

Posters

Geostatistical indexes for quantifying spatial information on hyperspectral imaging: an application for the evaluation of regression models.

Ana Herrero-Langreo^{1*}; Nathalie Gorretta²; Bruno Tisseyre²; Aoife Gowen¹; Jun-Li Xu¹; Gilles Chaix^{3,4}; Jean-Michel Roger²

1. UCD School of Biosystems and Food Engineering, University College of Dublin (UCD), Belfield, Dublin 4 Ireland

2. ITAP, Univ Montpellier, Irstea, Montpellier SupAgro, Montpellier, France

3. CIRAD, UMR AGAP, Montpellier, France

4. ESALQ, Departamento de Ciências Florestais, Universidade de São Paulo, São Paulo, Brazil.

* ana.herrero-langreo@ucd.ie

Keywords: Hyperspectral, Modelling, NIR, Geostatistics, Calibration, Image, Spatial, Overfitting, PLS

Hyperspectral (HS) images have the characteristic of containing both spectral and spatial information of a sample. Typically, spectral information can be related with chemical and physical properties through multivariate regression models. The application of these models onto HS images results in prediction maps, which provide an estimation of the modelled chemical information for each pixel of the image. This approach has wide applications in food processing industries for online monitoring of product quality and process control. One of the main difficulties derived from an imaging set up, is that the size of the pixels is usually much smaller than the area required to obtain a wet chemical reference. This means that, as opposed to point spectroscopy, the performance of the estimations cannot be evaluated by directly comparing observed and estimated values for each pixel. Moreover, the selection of regression model parameters, such as the number of latent variables (LV) in a partial least squares (PLS) model, cannot be assessed on a pixel basis either. Nonetheless, compared to point spectroscopy, HS imaging does provide information on the spatial distribution of the predicted values.

The objective of this work is to propose a quantitative approach to use spatial information of prediction maps for supporting the evaluation of regression models applied to HS images. This approach is based on the use of geostatistical indexes, which allow decomposing the total variance of the prediction maps into two components: non spatially structured and spatially structured variance, represented respectively by the nugget effect (C0) and the partial sill (C1). This strategy was tested in a simulated dataset and two real case studies. Geostatistical indices of the prediction maps were compared to model performance metrics for PLS models with increasing number of LV. As a result, this work establishes a connection between linear regression model performance estimates and the spatial decomposition of variance in prediction maps, when the ground truth to estimate is spatially structured.

The presented study [1], allowed to evaluate HS imaging models, not only from average estimations which can be compared to reference values, but also from the spatial structure of prediction maps. This approach does not require ground truth values and could be used as a source of information for supporting the choice of optimum calibration options, such as the number of LV, or the pre-treatments, complementing the traditional visual inspection of prediction maps with quantitative and objective metrics.

Further works could explore the effect of spatial structure at different scales in the data. This approach could also be applied to evaluate variations of other parameters in the model, such as different methods for spectral pretreatments.

[1] Herrero-Langreo A., Gorretta N., Tisseyre B., Gowen A., Jun-Li Xu, Chaix G., Roger JM. (2019), Using Spatial Information for Evaluating the Quality of Prediction Maps from Hyperspectral Images: A Geostatistical Approach. *Analytica Chimica Acta*. 1077 116–128. doi:10.1016/j.aca.2019.05.067.