



RESEARCH

Open Access



Integrated modelling approaches for estimating total organic carbon, total nitrogen, total phosphorus, and ash content in forest floor samples using ATR-FTIR spectroscopy

Ruth C. Martín-Sanz^{1*}, Marina Getino-Álvarez¹, Valentín Pando¹, Francisco Lafuente¹ and M. Belén Turrión¹

Abstract

Key message ATR-FTIR spectroscopy combined with statistical modelling provides accurate predictions of forest floor properties: total organic carbon, total nitrogen, total phosphorus and ash. Performance decreases in carbonate-rich samples, particularly for organic carbon and the carbon to nitrogen ratio; excluding such samples improves reliability. This makes ATR-FTIR a fast and sustainable complement to conventional chemical analyses for forest monitoring and nutrient cycling studies.

Context Organic forest layers and their decomposition are essential for nutrient cycling, soil fertility, and carbon sequestration in forest ecosystems. While infrared spectroscopy is widely applied to assess physical and chemical attributes in soil samples, its effectiveness in evaluating forest floor samples remains poorly studied.

Aims This study evaluated the efficacy of Attenuated Total Reflection Fourier Transform Infrared Spectroscopy (ATR-FTIR) to predict forest floor properties: total organic carbon, total nitrogen, and their ratio, total phosphorus, and ash content.

Methods Forest floor material from a European network of forest plots was analyzed. Organic layers of *Pinus sylvestris*, *L.*, *Quercus* spp., and *Fagus sylvatica* L. in monospecific and mixed forest stands were examined, differentiating between species and litter layers: undecomposed, intermediate decomposed, and humified. ATR-FTIR spectra (4000 to 400 cm⁻¹) were recorded and analyzed. A modified partial least-squares method and a Generalized Linear Mixed Model with cross-validation were used to develop equations and predictive models for forest floor properties, combining ATR-FTIR spectra with laboratory data.

Results Accurate explanatory and predictive models were developed for forest floor samples using ATR-FTIR spectroscopy, highlighting its potential as a complementary or alternative tool to traditional chemical analyses. Total organic carbon predictions were more accurate when carbonate-containing samples were excluded from calibration.

Conclusions ATR-FTIR provides a reliable alternative for assessing forest floor properties, supporting sustainable forest management by improving monitoring of nutrient cycling and carbon storage.

Handling editor: Erwin Dreyer

*Correspondence:

Ruth C. Martín-Sanz
ruthcristina.martin@uva.es



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords Forest floor decomposition, Forest soil organic layers, Mid-infrared spectroscopy, Mixed forest, Monospecific forest, ATR-FTIR spectral prediction

1 Introduction

Decomposition of forest organic layers is essential for ecosystem functioning, regulating nutrient cycles, supporting soil fertility (Austin et al. 2014), and contributing to carbon (C) storage. As a key part of the biogeochemical cycle, litter decomposition releases essential nutrients such as nitrogen (N) and phosphorus (P) (Krishna and Mohan 2017). Understanding forest floor dynamics supports sustainable forest management and global climate goals by enhancing C sequestration and ecosystem resilience (IPCC 2019; Mäkipää et al. 2023). Analyzing litter composition reveals nutrient availability and cycling rates, while changes in its characteristics can indicate ecosystem disturbances or shifts. This knowledge helps guide forest management decisions regarding harvesting, prescribed burns, and other interventions, minimizing ecological impacts (Bravo et al. 2012).

Traditional methods for determining total organic carbon (TOC), total nitrogen (TN), and total phosphorus (TP) in soil and plant materials are accurate but time-consuming, costly, and environmentally demanding, especially when applied to large datasets (Angelopoulou et al. 2020). They often involve hazardous reagents (e.g., dichromate or sulfuric acid), produce toxic waste (Demattê et al. 2019), or rely on scarce resources like helium used in dry combustion procedures (Anderson 2018; Ramírez et al. 2022). In contrast, Fourier-transform infrared (FTIR) spectroscopy offers a reliable, rapid, environmentally friendly, and non-destructive alternative for analyzing soil and litter (Soriano-Disla et al. 2014). FTIR is a vibrational spectroscopic technique that detects molecular bonds via infrared absorption, producing spectra that reflect the sample's chemical structure. In soil and litter studies, FTIR spectra are widely used to identify functional groups (e.g., aliphatic, aromatic, carbonyl, phenolic) and infer compounds such as lignin, cellulose, proteins, and lipids. This enables qualitative and semi-quantitative characterization of organic matter composition and decomposition processes. Combined with multivariate statistical techniques such as partial least squares regression (PLSR), FTIR can also predict chemical properties quantitatively (Margenot et al. 2023; Sanderman et al. 2020). This dual function—chemical characterization and rapid quantification—makes FTIR especially valuable for ecological and biogeochemical

studies. Among FTIR methods, attenuated total reflectance (ATR-FTIR) is ideal for solid samples, requiring minimal preparation and avoiding water interference (Viscarra Rossel et al. 2022; Huck 2016). Quantitative spectroscopy typically requires calibration using multivariate regression techniques that relate absorbance at selected wavelengths to a reference property measured by conventional methods (Janik et al. 1998). PLS regression is especially prominent due to its ability to maximize covariance between predictors and response variables, making it well-suited for high-dimensional, collinear datasets (Wold et al. 1984). Despite its predictive strength, PLS cannot account for hierarchical data structures or include random effects, which restricts its explanatory power. In contrast, Generalized Linear Mixed Models (GLMMs) incorporate both fixed and random effects, making them better suited for grouped data and categorical variables.

Numerous studies have shown FTIR's strong potential to predict soil properties and organic matter composition, particularly in mineral soils and peatland horizons (e.g., Margenot et al. 2023; Tinti et al. 2015; Xie et al. 2011). These authors studied TOC, TN, exchangeable Ca, Al or K, and available P, among other soil properties, achieving high calibration accuracy and standard errors similar to traditional laboratory methods (e.g., $\sim 1 \text{ g kg}^{-1}$ for organic C; see Margenot et al. 2023; Tatzber et al. 2010; Viscarra et al., (Viscarra Rossel, et al., 2006); Xie et al. 2011). However, FTIR spectroscopy in forest litter layers is less common (but see, for example, Turrión et al. 2025) and has usually shown lower accuracy, likely due to the higher chemical and structural heterogeneity across decomposition stages. Dynamic changes in compounds such as lignin, cellulose, proteins, lipids, and polyphenols may cause overlapping or shifting absorption bands, increasing spectral variability and reducing model performance. Tatzber et al. (2010) found that grouping samples by carbonate content or forest type improved TOC predictions, especially in organic layers, and grouping inorganic samples by parent material enhanced carbonate content estimates. These findings highlight the sensitivity of FTIR-based models to sample heterogeneity and the limitations of applying general models to chemically diverse forest floor materials. To address this gap,

modeling approaches must be refined to account for forest floor organic matter complexity.

We hypothesized that ATR-FTIR spectroscopy can be used to develop calibration and predictive models tailored to forest floor analysis. Such models could help identify litter components linked to soil organic matter (SOM) decomposition and guide research on how soil properties and environmental or site-specific factors influence forest floor C stocks under climate change. Therefore, this study aimed to assess the performance of two modeling approaches—Partial Least Squares (PLS), widely used in spectral analysis, and Generalized Linear Mixed Models (GLMM), which incorporate environmental factors—in predicting forest floor properties using ATR-FTIR spectroscopy. Target variables included TOC, TN, TP, TOC/TN ratio, and ash content. Additionally, we evaluated the impact of carbonates on spectral definition and

identified key functional groups driving property predictions in both model types.

2 Material and methods

2.1 Forest floor dataset

We collected organic soil layer samples of the forest floor from experimental forest plots belonging to a European network (COST Action FP 1206 EuMiXFOR), following a transect from Southwest to Northeast Europe (Fig. 1).

The selected forest stands were as follows: Valberzoso (Castilla y León, northern Spain); Sierra de Montgrony (Catalonia, northeastern Spain); Hagenauer Forest (Upper Bavaria, southern Germany); and Niepolomice Forest (southern Poland). To facilitate reading this work, these four areas are hereafter referred to as Montaña Palentina–Spain, Pyrenees–Spain, Germany, and Poland, respectively. The ages of all the stands ranged from 50 to 57 years, and no silvicultural intervention has been done in the last ten

