Evolutionary computing for the characterisation and quantification of microbial systems from hyperspectral data

Royston Goodacre* and Douglas B. Kell.

Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, Wales, UK
*Telephone: +44 (0)1970 621947 Telefax: +44 (0)1970 621947 E-mail: rrg@aber.ac.uk

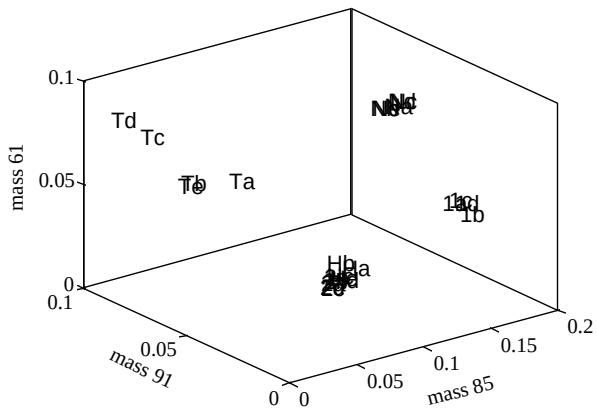
The whole-organism fingerprinting methods of pyrolysis mass spectrometry, diffuse reflectance-absorbance Fourier transform infrared spectroscopy, and Raman microscopy can be used for the characterisation of microorganisms and for the prediction of their chemical properties[1-3]. However, this is only possible in combination with modern supervised machine learning techniques, such as artificial neural networks (ANNs) and partial least squares (PLS).

Nevertheless, the interpretation of the calibration models from ANNs and PLS is often very difficult, and the information in terms of which masses in the mass spectrum or vibrations in infrared or Raman spectra are important are not readily available. ANNs and PLS are often perceived as 'black box' approaches to modelling spectra, and to allow the deconvolution of complex hyperspectral data it is necessary to develop a system that itself produces 'rules' that are readily comprehensible. Genetic programming (GP) is an evolutionary technique which uses the concepts of Darwinian selection to generate and optimize a desired computational function or mathematical expression[4]. GPs can also be used to classify biological systems but they evolved function trees (or mathematical rules) enabling the deconvolution of spectra.

Three examples will be presented…

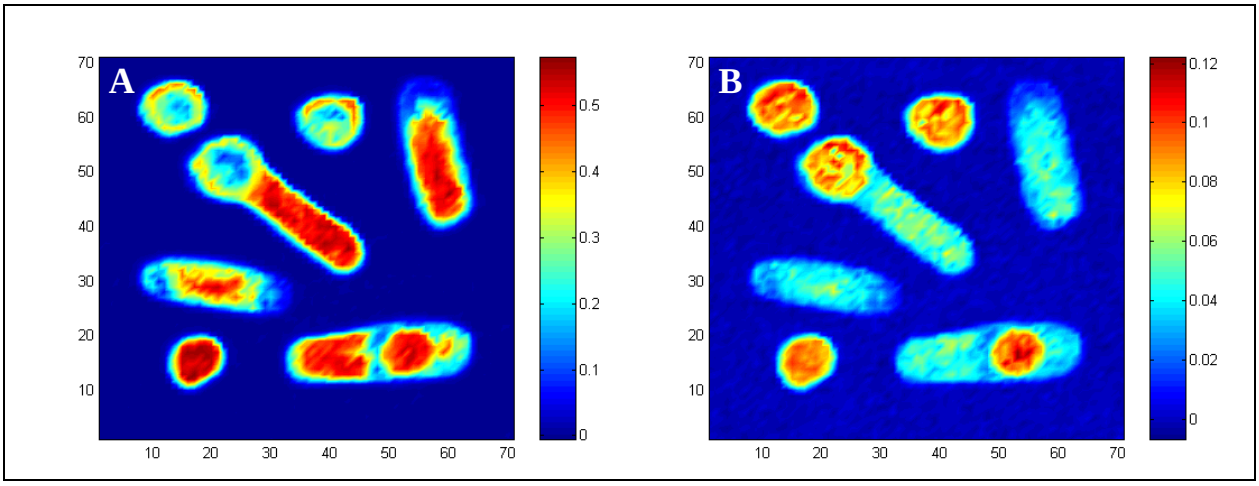
1. The identification of a group of bacteria which have been implicated in periodontitis, endodontic infections and dentoalveolar abscesses[5]:

Studies on the classification of oral bacteria belonging to the *Eubacterium* genus showed that specific masses from the breakdown products of polysaccharide (m/z 58), lipids (m/z 61), pentose anhydrosugar (m/z 85) and proteins (m/z 91) could be used to discriminate hospital oral abscess isolates successfully. Moreover, when three of these are plotted (see opposite) this was sufficient to classify these four bacteria.

