

Assessment of Agricultural Practices on Volcanic Ash Soils Assisted by Automated Interpretation of Mid-Infrared Spectra and Partial Least Squares Multivariate Statistical Approach

Z. Hernández, G. Almendros, J. Sanz, J.P. Pérez-Trujillo,
J.A. González-Pérez, and F.J. González-Vila

Abstract An attempt to automated the interpretation of mid-infrared (MIR) spectral profiles of humic acids (HAs) from volcanic ash soils was carried out with the aid of plots displaying the valuable importance for projection (VIP) scores computed by partial least squares (PLS) regression. A set of about 200 dependent variables from whole soils and HAs was examined in 30 uncultivated and agricultural soils with the PLS-VIP method in order extract objective information about the extent to which the MIR intensities may be useful to forecast quantitative and qualitative agroecological and biogeochemical factors, as well as to ascribe individual bands to structural or functional HA variables which were determined by independent multianalytical approaches.

Keywords Soil organic matter • Humic acids • Mid-infrared spectra • Partial least squares

Z. Hernández • J.P. Pérez-Trujillo
Department of Analytical Chemistry, University of La Laguna, 38206 La Laguna, Spain

G. Almendros
National Museum of Natural Sciences (Spanish National Research Council, CSIC),
28006 Madrid, Spain

J. Sanz
Institute of General Organic Chemistry, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

J.A. González-Pérez (✉) • F.J. González-Vila
IRNAS-CSIC, Avda. Reina Mercedes, 10, 41012 Seville, Spain
e-mail: jag@irnase.csic.es

Introduction

It is frequently considered that the structure of soil organic matter (OM), in particular the humic acids (HAs), are highly responsive to land-use change, soil agricultural practices, or the variability of local soil-forming factors, mainly vegetation and geological substrate.

In particular, the wide mineralogical variability provided by volcanic ash soils in the Canary Islands (Spain), between 27°60' and 28°35' N and 16°59' and 16°07' W, with a maximum altitude of 3.718 m, results in singular agroecosystems under homogeneous climatic conditions where humification processes may dramatically vary within few meters as a consequence of the massive transport of allochthonous (natural or anthropogenic) pyroclastic soil-forming minerals.

In these Anthrosols, short-range minerals are considered to display specific activity as regards the insolubilization of humic substances and their precursors and play an active role in physical encapsulation of particulate OM and in the formation of stable aggregates. These are formed with recalcitrant C forms that stick to mineral surfaces or flocculate short-range aluminum oxides, leading to “hybrid” biodegradation-resistant OM forms. In this scenario, it is presumable that soil mineralogical features play an outstanding role in the qualitative and qualitative variability of soil OM which, in turn, could be assessed by suitable routine spectroscopic approaches such as IR spectroscopy.

The exploratory analysis of soil characteristics which are reflected in the mid-infrared (MIR) spectral profile of the whole soil and isolated HAs, could be examined by taking advantage of the valuable importance for projection (VIP) values calculated during the partial least squares (PLS) model building. This will then represent a combined measure of the extent to which a variable contributes to describing both dependent (X) and independent (Y) data sets.

Consequently, PLS-VIP method would provide an objective criterion to identify MIR band patterns more responsive to soil dependent properties. When such properties correspond to structural characteristics of the HAs (e.g., aromaticity, N-content, oxygen-containing functional groups, yields of lignin-derived methoxyphenols after degradation. . .), the PLS-VIP approach could also be used to assist the interpretation of the MIR spectra.

PLS and its alternatives. From a chemometric viewpoint, among the most interesting possibilities in applied HA research, would be to relate structural features of HAs with structural or functional descriptors at higher organizational levels of the agroecosystem.

In general, and when the information is contained in large arrays of spectroscopic data, this analysis should not be approached by simple stepwise regression methods, because of its poor performance when collinearity exists among variables.

In these cases, it is considered that PLS regression is particularly suitable when the matrix of predictors includes more variables than observations – such in the case of spectral arrays – and when there is large multicollinearity or internal redundancy

in the matrix of predictors. In line with this, PLS is often recommended in an early stage of theoretical developments in order to test and validate exploratory models.