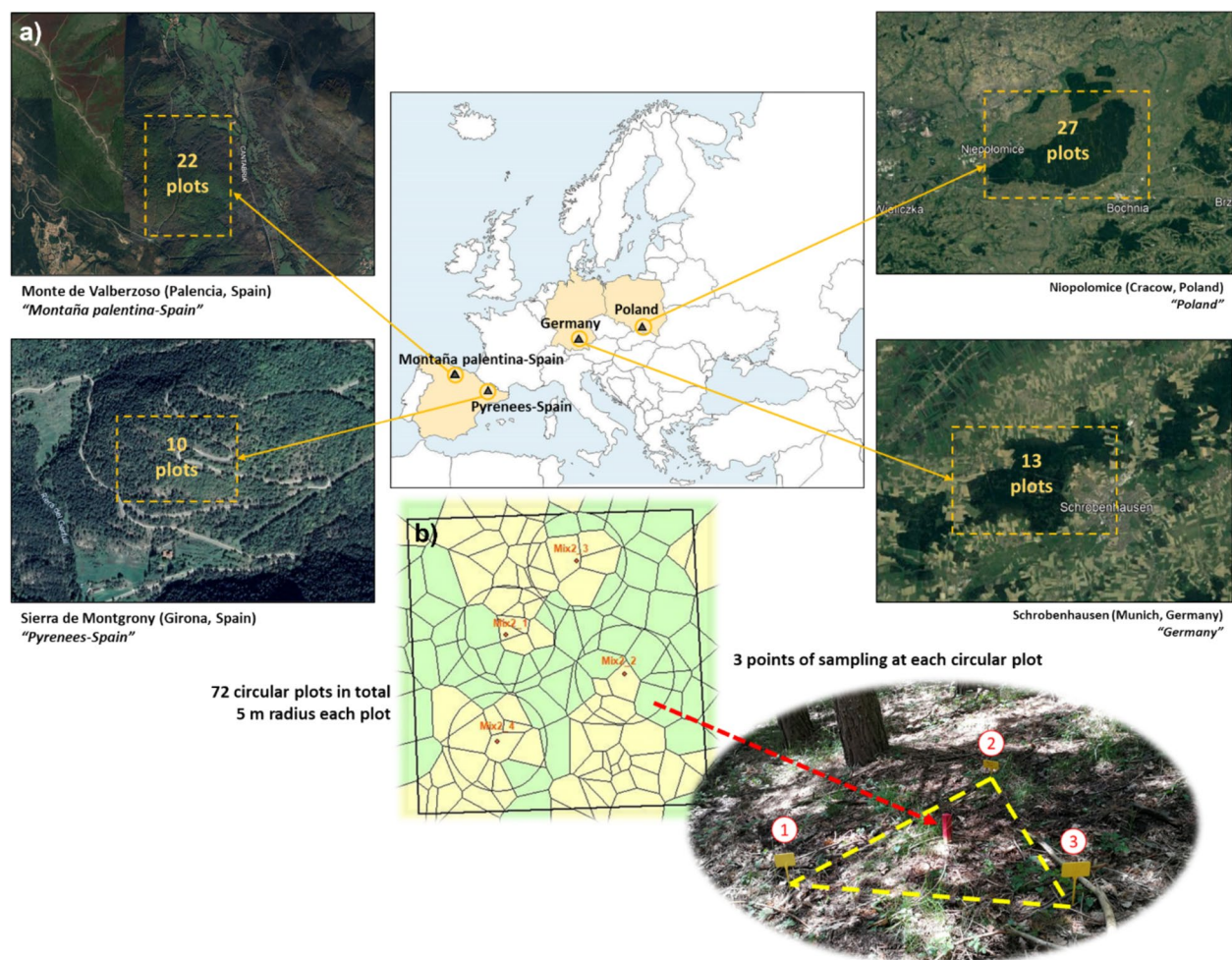


Fig. 1 Diagram of the study location and experimental setup. **a** Location of the study sites and number of plots sampled at each location. **b** Example of Voronoi diagrams to determine the percentage of species mixing and the number of litter collection units

years (Błońska et al. 2018; Getino-Álvarez et al. 2023). Mean annual temperature (MAT) and mean annual precipitation (MAP) range between 7.5 and 8.7 °C and between 670 and 1060 mm, respectively. Plot characteristics are summarized in Table 1

Denomination of sampled area	Location	Altitude (m)	Slope (%)	Geology	MAT (°C)	MAP (mm)	Soil type	Soil texture	Soil pH	Soil CO ₃ (%)
Montaña palentina	Spain	Valberzoso Forest, Palencia (Spain)	1140–1220	10–17	Sandstones, conglomerates and lutites	8.71	1150	Inceptisol	Sandy loam	4.1–6.40
Pyrenees	Spain	Sierra de Montgrony, Girona (Spain)	1080–1230	35–45	Sandstone, loams, and limestone	7.59	51	Entisol	Silt loam and clay loam	7.8–8.93
Germany	Hagenauer Forest, Neuburg-Schrobenhausen (Germany)	449–473	1	Tertiary sediments, coarse and fine sands	7.58	27	Spodosol	Sand and loamy sand	3.6–5.90	
Poland	Niepolomice Forest, Wieliczka (Poland)	191–231	0–4	Pleistocene sands	8.46	70	Spodosol	Sand and loamy sand	3.4–5.80	

Note: MAT, mean annual temperature; MAP, mean annual precipitation; soil type by Soil Survey Staff (2014)

1. Stand characteristics were obtained from the EuMiXFOR data set available in <https://doi.org/10.5061/dryad.8v04m>. To ensure consistent ecological contrasts and reduce confounding variability, the sampling design focused on pure and mixed stands of *Pinus sylvestris* L., *Fagus sylvatica* L., and *Quercus* sp., with Scots pine being the common species in all mixed stands. Rather than capturing the full diversity of European forests, this transect-based approach allowed us to evaluate model performance across meaningful gradients of species composition and forest floor conditions.

2.2 Experimental set-up

A total of 72 circular plots of 5 m radius were established: 22 in pure stands of Scots pine (*P. sylvestris*), 12 in pure stands of beech (*F. sylvatica*), 10 in pure stands of oak (*Quercus* sp.), 12 in mixed stands of pine and oak, and 16 in mixed stands of pine and beech (Fig. 1). Voronoi diagrams were used to estimate the mixing percentage by assigning areas of influence to each species present (Fig. 1). The main characteristics of the selected plots, including tree species, stand type, basal area, tree density, and biomass, are summarized in Table 2.

Forest floor material was collected between July and October of 2018 and 2019. Before collecting, coarse woody debris, such as large branches, was carefully removed from the surface. Leaf litter was then collected using a 25×25 cm wooden frame placed at three locations, each 1 m from plot center. The material collected from these three locations was combined to form a single litter collection unit per plot, serving as the basis for subsequent laboratory processing.

2.3 Litter sorting and sample preparation

In the laboratory, material from each litter collection unit was separated according to decomposition stage, tree species, and stand type, based on the hypothesis that both the chemical composition and the spectral response of forest floor material are affected by these factors. Specifically, the material was sorted by decomposition stage into the following: fresh, undecomposed litter (L), partially decomposed material with recognizable plant structures (F), and highly decomposed, humified material (H; Appendix A-Fig. 3). In monospecific plots, the three layers were assigned to the dominant species (pine, oak, or beech). In mixed plots, the L and F layers were separated by species, while the H layer was analyzed without species distinction due to its advanced state of decomposition. Additionally, for the F layer in mixed plots, a composite subsample without species separation was also analyzed to allow comparisons between combined-species and species-specific signals. This classification resulted in three samples per monospecific plot and six per mixed plot.

All litter fractions were dried at 60 °C for 48 h and then weighed (± 0.01 g) to calculate the total dry biomass (B) of each fraction per hectare (Mg ha^{-1}). Subsequently, samples were finely ground using a Retsch mixer mill MM 301 and stored in airtight containers for chemical and ATR-FTIR spectroscopic analysis.

2.4 Chemical analysis

Dry combustion using a LECO CNS928 elemental analyzer was used to determine total carbon (TC) and TN. Also, IC was determined by dissolving carbonates in excess HCl, followed by back titration of the remaining acid with NaOH (Bashour and Sayegh 2007). The difference between TC and inorganic carbon (IC) was used to calculate TOC, from which the TOC/TN ratio was subsequently derived. Only 42 samples from calcareous soils in the Pyrenees (Spain) contained IC. Ash content was calculated after calcination in a muffle furnace at 480 °C for 5 h. Subsequently, ashes were dissolved with an acidic solution ($\text{HCl} + \text{HNO}_3$), and the concentration of TP was determined using the colorimetric method of vanadomolybdophosphoric acid (Kitson and Mellon 1944).

2.5 Spectral measurements

Forest floor samples of L, F, and H layers were scanned with ATR-FTIR (Nicolet iS50 FT-IR-ATR). Sample reflectance was recorded at a resolution of 4 cm^{-1} and 64 scans per spectrum to ensure high-quality data acquisition. Background spectra were collected under identical conditions in ambient air before each sample measurement. The ATR crystal was cleaned with ethanol between

Table 1 Stands and soil characteristics

Denomination of sampled area	Location	Altitude (m)	Slope (%)	Geology	MAT (°C)	MAP (mm)	Soil Type	Soil texture	Soil pH	Soil CO ₃ (%)
Montaña palentina-Spain	Valberoso Forest, Palencia (Spain)	1140-1220	10-17	Sandstones, conglomerates and lutites	8.7	1150	Inceptisol	Sandy Loam	4.1-6.4	0
Pyrenees-Spain	Sierra de Montgrony, Girona (Spain)	1080-1230	35-45	Sandstone, loams, and limestone	7.5	951	Entisol	Silt Loam and Clay Loam	7.8-8.9	3.8-56.6
Germany	Hagenauer Forest, Neuburg-Schrobenhausen (Germany)	449-473	1	Tertiary sediments, coarse and fine sands	7.5	827	Spodosol	Sand and Loamy Sand	3.6-5.9	0
Poland	Niepolomice Forest, Wieliczka (Poland)	191-231	0-4	Pleistocene sands	8.4	670	Spodosol	Sand and Loamy Sand	3.4-5.8	0

MAT Mean Annual Temperature, MAP Mean Annual Precipitation; Soil Type by Soil Survey Staff (2014)

Table 2 Main characteristics of the sampled plots

Denomination of sampled area	Type of mass	N° plots	Tree species	% Psy	DBH (cm)	AB(m ² ha ⁻¹)	N (trees ha ⁻¹)	Dry Biomass (Mg ha ⁻¹)
Montaña palentina-Spain	Pure pine	7	<i>Pinus. sylvestris</i>	99-100	15.5-41.2	61.2-93.4	767-1407	5.9-19.3
	Pure oak	7	<i>Quercus petraea</i>	0	8.5-34.5	24.5-64.4	896- 2687	7.9-16.8
	Mixed pine + oak	8	<i>P. sylvestris</i> + <i>Q. petraea</i>	11-90	8.5-36.3	33.5-71.3	768- 1792	7.5-18.5
Pyrenees-Spain	Pure pine	3	<i>Pinus sylvestris</i>	100	13.3-31.8	16.3 -44.9	512-896	7.1-10.1
	Pure oak	3	<i>Quercus pyrenaica</i>	0	7.5-21.0	16.1-28.4	1408- 1920	4.4-12.5
	Mixed pine + oak	4	<i>P. Sylvestris</i> + <i>Q. pyrenaica</i>	11-81	7.9-29.8	25.7-36.3	640- 1152	4.7-9.8
Germany	Pure of pine	4	<i>Pinus sylvestris</i>	99-100	21.3-41.0	33.9-63.7	256-640	15.8-18.4
	Pure of beech	4	<i>Fagus sylvatica</i>	0	7.7-30.5	24.7-33.1	896-1408	12.2-19.4
	Mixed pine + beech	5	<i>P. sylvestris</i> + <i>F. sylvatica</i>	8-64	7.5-38.9	22.0-54.5	768-1280	10.5-14.2
Poland	Pure of pine	8	<i>Pinus sylvestris</i>	100	15.6-39.5	32.9-72.8	512-1408	3.0-21.1
	Pure of beech	8	<i>Fagus sylvatica</i>	0	8.6-39.0	23.6-41.4	384-1024	5.7-15.2
	Mixed pine + beech	11	<i>P. sylvestris</i> + <i>F. sylvatica</i>	14-87	7.4-40.6	19.8-54.6	512-1280	6.5-25.9

%Psy *Pinus sylvestris* percentage, DBH Diameter at breast height (measured at 1.3 m above the ground) in centimeters, AB Basal area per hectare, N Tree density, Dry Biomass Aboveground dry biomass per hectare, Mixed plots include a combination of species indicated in the "Tree species" column

samples. Special attention was given to spectral pre-processing (Zhang et al. 2024). Baseline correction was applied to all spectra to minimize slope variations caused by scatter and particle size differences (Griffiths et al. 2007; Farnet-Da Silva et al. 2017). An automated peak-picking tool (5% sensitivity) ensured consistent peak identification across all spectra (Appendix A Table 7). Spectra were processed in OMNIC software using baseline correction (connecting line method), smoothing (Savitzky-Golay method), and normalization (by dividing each peak height by the total peak height, ensuring comparability between samples; Tinti et al. 2015). Interpretation was based on standard peak assignments for organic matter (Tatzber et al. 2010; Tinti et al. 2015). Appendix A Table 7 summarizes peak positions, integration ranges, and their chemical assignments.