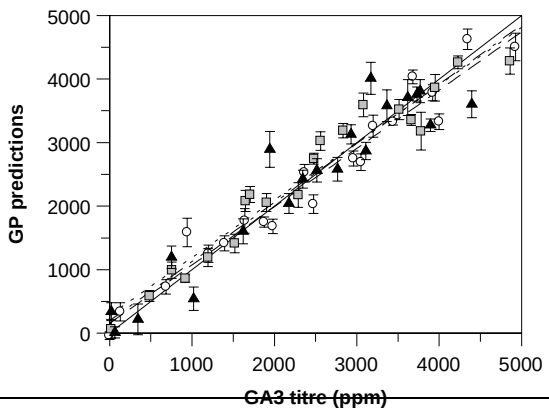


2. The detection of the dipicolinic acid biomarker in *Bacillus* spores[6]:

GP was used to determine the physiological state of *Bacillus* spp. (vegetative cells or spores) correctly. For PyMS it was found that m/z 105 was characteristic and is a pyridine ketonium ion ($C_6H_3ON^+$) obtained from the pyrolysis of dipicolinic acid (pyridine-2,6-dicarboxylic acid; DPA), a substance found in spores but not in vegetative cells. In addition, a pyridine ring vibration at 1447 - 1439 cm^{-1} from DPA was found to be highly characteristic of spores in FT-IR analysis. Thus, although the original datasets recorded hundreds of spectral variables from whole cells simultaneously, a simple biomarker can be used for the rapid and unequivocal detection of spores of these organisms. Shown below are chemical images of (A) amide I vibration at 1666 cm^{-1} and (B) pyridine vibration at 1443cm^{-1}. Vegetative and sporulated biomass from a *B. cereus* strain was applied to the surface of a 7cm by 7cm metal plate (~200 µg/cm^2, dry weight). Data were acquired at a resolution of 1 mm (therefore these maps are 71 by 71 pixels; 5041 spectra).



3. Monitoring complex industrial bioprocesses producing gibberellic acid[6]:



FT-IR and Raman spectroscopies were used in a quantitative fashion to analyse a diverse range of unprocessed fed-batch fermentations broths containing the fungus *Gibberella fujikuroi* producing the natural product gibberellic acid. To gain quantitative information in terms of the gibberellic acid titre and to to decipher the spectra GPs were employed. The results from the FT-IR and Raman (see opposite) studies showed that the models formed could be used to quantify the product in these industrial fermentations slightly better PLSR or ANNs, and were based on spectral features derived from the gibberellic acid 3 molecule itself.

[1]Naumann, D., Helm, D., Labischinski, H. & Giesbrecht, P. (1991). The characterization of microorganisms by Fourier-transform infrared spectroscopy (FT-IR). In *Modern techniques for rapid microbiological analysis*, pp. 43-96. Edited by W. H. Nelson. New York: VCH Publishers.

[2]Goodacre, R., Timmins, É. M., Burton, R., Kaderbhai, N., Woodward, A. M., Kell, D. B. & Rooney, P. J. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral, whole organism fingerprinting and artificial neural networks. *Microbiology* **144**, 1157-1170.

[3]Maquelin, K., Choo-Smith, L.-P., van Vreeswijk, T., Endtz, H. P., Smith, B., Bennett, R., Bruining, H. A. & Puppels, G. J. (2000). Raman spectroscopic method for identification of clinically relevant microorganisms growing on solid culture medium. *Analytical Chemistry* **72**, 12-19.

[4]Koza, J. R. (1992). *Genetic Programming: On the Programming of computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

[5]Taylor, J., Goodacre, R., Wade, W. G., Rowland, J. J. & Kell, D. B. (1998). The deconvolution of pyrolysis mass spectra using genetic programming: application to the identification of some *Eubacterium* species. *FEMS Microbiology Letters* **160**, 237-246.

[6]Goodacre, R., Shann, B., Gilbert, R. J., Timmins, É. M., McGovern, A. C., Alsberg, B. K., Kell, D. B. & Logan, N. A. (2000). The detection of the dipicolinic acid biomarker in *Bacillus* spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Analytical Chemistry* **72**, 119-127.

[7]McGovern, A. C., Broadhurst, D., Taylor, J., Gilbert, R. J., Kaderbhai, N., Winson, M. K., Small, D. A. P., Rowland, J. J., Kell, D. B. & Goodacre, R. (2000). Monitoring of complex industrial bioprocesses for metabolite concentrations using modern spectroscopies and machine learning: application to gibberellic acid production. *Biotechnology & Bioengineering*, submitted.