Forecasting soil agrochemical properties from ATR-MIR spectroscopy and partial least-squares regression analysis

Hernández, Z.[1]*, Recio-Vazquez, L.[2], Pérez-Trujillo, J.P.[1], Sanz, J.[3], Álvarez, A.[4], Carral, P.[4] Almendros, G.[2]

[1] Dpto. de Química Analítica, Universidad de La Laguna, 38206 La Laguna, S/C Tenerife, Spain
[2] MNCN-CSIC, Madrid. [3] IQOG-CSIC, Madrid. [4] UAM, Cantoblanco, Madrid, Spain. *E-mail: zulimar@mncn.csic.es

A valid characterization of agroecosystems requires explaining different aspects of its functioning through the determination of a large number of analytical variables. Nowadays, the use of spectroscopic techniques leads to a rapid and efficient field monitoring of anthropogenic activities on soil. In general non-destructive techniques are faster and less expensive than conventional wet chemical analytical methods, requiring small amounts of sample or no sample pre-processing [1].

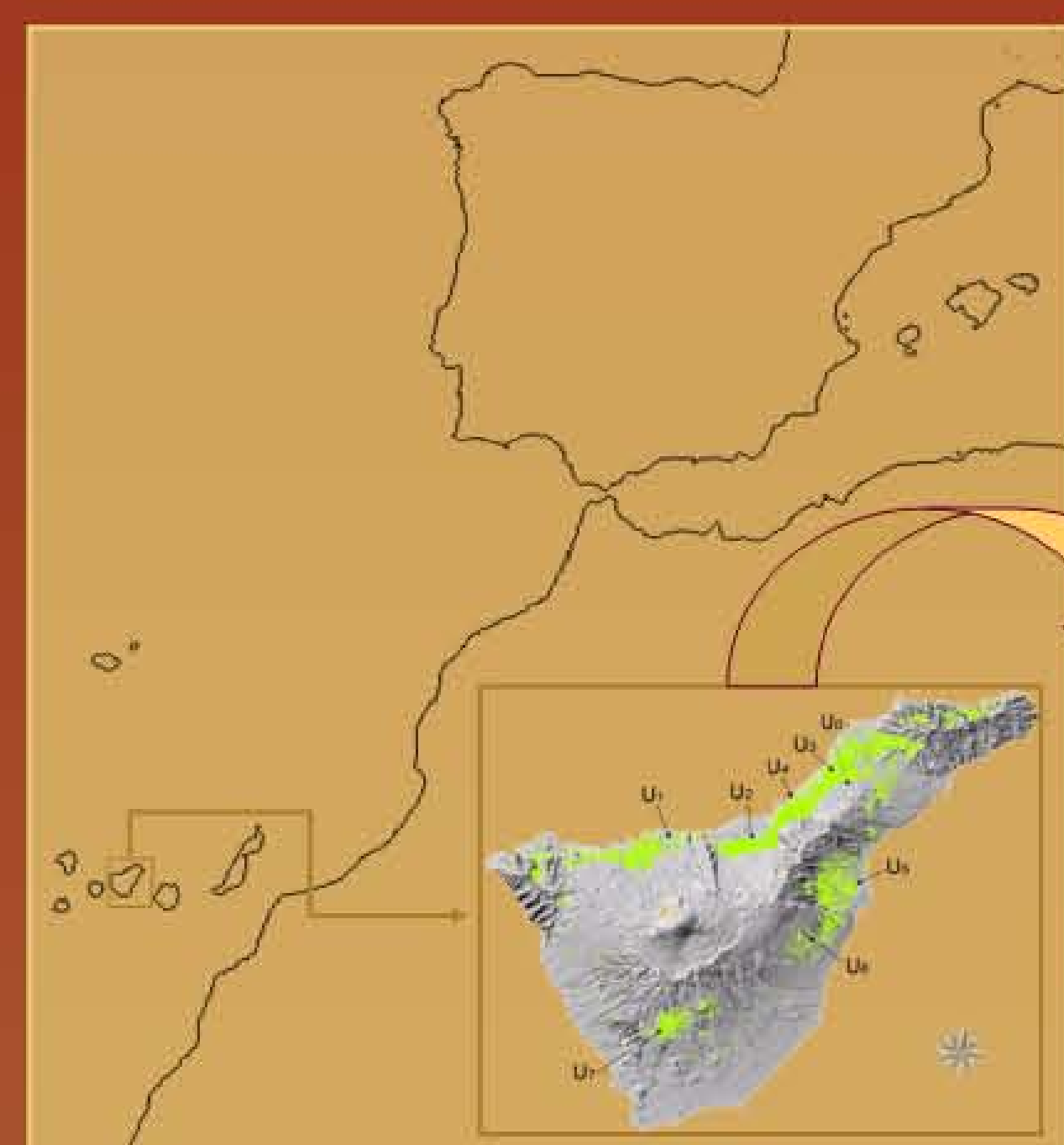## 1. Experimental design, field sampling and preparation of soil samples:

