/l $NaNO_3$	27.0493	0.0000	72.9507	0.0000
120 mg/l $NaNO_3$	32.4592	0.0000	87.5408	0.0000
20 mg/l NaCl	7.8678	0.0000	0.0000	12.1322
40 mg/l NaCl	15.7357	0.0000	0.0000	24.2643
60 mg/l NaCl	23.6035	0.0000	0.0000	36.3965
80 mg/l NaCl	31.4713	0.0000	0.0000	48.5287
100 mg/l NaCl	39.3392	0.0000	0.0000	60.6608
120 mg/l NaCl	47.2070	0.0000	0.0000	72.7930
20 mg/l KNO_3	0.0000	7.7345	12.2655	0.0000
40 mg/l KNO_3	0.0000	15.4689	24.5311	0.0000
60 mg/l KNO_3	0.0000	23.2034	36.7966	0.0000
80 mg/l KNO_3	0.0000	30.9379	49.0621	0.0000
100 mg/l KNO_3	0.0000	38.6724	61.3276	0.0000
120 mg/l KNO_3	0.0000	46.4068	73.5932	0.0000

The measurement of interest is the concentration of nitrate ions, with the pure deionised water and sodium chloride solutions serving as controls with 0 mg/l of nitrate. The inclusion of the sodium chloride solutions allows for a greater degree of confidence that results are specific to nitrate levels.

The microwave absorption spectrum was calculated by taking the S_{12} magnitude (difference between signal transmitted at port 1 and received at port 2) readings for each solution were imported into a 190×4001 data matrix in Matlab 2016a, which was used for all statistical work in this paper, the multilayer perceptron classifier model was produced using the Netlab toolbox [13].

A. Statistical analysis

Statistical modelling was carried out using both linear and non-linear methods, with both classification and regression to implement detection and calibration, respectively. The threshold for detection was taken to be a concentration of NO_3^- greater than 30 mg/l. Data was randomly assigned to training and test groups (N=95 for both groups). Regression models will be compared using percentage of variance explained in the test data (calculated as Pearson's r squared between target and fitted value), classification models will be compared using the ROC framework. The models to be used are shown in table 2.

TABLE II
Statistical Models Used by Type

	Regression	Classification
Linear	Least squares linear regression	Model: Binary logistic regression
Non-linear	Model: MLP for regression	Model: MLP for classification (Sigmoid)

In order to select the features which provide the greatest predictive power for nitrate concentration, a feature selection based on pairwise χ^2 tests and G tests for conditional independence was used, with stepwise forward search (SFS). Stepwise forward search does not test all possible models, instead aiming simply to optimise the objective function at each step of the process [11]. This is appropriate for spectral data as the total size of the hypothesis space for multiple tests of dependence vs. independence of each feature with a variable of interest, given 4001 features is of size $2^{4001} \approx 2.6 \times 10^{1204}$. The feature selection algorithm starts with an empty set of features and, at each stage, adds the feature most strongly associated with nitrate concentration (based on mutual information), only adding a feature if at least one of the remaining candidate features is significantly associated with nitrate given the features already included in the model, this process is summarised below:

1. Let T be the concentration of nitrate, F be the set of all candidate features and F' be the set of features included in the model.
2. For i=1:len(F):
    a. Test $T \perp F_i$ using a χ^2 test for pairwise independence and calculate the pointwise mutual information, $I(T; F_i)$
    b. If $T \perp F_i$, remove F_i from F
3. Append the feature with the greatest mutual information to F' and remove it from F
4. While F≠∅:
    a. For i=1:len(F):

i. Test $T \perp F_i \,|F'$ using a G test and calculate the conditional mutual information $I(T; F_i \,|F')$

b. If $T \perp F_i \,|F'$ remove F_i from F

c. Append the feature with the greatest mutual information to F' and remove it from F

Due to the large number of comparisons being made at each stage of this process, there is a risk of inflating the false positive rate, this was controlled using q-values as described by [14]. For each iteration through the set of candidate features in step 2 or 4, q-values were calculated and an association was considered to be significant if the corresponding q-value was less than or equal to 0.05.

III. Results

A. Classification

SFS (with S_{12} features binned by tercile and nitrate binned by above or below 30 mg/l for the χ^2 and G tests) identified 2 frequencies at approximately 4.1GHz and 4.6GHz at which S_{12} distinguished between those solutions in class 0 (less than 30 mg/l of NO_3^-) and class 1 (at least 30 mg/l of NO_3^-), summarized in table 3.

