Postprint of: Journal of Chromatography A (1533) 164-173 (2018)**GRAPHICAL STATISTICAL APPROACH TO SOIL ORGANIC MATTER RESILIENCE
USING ANALYTICAL PYROLYSIS DATA**

Gonzalo Almendros^{1*}, Zulimar Hernández², Jesús Sanz³, Sonia Rodríguez-Sánchez¹, Marco A. Jiménez-González¹, José A. González-Pérez⁴

¹ Museo Nacional de Ciencias Naturales (MNCN-CSIC), c/Serrano 115-B, Madrid (Spain)

² Universidad Autónoma de Madrid (UAM), c/ Francisco Tomás y Valiente 7, Madrid (Spain)

³ Instituto de Química Orgánica General, (IQOG-CSIC), c/Juan de la Cierva 3, Madrid (Spain)

⁴ Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS-CSIC), Av. Reina Mercedes 10, Seville (Spain)

* E-mail address: humus@mncn.csic.es (G. Almendros). Tel: +34 917452500

ABSTRACT

Pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS) of humic acids (HAs) from 30 agricultural soils from a volcanic island (Tenerife, Spain) was used to discern the molecular characteristics of soil organic matter (SOM) associated to resilience. For faster perceptual identification of the results, the yields of the pyrolysis products in the form of surface density plots were compared in an update of the Van Krevelen graphical statistical method. This approach, with respect to data reduction and visualization, was also used to collectively represent statistical indices that were obtained after simple and partial least squares (PLS) regression. The resulting plots illustrate different SOM structural domains (for example, carbohydrate- and lignin-derived and condensed lipid). The content of SOM and total mineralization coefficient (TMC) values can be well estimated from the relative abundance of 57 major pyrolysis compounds: SOM content and composition parallels the accumulation of lignin- and carbohydrate-derived structures (lignocellulosic material) and the depletion of condensed polyalkyl structures. In other words, in the volcanic ash soils that were studied, we found that the higher the amount of SOM, the lower its quality in terms of resilience. Although no cause-and-effect is inferred from this fact, it is evident that the resistance to biodegradation of the SOM is related to its molecular composition.

1. Introduction

There is a controversy about whether the resilience of soil organic matter (SOM) is intrinsic to its molecular composition or whether it depends on physical protection by soil minerals [1]. Several classical studies have suggested that SOM resistance to biodegradation may be explained by intrinsic factors, such as a chaotic macromolecular structure consisting of a

disordered arrangement of “building blocks” [2–4]. However, it is clear that some extrinsic factors may also play a role in the retention of organic C in soil [5]. Among these extrinsic factors there are the physical occlusion of SOM into micro-aggregates [6, 7] and the formation of organo-mineral complexes with fine clays or amorphous oxides [3, 4, 8, 9] with a substantial bearing on soil agro-ecological quality [10–12]. Notwithstanding, it is apparent that additional research is required in order to establish the extent to which soil carbon sequestration may depend on multiple interrelated factors [13–17].

Pyrolysis-gas chromatography-mass spectrometry (Py-GC-MS) involves a thermolytic degradation of macromolecules into small fragments that may be separated and identified by GC-MS [18–19]. This degradation technique is able to release meaningful structural fragments even from C–C bonded chemically recalcitrant materials [20]. In addition, SOM pyrolysis generates a wide range of products that can be related to its origin (e.g., methoxyphenols from lignin, furans from polysaccharides, N-molecules from proteins, etc. [21–22]). The technique has been used successfully to assess the SOM quality or maturity i.e., the extent to which plant biomacromolecules such as lignin are transformed into amorphous colloidal material of chaotic structure known as humic acids (HAs). In this material, structural domains consisting of plant or microbial macromolecules may remain to an extent that depends on its progressive stages of maturity [23–25]. In particular, progressive changes in the signature of lignin constituents during humification reflect the transformation of plant biomacromolecules into stabilized SOM [26].

In the present paper, a new method is described for using Py-GC-MS data that are generated from soils. A chemometric study of the chromatographic data from volcanic soils under agricultural management (Tenerife Island, Spain), including several predictive data mining analyses, was carried out in order to identify molecular proxies with potential use as soil carbon storage indexes. With this aim, total abundances of the major pyrolysis products were compared using the classical Van Krevelen graphical statistical method [27], updated as surface density values in the space defined by the compound-specific H/C and O/C atomic ratios [28]. These 3D Van Krevelen plots, which may be represented as surfaces or contour diagrams, were used in our approach to represent not only the sizes of the chromatographic peaks, but also statistical indices that were calculated by uni- or multivariate data analyses, or subtraction values between compound yields from different groups of samples for

comparison purposes. In all cases, the purpose was to identify, within the pyrolysis chromatographic data, molecular descriptors for SOM resilience.

2. Materials and Methods

2.1. Studied sites and soil samples

Tenerife (Spain) is a volcanic island that is one of seven in the Canary islands, located at 28.280 N and 16.290 W, with a maximum altitude of 3,718 m (Teide volcano). The geology of Tenerife island includes basaltic lava flows (predominant) and pyroclastic materials [29]. The present work analyzes 30 soils corresponding to vineyards subjected to different management practices (Table 1). The sampling points were selected with the intention to include the maximum C range and variability in quantitative environmental factors with a presumptive role in soil C storage. The main characteristics and agricultural system for each plot are summarized in Table 1. Further details have been reported elsewhere in the body of literature [30].

2.2. General analyses of soil

Soil samples were collected in triplicate from 100 m² plots from the upper 20–30 cm of the profile (Ap horizon), and immediately transported in a fridge to the laboratory to avoid microbial degradations. Before analysis, the samples were air-dried and homogenized to < 2 mm (fine earth); gravel (>2 mm) was weighed and discarded.

For pH measurements, a 1:2.5 soil-to-water mixture was used; N content was measured by micro-Kjeldahl digestion. Soil organic carbon (SOC) was determined by using the wet oxidation method [31]. Particle-size was measured using the hydrometer method [32].

2.3. Isolation and purification of humic acids

The isolation of SOM fractions was undertaken using alkaline reagents [33]. In brief, total extractable humic substances were isolated with 0.1 mol L⁻¹ Na₄P₂O₇ and 0.1 mol L⁻¹ NaOH. The first step is needed to remove Al associated with SOM and release the humic substances that are complexed with hydrous amorphous oxides, which are frequent in volcanic soils.

Then, the total humic extract was divided into (i) HAs, the acid-insoluble fraction, (ii) fulvic acids (fraction soluble at all pH conditions), and (iii) humin (non-extractable insoluble fraction). In order to remove the accompanying mineral content that was coextracted or incorporated in the course of the extraction, the HAs were dissolved in 0.5 mol L⁻¹ NaOH and centrifuged at 43,500 g; the centrifugation pellets were discarded and the sodium humates were re-precipitated with acid (6 mol L⁻¹ HCl). The HAs were finally dialyzed into cellophane bags (Visking® dialysis tubing, molecular weight cut-off 18,000 D) and desiccated at 40 °C.

2.4. Incubation experiment

The CO₂ released by the soils in the course of its biological transformation was determined under standard laboratory conditions. Soil samples (20 g) at 66% of water holding capacity (at atmospheric pressure) were incubated in Erlenmeyer flasks for 30 days at 27 °C. The CO₂ accumulated in the flasks was determined periodically (every 1 to 5 days, depending on the stage of the experiment) using a gas analyzer (Carmograph-12; Whöstoff, Germany). During each measurement, the atmosphere of the flasks was replaced with CO₂-free air passed through a soda lime filter. Data were expressed as total mineralization coefficients (TMC) in mg C kg C soil⁻¹ day⁻¹ [34].

2.5. Pyrolysis-gas chromatography-mass spectrometry

Analytical pyrolysis was performed in the HA fractions samples using a Pyrojector® (SGE Analytical Science, Melbourne, Australia). Approximately one mg HAs in a quartz capillary tube was introduced into a preheated furnace at 500 °C. The pyrolysis products were separated and analyzed in a GC-MS system (Finnigan Trace GC Ultra) with a Trace dual stage quadrupole (DSQ) mass spectrometer (Thermo Finnigan LLC., Austin TX, USA) fitted with a fused silica capillary non-polar general-purpose column HP-1 wall-coated, silphenylene polysiloxane, 30 m × 0.25 mm i.d. × 0.25 µm film thickness (Quadrex Corp., Woodbridge, CT, USA). Carrier gas was He and flow was adjusted to 1 mL min⁻¹. The total chromatographic time was 60 min. The initial temperature was kept to 50 °C for one minute, and then increased up to 100 °C at 30 °C min⁻¹, from 100 to 300 °C at 10 °C min⁻¹ and was isothermal at 320 °C.

