Spatial distribution of pyrogenic carbon in Iberian topsoils estimated by chemometric analysis of infrared spectra

by Jiménez-González, M.A., De la Rosa, J.M., Aksoy, E., Jeffery, S., Oliveira, B.R.F. and Verheijen, F.G.A.

DOI link to the version of record on the publisher's site

Jiménez-González, M.A., De la Rosa, J.M., Aksoy, E., Jeffery, S., Oliveira, B.R.F. and Verheijen, F.G.A. 2021. Spatial distribution of pyrogenic carbon in Iberian topsoils estimated by chemometric analysis of infrared spectra. *Science of The Total Environment*, 790, (148170).

Science of the Total Environment 790 (2021) 148170

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Spatial distribution of pyrogenic carbon in Iberian topsoils estimated by chemometric analysis of infrared spectra

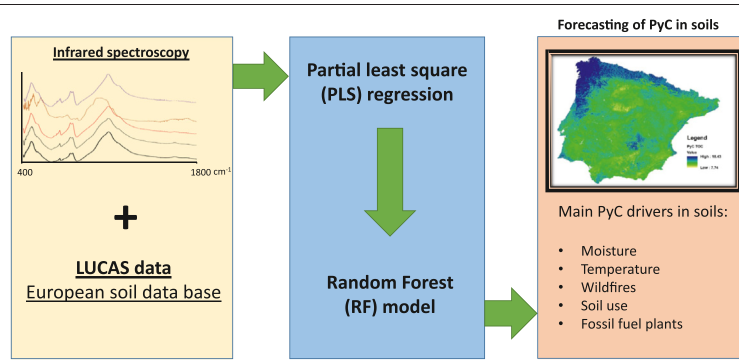
M.A. Jiménez-González [a,b], J.M. De la Rosa [c,*], E. Aksoy [d], S. Jeffery [e], B.R.F. Oliveira [f], F.G.A. Verheijen [f]

[a] HERCULES Laboratory, University of Évora, Largo Marquês de Marialva, 8, 7000-809 Évora, Portugal
[b] Department of Geology and Geochemistry, Autonomous University of Madrid, 28049 Madrid, Spain
[c] Instituto de Recursos Naturales y Agrobiología de Sevilla, Consejo Superior de Investigaciones Científicas (IRNAS-CSIC). Reina Mercedes Av., 10, 41012 Seville, Spain
[d] Via Luigi Vanvitelli, 2, 00153 Rome, Italy
[e] Agriculture and Environment Department, Harper Adams University, Newport, Shropshire TF10 8NB, United Kingdom
[f] Earth Surface Processes Team, Centre for Environmental and Marine Studies (CESAM), Dept. Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal

HIGHLIGHTS

- Soil PyC estimation by partial least square regression and random forest model on MIR spectra
- Forests accumulate more PyC than grasslands and agricultural soils.
- Notable influence of wildfires on topsoil PyC in the Iberian Peninsula
- Strong relationship between wildfire records, fossil fuel plants and PyC distribution

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 7 April 2021
Received in revised form 26 May 2021
Accepted 27 May 2021
Available online 1 June 2021

Editor: Manuel Esteban Lucas-Borja

Keywords:
Soil organic matter
Pyrogenic carbon
Black carbon
Carbon sequestration
Partial least squares regression
Random forest model

ABSTRACT

Understanding the global carbon (C) cycle is critical to accurately model feedbacks between climate and soil. Thus, many climate change studies focused on soil organic carbon (SOC) stock changes. Pyrogenic carbon (PyC) is one of the most stable fractions of soil organic matter (SOM). Accurate maps based on measured PyC contents are required to facilitate future soil management decisions and soil-climate feedback modelling. However, consistent measurements that cover large areas are rare. Therefore, this study aimed to map the PyC content and stock of the Iberian Peninsula, which covers contrasting climatic zones and has long-term data on wildfire occurrence. A partial least square (PLS) regression using the mid-infrared spectra (1800–400 cm^{-1}) was applied to a dataset composed of 2961 soil samples from the Iberian component of the LUCAS 2009 database. The values of PyC for LUCAS points were modelled to obtain a map of topsoil PyC by a random forest (RF) approach using 36 auxiliary variables. The results were validated through comparison with documented historical wildfire activity and anthropogenic energy production. A strong relationship was found between these sources and the distribution of PyC. Our study estimates that the accumulated PyC in Iberian Peninsula soils comprises between 3.09 and 20.39% of total organic carbon (TOC) in the topsoil. Forests have higher PyC contents than grasslands, followed by agricultural soils. The incidence of recurrent wildfires also has a notable influence on PyC contents. This study shows the potential of estimating PyC with a single, rapid, low cost, chemometric method using new or archived soil spectra, and has the ability to improve soil-climate feedback modelling. It also offers a possible tool for measuring, reporting and verifying soil C stocks, which is likely to be important moving forward if soils are used as sinks for C sequestration.