TABLE III
SELECTED FEATURES FOR CLASSIFICATION

Feature (S_{12} at given frequency)	Pairwise significance (χ^2)	PMI	Significance given other frequency (G)	CMI
4,147,502,848Hz	*	0.6943	*	0.2502
4,558,502,400Hz	*	0.6943	*	0.2502

*p<0.0001

Figure 1 shows spectra plots for four of the solutions used; deionised water, $NaNO_3$ 120mg/l, KNO_3 120mg/l and NaCl 120mg/l. In the normalised plot, clear separation between the nitrate containing solutions and the control solutions is visible at the two selected features. Greater absorption at 4.1GHz is associated with greater probability of higher nitrate levels and greater absorption at 4.6GHz associated is with lower nitrate levels.

The logistic regression and multi-layer perceptron models both perform well in separating the two classes, achieving AUROC of 0.9875 and 0.9871 respectively on the test data. Both models achieved 96% classification accuracy where a threshold is selected to maximise the positive predictive value (PPV); where 0.5 is used as the threshold for classification, the logistic regression model achieves 94% accuracy and the MLP achieves 89% accuracy. Further detail on the results and the decision boundaries given by the two models is given in the following subsections. The very similar performance, selected thresholds and decision boundaries produced suggest that using a non-linear method adds very little predictive power for this data and a logistic regression model would be more appropriate as a result.

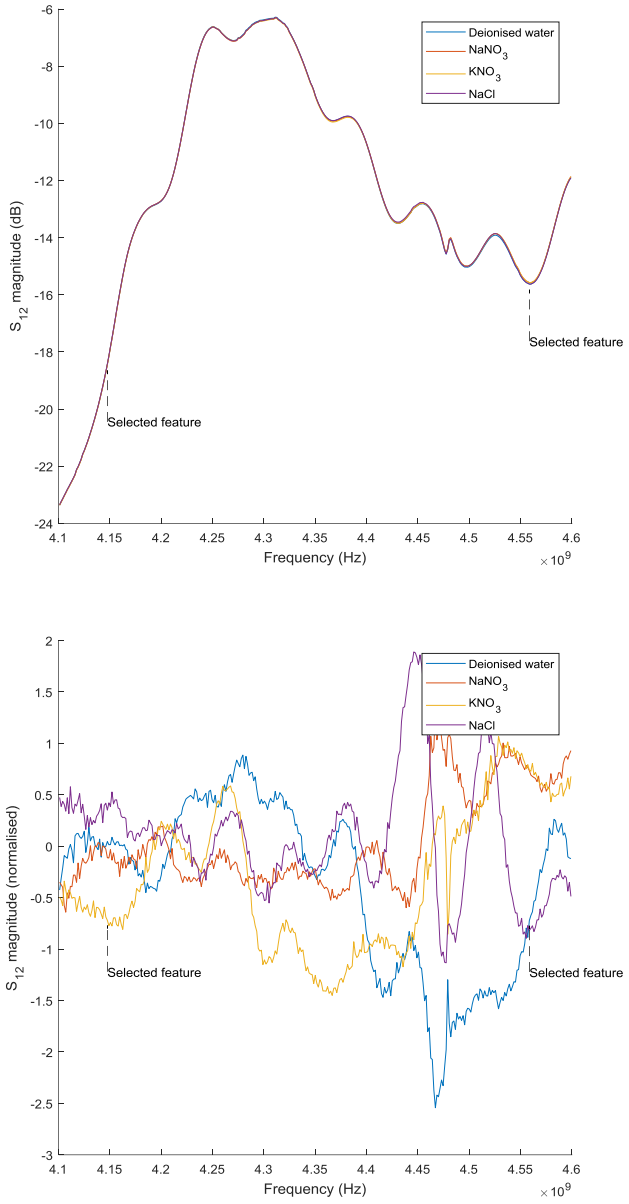


Fig. 1. Spectra plots at selected ranges. Shows average spectra plots for deionized water and the 120mg/l solutions of each substance both as actual S_{12} magnitude and normalized between 4.1 and 4.6 GHz.

1) Logistic regression

A logistic regression model was fitted to the training data, the model is outlined in table 4.

TABLE IV
LOGISTIC REGRESSION MODEL

Term	Beta	Significance
Constant	-1135.6523	0.4459
S_{12}@4147502848Hz	-183.3223	0.0162
S_{12}@4558502400Hz	143.8818	0.0053

Figure 2 shows the receiver operating characteristic (ROC) curves for predictions on both the training and test data using the logistic regression model and figure 3 shows scatter plots of the selected features on the training and test data with decision boundaries given by the logistic regression model.

For the test data, three classification thresholds were considered – the highest threshold which provided 100% sensitivity or recall (0.027), 0.5 and the point on the empirical ROC curve for the test data with the smallest Euclidean distance from the point [1,0] i.e. the threshold which provides an optimal compromise between sensitivity and false positive rate (0.622). The relevant parameters are summarised in table 5.