Tentative assignment and relative quantitation of pyrolysis compounds were undertaken by using: i) selected ion traces in reconstructed ion chromatograms for the major expected

compounds (i.e., the well-known, frequent SOM pyrolysis products), ii) comparison with own laboratory databases of chromatographic data (i.e., MS, retention time (RT), response factors, literature references, etc.) of pyrolysis compounds, based on previous injection either of commercial standards or purified preparations of plant and microbial biomacromolecules (e.g., lignin, suberins, polysaccharides, etc.) which were used to confirm the RTs of the peaks for the main series of homologues and iii) from their electron impact mass spectra (70 eV) using the NIST and Wiley spectral libraries for automatic library searching, in addition to “ad hoc” self-developed computer programmes for manual heuristic searching (mainly using reverse fitting) in the aforementioned specific databases of SOM pyrolysis compounds. The peak areas (as total area counts) of the different chromatographic peaks were integrated and calculated as total abundances.

In the case of most diagnostic pyrolysis compounds, no attempt for peak deconvolution was carried out, since the major problems of overlapped peaks corresponded to coelution with a well-defined alkyl series of alkanes, alkenes and fatty acids. The well-defined series was previously extracted as single ion traces using the ions at m/z 57, 55 and 73, respectively. Integration values of the peaks in the traces for these single ions were transformed into total ion counts via multiplying by suitable response factors that were calculated from independent injections of commercial standards. A similar semiquantitative routine, in the case of any other typical pyrolytic series (i.e., methoxyphenols, alkylbenzenes, catechols, etc.), was carried out by integrating the selected peaks in ion chromatograms at known RT. The total abundances as total ion chromatographic signal were calculated using conversion factors from commercial standards or from peak areas of pyrochromatograms from comparatively simpler materials of well-studied composition and low background effects (e.g., wood, cutin or lignin preparations). These materials were previously injected into the same columns to confirm the identities, MS patterns and empirical response factors (for total ions and for single ions) of each of the coinciding molecules.

2.6. Graphical representation for yields of pyrolysis compounds

Qualitative pyrolysis results were found to be very similar for all samples, differing mainly in the relative abundance of their pyrolysis compounds. To show, at a first sight, these semi-quantitative differences between samples, the integrated chromatographic data were represented as “surface density plots”, which are 3D interpolated surfaces in the space

defined by their H/C and O/C ratios. This classical Van Krevelen graphical statistical method [27] is usually applied to data from elemental analysis of organic materials such as coals or SOM fractions. This approach can be also applied to the individual compounds that are determined by GC-MS, which appear in the Van Krevelen plot as different chemical groups [28].

The chromatographic data were classified and represented by a procedure that was based on plotting quantitative values ($z(x,y)$) for individual compounds as density surfaces in the x,y space defined by their atomic H/C (y) and O/C (x) ratios, calculated from their empirical formulas (Table 2). By using self-developed computer programmes, the peak area data $z(x,y)$ were reallocated in a 2D data matrix (discrete grid size 50×50), in which the total abundance values $z(x,y)$ of the compounds were aggregated when coinciding in the H/C vs. O/C grid because of its similar (or equal, as in the case of olefins) atomic ratios. In a second stage, an interpolated surface is obtained by applying the moving average algorithm (i.e., averaging each cell value with its 4 neighbour cells). The resulting plots show a series of overlapped mountain-like broad peaks, or compound clusters, composed of several molecules with similar elemental composition, with some of them coinciding with the collective names of the chemical nomenclature (phenols, alkanes, etc.). Van Krevelen plots have the advantage of displaying the densities in the different regions of the H/C vs. O/C surface, which facilitates comparisons between samples. Another advantage is the possibility to infer biogeochemical processes [27] that may be involved in the stabilization of the SOM: dehydration, aromatization, dealkylation, decarboxylation, etc., which are related to the concentration of compound groups in specific regions of the atomic H/C vs. O/C plane.

2.7. Statistical data treatments

For multivariate data treatments, the dependent variables (DVs) chosen were i) SOC and ii) TMC, which are frequently used as surrogates for the potential C sequestration in soils. The independent variables (IVs) correspond to SOM molecular descriptors obtained by Py-GC-MS, i.e., the total abundance of 57 major normalized GC peaks following pyrolysis.

The relationships between the response variables, or dependent variables (DVs), i.e., SOC and TMC, and the normalized GC peaks (independent variables (IVs) or descriptors) were

examined by partial least squares regression (PLS). This is a linear multivariate model that is suitable for explaining complex relationships contained in matrices with collinear data and missing values, even when the number of IVs vastly exceeds the number of individuals [35], as is common in chromatographic analysis. A specific software for the chemometric analysis of spectroscopic data "ParLeS" [36], was used for different data pre-treatments (standardization of IVs as total or relative abundances, mean centring, and variance scale). Spurious models, due to overfitting when excessive number of PLS components or latent variables (LVs) are included in the model, were discarded by repeating PLS models with random values of DV. Also, Akaike information criterion (AIC) [37] and the root mean square error (RMSE) were considered in order to select the optimum predictive model.

In the second stage, it was assumed that the presence of significant relationships between DV and IVs means that the GC pattern contains relevant information that would explain the levels of the IV and hence requires an interpretation. This was carried out by three graphical approaches: (a) plotting the values of the variable importance in projection (VIP) scores calculated during the PLS regression. This index is considered to reflect the importance of each IV in the projection used in a PLS model and is often used for variable selection; (b) building up surface plots with the Pearson correlation coefficients between abundances of each IV (molecules) and the DV; (c) comparing IV patterns in soil subsets representing opposed levels of the DVs. For this purpose, the 30 soils were ordered in decreasing DV values and two subsets were considered for comparison, i.e., soils above vs. below the median of the DV, or soils 'behaving as C-sinks' vs. 'soils with 'poor C sequestration potential'. The same comparisons were undertaken between soils with extreme values of DV i.e., in the uppermost quartile (Q1) vs. the last quartile (Q4). The average molecular composition of the two subsets was calculated for the 57 compounds (IVs) and subtracted to obtain one surface density plot with positive and negative values, respectively, showing the compounds prevailing comparatively in C-rich and C-depleted soils [28].

2.8. Graphical representation of the statistical results

Both the molecular composition of each individual sample or sample average and the indices obtained by the above-mentioned three statistical approaches (a, b and c) were represented in the z axis as surface density plots or as contour plots for TOC and TMC values. Our objective was to obtain plots to illustrate structural domains of the SOM (carbohydrate-and

lignin-derived, condensed lipid, etc.) characterizing different scenarios of SOM accumulation or, in other words, of active organic matter mineralization. A secondary objective of subtracted density plots was to obtain simpler plots than those, which are more complex to interpret, representing the whole molecular assemblages released by pyrolysis. In fact, compounds (IVs) that were not relevant in the forecasting models i.e., with low VIP values, or not correlated with the DVs, or with the same concentration in samples at extreme DV values, should give indices with values equal or close to zero (basal plane), while not forming clusters or peaks.

3. Results

3.1. Analysis by pyrolysis-gas chromatography

The pyrolysis compounds released from soil HA are listed in Table 2. The main series that were detected corresponded to *n*-alkanes, *n*-alkenes, alkanolic acids (fatty acids), alkylbenzenes, methoxyphenols and alkylphenols. HAs also released known carbohydrate-derived constituents (methylfuran, furanone, furaldehydes and cyclic ketones) in addition to N-containing compounds (pyrroles, pyridines and indoles) from peptides and proteins (Fig. 1a). A dominance of lignin-derived units (ca. 40% total pyrolysis compounds) was found in some samples, mainly from Regosols with abundance of vitric materials (U5 and U6 sampling units) and soils with HAs incorporated into stable Al-humus organomineral complexes (U2 sampling unit). These lignin-derived compounds included conspicuous amounts of methoxyphenols, i.e., both guaiacyl (G)-lignin units (guaiacol and its methyl-, ethyl-, vinyl-, propenyl- and aceto- derivatives) and their corresponding syringyl (S) counterparts, in addition to vinylphenol (from grass lignins). On the other hand, the low relative abundances of lignin-derived pyrolysis compounds accompanied by high yields of methoxyl-lacking aromatic compounds (alkylbenzenes, alkylphenols and alkyl-naphthalenes) were characteristic of HAs from soils with non-andic characteristics (U3 and U7) which may indicate the occurrence of rapid microbial demethoxylation of lignin, or anthropogenic charcoal inputs, or both [38]. Finally, a high relative content of alkyl compounds (alkanes, olefins and fatty acids) was found in HAs from the U4 sampling unit, which could indicate the occurrence of a high microbial activity (also suggested by TMC data in Table 1).