* Corresponding author.
E-mail address: jmrosa@irnase.csic.es (J.M. De la Rosa).

https://doi.org/10.1016/j.scitotenv.2021.148170

1. Introduction

Soil organic matter (SOM) constitutes approximately 2/3 of the global terrestrial carbon (C) pool, which corresponds to an estimated 2060–2476 Pg of soil organic C provided by the Harmonized World Soil Database to a depth of 2 m (Batjes, 2016; Köchy et al., 2015). Therefore, the dynamics of organic carbon turnover in soils control a large part of the global C cycle. Soils have been a major source of atmospheric enrichment of carbon dioxide (CO_2) since the dawn of settled agriculture, about 10,000 years ago, releasing about 8×10^{14} kg C per year (Batjes, 2016). In particular, forest conversion to agriculture can release up to 75% of stored soil organic carbon as CO_2 (Schlesinger, 1997).

Black carbon, soot, elemental carbon, biochar and charcoal are terms used to describe a spectrum of chemically heterogeneous, aromatic, carbonaceous forms of pyrogenic carbon (PyC), i.e. compounds produced through the incomplete combustion of fossil fuels or biomass (Goldberg, 1985; Masiello, 2004). This material is ubiquitous in the atmosphere, ice, soils and sediments due to its widespread production and transmission (Masiello and Druffel, 1998). Every year, biomass and fossil fuel burning are estimated to release 7.5–17 Tg of PyC into the atmosphere and 56–123 Tg of PyC to the soil surface, while the flux of PyC from land to the oceans is 44–108 Tg of PyC, where sediments form a large PyC sink (Bird et al., 2015). Lehmann et al. (2008) reported a wide range of PyC/TOC contents in a continental-scale analysis of Australia (0–82%), demonstrating the significance of achieving accurate information about the distribution of PyC in soils for projections of future climate change. This need for knowledge has become even more important over the last decade. During this period of time the role of PyC as a soil ameliorant has been rediscovered, being called biochar when it is produced for this specific purpose. Numerous scientific studies propose the use of this form of PyC for soil C sequestration and the recovery of degraded soils. It has been reported as a long-term sink for atmospheric CO_2 with a potential contribution to mitigating climate change while also maintaining or improving soil health and productivity (Laird, 2008; Matthews, 2008; Lehmann and Joseph, 2009; Sohi et al., 2010; De la Rosa et al., 2018).

Small changes in the soil C pool can have large-scale effects both on agricultural productivity and on atmospheric greenhouse gas balance. As such, improved understanding of PyC behavior in the soil is required to allow accurate inclusion of PyC in both predictive and retrospective models of global carbon budgets. There are several procedures to quantify PyC in soils that are based on different extraction techniques (e.g. Hammes et al., 2007; Masiello, 2004), which can be divided into four main categories: thermal, chemical, optical and molecular markers (Schmidt and Noack, 2000; Hammes et al., 2007). However, PyC assessment is subject to serious experimental constraints (Hammes et al., 2007; De la Rosa et al., 2011; Santín et al., 2016) due to its physicochemical heterogeneity; PyC is not a single material, but a continuum of materials with diverse properties (Masiello, 2004).