Maximum positive predictive value, specificity and overall accuracy are obtained for a threshold of 0.622 improves accuracy, specificity, positive predictive value (PPV) and negative predictive value (NPV) without sacrificing sensitivity relative to a threshold of 0.500. This is our selected point on the ROC curve.

TABLE V
ROC PARAMETERS AT THREE THRESHOLDS – LOGISTIC REGRESSION

THRESHOLD	0.027	0.500	0.622
# True Positive	38	34	34
# False positive	9	2	0
# True negative	48	55	57
# False negative	0	4	4
Accuracy	90.53%	93.68%	95.79%
Sensitivity	100.00%	89.47%	89.47%
Specificity	84.21%	96.49%	100.00%
PPV	80.85%	94.44%	100.00%
NPV	100.00%	93.22%	93.44%

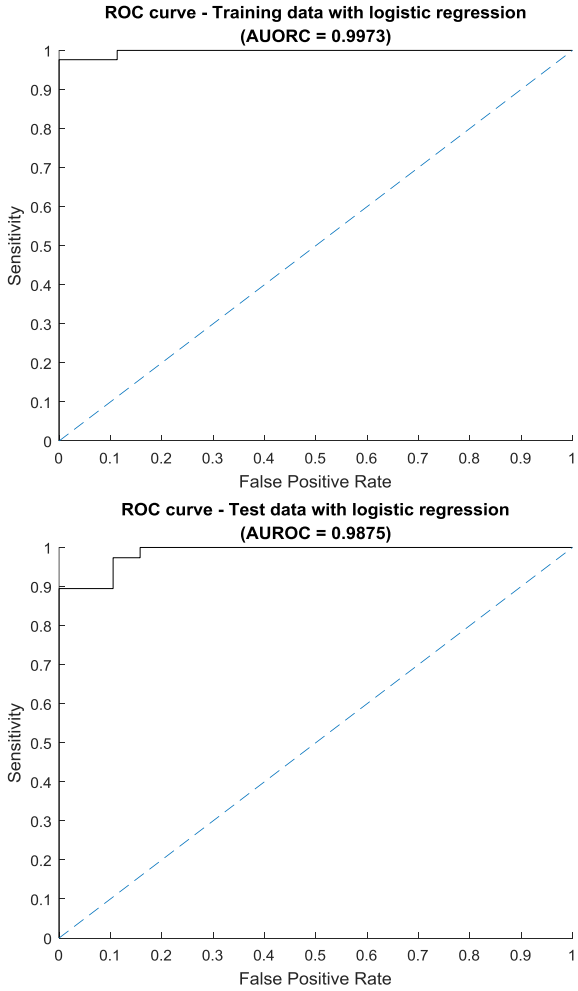


Fig. 2. ROC curves for logistic regression models. Classification performance is high for both training and test data. AUROC was 0.9973 for the training data and 0.9875 for the test data.

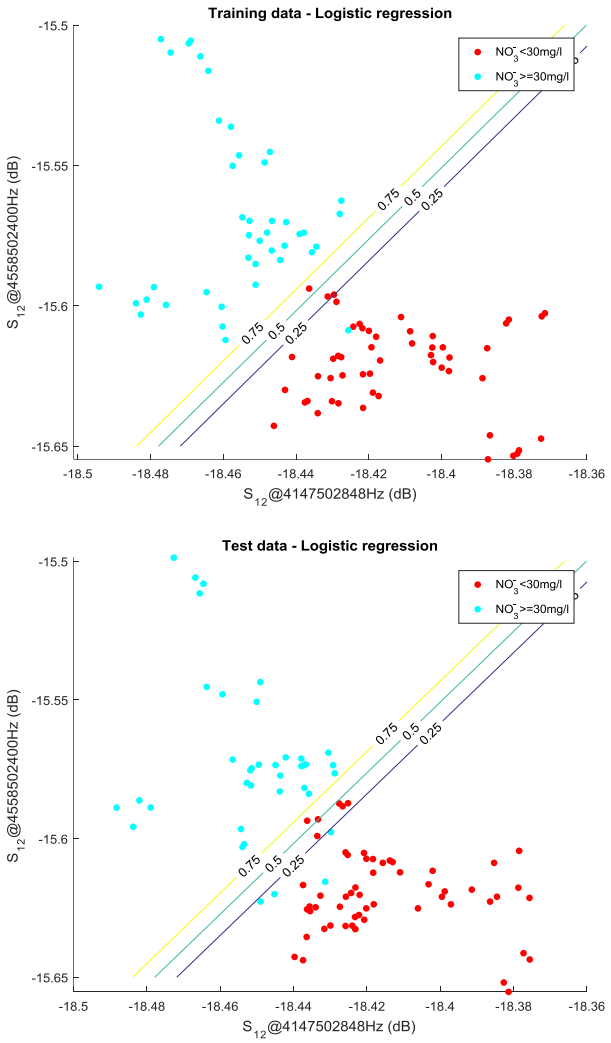


Fig. 3. Scatter plots with logistic regression decision boundaries. Contour lines show the predicted posterior probability for the logistic regression model. The two classes are linearly separable on the selected frequencies with a high level of accuracy.

