

Impact case study (REF3b)

Institution: University of York
Unit of Assessment: 10, Mathematical Sciences
Title of case study: Mathematical methods to improve food safety and traceability
1. Summary of the impact <p>Recent food crises show the importance of having effective means of food identification and analysis. Many tests have been developed to monitor food, but analysis of the resulting data is highly problematic. Mathematical techniques developed by Dr Julie Wilson at the University of York allow complex mixtures to be analysed and interpreted. They have enabled the Food and Environment Research Agency (Fera) to maximize the information available from food testing, resulting in improved food safety and authentication worldwide, and underpin the analytical testing services delivered by Fera. The techniques have been incorporated into a bespoke Matlab based solution which is now routinely used by Fera's <i>Chemical and Biochemical Profiling</i> section in the specialist testing services which Fera provides across the food storage and retail, agri-environment and veterinary sectors to over 7,500 customers in over 100 countries. In addition, the techniques are used in Fera's research, supporting around £8M worth of work to develop a wide range of global applications including the determination of disease-related biomarkers, contaminant detection, food traceability and the development of drought- and disease-resistant crop varieties.</p>
2. Underpinning research <p>Julie Wilson is a mathematician who began her career in number theory and, after a Royal Society University Research Fellowship 1999-2007 and an RCUK fellowship in chemoinformatics 2007-2010, has held a joint lectureship between the Departments of Mathematics and of Chemistry at the University of York since 2010. Wilson applies a wide range of mathematical and statistical techniques to a variety of scientific and technological problems, primarily in chemometrics. Wilson has been collaborating with Fera and its precursor the CSL (Central Science Laboratory) for around ten years, developing NMR data processing and chemometric techniques to analyse data from food safety and environmental studies. The research described was carried out at the University of York with data provided by Fera.</p> <p>The use of Nuclear Magnetic Resonance (NMR) methods allows the simultaneous identification of a wide range of small molecules, or metabolites, which provide characteristic "fingerprints" that detail the relative concentrations of compounds present in a sample. Each sample may produce thousands of data points, requiring peak modelling and other data reduction techniques. Chemometrics applies mathematical methods from statistics and pattern recognition to these metabolomic fingerprints, extracting relevant features that enable samples to be classified, anomalies recognized and markers for different biological states identified.</p> <p>Changes in experimental parameters such as temperature, pH and ionic strength result in unwanted shifts in peak position. It is common practice to accommodate small spectral shift changes by integrating the spectral data over regions of equal length. Uniform binning can dissect NMR resonances or assign multiple peaks to the same bin, adding to the variance and making data interpretation difficult. Wilson designed the adaptive binning algorithm [1] to allow variable-length bins, which correspond directly to peaks in the spectra and thus facilitate interpretation. As noise regions are excluded, the method significantly reduces variation within a biological class (for example, disease state) in comparison to fixed-width binning.</p> <p>Although the use of integrated peaks rather than individual data points reduces the number of variables, the search space in metabolomics studies is still prohibitively large for evolutionary computing methods such as Genetic Programming (GP). However, the advantage of GPs over standard multivariate analyses is that they do not involve a transformation of the variables, and thus produce results that are easier to interpret in terms of the underlying chemistry. Wilson therefore developed a two-stage GP algorithm [2] designed specifically for use with (the one-dimensional) ¹H NMR datasets. Computational efficiency is significantly improved by limiting the number of generations in the first stage and only submitting the most discriminatory variables to</p>

Impact case study (REF3b)

the second stage, in which the optimal classification solution is sought.

Efficient feature extraction is also required to allow two-dimensional NMR techniques, such as Heteronuclear Single Quantum Coherence (HSQC) and Heteronuclear Multiple Bond Correlation (HMBC), to be used in the analysis of complex mixtures. **Wilson's** feature extraction method uses a modified Lorentzian function to model peaks in ^1H - ^{13}C HSQC spectra [3] and provides elliptical footprints corresponding to peaks in the spectra. Integrating over these footprints for each spectrum provides a dramatically reduced set of variables that now allows metabolomic analyses to be performed with 2-D spectra.

In the specific case of phase-cycled HSQC, systematic noise needs to be removed before feature extraction. Despite its superior sensitivity, this technique has been limited by the presence of noise ridges, which can mask genuine peaks of low-concentration compounds. **Wilson's** Correlated Trace Denoising (CTD) algorithm [4] takes advantage of the systematic nature of this so-called t1 noise and, unlike other methods for t1 noise removal that have specific pre-requisites, CTD can be used regardless of complexity and the number of peaks in a spectrum, making it suitable for metabolomic studies.

3. References to the research

[1] R. Davis, A. Charlton, S. Oehlschlager and **J. C. Wilson**. Novel feature selection method for genetic programming using 1H NMR data. *Chemom. Intell. Lab. Syst.* **81** (2006) 50-59.