The semiquantitative information provided by the pyrograms, included in Table 2, was used to build up surface density plots (Fig. 2), showing a series of peaks or clusters of peaks corresponding to the major structural domains of the HA. Fig. 2 (a) depicts a model surface density plot calculated from the average pyrochromatograms of the 30 HA extracted from all of the studied soils. This can be seen as a model for the chemical composition of HA in soils that developed on volcanic ash materials. The different broad peaks or compound clusters suggest a balanced composition in the HAs structure, consisting of aliphatic and aromatic domains. Alkanes and olefins are clearly separated in the plot in terms of the H/C ratios, as well as the oxygen-containing alkanolic acids. Independent clusters are often formed for the short- and long-chain molecules, which is interpreted as indicative of origin *viz.* from the microbial metabolism (short chain) and greater levels of plant epicuticular waxes (long chain) [39]. Alkylbenzene structures were also distinguishable from polycyclic aromatic compounds in terms of their H/C ratio (e.g., H/C values < 1.1), which was also the case for alkylphenols vs. methoxyphenols. In the latter case, S-type compounds (syringyl-derived lignin structures with two methoxyl groups) are clearly distinguished from the G-type (guaiacyl-derived lignin structures with one methoxyl group) compounds. At the same time, the length of the residual alkyl chain of these phenols (from methyl to propenyl) is denoted in the surface height of the methoxyphenol region corresponding to different H/C values. All of these features are often informative about the extent of lignin demethoxylation and side-chain breakdown which is believed to depend on lignin origin and transformation extent [23, 40, 41]. The compounds clustering in the plot regions with comparatively high O/C ratios correspond to heterocyclic, carbohydrate-derived structures such as furaldehydes and furanones.

An additional interesting feature of the graphical interpretation of the Py-GC-MS data as surface density plots is that some biogeochemical processes involved in the humus formation can be reflected as shifts of both H/C and O/C scores of the main groups of pyrolysis compounds. In particular, progressive 'maturation' or 'stabilization' of the SOM is often described as the extent to which the atomic ratios of its structural constituents differ from those of the precursor biomacromolecules. For instance, lignocellulosic biomass (yielding methoxyphenols and carbohydrate-derived products), microbial biomass (with a conspicuous lipoproteic character), etc., are composed of major units with comparatively large H/C and O/C atomic ratios, compared to those of HAs. In general, humification

processes lead to changes in both atomic ratios of the pyrolysis compounds and increased abundances of the molecules clustering towards the origin of the coordinates [42].

Additional surface density plots shown in Fig. 2 (b,c) illustrate the composition of HA from two soils with extreme values for the DV SOC; b) with the lowest and c) with the highest SOC content. The differences are evident and, as discussed below, the former shows concentration of C–C-linked, O-lacking alkyl- and alkylbenzene structures, whereas, the latter shows a variety of hydroxyl- and methoxyl-containing cyclic structures, generally resembling pyrolysates from raw plant (lignocellulosic) litter.

3.2. Forecasting soil organic carbon concentration (SOC) and mineralization (TMC) by PLS regression

The results from the PLS regression analysis demonstrated that both DVs, i.e., SOC (informing on soil organic C storage) and TMC (informing on CO₂ release from soil) can be significantly estimated ($P < 0.05$) from the relative abundance of the 57 major pyrolysis compounds that are listed in Table 2. Significant correlations were found between the calculated and the experimental data for both DVs; SOC (Fig. 3a) and TMC (Fig. 3b). In the case of TMC, the best model incorporates information from the first 12 PLS components (LVs). Both criterion, RMSE and AIC, for selecting optimum models for SOC and TMC suggested that even the model with 12 LVs was not over-fitted; i.e., the relevant information is not obtained from noise or random values of minor GC peaks. This was confirmed independently by repeating the PLS with randomized DVs and obtaining that, with the same number of LVs, the PLS models were non-significant i.e., there was no correlation between observed vs. estimated values, that were calculated using the leave-one-out cross-validation method. The model for SOC (Fig. 3a) was much more significant than for TMC (Fig. 3b), which was probably due to the higher expected errors of the latter DV in sampling and experimental determinations. In the case of SOC, the first significant ($P < 0.01$) PLS model was achieved with only 2 LVs. Nonetheless, the model with 12 LVs was selected in order to have comparable VIP values for both, SOC and TMC. The compound types showing the highest variability in amount, depending on the levels of the DVs, are mapped in contour diagrams in Figs. 3c and 3d, showing opposite patterns for SOC and TMC, as discussed below.

3.3. Relevant pyrolysis products in forecasting soil organic carbon concentration (SOC) and mineralization (TMC)

VIP values have been used to identify compounds which differ significantly in concentration, in terms of DV. The VIP values are represented as bars below the chromatographic plot in Fig. 1a. However, the plot of VIPs is rather complex, which is not very useful for rapid identification of patterns or trend recognition, especially when several chromatograms should be compared. This information can be simplified, as in Fig. 4a, whereby the VIP values are not ordered by RT, as depicted in Fig. 1b, but by the atomic ratios of the corresponding compounds in the Van Krevelen space, and the individual values are averaged on a density surface. The 3D VIP plot suggests that compounds behaving as proxies for SOC accumulation, or bearing discriminant values in terms of SOC levels, correspond to a large extent to polyalkyl pyrolytic compounds (mainly olefins) and to molecules with comparatively large O/C and low H/C ratio, which are mainly methoxyphenols and anhydrosugar-type structures.

Notwithstanding, by definition VIP values are always positive, and for their interpretation it would be helpful to know whether the important compounds are those prevailing in pyrograms of the samples with high or low DVs, i.e., SOC contents. This could be undertaken with the aid of other indices that are calculated for the pyrograms (b & c approaches, see section on Materials and Methods). An example is illustrated in Fig. 4b, where the VIP values for SOC are compared with the surface representing the subtraction values of the abundances of pyrolysis compounds from samples with high and low SOC (Q1 and Q4 quartiles). The plot clearly reveals that the abundance of alkanes (and long-chain fatty acids), as well as of low H/C polycyclic compounds, is characteristic of samples with low SOC levels. This result was, somehow, already expected from the observation of the exploratory plots in Figs. 2b and 2c, that demonstrated the different composition of HAs from soils with contrasted SOM contents. The above confirmation that the PLS model is significant, and the observation of VIPs informing on which compounds are important in explaining the differences in SOC levels, was also a validation that simple comparisons based on averaged pyrograms, corresponding to samples with extreme values of the DV, are valid in this case study. This is also shown as contour diagrams in Figs. 3d and 3e in the case of SOC and TMC, showing similar patterns with the colours inverted, as expected for the above-indicated negative correlation between high C levels and low soil respiration in the studied soils.

Finally, in the case of SOC, Fig. 5 compares the visual patterns obtained with the VIP values (Fig. 4a), Pearson's r correlation coefficients (Fig. 4b) and the subtraction values between pyrograms of samples with high or low values of DVs (in this case, samples with SOC values above and below the median, Fig. 4c). For the b and c plots, the positive and negative values are shown in red and green, respectively. Direct comparison of these surfaces (at least b and c) coincides with noting that high SOC levels are associated to the dominance of alkyl peaks (mainly alkanedienes), accompanied by H- or alkyl-substituted benzenic structures, whereas, in C-depleted soils the major density of the surfaces was found in the areas for O-containing compounds, not only phenols and furans but also in the regions for long-chain ($> C_{12}$) fatty acids.