Assuming the above consideration, in this study, PLS-VIP approach was used to infer valuable information about three complementary features as follows:

1. To examine the potential of PLS for robust prediction dependent variables in whole soil or HAs, exclusively based in the information extracted from MIR spectral profiles.
2. To discuss the PLS-VIP plots for the different HA characteristics independently determined to be used in the selection of relevant MIR peaks (simultaneously ascribed to more or less specific functional groups) which really have an effect on the response.
3. To discuss the meaning of PLS-VIP plots in relation with the different whole soil characteristics, which in turn would explain shifts in HAs structural features as affected by the natural variability of different soil types, or induced by local management practices.

Material and Methods

Experimental design. In this study, a series of routine analytical variables of agroecological interest were determined in 30 soil samples mainly from vineyards and zones under natural vegetation. More than 200 soil physical and chemical characteristics were determined among hydrophysical, agrochemical, and biological properties, including concentration of the different soil OM fractions. The majority of structural information on soil OM was obtained from the isolated HAs fractions by visible spectroscopy and ^{13}C solid state nuclear magnetic resonance (^{13}C CPMAS NMR). In addition, data from the molecular characterization of HAs was obtained using analytical pyrolysis (Py-GC/MS).

Of the above-indicated set of variables, special attention was paid to surrogates of the performance of agroecological management practices, i.e., variables informing on the quantity and quality of soil OM forms. Concerning quantity, the interest was centered on the potential for soil C sequestration, reflected by total soil C and OM biodegradability as estimated from respirometric laboratory experiments. Concerning quality, most of the variables determined are informative about the extent to which plant and microbial OM was either preserved or transformed into humic substances, i.e., data from free particulate OM, HA concentration, and selected HA characteristics indicating maturity or diagenetic alteration, i.e., aromaticity, optical density, percentage of aromatic, *O*-alkyl and alkyl-amide carbons by ^{13}C NMR, optical density and atomic ratios, and pyrolysis yields of constituents with known precursors such as lignin (e.g., methoxyphenols) or other recalcitrant structural domains (e.g., alkylbenzenes and polyaromatic fragments).

PLS and PLS-VIP plots. PLS regression was carried out by simultaneously processing information from the individual variables in the above matrix with

soil and HA properties versus the spectral data matrix, i.e., the digital MIR spectra arrays. The PLS was performed with the ParLeS software (Viscarra Rossel 2008). The MIR spectral arrays were adjusted to 640 data points and transformed into MIR second derivative, which lead to a resolution-enhancement effect of interest in intrinsically broad spectral profiles typical in macromolecular heterogeneous material such as HAs. The matrix with spectral intensities was sequentially processed with supplementary soil dependent variables from the matrix of up to 200 soil and HA characteristics independently analyzed by standard analytical procedures (Hernández and Almendros, 2012). The VIP plots were interpreted only when the soil dependent variable processed in the successive program runs (e.g., amorphous oxides, pH, total organic carbon, aggregate stability, nitrogen. . .) was significantly ($P < 0.05$) predicted from MIR information.

Results and Discussion

In the first step, MIR spectroscopy and PLS analysis is used to predict the values of a series of characteristics of agroecological interest in soil and isolated HAs. In this step, routine soil variables, such as soil organic C, total N, amorphous material contents, heavy metals, and texture, were successfully forecasted ($P < 0.05$) with the PLS model.

Preliminary results show that the intensity of the $2,920\text{-cm}^{-1}$ MIR band shows high correlation with total OM content, regardless of the OM quality. The application of this spectroscopic technique also leads to valid forecasting of the amount of amorphous oxides.

In the present study, the different VIP plots straightforwardly illustrate the MIR peaks – or wavelength ranges – more important for prediction of the variables with agroecological or environmental relevance. All the variables resulting in a VIP score >0.8 were considered to be relevant in the model, and require interpretation.

References

- Hernández, Z., and G. Almendros. 2012. Biogeochemical factors related with organic matter degradation and C storage in agricultural volcanic ash soils. *Soil Biology and Biochemistry* 44: 130–142.
- Viscarra Rossel, R.A. 2008. ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems* 90: 72–83.