2) *Multi-layer Perceptron*

The multi-layer perceptron model (produced in Netlab) used cross entropy as its objective function, using a quasi-Newton optimisation algorithm with 20 hidden nodes and a prior weight decay of 0.001.

Table 6 summarises the results for the MLP model on the test data (again using the highest threshold with 100% sensitivity, the threshold closest to [0,1] on the ROC curve and 0.5) and figure 4 shows the ROC curves provided by the MLP on the training and test data. As with the logistic regression model, classification performance was high, with the best performing threshold providing 96% accuracy and the 0.5 threshold providing 89% accuracy. Figure 5 shows scatterplots on the selected features for training and test data with the decision boundaries given by the MLP.

TABLE VI
ROC PARAMETERS AT THREE THRESHOLDS – MULTI-LAYER PERCEPTRON

THRESHOLD	0.024	0.500	0.601
# True Positive	38	34	34
# False positive	10	3	0
# True negative	47	54	57
# False negative	0	4	4
Accuracy	89.47%	92.63%	95.79%
Sensitivity	100.00%	89.47%	89.47%
Specificity	82.46%	94.74%	100.00%
PPV	79.17%	91.89%	100.00%
NPV	100.00%	93.10%	93.44%

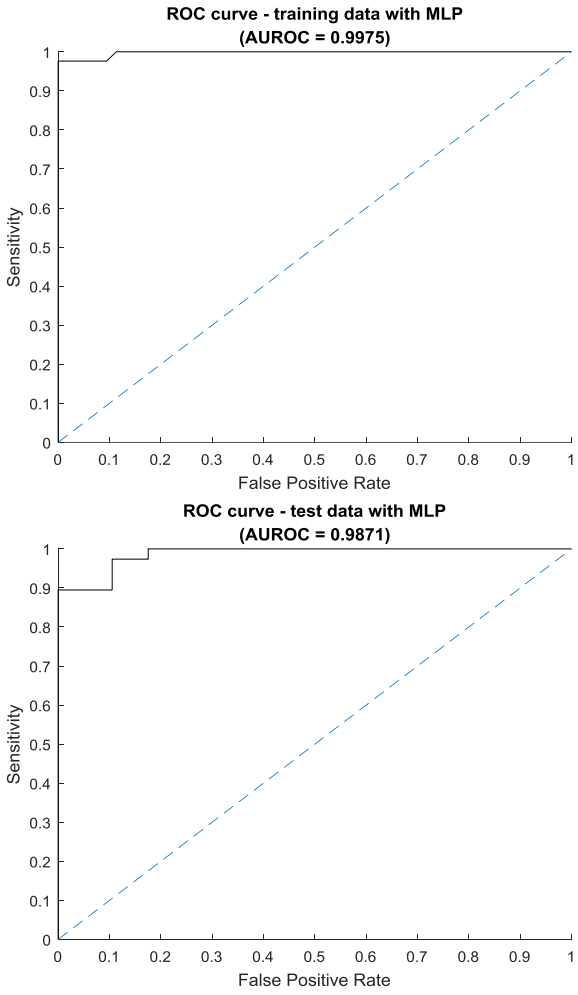


Fig. 4. ROC curves for multi-layer perceptron models. Classification performance is high for MLP models. AUROC is 0.9975 for the training data and 0.9871 for the test data.