2.6 Data analysis

Predictive models for forest floor properties (TOC, TN, TP, ash%, and TOC/TN) were developed by combining ATR-FTIR spectra with laboratory data. Two modeling approaches were tested: Partial Least Squares (PLS) regression with spectral data only and Generalized Linear Mixed Models (GLMM) incorporating spectral and categorical variables. PLS models were fitted using the PROC PLS and GLMMs with PROC GLIMMIX in SAS 9.4 (SAS codes are shown in Appendix B).

Due to the presence of carbonates in calcareous soils from Pyrenees–Spain, TOC and TOC/TN models were evaluated using both the full dataset (“all-samples”) and a subset excluding carbonate-containing samples (“no-carbonate-samples”) to test for improved performance. PLS addressed multicollinearity in predictors, while GLMMs included random effects and categorical variables such as litter type, decomposition layer, and site.

Data were randomly split into training (calibration) and testing (validation) subsets with consistent proportions. Model performance was evaluated using the coefficient of determination (R^2), Lin’s concordance correlation coefficient (LCCC), root mean square error (RMSE), mean absolute error (MAE), coefficient of variation (CV), and the ratio of prediction deviation (RPD), defined as the ratio between the standard deviation and the RMSE. Full modeling details are provided in Appendix C, and all data obtained in this study is available at <https://zenodo.org/records/16315149> (Martín-Sanz et al. 2025).

3 Results

3.1 Forest floor litter ATR-FTIR spectra

ATR-FTIR spectra of forest floor samples (Fig. 2) exhibited a consistent pattern with 19 distinct absorption peaks. Signals were mainly distributed across the regions 2918–2850 cm⁻¹, 1732–1693 cm⁻¹, 1646–1500 cm⁻¹, 1450–1300 cm⁻¹, 1238–1167 cm⁻¹, 1029 cm⁻¹, and around 898 cm⁻¹.

The 1646 cm⁻¹ peak was absent in pine litter, and variation occurred around 1450–1300 and 1238–1260 cm⁻¹. The 898 cm⁻¹ peak, associated with carbonate content, was entirely absent in samples from non-carbonate soils. Consequently, this wavenumber received a coefficient of zero in TOC and TOC/TN prediction models for those samples.

3.2 Descriptive statistics

Table 3 summarizes TOC, TN, TOC/TN, TP, and ash contents in the entire dataset and for the calibration and validation subsets. The values and variation ranged in the concentration of the studied variables between the complete dataset and the calibration and validation datasets were highly similar, both for all samples and for the

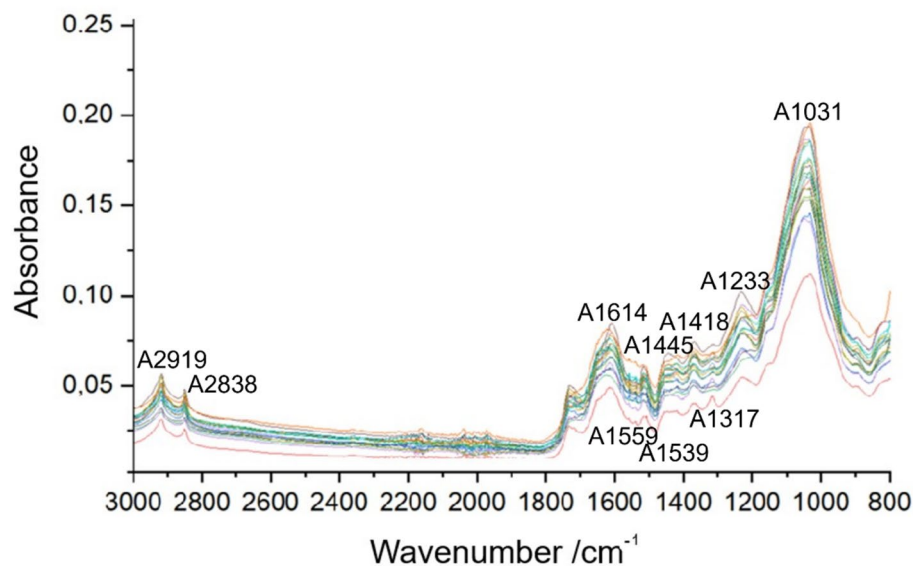


Fig. 2 ATR-FTIR spectra of forest floor samples highlighting major functional groups linked to organic matter composition

models with no-carbonate samples. This indicates that our data provided essential and adequate databases for building and validating prediction models.

3.3 Calibration and prediction of TOC, TN, TP, TOC/TN ratio and ash content in forest floor samples

3.3.1 Model performance and effect of carbonate samples

The performance of GLMM and PLS models for each response variable, using both the full data set and the data set excluding samples with carbonates, is shown in Table 4. Both models performed well, but the GLMM models performed better, with lower MAE and RMSE and higher R^2 values for all variables studied. Lin's concordance coefficient was also higher in the GLMM models, although following the McBride (2005) criterion, this coefficient was poor ($LCCC < 0.90$) for all variables except for TOC and ash, which were moderate ($LCCC = 0.90$ and 0.92 , respectively). RPD was higher in GLMM models. RPD for TOC, TN, and ash content in the GLMM was good ($RPD \geq 2$), meaning a high reliability of the models; being moderate for TOC/TN and TP ($1.4 \leq RPD < 2$). However, for the PLS models, only TOC and ash percentage showed good reliability according to RPD; TN and TOC/TN models showed a moderate RPD and poor for TP. By removing samples with carbonates, GLMM performance for TOC and TOC/TN improved substantially, but not equally in the PLS models (Table 4). As previously observed in the spectral data, the absence of the 898 cm^{-1} peak in non-carbonate samples likely contributed to the improved performance of models excluding such samples.

The predictive model results for both the training and validation datasets (GLMM and PLS) are shown in Tables 5 and 6, respectively. All performance, accuracy, and predictive capacity metrics (RMSE, MAE, RPD, R^2 , and LCCC) improved in the GLMM by removing the carbonate-containing samples, both for TOC and TOC/TN (Table 5). RPD improved from moderate to good for both TOC and TOC/TN, and the LCCC from poor to moderate for TOC. PLS models also improved for TOC across all metrics, but performance for TOC/TN slightly declined, though not significantly. This may be due to PLS relying on ATR-FTIR wavelengths to extract key factors to adjust the model according to the cross-validation criterion. By reducing the number of samples used in the model (by removing carbonate samples), the ATR-FTIR wavelengths differ from those initially included in the models with all data. This difference in the ATR-FTIR peaks (our response variable) affects the choice of factors in PLS. As a result, PLS used fewer factors (one instead of two), reducing accuracy.

Comparing the different metrics used to check performance, accuracy, and predictive capacity of these models, PLS models generally showed better predictive capacity than GLMM. Removing carbonate-containing leaf samples improves the performance of both models (GLMM and PLE), even if it reduces the sample size.

3.3.2 PLS regression coefficients

The coefficients from the PLS regression model for the 19 wavenumbers used to estimate TOC, TN, TP, TOC/TN ratio, and ash content are shown in Appendix D-Table 8. For TOC, the highest positive coefficients were observed

Table 3 Descriptive statistics of the studied variables for the explanatory models and their calibration and validation datasets

Full dataset											
TOC (%)						TP (mg/kg)					
Ent		Cal		Val		Ent		Cal		Val	
269		206		63		255		195		60	
43.67		43.59		34.59		65.31		64.81		66.94	
5.72		5.85		12.66		16.54		16.60		16.40	
19.98		24.22		18.94		23.46		23.46		27.84	
62.77		62.77		81.65		102.16		102.16		98.65	
45.18		45.24		31.22		66.84		66.45		67.23	
13.10		13.42		36.60		25.33		25.61		24.50	
Carbonate-free subset						Carbonate-free subset					
TOC (%)						TOC/TPN					
Ent		Cal		Val		Ent		Cal		Val	
227		181		46		227		181		46	
33.63		33.72		42.94		42.94		42.93		42.98	
5.58		5.71		10.57		10.57		10.97		10.57	
19.98		19.98		19.98		19.98		19.98		19.98	
51.16		51.16		59.85		59.85		59.85		59.85	
8.96		9.01		8.55		8.55		8.55		8.55	
84.22		83.87		85.91		85.91		83.87		85.91	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02		31.32	
36.60		36.08		38.53		36.60		36.08		38.53	
81.65		81.65		81.65		81.65		81.65		81.65	
31.22		31.02		31.32		31.32		31.02			

Table 4 Performance metrics comparison for GLMM and PLS models across target variables

	Generalized Linear Mixed Model (GLMM)						Partial Least Square model (PLS)						
	Full dataset			Carbonate-free subset			Full dataset			Carbonate-free subset			
	TOC (%)	TN (%)	TOC/TN	TP (mg/kg)	Ash (%)	TOC/TN	TOC (%)	TN (%)	TOC/TN	TP (mg/kg)	Ash (%)	TOC (%)	TOC/TN
<i>n</i>	269	269	269	255	255	227	269	269	269	255	255	227	227
CV (%)	5.7	13.2	18.7	16.9	32.3	4.7	6.6	21.7	19.6	33.8	30.0	5.5	23.2
RMSE	2.47	0.18	6.41	11.87	4.14	2.10	2.90	0.23	7.55	12.77	4.36	2.38	7.8
MAE	1.40	0.09	2.80	6.38	2.02	1.28	1.91	0.18	5.34	9.95	2.99	1.67	5.58
RPD	2.32	2.05	1.97	1.39	2.62	2.32	2.05	1.61	1.70	1.30	2.49	2.34	1.56
R ²	0.82	0.74	0.74	0.49	0.86	0.86	0.76	0.61	0.65	0.41	0.84	0.82	0.59
LCCC	0.90	0.86	0.85	0.68	0.92	0.92	0.86	0.75	0.77	0.56	0.88	0.90	0.74

n number of observations, CV coefficient of variation, RMSE root mean square error, MAE mean absolute error, RPD ratio of prediction to deviation (SD/RMSE), R² coefficient of determination, LCCC Lin's concordance correlation coefficient. For TOC and TOC/TN, results are shown for both the full dataset and the carbonate-free subset