Before sampling, a series of study areas were chosen considering the spatial variability and soil type. Soil samples were collected from the uppermost 20 cm, air-dried, homogenised to 2 mm and finely grounded with an agate mortar.

Figure 1. Above: Sampling plots in Tenerife Island (Spain). Right: Soil profiles, Andosol (U1) and Entisol (U5).

2. Routine analytical techniques applied to soil samples:

Conventional soil physical and chemical variables, such as organic matter (SOM), total N, amorphous materials, exchangeable bases, amorphous oxides or microbial activity were determined.

Figure 2. Right: Determination of SOM by wet oxidation method with $K_2Cr_2O_7$.

3. Spectroscopic analysis of soil samples:

Infrared spectra of the soil samples were obtained using a Fourier transform mid-infrared spectrophotometer with attenuated total reflectance (FT-IR/ATR) between 4000 to 600 cm^{-1}.
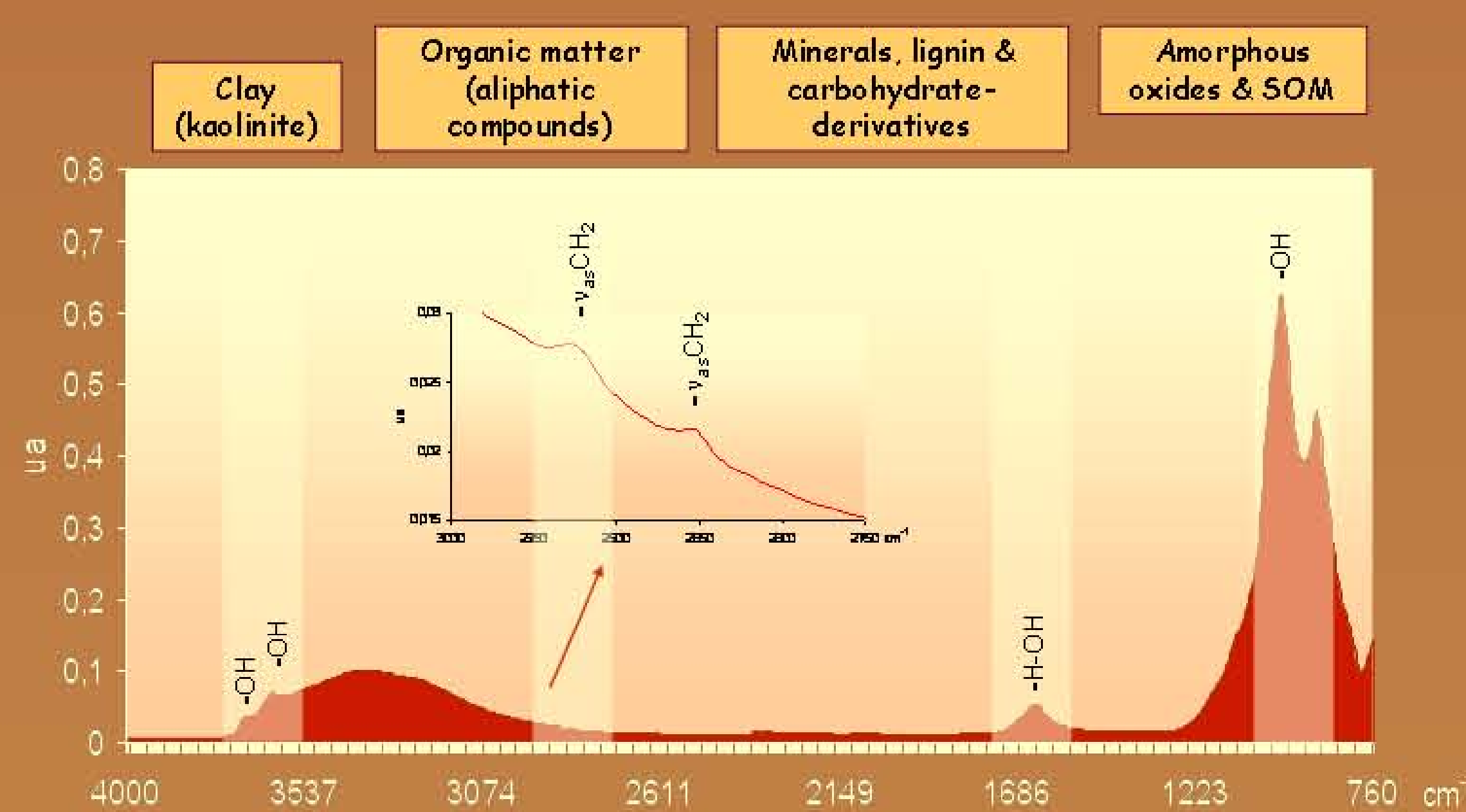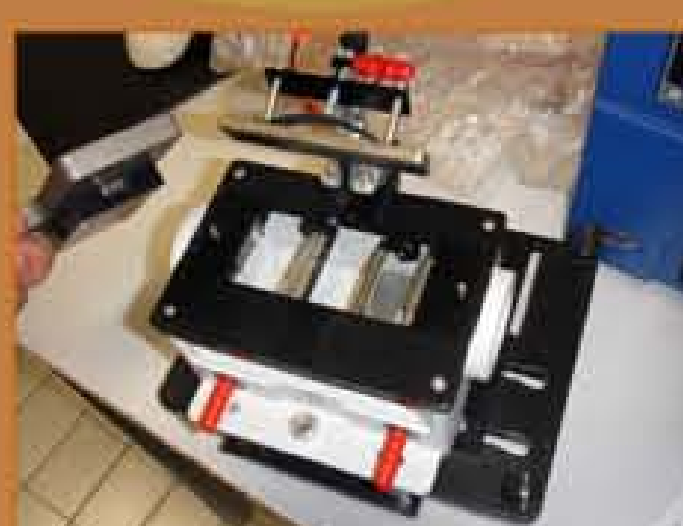
Sample compartment for ATR

Figure 3: FT-IR/ATR spectrum of a soil sample.

4. Application of partial least-squares (PLS) models:

In order to determine the extent to which spectral information may be used to predict physico-chemical information, the IR spectra data matrix was analyzed by PLS regression, using the application ParLeS [2].

CONCLUSIONS

After previous calibration using a spectral and analytical database of up to 200 soil samples, the application of FT-IR/ATR-PLS strategy lead to valid forecasting of important quantitative characteristics of volcanic ash soils, such as the amount of SOM, amorphous oxides, concentration of fungal pigments, etc in a short-time and cost-effective way. Such an approach is especially helpful to optimize field studies and to decide previously the number of samples to be collected during the sampling campaigns.