The isolation of PyC from natural matrices, such as soils and sediments, is particularly complex due to the presence of minerals, recalcitrant materials other than PyC, and potential interfering materials (De la Rosa et al., 2011). The use of strong chemical oxidation procedures to remove the interfering material has been shown to also oxidise PyC (Knicker et al., 2007) leading to underestimation, whereas hydrophobic materials may remain after the method applied leading to overestimation of PyC as a fraction of the soil. The use of benzenopolycarboxylic acids (BPCA) as molecular markers of PyC is one of the most widely published methods for estimating PyC in soils (Simeone et al., 2018). However, this method is, by definition, unable to detect the presence of highly condensed combusted material, i.e. it does not detect soot or highly condensed chars (Hammes et al., 2007). Spectroscopic methods can distinguish minerals and differentiate between types of carbon in soil. For instance, Simpson and Hatcher (2004) developed a method applying solid-state ^{13}C Nuclear Magnetic Resonance (NMR) spectroscopy to differentiate aryl-C and so minimize PyC overestimation. However, using this procedure is time demanding and expensive (De la Rosa et al., 2011). Given the need to quantify PyC in soils at large scales to effectively quantify it as a pool and investigate its spatial-temporal dynamics, De la Rosa et al. (2019) developed an analytical procedure that combines the improvement of mid infrared (IR) spectral data following processing (Fernández-Getino et al., 2013) and the calibration of IR bands with ^{13}C NMR spectra of soils and samples rich in PyC. This proxy allows a satisfactory quantification of PyC in soils across large scales (Nocita et al., 2015). This is necessary to quantify the efficacy of C sequestration strategies, such as the 4 PER 1000 Initiative (4 PER 1000, 2020), as well as the impact of other factors such as new soil management practises, or the increased occurrence of wildfires.

Approximately 2.5 million hectares (29% of Portugal's land area) were burned between 1980 and 2009 (ICNF, I.P.). In Spain, 7.1 million hectares (approximately 14% of the Spanish mainland) were burned between 1969 and 2009 (Ministry of Agriculture, Food and Environment of Spain, 2020). Most of these fires were concentrated in the northern and western regions of the Iberian Peninsula (Vázquez de la Cueva et al., 2006, Vázquez de la Cueva, 2012; Verde and Zêzere, 2010). After vegetation fires, considerable amounts of charred vegetation residues remain; with the conversion rate of fuel to PyC dependent on fuel characteristics and burning conditions (Santín et al., 2015). Residues of incomplete combustion of biomass are PyC, which are then transported by wind and water. Model calculations suggest that more than 80% of PyC deposition occurs on continental shelves (Santín et al., 2016). De la Rosa et al. (2011) estimated that PyC ranged from 4.4 to 14.4% of the TOC in marine sediments of the inner continental shelf of the Gulf of Cádiz. Rovira et al. (2009) showed that fires did not consistently increase PyC contents, measured by dichromate oxidation, of shrubland plots on old agricultural fields from different regions of Spain. However, according to Suman et al. (1997), it is estimated that as much as 90% of all forest fire residues remain in terrestrial environments. These large differences between estimations of PyC highlight the diversity of properties and processes governing the fate of PyC.

As there is a paucity of information on the quantity of PyC in soils, we aimed to investigate the PyC concentration in over 2900 soil samples of the Iberian Peninsula collected in the framework of the Land Use/Land Cover Area Frame Survey (LUCAS) project (sampling year 2009; European Soil Data Centre (ESDAC). This study aimed to measure topsoil PyC content in the Iberian Peninsula using spectroscopic data, and to apply machine learning/random forest to identify the environmental management/soil physical and chemical parameters that are the strongest predictors of PyC concentration in soils. The Iberian Peninsula, due to its geographical position, its geology and its variability in altitude, has an enormous variety of climates and ecosystems. These include arid and semi-arid, humid forests and high mountain climates. It is therefore an ideal scenario to represent different climatic zones, as the points were selected out of the main LUCAS grid for the collection of soil samples by a standardized sampling procedure (Tóth et al., 2013). This will allow to improve the current PyC map of the Iberian Peninsula (Reisser et al., 2016), improve our understanding of PyC dynamics, and provide a tool and process that can be applied to future data sets to allow us to develop a global understanding of the levels and dynamics of PyC within soils.

