Bayesian modelling of NIR spectra to quantify metals in agricultural soils

J. Morales[1], A.M. Mayoral[1], X. Barber[1], L. Galvez-Sola[2], C. Paredes-Gil[2], M.A. Bustamante[2], R. Moral[2].

[1] *Applied Statistical Unit, Operations Research Center, Miguel Hernández University*
[2] *Department of Agrochemistry and Environment, Miguel Hernández University*

Modern technology and instrumentation routinely delivers rich multidimensional data for which new statistical tools are needed. Such a case is the spectral decomposition of soil samples, pretended to be used in order to predict chemical properties of the soil, so avoiding complicated cost-time analysis. We count on agronomic data samples in which both chemical analysis and spectral decomposition have been measured. The goal is to chemically characterize soils by reducing dimensionality at their spectrum, which consists on NIR data at 2.200 different wavelengths. Also variation in time is considered for each sample, in order to predict variation in the chemical parameters and their associated NIR.
We resolve the problem from a regression setting, using P-splines. A Bayesian analysis will allow for a more flexible modeling that resolve both the dimension reduction problem at one time point (determination of the number of wavelengths knots to use for prediction) and also the dimension reduction at the time scale when inferring on the chemical variation of the soil (determination of the number of time points needed for prediction).