PLS in 6 steps

❖ Creating a data matrix where IR spectral data (640 var.) are independent variables and one analytical data is the dependent variable (17 vars.):

| Soil samples	Analytical variable pH value	IR spectra			
		4000	3996	3990	600 cm^{-1}
1	6.6	0.6376	0.6435	0.6462	...
2	6.6	0.6397	0.6428	0.6517	...
3	6.2	0.6870	0.6452	0.6463	...
4	5.3	0.6394	0.6453	0.6477	...
n	X_{i1}	Y_{i1}	Y_{j1}	Y_k	Y_{nm}

❖ Performing spectral data transformation using the second derivative:
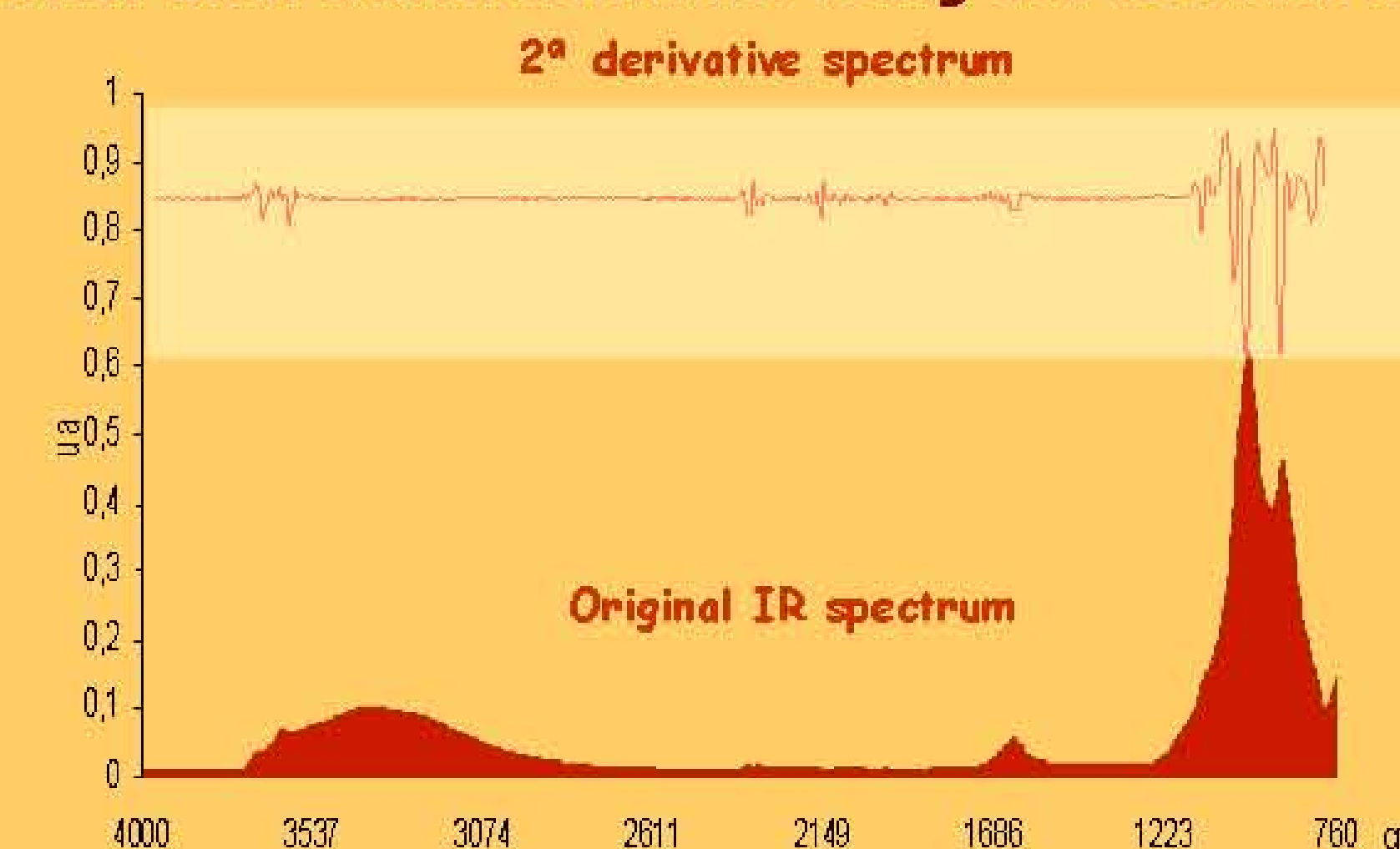
❖ Applying Principal Component Analysis (PCA) in order to calculate the number of eigenvectors explaining maximum variability:

91% cumulative variance was explained by five eigenvalues
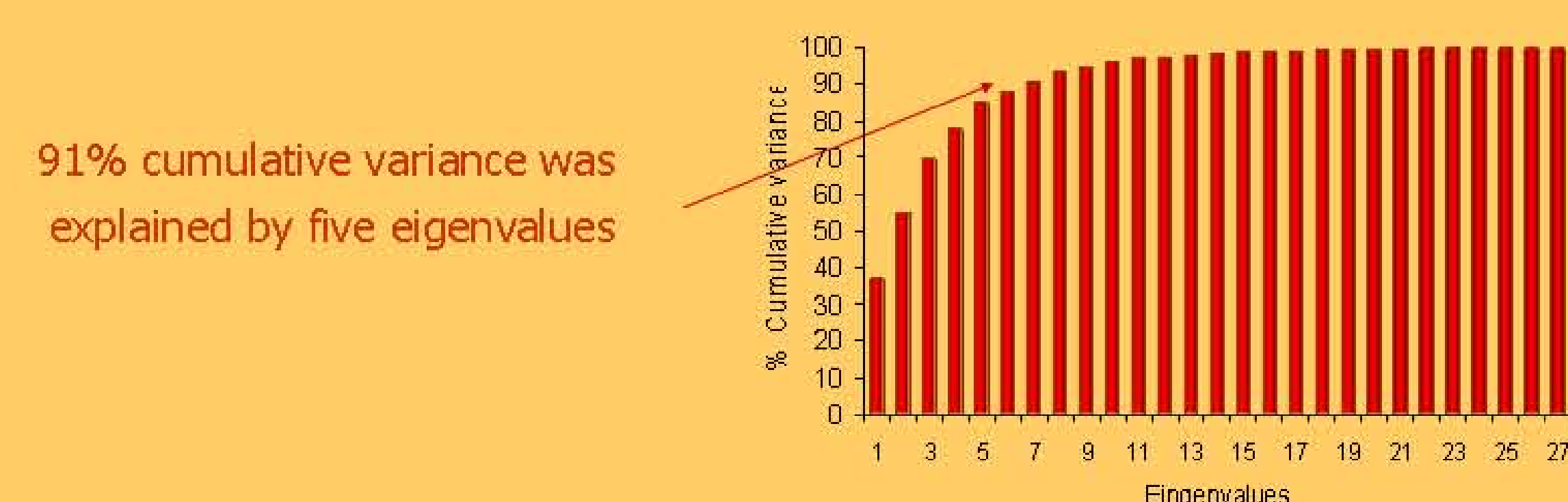
❖ Choosing the lineal regression explaining the maximum variability (R^2adj) with the lowest root mean squared error (RMSE) and obtaining a lineal regression model for each physico-chemical variable:
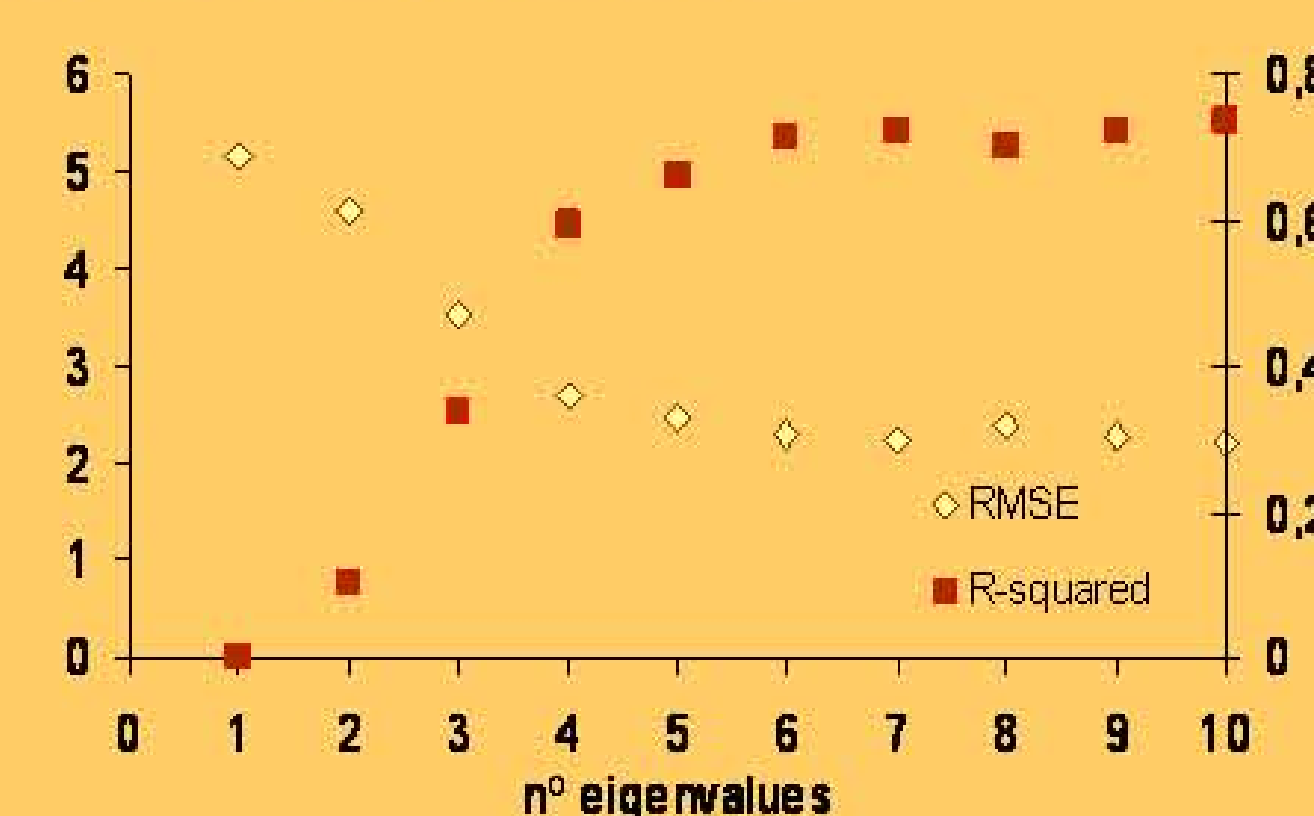