Table 5 Performance of GLMM for target variables in training and validation datasets

	Training data				Validation data							
	Full dataset				Carbonate-free subset				Full dataset			
	TOC (%)	TN (%)	TP (mg/kg)	Ash (%)	TOC (%)	TOC/TN	TP (mg/kg)	Ash (%)	TOC (%)	TN (%)	TP (mg/kg)	TOC/TN
<i>n</i>	206	206	195	195	181	181	60	60	63	63	60	46
CV (%)	6.07	12.78	18.76	36.20	4.90	15.52	18.13	24.56	5.06	16.21	18.13	16.16
RMSE	2.65	0.17	12.16	4.73	2.10	5.23	12.14	3.02	2.22	0.228	12.14	5.37
MAE	1.55	0.093	6.18	2.36	1.31	2.33	9.52	2.11	1.59	0.14	9.52	3.14
RPD	2.21	2.05	1.36	2.32	2.72	2.36	1.35	3.50	2.39	1.79	1.35	2.17
R ²	0.80	0.77	0.47	0.82	0.87	0.82	0.46	0.92	0.83	0.69	0.46	0.79
LCCC	0.88	0.87	0.68	0.90	0.93	0.90	0.65	0.96	0.90	0.82	0.65	0.88

n number of observations, CV coefficient of variation, RMSE root mean square error, MAE mean absolute error, RPD ratio of prediction to deviation (SD/RMSE), R² coefficient of determination, LCCC Lin's concordance correlation coefficient. For TOC and TOC/TN, results are shown for both the full dataset and the carbonate-free subset

Table 6 Performance of PLS regression models for target variables in the training and validation datasets

Training data				Validation data											
Full dataset				Carbonate-free subset					Full dataset					Carbonate-free subset	
	TOC (%)	TN (%)	TOC/TN	TP (mg/kg)	Ash (%)	TOC (%)	TOC/TN	TOC (%)	TN (%)	TP (mg/kg)	Ash (%)	TOC (%)	TOC/TN		
n	206	206	206	195	195	181	181	63	63	60	60	46	46		
CV (%)	6.6	17.3	21.9	20.2	34.9	6.2	23.9	6.4	17.1	17.0	30.0	5.9	19.8		
RMSE	2.90	0.24	7.59	13.08	4.57	2.66	8.05	2.86	0.24	11.35	3.69	2.54	6.57		
MAE	1.95	0.18	5.36	10.27	3.11	1.95	5.64	1.82	0.18	8.84	2.73	1.91	4.91		
RPD	2.04	1.52	1.65	1.27	2.40	2.15	1.53	2.11	1.69	1.45	2.87	2.00	1.79		
R ²	0.76	0.57	0.63	0.38	0.83	0.79	0.85	0.78	0.66	0.53	0.88	0.75	0.69		
CCC	0.77	0.63	0.66	0.48	0.77	0.80	0.68	0.86	0.76	0.66	0.91	0.86	0.77		

n number of observations, *CV* coefficient of variation, *RMSE* root mean square error, *MAE* mean absolute error, *RPD* ratio of prediction deviation, *R*² coefficient of determination, *LCCC* Lin's concordance correlation coefficient. For TOC and TOC/TN, results are shown for both the full dataset and the carbonate-free subset

at 1233, 1259, 1317, 1370, and 1445 cm^{-1} . These trends remained largely consistent between the full dataset and the subset excluding carbonate-affected samples.

For TN, the most relevant wavenumbers were 1204, 1418, 1512, 1539, and 1559 cm^{-1} . The wavenumbers contributing most to the TOC/TN ratio overlapped with those for TN but showed opposite signs, reflecting the expected inverse relationship. For TP, the strongest positive coefficients were observed at 1204, 1233, 1512, and 1539 cm^{-1} . Finally, ash content exhibited its highest coefficients at the same wavenumbers as TOC but with opposite signs, confirming the inverse relationship between the organic and inorganic fractions.

3.3.3 GLMM regression equations and significance

Using GLMM, 14 equations were developed for TOC, TN, the TOC/TN ratio, TP, and ash percentage. The coefficients for the 19 wavenumbers and intercepts for 13 sample groups defined by litter type (beech, pine, oak, pine + beech, pine + oak) and decomposition stage (L, F, and H) are provided in Appendix D-Table 9. All factors and their interactions included were statistically significant. In addition, a general intercept was also provided for cases where specific group information is unavailable. The coefficients illustrated the relative importance of each wavenumber across litter types and decomposition stages.

Additionally, Appendix D-Table 9 includes a comparison of coefficients for TOC and TOC/TN between the full dataset and the carbonate-free subset. Appendix D-Table 5 lists the p -values associated with each GLMM coefficient, identifying statistically significant contributions to prediction performance. For TOC, the most significant ($p < 0.001$) included 2919 cm^{-1} with a positive coefficient and with negative coefficients 1031, 1233, 1445, 1559, and 2838 cm^{-1} . These remained consistent across both datasets. In the carbonate-free subset, 1512 cm^{-1} gained significance, while 1559 cm^{-1} lost it. For TN, the wavenumber at 1370 cm^{-1} and 1539 cm^{-1} with positive coefficients, and those at 1445 cm^{-1} and 2919 cm^{-1} with negative coefficients were highly significant. For TOC/TN, significant peaks were found at 1418 cm^{-1} and 1445 cm^{-1} ($p < 0.001$), and at 1031 cm^{-1} and 1711 cm^{-1} in the carbonate-free subset. For TP, the wavenumbers with the greatest weight and $p < 0.001$ were 1445 and 1539 cm^{-1} , with additional significant contributions from 1370 to 1711 cm^{-1} ($p < 0.01$). Notably, 1445 cm^{-1} consistently exhibited significant weight ($p < 0.001$) for TOC, TN, TP, and the TOC/TN ratio. For ash content, significant wavenumbers ($p < 0.001$) included 1317, 1614, and 2919 cm^{-1} . Most wavenumbers significant for ash content did not show statistical

relevance with the TOC, TN, or TOC/TN, properties directly related to organic matter.

4 Discussion

4.1 Comparative strengths and limitations of PLS and GLMM models

We found differences in performance and explanatory power between the two modeling approaches. Our findings showed that GLMM models outperformed PLS models in predictive accuracy across all metrics, particularly for TOC, TN, and ash content. While PLS is commonly used for high-dimensional prediction (Sharma et al. 2019), GLMMs offered better explanatory power by accounting for data structure and handling non-normal data (Bolker et al. 2009; Turrión et al. 2025). Model performance was assessed using the classification proposed by Chang et al. (2001), widely used in soil chemometrics (Sanderman et al. 2020). Both methods effectively quantified TOC, TN, TOC/TN ratio, phosphorus, and ash content in forest litter. Excluding carbonate-containing samples randomly, both during calibration and prediction processes, substantially improved model performance, likely due to the reduced spectral interference and increased chemical homogeneity.

The use of MIR spectroscopy in forest litter is less common and typically yields lower accuracy than in mineral soils. While $R^2 > 0.90$ and $\text{RPD} > 2.7$ have been reported for TOC and TN in mineral soils, studies on litter show reduced performance, with R^2 around 0.70–0.75 and $\text{RPD} < 2.0$ (Tatzber et al. 2010; Zhang et al. 2024). In our study, TOC and TN predictions improved previous leaf litter-based results, though TOC/TN ratio predictions remained less reliable, consistent with findings by Barthès et al. (2008) and Ludwig et al. (2008), likely due to error propagation from the individual TOC and TN predictions. Model reliability was also affected by the narrow range of TOC (20–63%) and TN (0.6–2.3%) in our samples. Broader calibration ranges, as those used by Haghi et al. (2021) and Tatzber et al. (2010), generally yield stronger correlations and enhance model accuracy (Davies and Fearn 2005). Those studies explicitly state that their sample sets include mineral and organic soil layers, which accounts for the broader variability and the presence of low TOC contents. Future research should include samples with lower TOC values to improve generalizability.

Tailoring calibration datasets to specific chemical and spectral properties is essential. Tatzber et al. (2010) emphasized distinguishing between organic and mineral layers and grouping samples by vegetation and carbonate content. Guerrero et al. (2010) showed that combining local and regional samples improves N prediction in soils. These approaches are equally relevant for ATR-FTIR analysis of forest litter, where spectral and chemical complexity can compromise model performance.

4.2 Impact of carbonates on spectral predictive accuracy

Excluding forest floor litter samples from carbonate-rich soils significantly improved model performance for predicting TOC and TOC/TN ratio. This aligns with prior studies showing that carbonates interfere with soil organic carbon (SOC) detection by introducing overlapping spectral bands (e.g., Madari et al. 2006). In our case, carbonate-related signals appeared only in the H layer of litter from calcareous soils, where underlying mineral horizons contained up to 56.6% CO_3^{2-} (Table 1). These signals likely result from incorporation of carbonate-rich particles from the mineral soil due to mechanical mixing (e.g., soil fauna or runoff), atmospheric dust deposition, or capillary rise and carbonate precipitation at the soil surface under dry conditions (Jouquet et al. 2006; Grunewald et al. 2006). Their presence increases chemical heterogeneity, reducing model adequacy, while excluding them enhances dataset homogeneity and strengthens spectroscopic correlations (Tatzber et al. 2010; Xie et al. 2011). Although models developed using the full dataset, including samples with carbonates, yielded satisfactory results, removing carbonate-rich samples led to better statistical performance (higher R^2 , lower RMSE; Tables 4, 5 and 6). This highlights a trade-off between improved accuracy and broader applicability. However, with only 42 carbonate-affected samples, building a reliable separate model was not feasible. Therefore, expanding datasets to increase the representation of carbonate-rich litter is essential for developing specific calibrations tailored to calcareous environments. An alternative would be to wash litter samples to remove carbonates without reducing sample size. This approach may be feasible for the L horizon, composed of freshly fallen plant material. However, it is unsuitable for the more decomposed F and H layers, where organic matter is intermixed with mineral particles. Washing these layers could result in material loss or alteration, compromising analytical integrity. Given these limitations, we excluded carbonate-affected samples to maintain model consistency and reliability. Accordingly, we provide two equations for both PLS and GLMM models: one using the full dataset and another excluding carbonate-rich samples. If carbonate-free conditions are known, the latter should be used for optimal accuracy; otherwise, the full dataset model is recommended, though with slightly lower performance.