4. Discussion

Surface density plots built up on the Van Krevelen space were found useful for perceptually summarizing differences, that would otherwise be imperceptible, between Py-GC-MS quantitative data for a large number molecules. When these plots are used to represent statistical indexes or differences between pyrograms, or average pyrograms, the resulting surface is simplified, since only compound clusters reflecting those structures that are relevant for the process under study (in this case accumulation or mineralization of soil organic matter) are formed in the space defined by atomic ratios. In this way, the direct analytical pyrolysis of HAs that are extracted from volcanic ash soils, gives evidence that differences in molecular composition of SOM are related to differences in terms of its resistance to biodegradation.

The 3 chemometric approaches used in this work coincided in pointing out that the soils with high levels of SOC are those yielding, during pyrolysis, a series of compounds with an origin from lignin- and carbohydrate-like structures, whereas, the pyrograms of C-depleted soils show characteristically high yields of compounds derived from condensed polyalkyl structures, mainly α , ω -alkanedienes. In other words, in the studied volcanic ash soils, the larger the quantity of SOC, the lower its quality would be in terms of degradability and vice versa.

Although no cause-and-effect is inferred, here we also show that relevant soil variables, such as total amount of soil C (SOC) and its susceptibility to biodegradation (TMC), can be effectively estimated using only a limited series of pyrolytic descriptors that inform on HA structure.

Finally, we believe that the findings that are described in this paper are relevant and they provide new tools for studying soil organic carbon dynamics that, in turn, may help in deciphering relevant environmental issues, such as the effect of soil organic matter in the global C cycle and in climate change.

5. Conclusions

Samples of SOM that present a different resistance to biodegradation also have a different molecular composition. The three approaches used in this exploratory survey on molecular surrogates of soil C sequestration, viz., a) VIPs, b) Pearson's correlation indices, or c) subtraction values between average pyrochromatograms corresponding to samples with extreme values of the DVs under study, coincided in pointing out that the SOM levels paralleled the accumulation of lignin- and carbohydrate-derived structures, and the depletion of condensed polyalkyl structures. The above findings highlight the importance of the quality, as opposed to the quantity, of the SOM in volcanic ash soils. Practically, the same pattern, but with swapped signs, was obtained for the Py-GC-MS compounds with higher loadings in the estimation of the CO₂ release from the soil samples under laboratory conditions (TMC). Comparing the approaches (a–c) in terms of workload and time consumed, although PLS (approach a) is frequently invoked to have outstanding potential to extract underlying information in multivariate systems, in our case the main utility of PLS was the selection of the LVs which better estimate both the accumulation (SOC) and biodegradability (TMC) of the organic matter in soils. In fact, in this approach, depending on the number of LVs, IVs with VIPs > 1 were not systematically correlated ($P < 0.05$) with the corresponding DV. In the case of Pearson's correlation coefficients (approach b), the control of possible outliers is required to ensure further reliable results. Approach (c) was found to be the most simple and intuitive.

Acknowledgments

Financial support by Spanish MINECO projects (CGL2013-43845-P and FEDER co-financed CGL2016-78937-R INTERCARBON) is acknowledged. Marco A. Jiménez-González thanks the Spanish Ministry of Economy and Competitiveness (MINECO) for funding his pre-doctoral FPI fellowships (BES-2014-069238). The authors would like to thank the three anonymous reviewers for their helpful comments that greatly contributed to improving the final version of the article.

References

- [1] M.W.I. Schmidt, M.S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I.A. Janssens, M. Kleber, I. Kögel-Knabner, J. Lehmann, D.A.C. Manning, P. Nannipieri, D.P. Rasse, S. Weiner, S.E. Trumbore, Persistence of soil organic matter as an ecosystem property, *Nature* 478 (2011) 49–56. DOI: 10.1038/nature10386.
- [2] J.M. Oades, The retention of organic matter in soils, *Biogeochemistry* 5 (1988) 35–70. DOI: 10.1007/BF02180317.
- [3] J.A. Baldock, J.O. Skjemstad, Role of the soil matrix and minerals in protecting natural organic materials against biological attack, *Org. Geochem.* 31 (2000) 697–710. DOI: 10.1016/S0146-6380(00)00049-8.
- [4] B. Marschner, S. Brodowski, A. Dreves, G. Gleixner, A. Gude, P.M. Grootes, U. Hamer, A. Heim, G. Jandl, R. Ji, K. Kaiser, K. Kalbitz, C. Kramer, P. Leinweber, J. Rethemeyer, A. Schäffer, M.W.I. Schmidt, L. Schwark, G.L.B. Wiesenberger, How relevant is recalcitrance for the stabilization of organic matter in soils?, *J. Plant Nutr. Soil Sci.* 171 (2008) 91–110. DOI: 10.1002/jpln.200700049.
- [5] J.D. Jastrow, J.E. Amonette, V.L. Bailey, Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration, *Clim. Chang.* 80 (2007) 5–23. DOI: 10.1007/s10584-006-9178-3.
- [6] N. Chouliaras, J.C. Vedy, F. Jacquin, J.M. Portal, Fractionnement et caractérisation de la matière organique dans les rendzines. *Bull. I. N. Polyt. Nancy* 17 (1975) 65–74.
- [7] G. Almendros, F.J. González-Vila, Degradative studies on a soil humin fraction, Sequential degradation of inherited humin, *Soil Biol. Biochem.* 19 (1987) 513–520. DOI: 10.1016/0038-0717(87)90093-9.

- [8] B.P. Warkentin, T. Maeda, in: B.K.G. Theng (Ed.), *Soils with Variable. Charge*, New Zeland Society of Soil Science, Lower Hutt, New Zeland, 1980, pp. 281–301.
- [9] J.A. Baldock, P.N. Nelson, Soil organic matter. In: M.E. Sumner (Ed.), *Handbook of Soil Science*. CRC Press, Boca Raton, FL, 2000, pp. 25–84.
- [10] R. Lal, J.M. Kimble, R.P. Follet, in: R. Lal, J.M. Kimble, R.P. Follet (Eds.), *Soils Processes and the Carbon Cycle*, CRC Press, Boca Raton, FL, 1998, pp. 595–604.
- [11] J.A. González-Pérez, C.D. Arbelo, F.J. Gónzalez-Vila, A. Rodríguez Rodríguez, G. Almendros, C.M. Armas, O. Polvillo, Molecular features of organic matter in diagnostic horizons from andosols as seen by analytical pyrolysis, *J. Anal. Appl. Pyrol.* 80 (2007) 369–382. DOI: 10.1016/j.jaap.2007.04.008.
- [12] Y. Huang, W. Sun, Changes in topsoil organic carbon of croplands in mainland China over the last two decades, *Chin. Sci. Bull.* 51 (2006) 1785–1803. DOI: 10.1007/s11434-006-2056-6.
- [13] C. Pan, H. Zhao, X. Zhao, H. Han, Y. Wang, J. Li, Biophysical properties as determinants for soil organic carbon and total nitrogen in grassland salinization, *PLOS ONE*, 8 (2013) DOI: 10.1371/journal.pone.0054827.
- [14] R.S. Swift, Sequestration of carbon by soil, *Soil Sci.* 166 (2001) 858–871. DOI: 10.1097/00010694-200111000-00010.
- [15] H. Yuan, T. Ge, M. Kleber, C. Chen, A.G. O'Donnell, J. Wu, Significant role for microbial autotrophy in the sequestration of soil carbon, *Appl. Environ. Microbiol.* 78 (2012) 2328–2336. DOI: 10.1128/AEM.06881-11.
- [16] F. Wenting, A.F. Plante, A.K. Aufdenkampe, J. Six, Soil organic matter stability in organo-mineral complexes as a function of increasing C loading, *Soil Biol. Biochem.* 69 (2014) 398–405. DOI: 10.1016/j.soilbio.2013.11.024.
- [17] M.V. Lützow, I. Kögel-Knabner, K. Ekschmitt, E. Matzner, G. Guggenberger, B. Marschner, H. Flessa, Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions—a review, *Eur. J. Soil Sci.* 57 (2006) 426–445. DOI: 10.1111/j.1365-2389.2006.00809.
- [18] P. Leinweber, H.R. Schulten, Advances in analytical pyrolysis of organic matter, *J. Anal. Appl. Pyrol.* 49 (1999) 359–383. DOI: 10.1016/S0165-2370(98)00082-5.
- [19] F.J. González-Vila, G. Almendros, Analytical pyrolysis as a tool for the characterization of natural macromolecular organic matter. Reappraisal of its applications in agricultural and environmental sciences. *PITTCON 2011. Conference and Expo. The Pittsburgh*