❖ Identifying the most diagnostic FT-IR bands based on its variable importance for projection (VIP):
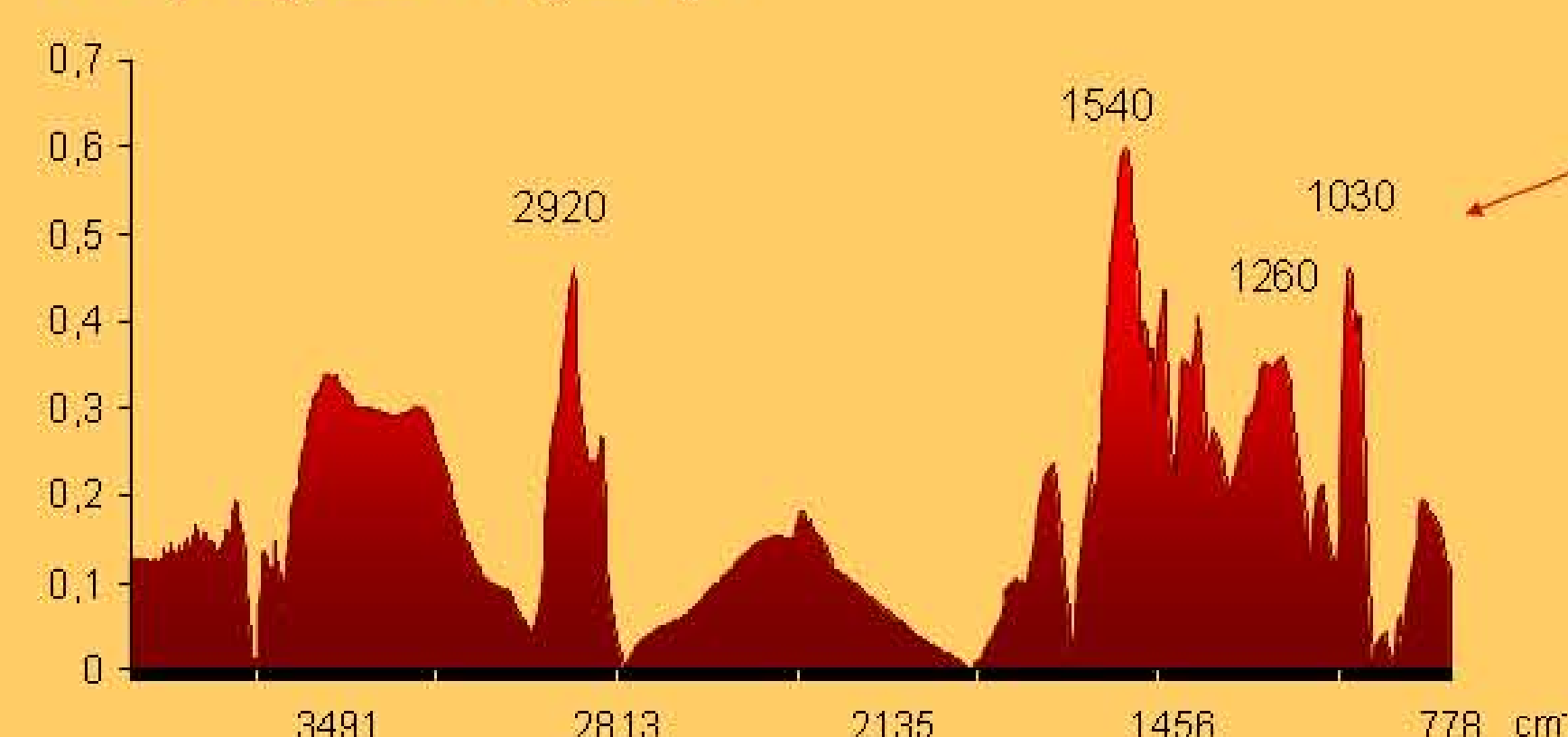
In the case of the SOM, this variable is explained mainly by bands at 2920 cm^{-1} (alkyl C–H stretching), 1540 cm^{-1} (amide bands), 1260 cm^{-1} (guaiacyl lignin units) and 1030 cm^{-1}.