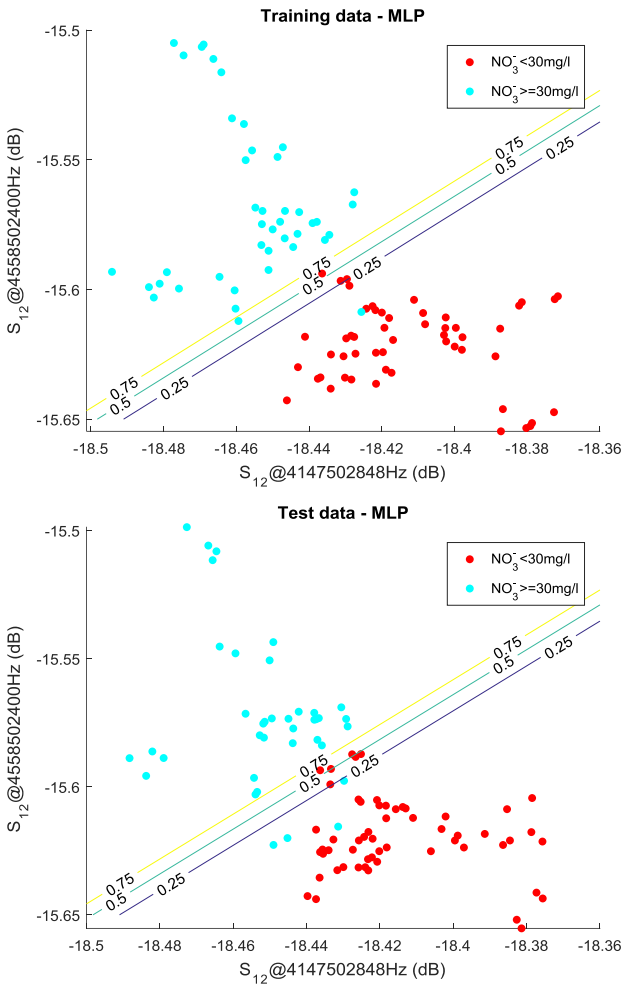


Fig. 5. Scatter plots with MLP decision boundaries. Contour lines show the predicted posterior probability for the multi-layer perceptron model. The resulting model is nearly indistinguishable from that produced by linear methods.

B. *Regression models*

For the regression models, both S_{12} measurements and nitrate concentration were binned by terciles of the training data for SFS and two useful frequencies at approximately 2.4GHz and 4.5GHz were identified, S_{12} at these frequencies was used to fit a least squares linear regression model and a multi-layer perceptron model. Figure 6 shows scatter plots of the two selected features vs. nitrate concentration in the training data only.

TABLE VII
SELECTED FEATURES FOR REGRESSION

FEATURE (S_{12} AT GIVEN FREQUENCY)	PAIRWISE SIGNIFICANCE (χ^2)		PMI	SIGNIFICANCE GIVEN OTHER FREQUENCY (G)		CMI
2,389,505,536Hz		*	0.4842		*	0.5618
4,495,502,336Hz		*	0.2653		*	0.3430

*p<0.0001

The performance of both the linear and MLP models was similar, Figure 7 shows the surfaces produced by the MLP and linear regression models over the domain of the selected features.

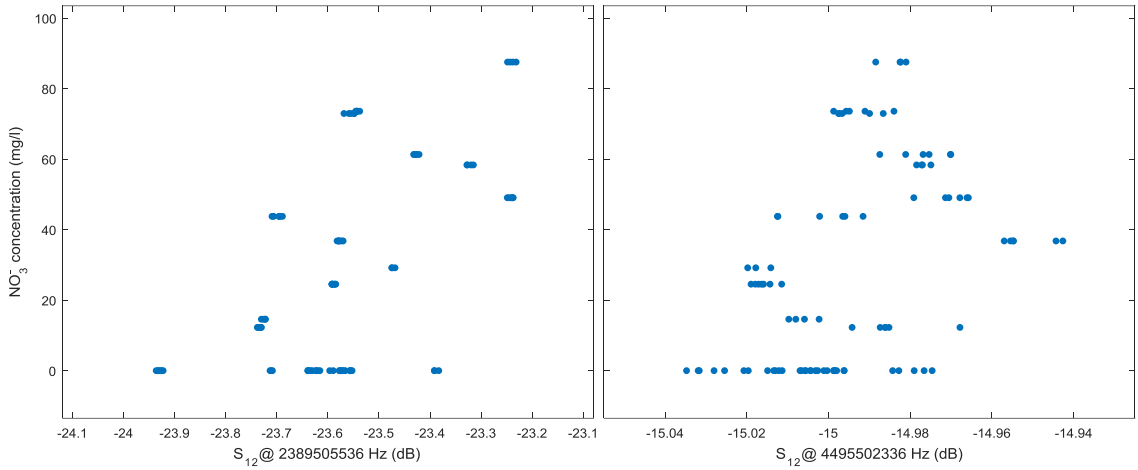


Fig. 6. Scatter plots of S_{12} at the selected frequencies vs. nitrate concentration.