- Conference on Analytical Chemistry and Applied Spectroscopy, Inc., March 1–18,
(2011): 1790–1795. Georgia World Congress Center, Atlanta, GA, USA
- [20] G. Almendros, M.E. Guadalix, F.J. González-Vila, F. Martin, Preservation of aliphatic
macromolecules in soil humins, *Org. Geochem.* 24 (1996) 651–659. DOI: 10.1016/0146-
6380(96)00056-3.
- [21] J.M. De La Rosa, J.A. González-Pérez, P.G. Hatcher, H. Knicker, F.J. González-Vila,
Determination of refractory organic matter in marine sediments by chemical oxidation,
analytical pyrolysis and solid-state ^{13}C nuclear magnetic resonance spectroscopy, *Eur. J.*
Soil Sci. 59 (2008) 430–438. DOI: 10.1111/j.1365-2389.2007.00979.x.
- [22] J. Kaal, J.A. Baldock, P. Buurman, K.G.J. Nierop, X. Pontevedra-Pombal, A. Martínez-
Cortizas, Evaluating pyrolysis–GC/MS and ^{13}C CPMAS NMR in conjunction with a
molecular mixing model of the Penido Vello peat deposit, NW Spain, *Org. Geochem.* 38
(2007) 1097–1111. DOI: 10.1016/j.orggeochem.2007.02.008.
- [23] P. Tinoco, G. Almendros, F.J. González-Vila, Impact of the vegetation on the lignin
pyrolytic signature of soil humic acids from Mediterranean soils, *J. Anal Appl Pyrolysis*
64 (2002) 407–420. DOI: 10.1016/S0165-2370(02)00041-4.
- [24] G. Almendros, Humic substances. In: W. Cheswort (Ed.), *Encyclopedia of Soil Science*,
Springer, Dordrecht, 2008 pp. 97–99.
- [25] I. Miralles, A. Piedra-Buena, G. Almendros, F.J. González-Vila, J.A. González-Pérez,
Pyrolytic appraisal of the lignin in soil humic acids: Assessment of its usefulness as
carbon sequestration marker, *J. Anal. Appl. Pyrolysis* 113 (2015) 107–115. DOI:
10.1016/j.jaap.2014.11.010.
- [26] I. Kögel, F. Ziegler, W. Zech, Lignin signature of subalpine Rendzinas (Tangel- and
Moderrendzina) in the Bavarian Alps, *J. Plant Nutr. Soil Sci.* 151 (1988) 15–20. DOI:
10.1002/jpln.19881510105.
- [27] D.W. Van Krevelen, Graphical-statistical method for the study of structure and reaction
processes of coal, *Fuel* 29 (1950) 269–284.
- [28] G. Almendros, P. Tinoco, J.M. De la Rosa, H. Knicker, J.A. González-Pérez, F.J. González-
Vila, Selective effects of forest fires on the structural domains of soil humic acids as
shown by dipolar dephasing ^{13}C NMR and graphical-statistical analysis of pyrolysis
compounds, *J. Soils Sediments* (2016) (article in press) DOI: 10.1007/s11368-016-1595-
y.

- 527 [29] Z. Hernández, G. Almendros, Humus formation in vineyards on volcanic ash soils from
528 the Island of Tenerife (Spain), IHSS15 Field Trip Guide (XV Meeting of the International
529 Humic Substances Society). Cabildo Insular de Tenerife, Área de Sostenibilidad,
530 Territorio y Medio Ambiente, Tenerife, 2010.
- 531 [30] Z. Hernández, G. Almendros, Biogeochemical factors related with organic matter
532 degradation and C storage in agricultural volcanic ash soils, *Soil Biol. Biochem.* 44
533 (2012) 130–142. DOI: 10.1016/j.soilbio.2011.08.009.
- 534 [31] A. Walkley, I.A. Black, An examination of the Degtjareff method for determining soil
535 organic matter, and a proposed modification of the chromic acid titration method, *Soil*
536 *Sci.* 37 (1934) 29–38.
- 537 [32] Soil Survey Staff, *Soil Survey Laboratory Manual Methods*, Soil Survey Laboratory and
538 Investigation Report No 42, U.S. Government Printing Office, Washington, DC, 2004.
- 539 [33] P. Duchaufour, F. Jacquin, Comparaison des processus d’humification dans les
540 principaux types d’humus forestiers, vol. 1, *Bulletin AFES*, 1975, pp. 29–36.
- 541 [34] G. Almendros, J. Dorado, Molecular characteristics related to the biodegradability of
542 humic acid preparations. *Eur. J. Soil Sci.* 50 (1999) 227–236. DOI: 10.1046/j.1365-
543 2389.1999.00240.x.
- 544 [35] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics.
545 *Chemometr. Intell. Lab.* 58 (2001) 109–130. DOI: 10.1016/S0169-7439(01)00155-1.
- 546 [36] R.P. Viscarra Rossel, ParLeS: Software for chemometric analysis of spectroscopic data,
547 *Chemometr. Intell. Lab.* 90 (2008) 72–83. DOI: 10.1016/j.chemolab.2007.06.006.
- 548 [37] H. Akaike, (1974). A new look at the statistical model identification, *IEEE Transactions*
549 *on Automatic Control* 19 (1974) 716–723. DOI: 10.1109/TAC.1974.1100705.
- 550 [38] Z. Hernández, G. Almendros, P. Carral, A.M. Álvarez, H. Knicker, J.P. Pérez Trujillo,
551 Influence of non-crystalline minerals in the total amount, resilience and molecular
552 composition of the organic matter in volcanic ash soils (Tenerife Island, Spain), *Eur. J.*
553 *Soil Sci.* 63 (2012) 603–615. DOI: 10.1111/j.1365-2389.2012.01497.x.
- 554 [39] B.R.T. Simoneit, J.O. Grimalt, T.G. Wang, R.E. Cox, P.G. Hatcher, A. Nissenbaum, Cyclic
555 terpenoids of contemporary resinous plant detritus and fossil woods, ambers and
556 coals, *Org. Geochem.* 10 (1986) 877–889. DOI: 10.1016/S0146-6380(86)80025-0.
- 557 [40] M.A. Jiménez-González, A.M. Álvarez, P. Carral, F.J. González-Vila, G. Almendros, The
558 diversity of methoxyphenols released by pyrolysis-gas chromatography as predictor of

559 soil carbon storage, J. Chromatogr. A. 1508 (2017) 130–137. DOI:
560 10.1016/j.chroma.2017.05.068.
561 [41] C. Saiz-Jiménez, J.W. de Leeuw, Lignin pyrolysis products – their structures and their
562 significance as biomarkers, Org. Geochem. 10 (1986) 869–876. DOI: 10.1016/S0146-
563 6380(86)80024-9.
564 [42] M. Schnitzer, S.U. Khan, Humic substances in the environment, Dekker, New York,
565 1972.

Fig. 1. a) Pyrochromatogram of humic acid (HA) from volcanic ash soil (sample G3, Table 1). The structure of some major compounds is shown on the peaks; b) Variable importance in projection (VIP) of the different peaks, calculated for a Partial Least Squares (PLS) model that was used to predict soil C concentration in terms of the total abundances of the 57 major pyrolysis compounds. Numbers on the peaks refer to Table 2.

Fig. 2. Surface density map displaying cumulative abundances of humic acid (HA) pyrolysis compounds represented in the space defined by their H/C and O/C atomic ratios in a Van Krevelen diagram. Plot a) shows a representative surface for an ideal HA sample, calculated as the average composition of the 30 studied soils (Table 2), which had developed on volcanic ash materials and similar vegetation (vineyard). Plots b) and c) correspond to two individual samples: sample G3 (Haplic Ustarent) which is the sample having the greatest soil organic carbon (SOC) concentration, and M1 (Typic Ustorthent) with the lowest SOC concentration.