4.3 Spectral interpretation of key functional groups

Infrared spectroscopy is effective for characterizing SOM due to its sensitivity to C–H, C–O, and C–N functional groups. Our results highlight the utility of ATR-FTIR with PLS in predicting leaf litter components by identifying key functional groups in organic matter analysis. For TOC, strong positive coefficients appeared at 1233 and 1259 cm^{-1}

(C–O stretching, O–H bending in carboxylic structures; Calderón et al. 2011), at 1317 and 1370 cm^{-1} , attributed to phenolic and aliphatic sites (Parker 1971), and at 1445 cm^{-1} reflecting lignin rings and aliphatic C–H bending (Parker 1971). Peaks at 2838 and 2919 cm^{-1} (aliphatic C–H stretching; Artz et al. 2008) also showed high positive coefficients. The 1559 cm^{-1} band appeared only in the full dataset, suggesting it is linked to carbonate content in the samples, as carbonate bands typically occur around 1500 cm^{-1} (Tinti et al. 2015). This divergence between models may reflect differences in how each approach captures spectral information and emphasizes distinct features of the dataset.

For TN, significant coefficients were at 1204 cm^{-1} (amide III, C–O/O–H in COOH, and P–O in aryl compounds; Tatzber et al. 2010; Tinti et al. 2015), 1418 cm^{-1} (O–H bending, C–N stretching, and CH_2 scissoring; Verchot et al. 2011), and 1512, 1539, and 1559 cm^{-1} (aromatic C=C, amide II, and carboxyl C; Janik et al. 2007; Tinti et al. 2015). These regions indicate nitrogenous functional groups such as amides and nitrogen-containing aromatic compounds, including aromatic amines and possible degradation products of amino acids (e.g., tryptophan or pyrrole-like structures), which are common in proteins and other N-rich substances in leaf litter. Notably, the same bands governed the TOC/TN ratio, but with reversed signs, reflecting the inverse C–N relationship.

For TP, the model detects both mineral and organically bound forms, including inorganic phosphates and typical organic P compounds such as phospholipids, nucleic acids, phosphate sugars (e.g., phytate), or ATP, which contain P–O and P=O bonds detectable in this spectral region. Key features included 1512 and 1539 cm^{-1} (aromatic vibrations, amide II, and lignin skeletal modes; Janik et al. 2007; Tinti et al. 2015) and 1204 and 1233 cm^{-1} (C–O and O–H in carboxylic groups, and P–O/P=O bonds; Tatzber et al. 2010; Tinti et al. 2015). The strong association of TP with TOC-related bands reflects the predominance of organic P in forest floor litter. Lastly, the highest coefficients for ash content coincide with those of TOC but with opposite signs, reflecting their inverse relationship.

Overall, PLS and GLMM models revealed partial agreement for variables closely associated with organic matter, particularly TOC and TN. For TOC, both models identified 1233 cm^{-1} (C–O and O–H in carboxylic groups, and P–O/P=O bonds) and 1445 cm^{-1} (CH and NH of amide II; aliphatic C–H deformation; Tatzber et al. 2010; Tinti et al. 2015) as relevant. For TN, 1539 cm^{-1} was significant in both models, confirming consistent detection of N-containing functional groups. In contrast, for TP and ash content, the overlap between models was limited. This divergence is expected given the complex chemical nature of these variables: TP includes organic and inorganic forms, while ash content reflects mineral content. These results underline the

complementary strengths of both methods: PLS captures dominant spectral patterns across all samples, while GLMM enhances interpretability by grouping and accounting for the hierarchical structure of the study design, including site as a random effect and fixed effects related to litter species, litter layer, and their nesting. This allows GLMM to detect site- or litter-specific responses that may be masked in PLS.

5 Conclusions

ATR-FTIR spectroscopy shows strong potential for predicting key chemical properties of forest floor organic layers, offering a rapid and environmentally friendly complement to conventional analyses. While elemental analyzers remain more widely available, ATR-FTIR is

particularly advantageous when resources are limited or large numbers of samples need to be screened. Importantly, its accuracy depends on the sample composition, with carbonate-rich conditions reducing predictive performance due to interference with spectral signals. These findings highlight the need for specific calibration strategies, possibly through larger and more diverse calibration datasets, approaches to handle chemical heterogeneity, or site- and species-specific models, to enhance robustness and applicability. Future research should also explore the integration of FTIR-based approaches into decision-support tools for sustainable forest management programs, supporting and enhancing assessments of nutrient cycling or carbon stock monitoring.

Appendix A

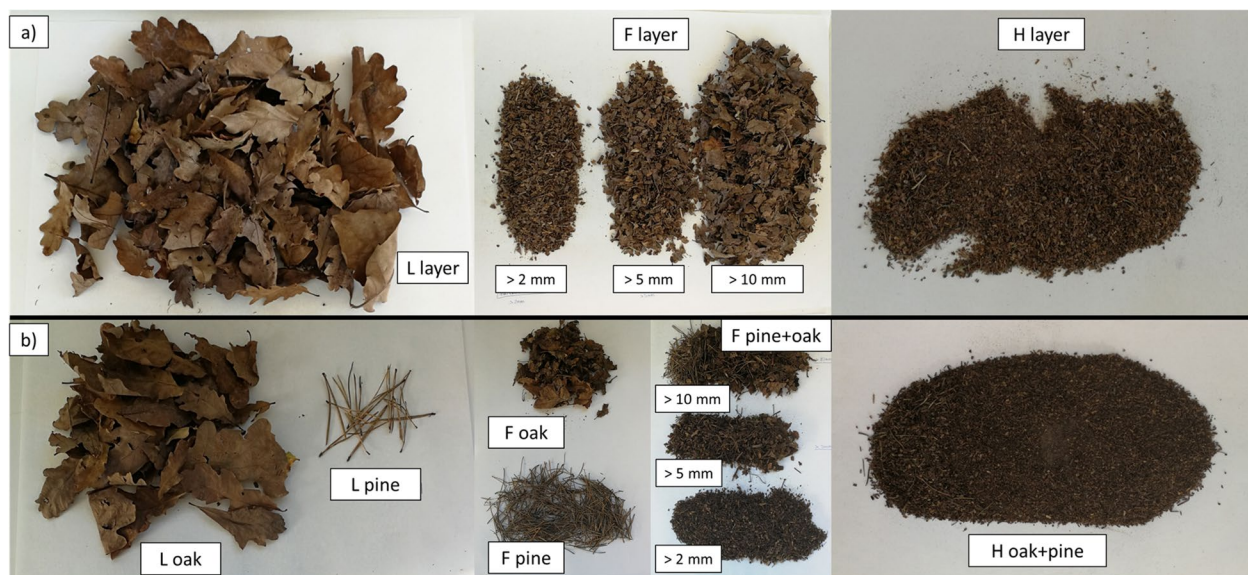


Fig. 3 Example of the different states of decomposition of leaf litter in: **a)** pure oak plot, where three layers can be distinguished (L, F and H) and **b)** mixed pine-oak plot, where five layers can be distinguished (L of pine, L of oak, F of pine, F of oak, F of pine-oak and H)

Table 7 Conventional band assignments used for the evaluation of the recorded infrared spectra of organic layers and associated peaks (Tatzber et al. 2010; Tinti et al. 2015)

Peak (cm ⁻¹)	Range of bands (cm ⁻¹)	Conventional band assignments inside the integrated bands
2919	2989–2877	Aliphatic CH stretch
2838	2873–2836	Aliphatic CH stretch
1729	1775–1711	C=O stretch
1711	1775–1711	C=O stretch
1646	1691–1642	C=O stretch (amide I) of tertiary amides (1680–1630 secondary, 1650 primary); C=N stretching of unsaturated linkages containing nitrogen (open chain and unconjugated cyclic); amide II of primary amides (CO–NH ₂), ureas
1614	1642–1569	Amide II of primary amides; carboxylates and amides, R–O–NO ₂
1559	1569–1488	Aromatic skeletal vibrations; amide II and aromatic C=C; aromatic skeletal (lignin)
1539	1569–1488	Aromatic skeletal vibrations; amide II and aromatic C=C; aromatic skeletal (lignin)
1512	1569–1488	Aromatic skeletal vibrations; amide II and aromatic C=C; aromatic skeletal (lignin)
1445	1479–1444	CH and NH of amide II; aliphatic C–H deformation
1418	1444–1408	Malonic and/or benzoic acids, C–H deformation and C–O stretching of phenolic groups, amide III of primary amides, carbonates
1370	1403–1354	C–O of phenolic OH, COO ⁻ and O–H; C(CH ₃) ₃ , CH ₃ bend, 37 NO ₃ , CO–CH ₃
1317	1342–1307	Malonic acids; R–SO ₂ –OR' (sulfonamides); C–N of aromatic amines
1259	1293–1256	Benzoic acids, C–O of aryl ethers; R–O–NO ₂ , P=O, C–O of phenolic groups
1233	1256–1198	C–O and OH of COOH, C–O of aryl ethers and phenols; acetates, P–O aryl
1204	1256–1198	C–O and OH of COOH, C–O of aryl ethers and phenols; acetates, P–O aryl
1154	1185–1144	C–OH of aliphatic OH; carbohydrates, R–SO ₂ –OR' (sulfonamides)
1031	1056–945	C–O stretch (1030–1080 = carbohydrate region); Si–OH, RHC=CH ₂ , sulfones, P–O aryl
898	899–866	Carbonate

Appendix B. SAS codes

B1. SAS code for the PLS models

```
proc pls data=datos method=pls nfac=10 cv=one cvtest plots=all
details;
model y=A1031--A2919/solution;
output out=datos1 predicted=pred xscore=V yresidual=res;
ods output XLoadings=ejesTOC;
run;
```