[doi:10.1016/j.chemolab.2005.09.006](https://doi.org/10.1016/j.chemolab.2005.09.006)

*[2] R. A. Davis, A. J. Charlton, J. Godward, S. A. Jones, M. Harrison and **J. C. Wilson**. Adaptive Binning: An Improved Binning Method for Metabolomics Data Using the Undecimated Wavelet Transform. *Chemom. Intell. Lab. Syst.* **85** (2007) 144-154. [doi:10.1016/j.chemolab.2006.08.014](https://doi.org/10.1016/j.chemolab.2006.08.014)

*[3] J. S. McKenzie, A. J. Charlton, J. Donarski, **J. C. Wilson**. Peak Fitting in 2D 1H-13C HSQC NMR Spectra for Metabolomic Studies. *Metabolomics*, **6** (2010) 574-582. [doi: 10.1007/s11306-010-0226-7](https://doi.org/10.1007/s11306-010-0226-7)

*[4] S. Poulding, A. J. Charlton, J. Donarski and **J. C. Wilson**. Removal of t1 Noise from 2D 1H-13C HSQC NMR Spectra by Correlated Trace Denoising, *J. Mag. Res.* **189** (2007) 190-199.

[doi:10.1016/j.jmr.2007.09.004](https://doi.org/10.1016/j.jmr.2007.09.004)

Chemometrics and Intelligent Laboratory Systems publishes 'novel developments in techniques ... characterized by ... statistical and computer methods'. *Metabolomics* is the official journal of the Metabolomics Society, and 'publishes ... the most significant current research'. The *Journal of Magnetic Resonance* publishes 'significant theoretical and experimental results' in 'all aspects of magnetic resonance'. All three are respected international peer-reviewed journals.

All research and algorithm development was carried out at the University of York by **Wilson** and her students Richard Davis, James McKenzie and Simon Poulding; other authors above are Fera scientists, who provided the data and integrated the techniques into Matlab software.

4. Details of the impact

The purpose of the Food and Environment Research Agency (Fera) is "to support and develop a sustainable food chain, a healthy natural environment, and to protect the global community from biological and chemical risks" [5]. It has over 7,500 government and commercial customers and provides services to customers in over 100 countries. As a government agency dealing with food safety and environmental issues, Fera is immediately involved in disease outbreaks, such as foot and mouth disease in cattle, and food contamination threats around the world. As a result of **Wilson's** work on chemometric methods, Fera scientists are now able to apply these underpinning mathematical techniques to a wide range of applications, allowing them to offer a more effective service to their customers and respond more rapidly to such outbreaks and threats.

Wilson's work with Fera began at the initiation of their metabolomics programme, giving them access to state of the art chemometric algorithms. This has resulted in Fera securing projects

Impact case study (REF3b)

totalling a value of £8M to date from Defra, the Food Standards Agency, the European Commission and BBSRC [6]. Applications are wide ranging with examples including the determination of disease-related biomarkers, contaminant detection, food traceability and the development of drought and disease resistant crop varieties. Many applications require a non-targeted approach, which relies on the ability to identify consistent differences between groups (for example between diseased and healthy animals). The new feature extraction methods significantly reduce the within-class variance that can mask these differences, thus revealing biochemical signals that might otherwise have been missed.

As part of their programme Fera have invested in the development of *Metabolab*, a bespoke, modular Matlab based software package [7], which incorporates **Wilson's** algorithms and also allows them to quickly and flexibly implement new algorithms as they emerge from research. *Metabolab* allows the techniques to be used by non-experts, and the software is now used routinely in the Chemical and Biochemical Profiling section at Fera for the processing of metabolomic datasets. Designed to efficiently process the extremely large data sets typically required to analyse two-dimensional spectra, this software gives Fera the competitive advantage of being able to use the new algorithms make use of highly resolved, and therefore more informative, 2-D NMR experiments for routine metabolomics studies.

Disease-related Biomarkers:

Using a simple blood test the chemometric techniques can be used to identify biomarkers to detect diseased, and therefore infectious, animals before physical signs are apparent. This is used to identify and distinguish many high profile diseases such as BSE in cattle, TB in badgers, foot and mouth disease, and various plant diseases. Dr Adrian Charlton, Head of Chemical and Biochemical Profiling at Fera, leads a research team that provides novel solutions to problems of food contamination and authentication. He says "Of particular note was the contribution that the adaptive binning and 2 stage GP algorithms made to the delivery of a £1.7M project for the Food Standards Agency (FSA), investigating the determination of novel biomarkers of BSE and scrapie" [6]. Bovine spongiform encephalopathy (BSE), commonly known as mad cow disease, is characterized by spongy degeneration of the brain in cattle with a variant in humans called Creutzfeldt–Jakob disease (CJD). As part of this project, a workshop was hosted by the University of York to advise the project team, with scientists from Fera, the Veterinary Laboratories Agency (VLA) and the Institute for Grasslands and Environmental Research (IGER), on the correct implementation of **Wilson's** novel approaches as well as other multivariate analysis techniques for application into other areas.