Fig. 3. Observed vs. predicted values for soil organic carbon (SOC, a) and total mineralization coefficient (TMC, b) of volcanic ash soils, calculated by partial least squares (PLS) regression using as predictors the total abundances of the major 57 pyrolysis compounds. Differences in the yields of pyrolysis compounds in terms of the different levels of SOC and TMC are shown in c), corresponding to a surface density map that is represented as a contour diagram, representing in the H/C vs. O/C space the average abundances of compounds from two sample sets: from soils with high SOM concentration (uppermost quartile of the SOC distribution, Q1), and from soils with low SOC concentration (quartile 4, Q4). The opposite pattern is observed for the soil TMC (d): positive values are shown in green colours, whereas negative values are shown in red.

Fig. 4. Comparison of surface density maps illustrating the importance of different types of pyrolysis compounds of humic acids (HAs) from vineyard soils in terms of the progressive levels of soil organic carbon (SOC): a) Variable importance in projection, b) difference between the average abundances of compounds in the pyrograms from two samples sets: from soils with high SOM concentration (uppermost quartile of the SOC distribution, Q1) and soils with low SOC concentration (quartile 4, Q4). Negative values are shown in reddish colours.

Fig. 5. Comparison of surface density maps showing indices illustrating the importance of different types of pyrolysis compounds of humic acids (HAs) from vineyard soils in terms of the progressive levels of organic carbon of the corresponding soils (SOC): a) Variable importance in projection (VIP), b) Pearson's r correlation coefficients between the SOC and the abundances of the individual pyrolysis compounds, c) comparing soil subsets that represent opposed levels of the SOC: The 3024 soils were ordered in decreasing DVs values and two subsets were considered, above vs. below the median of the SOC values, i.e., soils 'behaving as C-sinks' vs. 'soils with 'poor C sequestration potential'. Differences between subsets were plotted as subtraction values. In the case of plots b) and c), positive values are shown in green colours, and negative values are shown in red.

Table 1 Location, classification and general analytical characteristics of soils on volcanic ash plots in Tenerife Island (Spain).

Sampling unit	Plot	Altitude m asl	Soil type (USDA)	Clay g 100 g ⁻¹	pH	SOC ^b g 100 g ⁻¹	Soil C/N	TMC mg CO ₂ kg C soil ⁻¹ day ⁻¹
U ₁	G ₁	351	Haplic Ustarent	26	4.5	4.5	16	52.5
	G ₂	357	Haplic Ustarent	28	5.5	4.6	17	30.5
	G ₃	359	Haplic Ustarent	27	5.6	6.1	17	48.5
	C _G	364	Lithic Ustivitrant	10	6.3	4.9	22	84.8
U ₂	O ₁	364	Typic Haplanthrept	26	4.5	3.9	15	62.2
	O ₂	364	Typic Haplanthrept	30	4.4	3.3	14	56.3
	O ₃	355	Typic Haplanthrept	28	4.3	3.0	14	62.7
	O ₄	357	Typic Haplanthrept	27	4.3	3.0	13	62.7
U ₃	S ₁	540	Typic Haplanthrept	48	6.3	1.6	12	125.7
	S ₂	553	Typic Haplanthrept	46	6.1	2.1	12	71.2
	S ₃	542	Typic Haplanthrept	44	6.8	2.3	13	125.4
	S ₄	612	Typic Haplanthrept	52	6.0	1.7	12	120.9
	C _S	557	Andic Dystrustept	46	6.0	2.3	16	218.0
U ₄	M ₁	168	Typic Ustorthent	34	7.1	1.0	10	88.8
	M ₂	168	Typic Ustorthent	32	7.9	1.1	11	82.4
	M ₃	166	Typic Ustorthent	30	8.1	1.2	11	73.6
	M ₄	158	Typic Ustivitrant	37	6.1	3.5	12	61.2
	C _M	150	Typic Ustorthents	28	8.0	1.6	12	256.3
U ₅	H ₁	190	Vitrandid Torriorthent	18	7.1	2.9	14	81.0
	H ₂	185	Vitrandid Torriorthent	16	7.6	2.2	13	122.7
	H ₃	180	Vitrandid Torriorthent	16	6.9	2.3	12	91.4
	C _H	175	Vitrandid Torriorthent	14	8.2	1.0	13	140.8
U ₆	F ₁	890	Vitrandid Ustorthent	10	7.4	2.3	11	150.5
	F ₂	880	Vitrandid Ustorthent	10	7.3	2.8	11	87.9
	F ₃	875	Vitrandid Ustorthent	11	7.4	2.5	12	76.4
	C _F	882	Vitrandid Ustorthent	17	7.3	1.1	14	155.1
U ₇	V ₁	1256	Typic Haplanthrept	29	5.5	1.3	11	149.2
	V ₂	1257	Typic Haplanthrept	28	5.7	1.4	13	132.5
	V ₃	1257	Typic Haplanthrept	30	5.5	1.3	13	106.4
	C _V	1469	Humic Lithic Dystroxerept	18	6.0	12.1	27	171.6
LSD				3.1	0.4	0.5	0.8	24.3

SOC: soil organic carbon; TMC= total mineralization coefficient.

LSD: least significant difference between plots with a *P*-value < 0.05.

Table 2 Compounds^a released from soil humic acids (HAs) by analytical pyrolysis (Py), separated by gas chromatography (GC) and identified by mass spectrometry (MS). The compounds were used as descriptors (independent variables, IVs) in partial least squares (PLS) models to forecast soil organic carbon (SOC) levels and CO₂ release from soils (total mineralization coefficient, TMC).