❖ Extrapolating the corresponding physico-chemical variable from FT-IR spectral data and using the value of the residues to decide whether the lineal regression is acceptable:

$y = 0.9275x + 0.6364$
$R^2 = 0.9166$
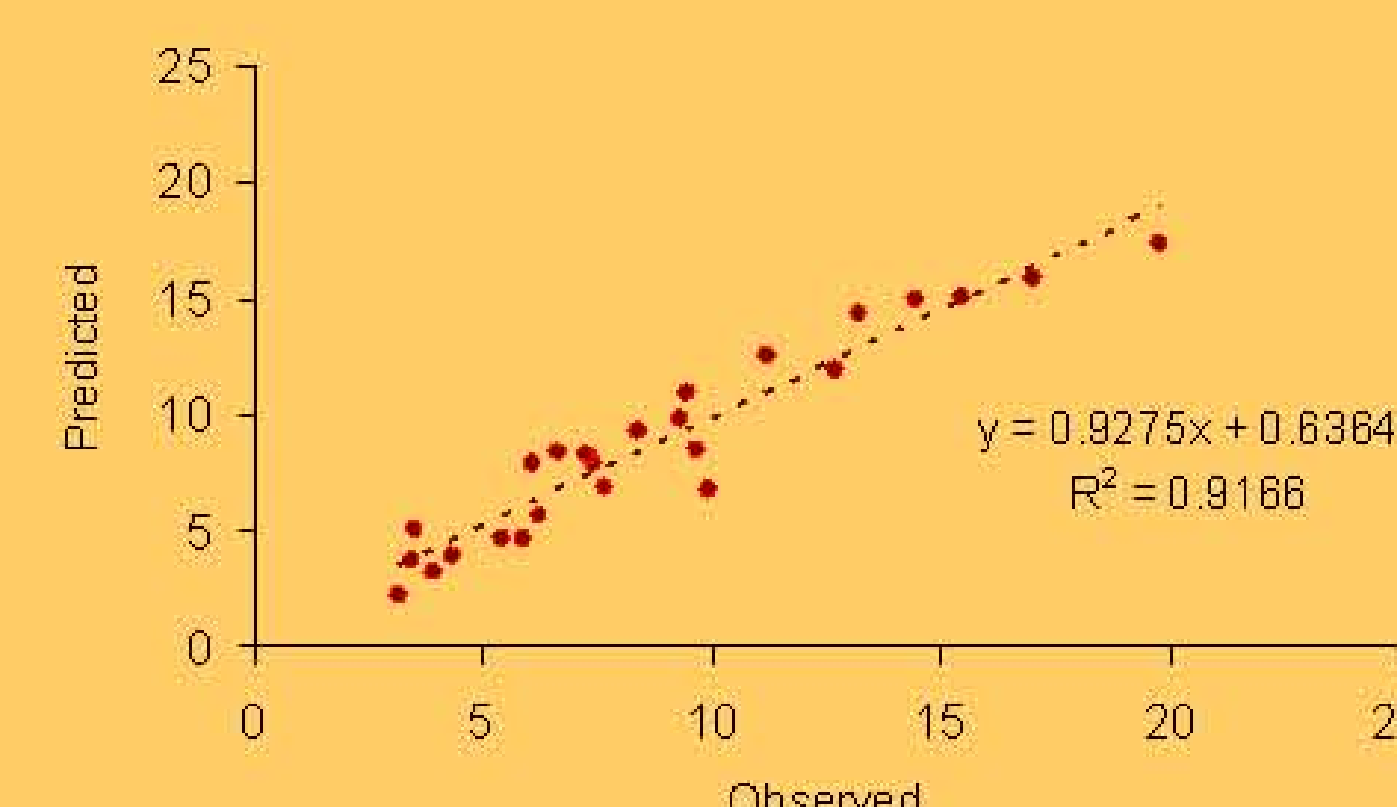
References:
[1] Janik L.J., Merry R.H. & Skjemstad J.O. 1998. Aust. J. Exper. Agric. 38, 681–696.
[2] Viscarra Rossel R.A. 2008. Chemometr. Intell. Lab. Syst. 90: 72–83.