B2. SAS code for GLMM done for variables with beta probability distribution

```

proc glimmix data=datos plots=all maxopt=1000 pconv=0.1;
class site litter layer id;
model y=layer litter(layer) A1031--A1614 A1711--A2919/dist=gamma
link=log chisq solution;
random site/subject=id;
random _residual_/subject=id group=site*layer;
lsmeans layer litter(layer) /bylevel om cl pdiff ilink
plot=meanplot(ilink) lines;
output out=datos4 predicted(ilink)=pred lcl(ilink)=predl
ucl(ilink)=predu predicted(noblup ilink)=predmarginal lcl(noblup
ilink)=predml ucl(noblup ilink)=predmu student=sresid
student(noblup)=sresmarginal;
run;

```

B3. SAS code for GLMM done for variables with gamma probability distribution

```

proc glimmix data=datos plots=all maxopt=1000 pconv=0.1;
class site litter layer id;
model y=layer litter(layer) A1031--A1614 A1711--A2919/dist=gamma
link=log chisq solution;
random site/subject=id;
random _residual_/subject=id group=site*layer;
lsmeans layer litter(layer) /bylevel om cl pdiff ilink
plot=meanplot(ilink) lines;
output out=datos4 predicted(ilink)=pred lcl(ilink)=predl
ucl(ilink)=predu predicted(noblup ilink)=predmarginal lcl(noblup
ilink)=predml ucl(noblup ilink)=predmu student=sresid
student(noblup)=sresmarginal;
run;

```

Appendix C. Detailed modelling procedures**C1. Partial Least Squares regression (PLS)**

Partial Least Squares approach is a robust multivariate linear regression method used to construct predictive models when there are many predictor variables that are highly

collinear. It finds a linear regression model after projecting the response and predicting variables in a new space, where each axis explains the maximum variance possible of the response variable. In this way, this method is similar to a Principal Component Analysis (PCA) followed by a

Multiple Linear Regression, but with the advantage that the variance of the dependent variable is considered. PLS has several advantages in comparison to other statistical methods used for similar purposes. The latent variables reduce the dimension of the original spectra and capture the maximum covariance between the observed values and the spectra. The PLS approach aims to extract the latent variables of the spectral data and develop a model with an optimal number of latent variables and unaffected by multicollinearity (Abdi 2010). For each PLS model, we chose the number of axes that minimized the predicted residual sum of squares calculated following the standard “leave-one-out” cross-validation given by the SAS PROC PLS procedure.

C2. Generalized Linear Mixed Models (GLMM)

Generalized linear mixed models are an extension of linear models that allow the response variable to be not normal, permit the use of a link function and the linear predictor can contain random effects. The main advantage of GLMM *vs.* PLS is its ability to include categorical factors in the model. In addition to the ATR-FTIR peak values, our GLMM model used “site” (a factor with four levels: 1=Montaña palentina-Spain, 2=Poland, 3=Germany, 4=Pyrenees-Spain) as a random effect, “leaf litter species” (a factor with five levels: 1=beech, 2=oak, 3=pine, 4=pine+beech, 5=pine+oak) as an inter-subject fixed factor, and “layer type” (a factor with three levels: 1=L fresh litter, 2=F fragmented litter, and 3=H humified litter) as an intra-subject fixed factor. We assumed that litterfall type (leaf litter species factor) is nested in each layer type, thus originating 13 levels in the factor: L-Beech, L-Oak, L-Pine, F-Beech, F-Oak, F-Pine, F-Pine+Beech, F-Pine+Oak, H-Beech, H-Oak, H-Pine, H-Pine+Beech and H-Pine+Oak. The sampling points were considered as experimental units. The probability distribution considered for the response variable was a beta distribution, and the variance structure considered for the within-subjects factor was Banded Main Diagonal. The logit function was used as the link function. As an exception, for the TOC/TN ratio model the probability distribution used was the gamma and the link function was the log.

Therefore, denoting by $\mu_{k(j);x_m}$ the expected value of the proportion of the dependent variable in each case in the k -type litterfall at layer type j and with values x_m ($m=1,2,\dots,19$) for the ATR-FTIR spectra peaks. We divided variables data by 100 for TOC, TN and ash percentage, and by 10^6 for TP. Thus, the model formulation was as follows:

$$\log\left(\frac{\mu_{k(j);x_m}}{1 - \mu_{k(j);x_m}}\right) = b_{0(k(j))} + \sum_{m=1}^{19} b_m x_m + \delta_l \quad (1)$$

$$\text{with } b_{0(k(j))} = \mu + \beta_j + \gamma_{k(j)} \quad (2)$$

and where: μ is the model’s general intercept; β_j is the effect of layer type j on the intercept model; $\gamma_{k(j)}$ is the effect of litterfall type k within layer type j on the intercept of the model; b_m is the linear effect of the ATR-FTIR peak m and δ_l is the random effect of the site l , with $\delta_l \sim N(0, \sigma_\delta^2)$, being σ_δ^2 the variance of the random effect. We suppose that δ_l are independent of each other.

For model errors we considered a $N(0, \sigma_{jl}^2 + \sigma^2)$ distribution, where σ_{jl}^2 is the within-subjects variance at layer type j at site l and σ^2 is the residual inter-subjects’ random variance of the model. The random errors are independent of each other and independent of the random effects δ_l .

Therefore, we estimated 11 variance parameters (σ_δ^2 , σ^2 and 9 σ_{jl}^2 parameters) by maximizing the residual pseudo-likelihood using the global Quasi-Newton optimization technique. From the obtained values, we then estimated the coefficients $b_{0(k(j))}$ and b_m of the final equation provided by the linear mixed model. Comparison of least squares means of factor levels was done using simple pairwise *t* tests. The normality assumptions for the random effect and the model residuals were tested using the Kolmogorov-Smirnov test of the conditional studentized residuals and the marginal studentized residuals, together with their respective normal probability plots. The predicted values of the model $\hat{\mu}_{k(j);x_m}$ were obtained using the inverse of the link function, that is, by the expression:

$$\hat{\mu}_{k(j);x_m} = \frac{1}{1 + \exp(-b_{0(k(j))} - \sum_{m=1}^{19} b_m x_m)} \quad (3)$$

These predicted values were 100 $\hat{\mu}_{k(j);x_m}$ for TOC, TN and ash percentage and 1000000 $\hat{\mu}_{k(j);x_m}$ for TP.

Importantly, for TOC/TN ratio the equation model was slightly different:

$$\log(\mu_{k(j);x_m}) = b_{0(k(j))} + \sum_{m=1}^{19} b_m x_m + \delta_l \quad (4)$$

being $\mu_{k(j);x_m}$ the expected value of TOC/TN. In addition, the predicted values for this model were obtained using this expression:

$$\hat{\mu}_{k(j);x_m} = \exp\left(-b_{0(k(j))} - \sum_{m=1}^{19} b_m x_m\right) \quad (5)$$

C3. Model Validation

The dataset was randomly split into 210 observations for training (calibration dataset, 77%) and 64 observations for testing (validation dataset, 23%). This split was consistent for all dependent variables, and for TP and ash %, the same percentage was maintained despite missing

values, necessitating a reconsideration of the allocation of data. When modelling with the no-carbonate-samples, the dataset was randomly split into 181 observations for training (80%) and 46 observations for testing (20%), for both TOC and TOC/TN.

C4. Reported metrics to assess models' performance

- The coefficient of determination R^2 of the linear regression between predicted and observed values:

$$R^2 = 1 - \frac{\sum_i^n (y_i - y_p)^2}{\sum_i^n (y_i - \bar{y})^2} \quad (6)$$

where n = number of observations, y_i = actual value, y_p = predicted value and $\bar{y}$ = mean value of y .

- The Lin's concordance correlation coefficient LCCC (Lin 1989) between predicted and observed values. It measures the linear correlation between the predicted and true values but penalized by a bias term, allowing assessing the agreement between the observed and predicted values with respect to the 1:1 line (Lawrence and Lin 1989). This measure takes values between -1 and 1, and it measures the concordance of the data to the 45° line.

$$LCCC = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\bar{x} - \bar{y})^2} \quad (7)$$

where n = number of observations, ρ = correlation coefficient between the observed and predicted values, σ_x and σ_y = SD of observed and predicted values, respectively, $\bar{x}$ and $\bar{y}$ mean values for observed and predicted values, respectively.

- The root-mean-square error (RMSE) that represents the standard deviation of the prediction errors (or residuals), indicating how far the predicted values are, on average, from the observed actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_p)^2} \quad (8)$$

where n = number of observations, y_i = actual value and y_p = predicted value.

- The mean absolute error (MAE) of the residuals measures the average of the absolute differ-

ences between the actual values and the model's predictions. The MAE indicates the average prediction error magnitude, disregarding the direction of the error (positive or negative). By using absolute values, it ensures all errors contribute positively to the metric.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_p| \quad (9)$$

where n = number of observations, y_i = actual value and y_p = predicted value.

- The coefficient of variation (CV) that is the ratio of the RMSE by the mean.
- The ratio of prediction deviation (RPD) is the ratio between the standard deviation and the RMSE:

$$RPD = \frac{SD}{RMSE} \quad (10)$$

A higher R^2 value indicates a higher correlation and better fit between the predicted and observed values. We followed (McBride 2005) to interpret Lin's CCC: $LCCC < 0.90$ poor, $0.90 \leq LCCC < 0.95$ moderate, $0.95 \leq LCCC < 0.99$ substantial, and $LCCC \geq 0.99$ excellent). RMSE and MAE are essential metrics for assessing regression model accuracy, as both are expressed in the same units as the predicted variable, facilitating interpretation. RMSE, sensitive to large errors due to its use of squared differences, is particularly useful when outliers significantly affect model performance. In contrast, MAE, based on average absolute differences, is less influenced by outliers and emphasizes consistency in error magnitude. Together, these metrics provide complementary insights, with lower values for both metrics indicating better predictive accuracy and reliability of the model. Finally, RPD measures how well a predictive model can explain the variability of the actual data compared to the error it makes in predictions. It compares the "natural spread" of the observed data with the average error of the models' predictions. Thus, the higher the RPD value, the higher the reliability and robustness of the model. According to (Chang et al. 2001): $RPD < 1.4$ indicates poor predictive ability, $1.4 \leq RPD < 2$ suggests moderate predictive ability and $RPD \geq 2.0$ indicates good predictive ability. All these metrics were reported for models trained using all the data and for models trained only on the training set.