No.	Compound	H/C	O/C	G ₁	G ₂	G ₃	G ₆	O ₁	O ₂	O ₃	O ₄	S ₁	S ₂	S ₃	S ₄	C _S	M ₁	M ₂	M ₃	M ₄	C _M	H ₁	H ₂	H ₃	C _H	F ₁	F ₂	F ₃	C _F	V ₁	V ₂	V ₃	C _V
1	C ₁ -propanal	2.000	0.250	3.9	3.0	4.6	3.4	4.9	2.3	2.7	2.2	2.1	2.0	1.8	1.0	2.3	1.2	1.4	1.4	2.5	3.9	1.6	9.6	1.6	2.0	2.5	3.0	3.0	1.8	1.2	3.0	2.2	1.1
2	C ₁ -furan	1.200	0.200	6.7	9.6	4.5	10.0	5.7	3.4	2.5	1.6	0.0	0.0	5.1	6.9	0.0	1.5	1.8	1.8	0.6	8.0	0.4	1.0	1.6	4.8	2.3	4.3	5.6	1.1	0.4	1.0	2.2	0.0
3	Pentanal I ^b	2.000	0.200	4.8	2.5	3.9	3.9	5.6	3.5	3.2	3.1	3.0	3.1	4.0	5.8	3.0	1.9	2.9	3.7	4.3	5.2	2.1	0.6	2.4	3.0	3.3	3.9	4.0	6.8	3.0	3.9	4.2	1.6
4	Pentanal II	2.000	0.200	2.7	2.0	2.7	2.3	3.6	2.3	2.8	2.4	2.4	0.0	3.0	2.0	2.6	1.8	2.4	2.9	2.4	3.3	2.0	1.0	2.7	2.2	3.4	3.2	3.0	5.7	1.9	2.6	3.4	1.7
5	Benzene	1.000	0.000	1.0	0.9	0.4	0.7	0.8	0.9	0.4	0.5	4.5	3.7	0.6	2.0	0.4	0.0	1.2	0.0	0.5	1.3	0.6	0.3	1.0	0.0	0.7	0.3	0.9	0.0	0.0	0.8	0.0	0.0
6	Butanal	2.000	0.250	1.0	1.0	2.2	2.5	2.4	1.1	0.4	0.4	3.5	0.6	1.0	2.2	1.2	1.8	0.4	1.4	0.9	1.0	0.8	1.5	0.9	0.0	0.2	0.2	0.4	2.0	0.7	1.2	2.3	0.0
7	Hexanal	2.000	0.167	2.9	3.2	2.7	1.3	3.4	2.2	2.4	2.1	5.3	2.0	2.6	1.3	3.1	2.6	2.1	0.8	2.1	1.8	1.5	1.0	2.3	0.0	3.0	1.8	3.6	0.0	0.5	2.8	0.0	2.2
8	Pyrrrole	1.250	0.000	4.7	4.8	2.4	5.0	1.3	4.6	2.2	1.0	15.4	4.9	4.9	2.0	3.1	2.9	2.0	2.9	3.2	1.2	3.1	3.7	1.2	0.0	4.9	4.5	2.6	7.1	2.2	3.2	3.3	0.5
9	Pyridine	1.000	0.000	2.0	1.8	5.0	0.7	1.9	2.4	1.6	1.5	3.1	4.3	2.1	1.0	2.0	3.0	1.8	2.9	0.9	6.0	2.0	3.1	2.3	0.0	0.6	1.4	2.4	1.5	1.6	2.9	0.6	1.5
10	Toluene	1.143	0.000	8.3	11.8	11.8	4.3	3.2	9.6	6.1	3.9	6.8	3.1	8.8	6.2	6.3	3.7	4.9	2.2	6.4	8.9	7.9	2.4	4.4	2.5	5.2	4.0	3.2	6.4	3.3	8.1	4.7	3.0
11	Furanone	1.000	0.500	5.0	5.4	5.4	4.4	2.7	4.3	3.2	4.4	1.1	0.6	6.6	1.7	5.9	0.5	4.1	2.2	2.8	5.5	3.6	1.2	3.6	0.0	4.9	4.6	5.9	0.7	1.0	5.3	3.1	1.9
12	Furaldehyde	0.800	0.400	2.6	1.8	3.7	2.2	3.2	2.0	1.0	1.0	5.3	2.3	2.2	2.6	1.7	2.1	1.5	3.8	2.8	1.7	1.1	1.9	1.3	0.0	0.7	1.7	1.8	2.6	1.5	2.5	1.5	1.2
13	C ₁ -pyrrole I	1.750	0.000	1.1	1.2	1.0	0.9	0.8	1.1	0.3	0.5	0.0	0.0	1.0	2.1	0.9	0.3	0.9	1.3	0.8	1.5	1.2	1.0	1.5	0.0	0.8	0.6	0.4	0.0	1.0	1.0	0.6	0.6
14	C ₁ -pyrrole II	1.750	0.000	0.8	0.4	0.3	0.4	0.5	0.8	0.7	1.1	0.0	0.0	0.7	1.3	0.6	0.5	0.7	1.4	0.5	0.7	0.5	0.8	0.5	0.0	0.4	0.7	0.3	0.0	0.3	0.9	0.3	0.4
15	Olefin C ₅	2.000	0.000	0.5	0.2	0.3	0.7	0.7	0.5	0.7	0.5	0.0	0.0	0.5	0.0	0.4	0.0	0.9	0.7	0.5	0.6	0.2	0.4	0.7	0.0	0.5	0.5	0.7	0.0	0.4	0.5	0.3	0.3
16	Fatty acid C ₆	2.000	0.333	0.0	0.3	0.0	0.0	0.0	0.8	0.5	0.4	0.0	0.0	0.0	0.2	1.5	0.0	0.8	1.2	0.7	0.6	0.4	1.0	0.3	0.0	0.0	0.0	0.0	0.0	0.6	1.3	0.0	0.9
17	Xylene	1.250	0.000	1.0	3.8	0.0	1.3	0.0	0.2	0.9	1.3	1.7	0.0	4.7	3.8	1.5	0.7	0.8	1.8	2.9	0.6	1.6	0.6	2.3	0.0	0.5	0.6	0.0	0.5	4.0	1.6	3.1	0.5
18	Styrene	1.000	0.000	0.8	0.0	0.0	0.0	0.0	0.0	0.2	0.3	3.0	0.0	0.0	0.7	0.1	3.1	0.4	0.5	0.3	0.5	0.8	0.3	1.2	0.0	0.0	1.2	0.0	13.2	0.9	0.4	1.1	0.4
19	C ₁ -furaldehyde	1.000	0.333	1.9	1.2	2.3	1.7	1.9	1.1	0.6	0.2	1.2	1.2	1.5	1.7	1.8	1.4	1.2	2.4	2.8	0.8	0.7	0.9	0.6	0.0	0.0	1.1	1.0	0.0	1.2	1.3	1.1	0.9
20	Phenol	1.000	0.167	4.2	4.5	7.6	11.2	5.9	4.4	3.3	3.2	11.2	5.5	7.1	2.1	5.4	1.5	2.1	1.1	2.6	3.7	12.8	11.3	11.0	0.0	5.1	5.7	4.0	6.3	1.0	3.8	2.1	4.1
21	Pyrrrolcarboxaldehyde	1.000	0.200	1.3	0.7	0.6	0.0	0.0	0.8	0.0	0.0	2.1	0.7	0.7	0.8	0.0	0.5	0.7	1.0	0.5	0.0	0.0	0.7	0.7	0.0	0.0	0.0	0.0	0.0	0.9	0.9	0.0	0.5
22	C ₁ -Phenol	1.143	0.143	0.5	0.0	0.0	0.0	0.0	0.0	0.2	0.6	2.0	0.0	0.0	0.3	0.0	0.3	0.0	0.0	0.2	0.4	0.5	1.0	0.4	0.0	0.0	0.0	0.0	0.0	0.9	0.4	0.0	0.6
23	Guaiacol (G)	1.143	0.286	7.0	4.2	6.6	9.7	16.1	14.6	18.2	19.6	0.0	0.7	7.5	0.0	0.0	1.1	6.2	4.0	2.6	7.3	17.0	14.7	15.3	0.0	11.2	13.2	12.3	11.5	3.5	5.5	6.9	14.5
24	C ₂ -Phenol I	1.143	0.143	0.5	0.5	1.9	2.2	9.1	0.8	4.2	1.7	0.0	0.0	2.8	3.1	6.5	0.0	1.1	0.0	0.6	0.8	0.9	0.7	0.9	0.0	4.0	2.5	1.7	0.0	0.0	1.2	0.0	1.7
25	Pyrrolidinedione	1.250	0.500	1.7	1.7	1.9	0.0	0.9	1.4	0.9	0.6	1.2	0.5	0.0	0.0	0.0	0.0	1.7	1.2	1.1	0.7	0.5	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0
26	C ₁ -indane	1.200	0.000	0.4	0.3	0.7	0.0	0.5	0.6	0.0	0.0	1.1	1.3	0.0	0.0	0.5	0.0	0.4	0.0	0.3	0.5	0.2	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.5	0.0	1.9	0.0
27	Olefin C ₈	2.000	0.000	0.4	1.8	0.6	0.0	0.7	0.4	0.1	0.8	0.0	0.0	0.2	0.4	0.5	0.0	0.5	0.0	0.3	0.5	0.3	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.3	0.4	0.0	0.8
28	Alkane C ₈	2.250	0.000	2.0	1.8	1.4	0.0	0.5	2.7	1.1	1.3	0.0	0.0	0.0	0.5	1.1	0.0	1.4	0.0	0.6	1.1	1.1	0.0	0.6	0.0	0.0	0.0	0.0	0.0	1.0	2.3	0.0	1.1
29	C ₂ -Phenol II	1.250	0.125	0.0	0.1	0.0	0.0	0.3	0.4	0.5	0.5	0.0	0.0	0.4	1.3	0.2	0.0	0.5	0.0	0.4	0.2	0.4	0.7	0.0	0.0	0.8	1.1	0.6	0.0	1.1	0.5	0.5	1.0
30	C ₂ -Phenol III	1.250	0.125	0.2	0.1	0.8	1.1	0.0	0.3	0.3	0.7	0.0	0.0	0.8	0.7	0.7	0.0	0.5	0.7	0.2	1.1	0.8	0.4	0.7	0.0	0.5	0.5	0.3	0.0	0.0	0.3	0.0	1.4
31	Methylguaiacol (MG)	1.250	0.250	0.3	0.2	0.2	0.2	0.5	0.9	3.3	4.1	0.0	0.0	0.5	0.0	0.0	0.0	0.6	0.0	0.0	0.4	1.7	1.2	1.3	0.0	2.8	2.4	1.9	0.0	0.4	0.3	1.2	4.1
32	Benzothiazole	0.714	0.000	0.0	0.6	0.4	0.3	0.0	0.0	0.2	0.0	1.4	1.7	0.7	5.8	0.4	3.0	0.3	1.7	1.6	0.3	1.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.5	0.2
33	Olefin C ₉ I	2.000	0.000	0.3	0.1	0.0	0.3	0.4	0.4	0.2	0.3	0.0	0.0	0.2	0.5	0.5	0.3	0.7	0.7	0.6	0.4	0.2	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.7	0.4	0.6	0.5
34	Vinylphenol	1.333	0.222	0.1	0.0	0.0	2.1	0.4	1.0	1.1	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.4	0.5	0.4	1.2	0.5	0.0	2.2	2.0	3.0	0.0	0.5	0.2	0.7	0.9
35	Alkane C ₉ II	2.222	0.000	0.8	0.9	0.4	0.8	0.3	0.3	0.5	0.4	0.0	0.0	0.4	0.4	0.8	0.5	0.5	0.5	0.5	0.4	0.7	0.4	0.3	0.0	0.0	0.0	0.0	0.0	0.3	0.9	0.5	0.4
36	Alkane C ₉ III	2.222	0.000	0.4	0.0	0.0	0.0	0.3	0.4	0.4	0.3	0.0	0.0	0.3	0.0	0.4	0.0	0.4	0.7	0.5	0.7	0.3	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.3
37	Ethylguaiacol (EG)	1.111	0.222	0.1	0.0	0.0	0.0	0.6	1.0	2.1	2.5	0.0	0.0	0.8	0.0	0.0	0.0	0.9	0.0	0.0	0.2	1.0	1.1	0.6	0.0	1.9	2.5	1.3	0.0	0.4	0.5	0.5	3.6
38	Vinylguaiacol (VG)	1.200	0.200	0.5	0.0	0.3	0.6	1.2	1.3	5.2	4.9	0.0	0.0	0.6	0.0	0.0	0.0	1.5	0.9	0.2	1.2	1.4	1.8	2.0	0.0	7.9	5.1	7.2	2.7	0.7	0.6	1.7	3.1
39	Olefin C ₁₀	2.000	0.000	0.2	0.3	0.1	0.0	0.0	0.3	0.3	0.4	0.0	0.0	0.2	0.5	0.4	0.0	0.7	0.7	0.4	0.3	0.2	0.2	0.4	0.0	0.1	0.3	0.0	0.1	0.4	0.5	0.6	0.8
40	Alkane C ₁₀	2.200	0.000	0.9	1.0	0.6	0.6	0.5	1.1	0.6	0.7	0.0	0.0	0.5	0.4	1.3	1.1	0.8	0.8	0.7	0.7	0.7	0.6	0.5	0.0	0.3	0.8	0.0	0.8	0.4	0.9	0.7	1.2
41	Syringol (S)	1.250	0.375	0.5	0.2	0.2	0.3	0.7	1.3	3.8	3.7	1.4	0.4	0.3	0.0	0.0	0.0	1.4	0.9	0.1	1.2	0.9	1.1	1.5	0.0	6.5	4.0	7.3	1.0	0.2	0.5	0.7	1.6
42	Olefin C ₁₁	2.000	0.000	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	0.3	0.6	0.5	0.8	0.7	1.5	0.4	0.3	0.2	0.2	0.									