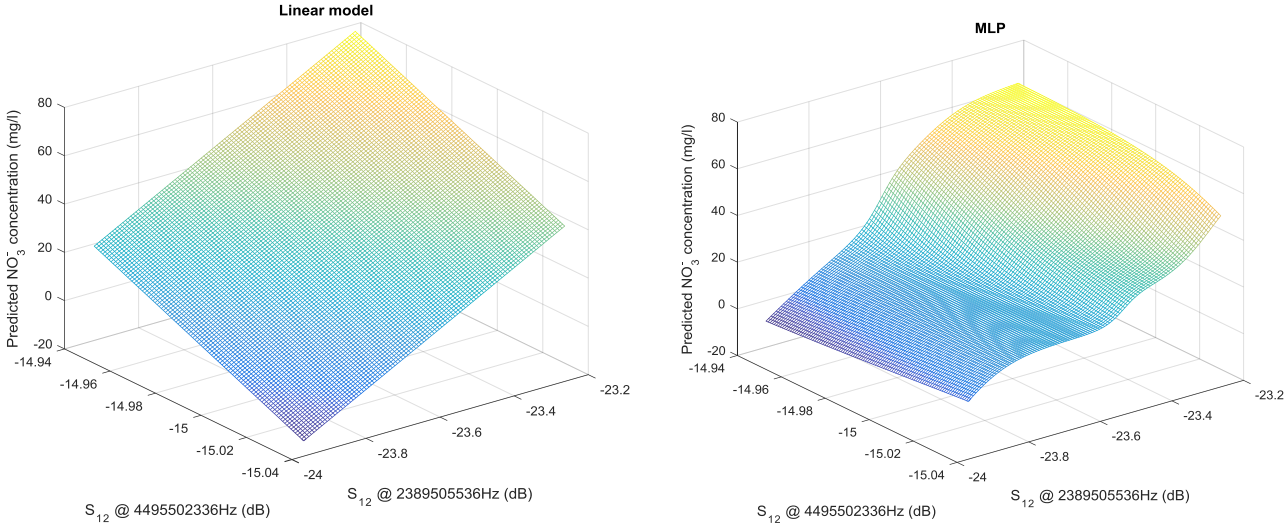


Fig. 7. Surfaces produced by the two regression models. Shows the surface fitted across the domain of the selected features for the linear regression (left) and MLP (right) models. Both show a generally monotonic relationship between the selected features and nitrate concentration, though there is some curvature to the MLP.

1) Least Squares Linear Regression

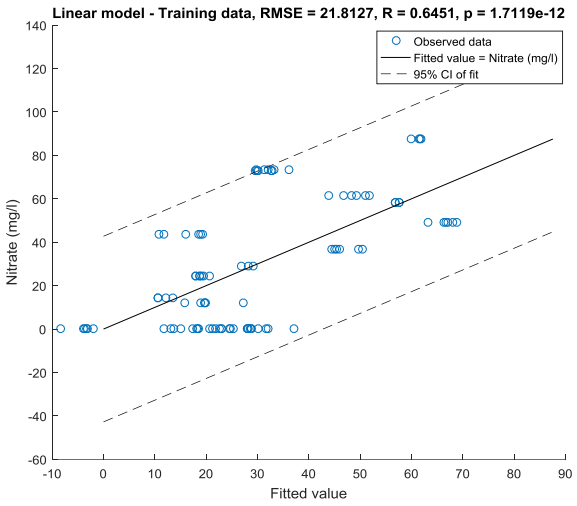
The linear regression model using the two selected features is summarised in table 8. For both features, greater absorption is associated with higher nitrate concentration. The model performed better than chance, but only explained 42% of the variance in nitrate concentration in the training data and 45% of the variance in the nitrate concentration in the test data. Figure 8 shows the performance of the linear regression model on both test and training data.

TABLE VIII
LINEAR REGRESSION MODEL

TERM (S_{12}@)	COEFFICIENT	STD. ERR OF COEFFICIENT	SIGNIFICANCE
Constant	8228.5000	1717.3000	*
2389505536Hz	82.2910	13.8420	*
4495502336Hz	417.4500	119.0300	0.0007

*p<0.0001

- Number of observations: 95, Error degrees of freedom: 92
- Root Mean Squared Error: 22.2
- R-squared: 0.416, adjusted R-Squared 0.403
- F-statistic vs. constant model: 32.8, p-value = 1.78×10^{-11}