Appendix D

Table 8 PLS regression model coefficients for each wavenumber (PEAK) and studied variable obtained for the full dataset and the carbonate-free subset

PEAK	b_m		Carbonate-free subset				
	Full dataset						
	TOC	TN	TOC/TN	TP	Ash	TOC	TOC/TN
A898	75.013	-1.392	65.855	-93.115	-52.066	0.000	0.000
A1031	-45.839	-1.598	-7.659	-33.535	81.407	-37.464	-13.765
A1154	-124.283	1.765	-114.314	131.042	250.247	15.119	-154.343
A1204	40.234	14.167	-234.922	404.965	-2.898	70.152	-70.344
A1233	216.873	5.437	33.040	239.929	-421.343	242.734	130.534
A1259	133.515	0.755	96.399	-11.956	-273.048	129.169	140.862
A1317	123.679	-4.747	101.174	-91.470	-145.501	386.958	170.269
A1370	115.160	-1.929	63.551	-75.448	-215.417	235.760	190.819
A1418	-26.948	9.045	-221.235	140.898	76.074	-161.733	-89.187
A1445	130.638	-4.926	250.793	-71.419	-264.867	4.965	270.557
A1512	31.453	14.794	-354.749	535.399	-124.734	220.502	-249.081
A1539	-10.293	11.805	-317.746	374.210	26.886	118.662	-227.208
A1559	-11.947	6.623	-244.859	210.135	59.878	-95.989	-230.905
A1614	43.236	0.692	-110.440	185.524	33.363	-123.121	-98.577
A1646	31.933	2.530	-6.984	67.485	-86.026	38.878	0.031
A1711	-53.962	1.028	83.836	-171.019	42.546	-231.155	214.851
A1729	14.522	-4.397	191.809	-196.716	-71.216	34.207	186.940
A2838	73.049	-7.271	303.301	-232.472	-212.046	117.047	221.687
A2919	49.812	-6.305	237.486	-198.633	-136.060	51.099	171.217
Intercept	10.799	-0.525	57.579	-7.206	67.933	-10.389	15.190

Note: The coefficients (b_m) and intercept represent the contribution to the model's predictions of each wavenumber and baseline value, respectively

Table 9 GLMM coefficients obtained for each wavenumber and intercepts for each group across the studied variables

Layer+Litter	$b_{0(i/jk)}$	PEAK	b_m												
				Full dataset					Carbonate-free subset						
				TOC	TN	TOC/TN	Ash	TP	TOC	TN	TOC/TN	Ash	TP	TOC	TOC/TN
F+L+Beech	2.269	-4.648	4.117	-7.113	4.471	-1.185	0.641	A988	-1.806	-0.307	1.158	-3.010	-5.051	0.000	0.000
L+Oak	2.128	-4.450	3.831	-7.048	4.602	-1.132	0.360	A1031	-7.058	-1.786	-0.387	-2.602	-0.104	-3.209	3.916
L+Pine	2.353	-4.231	3.780	-6.580	4.000	-1.088	0.340	A1154	-2.821	-6.849	-0.655	-13.150	5.944	-1.374	0.943
F+Beech	3.228	-4.435	3.927	-6.786	4.661	-1.212	0.474	A1204	-7.734	13.860	-12.138	18.979	24.195	-2.275	-14.689
F+Oak	2.125	-4.310	3.773	-6.776	4.763	-1.333	0.293	A1233	13.501	7.127	4.246	-7.792	-39.336	14.332	15.203
F+Pine	2.301	-4.319	3.920	-6.678	4.455	-1.137	0.492	A1259	2.796	-1.806	2.761	-4.267	-28.233	7.045	6.130
F+PineBeech	2.183	-4.246	3.746	-6.871	5.167	-1.237	0.331	A1317	3.820	-8.138	4.544	-5.495	-38.907	12.127	13.118
F+PineOak	2.217	-4.243	3.817	-6.671	4.688	-1.238	0.370	A1370	14.101	55.253	3.370	45.388	-87.621	23.667	8.828
H+PineBeech	2.258	-4.298	3.801	-6.824	4.726	-1.159	0.390	A1418	0.475	-4.241	-38.700	-21.284	50.100	-5.390	-47.889
H+PineOak	2.256	-4.170	3.739	-6.693	4.424	-1.141	0.275	A1445	-33.289	-58.358	47.607	-46.102	29.982	-32.084	51.049
H+Beech	2.300	-4.262	3.821	-6.865	4.547	-1.122	0.394	A1512	5.292	5.878	-9.726	7.529	-19.793	11.537	-6.581
H+Oak	2.218	-4.223	3.741	-6.883	4.874	-1.195	0.332	A1539	3.903	15.498	-8.341	19.801	-27.292	4.508	-4.580
H+Pine	2.218	-4.223	3.741	-6.883	4.874	-1.195	0.332	A1559	-14.594	-1.129	-2.649	-7.702	26.634	-6.015	1.400
General Int.	2.251	-4.322	3.846	-6.796	4.580	-1.184	0.410	A1614	-5.598	-1.922	-0.540	-5.928	19.886	-3.663	3.932
								A1646	-3.660	3.341	-3.351	1.693	-16.894	0.000	0.000
								A1711	-2.598	-14.078	12.693	-18.751	-8.526	-1.993	23.054
								A1729	-8.277	12.338	-17.042	2.360	22.184	-4.358	-15.729
								A2838	-24.781	12.788	-12.344	6.442	14.214	-40.174	-2.944
								A2919	16.650	-28.319	19.142	-22.663	-54.070	34.595	15.054

Note: Intercepts (b_0) are shown for each combination of decomposition layer and litter type. For TOC and TOC/TN analysis was performed using two datasets: all available samples and a subset excluding carbonate-containing samples. Predictions are calculated using the following equations:

- $\hat{y} = \exp(b_{0ijk} + \sum_{m=1}^{19} b_{mxm})$ for TOC/TN
- $\hat{y} = 100(1 + \exp(-b_{0ijk}) - \sum_{m=1}^{19} b_{mxm}))^{-1}$ for TOC, TN and Ash
- $\hat{y} = 1000000(1 + \exp(-b_{0ijk}) - \sum_{m=1}^{19} b_{mxm}))^{-1}$ for TP

The coefficients $\text{b0}_{(ijk)}$ represent the intercepts for each group, and bm are the coefficients associated with specific wavenumbers, and xm are the values corresponding to the respective wavenumbers. The General intercept (General Int.) represents the global intercept when combining data across all groups, providing an overall baseline for the model predictions

Table 10 *P*-values for the coefficients of the GLMM equations for each wavenumber (PEAK) and variable

PEAK	b_m						
	Full dataset					Carbonate-free subset	
	TOC	TN	TOC/TN	TP	Ash	TOC	TOC/TN
A898	0.2005	0.7988	0.6368	0.0352	0.2808	-	-
A1031	<.0001	0.1689	0.8742	0.1018	0.9833	<.0001	<.0001
A1154	0.3246	0.1284	0.8845	0.0134	0.5240	0.6419	0.8288
A1204	0.1169	0.0814	0.1307	0.0867	0.1712	0.6527	0.0940
A1233	0.001	0.2847	0.5646	0.4581	0.0213	0.0057	0.0851
A1259	0.3955	0.6602	0.4845	0.3263	0.0073	0.0339	0.0860
A1317	0.2638	0.0977	0.4479	0.5478	0.0002	0.0111	0.1140
A1370	0.1038	<.0001	0.7607	0.0027	0.0039	0.0089	0.4552
A1418	0.9117	0.6238	0.0002	0.0215	0.0033	0.4574	<.0001
A1445	<.0001	<.0001	<.0001	0.0004	0.1615	<.0001	<.0001
A1512	0.0448	0.3264	0.1029	0.2462	0.0976	0.0006	0.2739
A1539	0.1075	0.0029	0.0984	0.0009	0.0049	0.2148	0.3368
A1559	0.001	0.8181	0.5813	0.1360	0.0115	0.2534	0.7784
A1614	0.0142	0.5144	0.8875	0.1540	0.0002	0.3019	0.4518
A1646	0.0208	0.0174	0.1790	0.3467	0.0018	-	-
A1711	0.4326	0.0087	0.0491	0.0030	0.4041	0.5832	0.0007
A1729	0.0114	0.0612	0.0160	0.7539	0.0927	0.3089	0.0287
A2838	0.0002	0.2952	0.3800	0.6432	0.5020	<.0001	0.8466
A2919	<.0001	0.0024	0.0682	0.0324	0.0006	<.0001	0.1933

Note: total organic carbon (TOC), total nitrogen (TN), TOC/TN ratio, total phosphorus (TP), and ash content. P-values indicate the statistical significance of each coefficient (bm) for predicting the respective variable. Significant values (i.e., $p < 0.001$) are in bold case, providing insights into the contribution of each wavenumber to the model's predictions

Acknowledgements

This research has been possible thanks to the people who participated in the EuMIXFOR project, establishing the experimental sites that were sampled in this study. We would like to thank the invaluable help of all the people who have participated in field work, laboratory analysis, and statistical and forestry assistance. Specifically, we would like to thank Carmen Blanco and Juan Carlos Arranz from the University of Valladolid for their advice in the laboratory analyses, and Andrés Bravo-Núñez (University of Valladolid and Agresta S. Coop.) for his help with statistical analysis.

Authors' contributions

Conceptualization: María-Belén Turrión and Francisco Lafuente; Methodology: María-Belén Turrión, Francisco Lafuente and Ruth C. Martín-Sanz; Formal analysis and investigation: María-Belén Turrión, Francisco Lafuente, Ruth C. Martín-Sanz, Marina Getino-Álvarez and Valentín Pando; Writing - original draft preparation: María-Belén Turrión and Ruth C. Martín-Sanz; Writing - review and editing: María-Belén Turrión, Ruth C. Martín-Sanz and Francisco Lafuente; Funding acquisition: Marina Getino-Álvarez and María-Belén Turrión; Supervision: María-Belén Turrión and Francisco Lafuente. The authors read and approved the final manuscript.

Data availability

Full modeling details are provided in Appendix C and all data obtained in this study is available at <https://zenodo.org/records/16315149> (Martín-Sanz et al. 2025)

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Instituto Universitario de Investigación en Gestión Forestal Sostenible (iuFOR), University of Valladolid, Palencia, Spain.