43	Alkane C ₁₁	2.182	0.000	0.9	0.9	0.3	0.5	0.3	0.8	0.6	0.5	0.0	0.0	0.6	0.7	1.0	1.0	0.8	1.0	0.7	0.6	0.6	1.0	0.6	0.0	0.2	0.7	0.2	0.5	0.7	0.9	0.8	1.3
44	Acetoguaiacone (AG)	1.111	0.333	0.3	0.0	0.4	0.0	0.9	1.0	3.3	3.0	0.0	0.0	0.3	0.0	0.0	0.0	0.7	0.0	0.0	0.2	1.0	1.4	1.0	0.0	0.9	0.6	1.3	0.0	0.5	0.2	0.4	1.5
45	Alkane C ₁₂ I	2.167	0.000	1.3	1.3	0.7	0.0	0.6	0.1	0.3	0.4	0.0	0.0	0.2	0.0	0.2	0.0	0.8	0.0	0.0	0.6	0.0	0.0	0.5	0.0	0.3	0.3	0.0	0.0	0.0	0.2	0.8	0.3
46	Olefin C ₁₂	2.000	0.000	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.5	0.0	0.0	0.2	0.3	1.1	1.4	0.9	1.8	0.7	0.5	0.3	0.2	0.4	0.0	0.1	0.3	0.4	0.3	1.1	0.7	0.9	0.0
47	Alkane C ₁₂ II	2.167	0.000	0.0	0.0	0.0	0.7	0.0	1.2	0.6	0.8	0.0	0.0	0.9	1.0	2.6	3.4	1.0	1.8	1.2	0.9	0.7	0.6	0.7	0.0	0.5	1.8	0.5	1.8	0.6	1.3	1.1	1.9
48	Vinylsyringol (VS)	1.200	0.300	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.3	0.0	0.2	0.1	0.0	2.1	1.2	2.3	0.0	0.4	0.3	0.4	0.2
49	Olefin C ₁₃	2.000	0.000	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.6	0.3	0.0	0.5	0.8	0.4	0.2	0.1	0.0	0.2	0.0	0.0	0.4	0.0	0.0	0.6	0.3	0.6	0.7
50	Alkane C ₁₃	2.154	0.000	0.6	0.3	0.3	0.3	0.0	0.5	0.4	0.5	0.0	0.0	0.4	0.4	0.6	0.0	0.6	0.8	0.5	0.4	0.4	0.2	0.4	0.0	0.4	0.7	0.2	0.0	0.7	0.6	0.8	1.1
51	Olefin C ₁₄	2.000	0.000	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.0	0.0	0.0	0.4	0.3	0.7	0.7	0.5	0.3	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.6	0.7	0.4	0.7	0.9
52	Alkane C ₁₄	2.143	0.000	0.6	0.8	0.0	0.6	0.0	0.4	0.4	0.5	0.0	0.0	0.4	0.6	0.9	1.5	0.8	1.4	0.9	0.6	0.4	0.2	0.8	0.0	0.5	0.0	0.3	0.2	0.6	0.7	1.1	1.3
53	Olefin C ₁₅	2.000	0.000	0.2	0.3	0.0	0.0	0.0	0.0	0.2	0.4	0.0	0.0	0.0	0.2	0.8	0.0	0.7	1.4	0.4	0.3	0.1	0.3	0.3	0.0	0.0	0.0	0.0	0.0	1.2	0.5	1.1	0.9
54	Alkane C ₁₅ I	2.133	0.000	0.6	0.0	0.5	2.0	0.0	0.3	0.5	0.6	0.0	0.0	0.4	0.4	0.8	0.0	1.0	0.8	0.5	0.7	0.3	0.2	0.7	0.0	0.0	0.0	0.0	0.0	1.3	0.7	0.9	1.0
55	Alkane C ₁₅ II	2.133	0.000	1.1	0.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	12.5	1.8	0.6	0.0	0.0	0.0	0.2	0.5	0.0	0.3	0.4	0.4	0.0	0.5	0.8	0.7	0.4
56	Alkane C ₁₆	2.125	0.000	0.4	0.7	0.0	0.0	0.0	0.4	0.4	0.7	0.0	0.0	0.0	0.0	0.2	2.6	0.0	0.0	0.0	0.3	0.3	0.0	0.6	0.0	0.4	0.2	0.4	0.0	0.0	0.9	0.0	0.9
57	Fatty acid C ₁₆	2.000	0.125	2.5	0.8	0.0	1.5	0.0	0.0	2.1	3.3	11.4	10.9	0.0	3.7	0.0	4.8	10.0	2.7	12.5	2.8	0.9	1.9	0.5	32.3	0.0	0.0	3.8	0.0	1.2	5.4	3.6	3.2

^a Total abundance, percentage of total chromatographic area (total ion trace). The sum of the compound does not exactly amount to 100 because some chromatographic peaks were not used as IVs for the models (i.e., those with less than 0.3% of total chromatographic areas, or those exclusively found in a few (i.e., 3–5) samples). Atomic ratios H/C and O/C are shown in addition.

^b Roman numerals indicate different isomers.

Research highlights

- Soil organic C content and biodegradation rates can be predicted from Py-GC-MS data.
- Soil organic matter of different resilience has also different molecular structure.
- Soils with high C content are rich in methoxyphenols and low in alkyl structures.
- Pyrochromatograms converted in surface density plots inform about soils' C storage.