2. Methods and materials

2.1. LUCAS spectral database

The LUCAS topsoil sampling survey was carried out in 2009. The samples were analyzed by a single laboratory for the main properties (coarse fragments, particle size distribution (% clay, silt and sand content), pH in $CaCl_2$ and H_2O, organic carbon, carbonate content, phosphorous content, total nitrogen content, extractable potassium content, cation exchange capacity and multispectral properties) of topsoil in 23 Member States of the European Union (EU) based on standard sampling and analytical procedures. Approximately 20,000 points were selected to collect around 0.5 kg of topsoil (0–20 cm). The infrared (IR) spectra of topsoil samples from the Iberian Peninsula (Portugal and Spain) were extracted from the 2009 LUCAS database (Tóth et al., 2013; Orgiazzi et al., 2018). The dataset selected for this study was composed of 3081 spectra. This work focused on the region between 400 and 1800 cm^{-1} of the IR spectra (2801 spectral points) where most of the diagnostic bands of minerals and aromatic groups appear (De la Rosa et al., 2019). The region between 1800 and 4000 cm^{-1} was not considered as the IR signals corresponding to alkyl-C and -OH are not of interest for the quantification of PyC according to the predictive model based on measured contributions in the aryl-C by ^{13}C NMR spectroscopy (De la Rosa et al., 2019). In addition to the expected IR bands corresponding to aryl C, other bands inform about the patterns of oxygen-containing functional groups and the mineralogical composition characteristic of the soils with greater black carbon storage capacity. De la Rosa et al. (2019) showed with the variables of importance (VIPs) traces of PyC rich samples that aromatic bands at 1620 and 1510 cm^{-1} are the most important in the prediction model for PyC-rich samples. Each sample's IR spectrum was normalized by equalling the sum of the intensities of the 2801 points of the 1800–400 cm^{-1} region of the spectra, to 100 in Microsoft Excel. This procedure was carried out both on the spectra of the LUCAS samples and on the 42 samples used to generate the model, to make them comparable.

2.2. PLS model from IR spectra

The prediction of the PyC content from soil IR spectra was carried out based on the model generated by De la Rosa et al. (2019). In this study, 42 different soils (representing *Histic Humaquepts*, Leptosols Cambisols and Anthrosols; World Reference Base for Soil Resources, 2014) were analyzed to build the partial least squares (PLS) regression model using the software ParLeS (Viscarra Rossel, 2008). Thus, the selected samples were milled, homogenized oxidized with potassium dichromate (60 °C) and subsequently analyzed by both Fourier Transform-Infrared (FT-IR) and ^{13}C NMR spectroscopies. All spectra used to build and verify the model were acquired on the same equipment, which reduced sources of uncertainty (Dangal and Sanderman, 2020). The FT-IR were acquired from Potassium bromide (KBr) pellets containing 1 mg of powdered sample and 100 mg of KBr scanned by a IR JASCO 4100 spectrometer (Jasco Corporation, Tokio, Japan) with 60 scans per sample at a resolution of 2 cm^{-1}. The solid-state ^{13}C NMR spectra were obtained with a Bruker Avance III HD 400 MHz (Bruker Biospin, Rheinstetten, Germany), into 4 mm zirconium rotors with KEL-F caps, rotated with a speed of 14 kHz. The number of scans needed per sample ranged from 100,000 to 170,000. The integration of the signal intensity of each C group identified by ^{13}C NMR spectra was performed by using MestreNova 10 Software (Santiago de Compostela, Spain). Among the various existing methodologies for the detection and quantification of PyC in soils, ^{13}C NMR spectroscopy was chosen despite its high cost per sample and long measurement time to build the model. This is because both ^{13}C NMR and FT-IR rely on spectroscopic properties and the results obtained are based on the same chemical property of the soil samples, the presence of aryl-C.

The determination of TOC contents of soils was performed in triplicate by dry combustion (975 °C) after the removal of carbonates (HCl; 1 M) with a Flash 2000 elemental micro–analyser (Thermo Scientific, Bremen, Germany).