Food Traceability:

At European level, the genetic programming approaches developed by **Wilson** were used to underpin a €15M FP6 project (TRACE) to 'provide consumers with added confidence in the authenticity of European food through complete traceability along entire fork-to-farm food chains' [8]. The project was coordinated by Fera and utilized a range of analytic tools that make use of the computational techniques developed by **Wilson's** team at the University of York [9]. The methods enable molecular fingerprinting to be used to determine the origin of food. Products to which TRACE's methods have been applied include European mineral water, cereals, honey, meat and chicken [5,8,10]. For example, Corsican honey is the only one produced in France that carries the prestigious Appellation of Controlled Origin designation (AOC label). As a result of the new methods, it is now possible to use a number of chemical markers to make fine geographical distinctions between different origins and content of honey, with widely differing prices [10].

The research has been featured repeatedly in *New Scientist* [11] and the results from the TRACE project have been disseminated in over 200 presentations and workshops worldwide to an enormous range of participants from industry [8]. In 2012 **Wilson** was an invited speaker at *New developments in food science: realising the potential of 'omics' technologies*, the 13th annual joint symposium in 2012 of FERA and the US Joint Institute for Food Safety and Applied Nutrition. The meeting's sponsors included Agilent, Thermo Scientific, AB Sciex and Waters.

Contaminant Detection:

The methods have also enabled significant improvements in procedures for the detection of

Impact case study (REF3b)

contaminants. In this case the differences from what is considered normal need to be recognized, as any extraneous variance could result in false negatives. Some toxins can be lethal at extremely low concentrations. The new techniques allow compounds that may only occur at lower concentration, and which may have been obscured in variance-based multivariate analyses, to be identified (e.g. melamine in milk and infant formula). Furthermore, the variables relate to peaks in the spectra rather than individual data points, thereby making it easier to interpret the results and thus identify the chemical compounds responsible.

Disease Resistant Crops:

Most recently Fera have won a €3M, 5 year project from the European Commission (ABSTRESS) [12], which is further exploiting and continuing to develop the technologies arising from the collaboration with **Wilson**. The project aims to identify the processes in plant biochemistry associated with the way drought and disease combine to make matters much worse than either alone. Building on the information available from chemometric techniques researchers are developing novel principles and techniques that can be used to significantly reduce the time taken to produce new crop varieties in support of commercial plant breeding. This should produce new crop varieties that are more able to withstand the challenges commonly associated with climate change, such as extreme weather and changing incidence of pests and diseases. Although the University of York is not a partner in ABSTRESS, Fera have sponsored an EngD studentship, co-supervised by **Wilson**, on the integration of data from the different –omics technologies being used in the project.

In addition to the EngD, the collaboration with Fera has led to funding for two PhD students: Richard Davis held an EPSRC CASE studentship with Fera (then CSL) and James McKenzie had Fera seedcorn funding.

5. Sources to corroborate the impact

[5] <http://www.fera.defra.gov.uk/> (accessed 15/10/2012). Corroborates claim of Fera's purpose.

[6] E-mail provided by Head of Chemical and Biochemical Profiling at FERA. Corroborates the value to Fera of project funding and the contribution of the methods developed by **Wilson** in all projects mentioned.

[7] *Metabolab* software. Corroborates the claim that adaptive binning and the two-stage GP have been incorporated into the software.

[8] TRACE events archive: <http://trace.eu.org/archive/events.php> (accessed 24/09/2013). Corroborates the extent to which TRACE results have been disseminated and the level of industrial contacts.

[9] Fera Annual Review 2011-12, p10 <http://fera.co.uk/news/documents/feraAnnualReview1112.pdf> Corroborates Fera's involvement in food safety and traceability studies.

[10] J. A. Donarski, S. A. Jones, A. J. Charlton. Application of Cryoprobe 1H Nuclear Magnetic Resonance Spectroscopy and Multivariate Analysis for the Verification of Corsican Honey. *J. Agr. Food Chemistry* 56 (2008) 5451; J. A. Donarski, S. A. Jones, M. Harrison, M. Driffield, A. J. Charlton. Identification of botanical biomarkers found in Corsican honey. *Food Chemistry* 118 (2010) 987-994. Corroborates use of methods in published research, and application to Corsican honey.

[11] K. Ravilious, "Buyer beware; When you shell out for a premium food how do you know you're getting what you pay for?", *New Scientist*, 11th November 2006, p40-43; M. Inman, "Fifty ways to interrogate your dinner; To check the credentials of the food you would like to eat, just take your cellphone to the supermarket and snap the barcode", *New Scientist*, 13th June 2009, p18-19. Corroborates the reporting of food safety and traceability issues.

[12] <http://www.abstress.eu/> (accessed 15/10/2012). Corroborates Fera's coordination of the ABSTRESS project.