BAYESIAN WAVELENGTH SELECTION IN MULTICOMPONENT ANALYSIS

P. J. Brown[*], M. Vannucci[†] and T. Fearn[‡]

Abstract

Multicomponent analysis attempts to simultaneously predict the ingredients of a mixture. If near-infrared spectroscopy provides the predictor variables then modern scanning instruments may offer absorbances at a very large number of wavelengths. Although it is perfectly possible to use whole spectrum methods (eg. PLS, ridge and principal component regression), for a number of reasons it is often desirable to select a small number of wavelengths from which to construct the prediction equation relating absorbances to composition. This paper considers wavelength selection with a view to using the chosen wavelengths to simultaneously predict the compositional ingredients, and is therefore an example of multivariate variable selection. It adopts a binary exclusion/inclusion latent variable formulation of selection and uses a Bayesian approach. Problems of search of the vast number of possible selected models are overcome by a Metropolis sampling technique, using swapping, deleting and adding moves inspired by Genetic Algorithms.

KEYWORDS: Multivariate Regression; Bayesian Wavelength Selection; Metropolis Algorithm; NIR Spectroscopy; Multicomponent Analysis; Selection Bias; Model Averaging.

[*]University of Kent at Canterbury, Institute of Mathematics and Statistics, Cornwallis Building, Canterbury, CT2 7NF, UK. FAX 01227-827932

[†]University of Kent at Canterbury

[‡]University College London, UK