The PyC contents of the 42 samples – i.e. the aryl-C contents as previously determined by ^{13}C NMR spectroscopy (De la Rosa et al., 2019) were used to generate a prediction model using the mid-IR spectra of these samples in the 1800–400 cm^{-1} region, which is composed of 2801 points. The normalized intensities (see Section 2.1) of these points of the IR spectra were used as descriptors (independent variables) to predict the PyC contents (dependent variable) using PLS regression. The selection of the number of latent variables to obtain the PLS model was based on two complementary criteria: i) the root mean squared error; and ii) the Akaike's (1974) information criterion, which gives relative information about the quality of the model. Finally, observing these criteria, the best model was obtained using 16 latent variables. The cross-validation plot for the PLS model using the experimental values of SOC and the values predicted by the model showed a significant correlation between them ($R^2 = 0.68$; Fig. 1.a). This high correlation indicated a successful PLS forecasting of PyC in the different samples by using the mid-IR spectra in the 1800–400 cm^{-1} range. This led to significant ($P < 0.05$) cross-validation coefficients for PyC, determined as the aryl-C content. The PyC content of the 42 samples used to build the model ranged from 1.41 to 28.01% of the TOC contents, which is within the usual PyC range in soils (Forbes et al., 2006). The normalized intensities of the 1800–400 cm^{-1} region of the IR spectral data of LUCAS (2009) were introduced in the model to estimate the PyC content of the 3081 topsoil samples from the Iberian Peninsula (see Section 2.3).

2.2.1. Validation against measured PyC contents in Iberian soils

For the validation of the model, an extra set of eleven soil samples not used to generate the model with a wide range of aryl-C content measured by ^{13}C NMR spectroscopy, was selected. The PyC content in these soils was determined in the same conditions than the soils used to generate the PLS model. Finally, the PyC contents predicted by the model from the IR spectra of these soils were compared with the experimental values determined by ^{13}C NMR spectroscopy and TOC analysis. The results of the validations showed that the values predicted were similar to the values determined in the lab for these soil samples (Fig. 1.b).

2.3. Data preparation and processing

We considered three rules for preparing the dataset: i) normal distribution; ii) exclude samples without SOC; iii) exclude soil samples with low TOC (<0.4%) as well as high $CaCO_3$ (>5%) contents, because this is known to mask the bands used in the PLS model (De la Rosa et al., 2019). A histogram of the predicted values showed one near normal distributed population with a mode around 11.6 PyC/TOC (%), and a second left-tailed population with a mode around 2.5 PyC/TOC (%). After discarding the second, skewed population (cut-off at 3.0; $n = 42$), the remaining population had a near-normal distribution with a mode of 12.13% (Fig. S2). Thirty-one samples with 0%SOC were removed from the dataset. In addition, fifty-seven samples with SOC contents <0.4% and $CaCO_3$ contents >5% were removed from the dataset. The final dataset used for this study has 2961 samples.

2.4. Random forest model

Estimated PyC values for the LUCAS spectra were modelled to predict the geospatial distribution of topsoil pyrogenic carbon relative to total SOC by using a random forest (RF) algorithm and 36 auxiliary variables (Table 1, with further information in Annex 1). All auxiliary layers, which are in raster format, were pre-processed; prepared in the same resolution (100 m), with the same projection system (LAEA) and extent before building the model.

The use of RF models, an ensemble machine learning technique, i.e. using an algorithm that combines predictions from multiple models, was first proposed by Breiman (2001) by combining classification and regression tree (Breiman et al., 1984) and bagging (Breiman, 1996). We used “RandomForest”, “sp” and “caret” R packages, and defined the number of trees to be built in the forest as 1000 (ntree = 1000) as required by the model (Díaz-Uriarte and de Andrés, 2006). In our RF model, out-of-bag (OOB), i.e. a random subset of data not used in the tree-building process, samples were predicted and root mean square error (RMSE) (Zong et al., 2020), coefficient of determination (R^2), and bias were calculated. The relative importance of the predictors was also obtained. We calculated cross-validation by randomly dividing the entire dataset into training and validation subsets as 70% and 30% of the data.