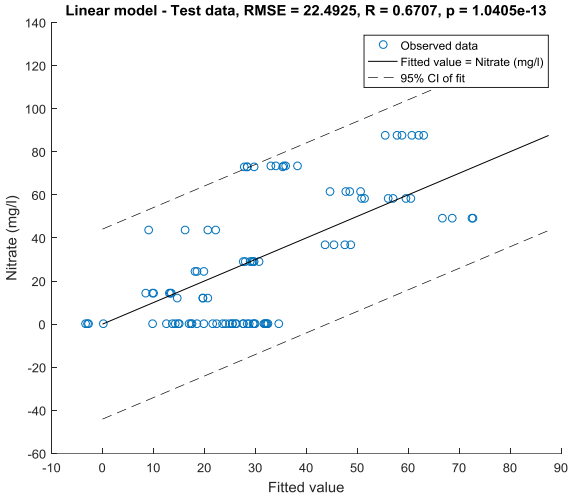
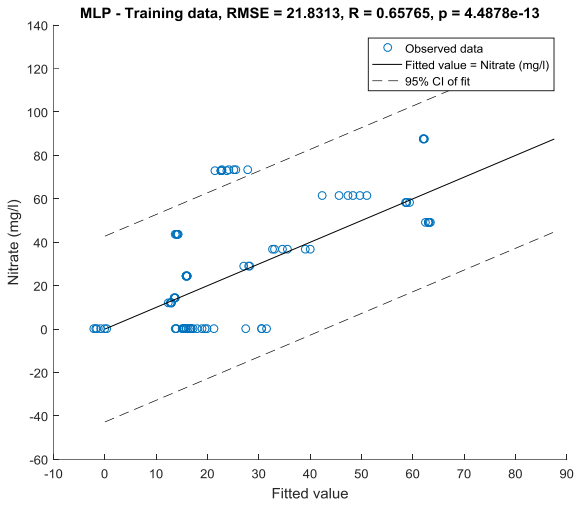


Fig. 8. Scatter plot of fitted value from linear regression vs. Nitrate concentration.



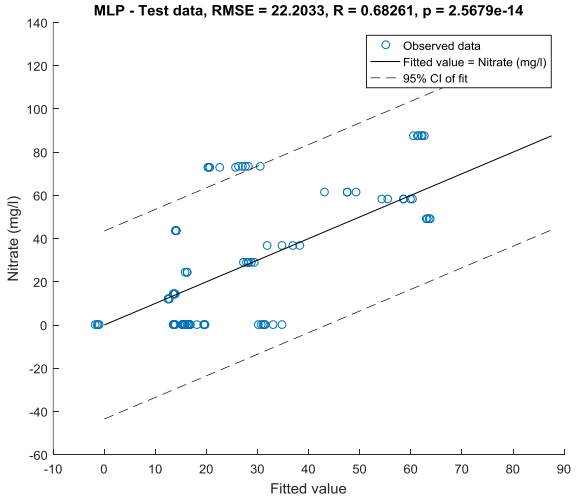


Fig. 9. Scatter plot of fitted value from multi-layer perceptron vs. Nitrate concentration.

2) Multi-layer perceptron

For the MLP, the selected features were first normalised using a Z transform, the parameters used were:

- Number of hidden nodes: 20
- Learning rate: 0.001
- Momentum: 0.100
- Weight decay: 0.005
- Maximum number of epochs: 50,000

The model explained 43% of the variance in nitrate concentration for the training data and 47% of the variance in nitrate concentration for the test data. Figure 9 shows the performance of the multi-layer perceptron for both training and test data.

IV. Discussion

We propose a novel feature selection methodology to identify a minimal set of predictive measurements for the detection and estimation of nitrate concentration in deionised water from microwave absorption spectra. This methodology can contribute to narrowing down the range of investigated frequencies for future experiments and for the identification of frequencies at which measurements could be made using a smaller, less expensive device in future practical applications.

It appears to be relatively straightforward to identify a small number of frequencies to accurately classify the samples on the chosen concentration threshold of 30mg/l. The calibration task is more challenging, with a high level of noise in the data making precise measurements difficult and accuracy poor in comparison to existing ion selective electrode and optical sensor technology.

Using non-linear methods does not seem to improve performance for these data to any great extent. Both the linear (logistic regression) and non-linear (MLP) classification models achieved 96% accuracy at their optimum threshold, the improvement in performance from using an MLP is very small.

This is likely a result of the simplicity of the samples considered (each of the salt solutions contained only a single solute). More complex, real world solutions (such as samples from raw river water) may require the use of non-linear methods as some contaminants may have similar effects on the same part of the EM spectrum, this is outside of the scope of this paper as an early proof of concept, but will inform future work from our group.

REFERENCES

[1] S. Fields, "Global Nitrogen: Cycling out of Control," *Environmental Health Perspectives*, vol. 112, no. 10, pp. A557-A563, 2004.

[2] D. B. Donald, M. J. Bogard, K. Finlay *et al.*, "Phytoplankton-specific response to enrichment of phosphorus-rich surface waters with ammonium, nitrate, and urea," *PloS one*, vol. 8, no. 1, pp. e53277, 2013.

[3] M. H. Ward, T. M. deKok, P. Levallois *et al.*, "Workgroup Report: Drinking-Water Nitrate and Health-Recent Findings and Research Needs," *Environmental Health Perspectives*, vol. 113, no. 11, pp. 1607-1614, 2005.