Received: 7 April 2025 Accepted: 30 August 2025

Published online: 11 December 2025

References

- Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdiscip Rev Comput Stat 2:97–106. <https://doi.org/10.1002/wics.51>
- Anderson ST (2018) Economics, helium, and the U.S. federal helium reserve: summary and outlook. Nat Resour Res 27:455–477. <https://doi.org/10.1007/s11053-017-9359-y>
- Angelopoulou T, Balafoutis A, Zalidis G, Bochtis D (2020) From laboratory to proximal sensing spectroscopy for soil organic carbon estimation—a review. Sustainability 12:443. <https://doi.org/10.3390/su12020443>
- Artz RR, Chapman SJ, Robertson AJ et al (2008) FTIR spectroscopy can be used as a screening tool for organic matter quality in regenerating cutover peatlands. Soil Biol Biochem 40:515–527. <https://doi.org/10.1016/j.soilbio.2007.09.019>
- Austin AT, Vivanco L, González-Arzac A, Pérez LI (2014) There's no place like home? An exploration of the mechanisms behind plant litter—decomposer affinity in terrestrial ecosystems. New Phytol 204:307–314. <https://doi.org/10.1111/nph.12959>
- Barthès BG, Brunet D, Hien E et al (2008) Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples. Soil Biol Biochem 40:1533–1537. <https://doi.org/10.1016/j.soilbio.2007.12.023>
- Bashour SII, Sayegh AH (2007) Methods of analysis for soils of arid and semi-arid regions. Food and Agriculture Organization of the United Nations, Rome

- Błońska E, Gruba P, Klammerus-Iwan A et al (2018) What characteristics of soil fertility can improve in mixed stands of scots pine and European beech compared with monospecific stands? *Commun Soil Sci Plant Anal* 49:237–247. <https://doi.org/10.1080/00103624.2017.1421658>
- Bolker BM, Brooks ME, Clark CJ et al (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24:127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Bravo F, Rodríguez F, Ordóñez C (2012) A web-based application to simulate alternatives for sustainable forest management: SIMANFOR. *For Syst* 21:4–8. <https://doi.org/10.5424/fs/2112211-01953>
- Calderón FJ, Mikha MM, Vigil MF et al (2011) Diffuse-reflectance mid-infrared spectral properties of soils under alternative crop rotations in a semi-arid climate. *Commun Soil Sci Plant Anal* 42:2143–2159. <https://doi.org/10.1080/00103624.2011.596243>
- Chang C-W, Laird DA, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. *Soil Sci Soc Am J* 65:480–490. <https://doi.org/10.2136/sssaj2001.652480x>
- Davies AMC, Fearn T (2005) Back to basics: the principles of principal component analysis. *Spectrosc Eur* 16:20–23. <https://www.spectroscopyeurope.com/td-column/back-basics-principles-principal-component-analysis>
- Demattê JAM, Dotto AC, Bedin LG, Sayão VM, Souza AB (2019) Soil analytical quality control by traditional and spectroscopy techniques: constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* 337:111–121. <https://doi.org/10.1016/j.geoderma.2018.09.010>
- Farnet-Da Silva AM, Ferré E, Dupuy N et al (2017) Infra-red spectroscopy reveals chemical interactions driving water availability for enzyme activities in litters of typical Mediterranean tree species. *Soil Biol Biochem* 114:72–81. <https://doi.org/10.1016/j.soilbio.2017.06.026>
- Getino-Álvarez M, San-Martín R, Pretzsch H et al (2023) Assessing soil C stock and C to N ratio of soil organic matter under mixed pine-beech forests at different scales. *Eur J Forest Res* 142:1081–1098. <https://doi.org/10.1007/s10342-023-01578-5>
- Griffiths PR, de Haseth JA, Winefordner JD (2007) Fourier transform infrared spectrometry. Wiley
- Grunewald G, Kaiser K, Jahn R, Guggenberger G (2006) Organic matter stabilization in young calcareous soils as revealed by density fractionation and analysis of lignin-derived constituents. *Org Geochem* 37:1573–1589. <https://doi.org/10.1016/j.orggeochem.2006.05.002>
- Guerrero C, Zornoza R, Gómez I, Mataix-Beneyto J (2010) Spiking of NIR regional models using samples from target sites: effect of model size on prediction accuracy. *Geoderma* 158:66–77. <https://doi.org/10.1016/j.geoderma.2009.12.021>
- Haghi RK, Pérez-Fernández E, Robertson AHJ (2021) Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: a comparison of near infrared and mid-infrared spectroscopy. *Geoderma* 396:115071. <https://doi.org/10.1016/j.geoderma.2021.115071>
- Huck C (2016) Recent developments in solid-phase extraction for near and attenuated total reflection infrared spectroscopic analysis. *Molecules* 21:633. <https://doi.org/10.3390/molecules21050633>
- Intergovernmental Panel on Climate Change (IPCC). (2019) 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.
- Janik LJ, Merry RH, Skjemstad JO (1998) Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust J Exp Agric* 38:681–696. <https://doi.org/10.1071/EA97144>
- Janik LJ, Skjemstad JO, Shepherd KD, Spouncer LR (2007) The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Soil Res* 45:73–81. <https://doi.org/10.1071/SR06083>
- Jouquet P, Dauber J, Lagerlöf J, Lavelle P, Lepage M (2006) Soil invertebrates as ecosystem engineers: intended and accidental effects on soil and feedback loops. *Appl Soil Ecol* 3:153–164. <https://doi.org/10.1016/j.apsoil.2005.07.004>
- Kitson RE, Mellon MG (1944) Colorimetric determination of phosphorus as molybdovanadophosphoric acid. *Ind Eng Chem Anal Edition* 16:379–383
- Krishna MP, Mohan M (2017) Litter decomposition in forest ecosystems: a review. *Energy Ecol Environ* 2:236–249. <https://doi.org/10.1007/s40974-017-0064-9>
- Lawrence I, Lin K (1989) A concordance correlation coefficient to evaluate reproducibility. *Biom* 255–268. <https://doi.org/10.2307/2532051>
- Ludwig B, Nitschke R, Terhoeven-Urselmans T et al (2008) Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. *J Plant Nutr Soil Sci* 171:384–391. <https://doi.org/10.1002/jpln.200700022>
- Madari BE, Reeves JB, Machado PLOA et al (2006) Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma* 136:245–259. <https://doi.org/10.1016/j.geoderma.2006.03.026>
- Mäkipää R, Abramoff R, Adamczyk B et al (2023) How does management affect soil C sequestration and greenhouse gas fluxes in boreal and temperate forests? – a review. *For Ecol Manage*. <https://doi.org/10.1016/j.foreco.2022.120637>
- Margenot AJ, Parikh SJ, Calderón FJ (2023) Fourier-transform infrared spectroscopy for soil organic matter analysis. *Soil Sci Soc Am J* 87:1503–1528. <https://doi.org/10.1002/saj2.20583>
- Martín-Sanz RC, Getino-Álvarez M, Pando V, Lafuente F, Turrión MB (2025) Elemental Physicochemical and ATR-FTIR Spectral Parameters of Forest Floor in Pure vs. Mixed Stand. Zenodo. <https://doi.org/10.5281/zenodo.16315149>
- McBride GB (2005) A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. Technical Report, National Institute of Water & Atmospheric Research, Hamilton, New Zealand. NIWA Client Report: HAM2005–062
- Parker FS (1971) Applications of infrared spectroscopy in biochemistry, biology, and medicine. Springer, New York
- Ramírez PB, Mosier S, Calderón F, Cotrufo MF (2022) Using mid-infrared spectroscopy to optimize throughput and costs of soil organic carbon and nitrogen estimates: an assessment in grassland soils. *Environments* 9:149. <https://doi.org/10.3390/environments9120149>
- Sanderman J, Savage K, Dungal SRS (2020) Mid-infrared spectroscopy for prediction of soil health indicators in the United States. *Soil Sci Soc Am J* 84:251–261. <https://doi.org/10.1002/saj2.20009>
- Sharma P, Sarstedt M, Shmueli G, et al (2019) PLS-based model selection: The role of alternative explanations in information systems research. *J Assoc Inf Syst* 4:346–397. <https://doi.org/10.17705/1jais.00538>
- Soriano-Disla JM, Janik LJ, Viscarra Rossel RA et al (2014) The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl Spectrosc Rev* 49:139–186. <https://doi.org/10.1080/05704928.2013.811081>
- Tatzber M, Mutsch F, Mentler A et al (2010) Determination of organic and inorganic carbon in forest soil samples by mid-infrared spectroscopy and partial least squares regression. *Appl Spectrosc* 64:1167–1175. <https://doi.org/10.1366/000370210792973460>
- Tinti A, Tugnoli V, Bonora S, Francioso O (2015) Recent applications of vibrational mid-Infrared (IR) spectroscopy for studying soil components: a review. *J Cent Eur Agric* 16:1–22. <https://doi.org/10.5513/JCEA01/16.1.1535>
- Turrión MB, Martín-Sanz RC, Getino-Álvarez M, Pando V, Lafuente F (2025) Interactions between deciduous and coniferous litter in decomposition processes: insights from ATR-FTIR spectroscopy and elemental analyses. *For Ecol Manage* 595:123038. <https://doi.org/10.1016/j.foreco.2025.123038>
- Verchot LV, Duta L, Shepherd KD, Albrecht A (2011) Organic matter stabilization in soil aggregates: understanding the biogeochemical mechanisms that determine the fate of carbon inputs in soils. *Geoderma* 161:182–193. <https://doi.org/10.1016/j.geoderma.2010.12.017>
- Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik JO, Skjemstad LJ (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1–2):59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>
- Viscarra Rossel RA, Behrens T, Ben-Dor E et al (2022) Diffuse reflectance spectroscopy for estimating soil properties: a technology for the 21st century. *Eur J Soil Sci* 73:e13271. <https://doi.org/10.1111/ejss.13271>
- Wold S, Ruhe A, Wold H, Dunn IWJ (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J Sci Stat Comput* 5:735–743. <https://doi.org/10.1137/0905052>
- Xie HT, Yang XM, Drury CF et al (2011) Predicting soil organic carbon and total nitrogen using mid-and near-infrared spectra for Brookston clay loam soil in southwestern Ontario, Canada. *Can J Soil Sci* 91:53–63. <https://doi.org/10.4141/cjss10029>
- Zhang J, Zhang J, Zhong Z (2024) Atr-ftir and NIR spectra combined with chemometrics and convolutional neural networks for identification of polygonati rhizome. *J Mol Struct*. <https://doi.org/10.1016/j.molstruc.2024.139449>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.