[4] IARC Working Group on the Evaluation of Carcinogenic Risk to Humans, "Ingested nitrate and nitrite and cyanobacterial peptide toxins," *IARC monographs on the evaluation of carcinogenic risks to humans*, vol. 94, 2010.

[5] N. Espejo-Herrera, K. P. Cantor, N. Malats *et al.*, "Nitrate in drinking water and bladder cancer risk in Spain," *Environmental Research*, vol. 137, pp. 299-307, 2//, 2015.

[6] N. Al-Dasoqi, A. Mason, R. Alkhaddar *et al.*, "Real-time non-destructive microwave sensor for nutrient monitoring in wastewater treatment.", *Journal of Physics: Conference Series*, p. 012043.

[7] O. Korostynska, I. Nakouti, A. Mason *et al.*, "Planar electromagnetic wave sensor for instantaneous assessment of pesticides in water." pp. 942-947.

[8] DEFRA, "Observatory monitoring framework - indicator fact sheet, Indicator DA3: Nitrate and phosphate levels in rivers," DEFRA, ed., 2012.

[9] O. Korostynska, A. Mason, and A. Al-Shamma'a, "Monitoring of nitrates and phosphates in wastewater: current technologies and further challenges," *International journal on smart sensing and intelligent systems*, vol. 5, no. 1, pp. 149-176, 2012.

[10] Sea-Bird Costal. "SUNA V2 UV Nitrate Meter," 4th of November, 2016; http://sea-birdcoastal.com/suna.

[11] H. Liu, and H. Motoda, "Less Is More," *Computational Methods of Feature Selection*, Data Mining and Knowledge Discovery H. Liu and H. Motoda, eds., pp. 3-17: Chapman & Hall/CRC, 2002.

[12] M. Soltani, and M. Omid, "Detection of poultry egg freshness by dielectric spectroscopy and machine learning techniques," *LWT - Food Science and Technology*, vol. 62, no. 2, pp. 1034-1042, 7//, 2015.

[13] I. Nabney, *NETLAB: algorithms for pattern recognition*: Springer Science & Business Media, 2002.

[14] Y. Benjamini, and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the royal statistical society. Series B (Methodological)*, pp. 289-300, 1995.

AUTHOR BIOGRAPHIES

Sean Cashman has a BSc (2015) in Mathematics from Liverpool John Moores University (LJMU). Currently he is working towards a PhD titled "Applications of feature selection and statistical modelling in microwave spectroscopy for the development of a waterway sensor platform" at LJMU.

Dr. Olga Korostynska has a B.Eng (1998) and M.Sc (2000) in Biomedical Engineering from National Technical University of Ukraine (KPI) and PhD (2003) in Electronics and Computer Engineering from the University of Limerick, Ireland. Currently she is a Senior Lecturer in Advanced Sensor Technologies in Faculty of Engineering and Technology, Liverpool John Moores University, UK. Previously she was awarded Marie Curie Fellowship to develop novel microwave sensors for real-time water quality monitoring. She was also an engineer in the National Telecommunication Institute in Ukraine; then a Postdoctoral Researcher in the University of Limerick, working on a number of projects, including those funded by IRCSET, EI and EU FP7 and also was a Lecturer in Physics in Dublin Institute of Technology, Ireland. She has published a book, 14 book chapters and over 200 scientific papers in peer-reviewed journals and conference proceedings.

Paulo Lisboa is Head of Department of Applied Mathematics and Head of the Research Centre for Data Science at Liverpool John Moores University. He has over 250 refereed publications, with awards for citations. His research focus is on rigorous methods to interpret complex models for validation by subject area experts. He chairs the Medical Data Analysis Task Force and co-chairs the Big Data Task Force in the Data Mining Technical Committee of the IEEE Computational Intelligence Society. He has editorial and peer review roles in journals and funding bodies including co-chairmanship of the H2020 Advisory Group for Health, Demographic Change and Wellbeing.

Prof. Andrew Shaw graduated from Liverpool University with a BEng Hons in Electrical and Electronic engineering, MSc (Eng) in Materials science and a Ph.D. titled "the realisation of an industrial free electron laser", completed in 1995. He worked as a postdoctoral researcher at the university for 8 years working on industrial microwave applications for both material processing and sensor technologies. In 2003 he became a lecturer at Liverpool University within the electrical engineering department whilst researching the use of RF communications for subsea communications as part of MoD funded project and later as an EU funded project. In 2005 he joined Liverpool John Moores University, at first with the General Engineering Research Institute as a Senior Lecturer and then as Head of the Electrical Engineering department within the School of Engineering. He is now a Reader in Environmental and Sustainable Technologies, head of the BEST Research Institute and head of the RF and Microwave research group