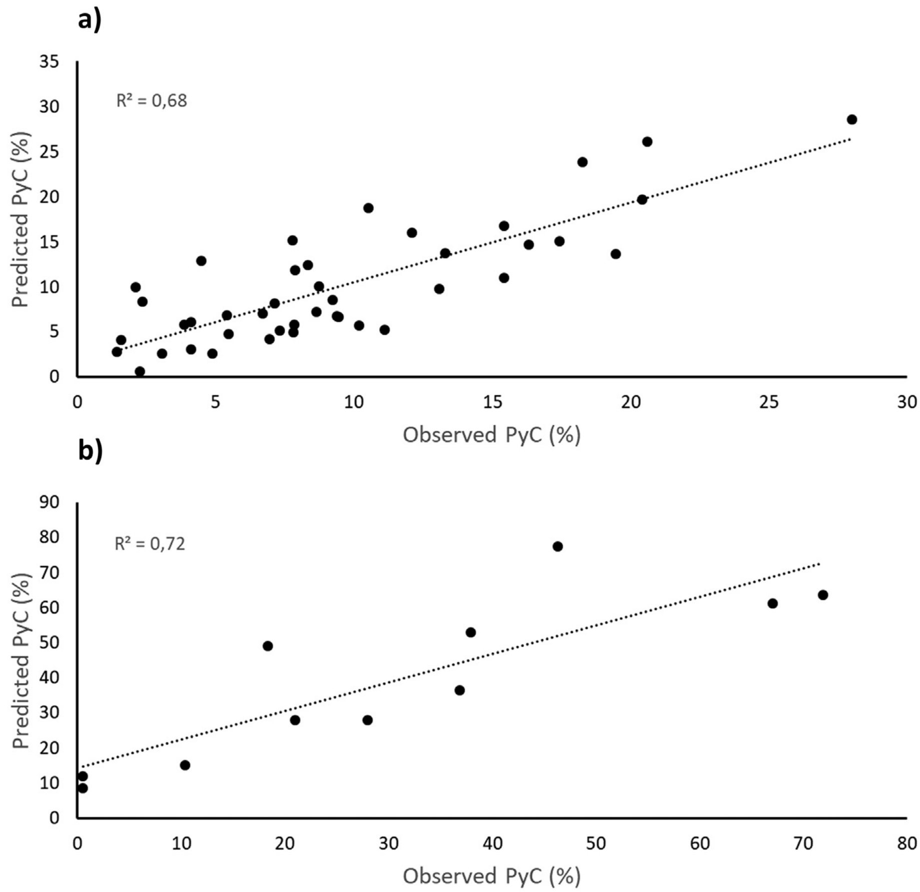


Fig. 1. a) Cross-validation plots (experimental vs predicted values) corresponding to partial least squares (PLS) models to predict the pyrogenic carbon (PyC). b) Experimental and predicted values of the samples used to validate the model.

Table 1
Auxiliary variables used in the RF approach.

Variable	Description	Resource
EX1MOD5	Mean monthly MODIS EVI January–February	SoilGrid covariates
EX2MOD5	Mean monthly MODIS EVI March–April	SoilGrid covariates
EX3MOD5	Mean monthly MODIS EVI May–June	SoilGrid covariates
EX4MOD5	Mean monthly MODIS EVI July–August	SoilGrid covariates
EX5MOD5	Mean monthly MODIS EVI September–October	SoilGrid covariates
EX6MOD5	Mean monthly MODIS EVI November–December	SoilGrid covariates
TMDMOD3	Mean annual LST (daytime) MODIS	SoilGrid covariates
TMNMOD3	Mean annual LST (nighttime) MODIS	SoilGrid covariates
PRSCHE3	Total annual precipitation at 1 km	SoilGrid covariates
B07CHE3	Temperature Annual Range [°C] at 1 km	SoilGrid covariates
VW1MOD1	Monthly MODIS Precipitable Water Vapor January–February	SoilGrid covariates
VW2MOD1	Monthly MODIS Precipitable Water Vapor March–April	SoilGrid covariates
VW3MOD1	Monthly MODIS Precipitable Water Vapor May–June	SoilGrid covariates
VW4MOD1	Monthly MODIS Precipitable Water Vapor July–August	SoilGrid covariates
BDRICM	Depth to bedrock (R horizon) up to 200 cm	SoilGrid soil properties
BLDFIE	Bulk density (fine earth)	SoilGrid soil properties
CECSOL	Cation Exchange Capacity of soil	SoilGrid soil properties
OCSTHA	Soil organic carbon stock	SoilGrid soil properties
PHIHOX	pH index measured in water solution	SoilGrid soil properties
TAXNWRB	World Reference Base legend	SoilGrid soil properties
TEXMHT	Texture class (USDA system)	SoilGrid soil properties
CLC2018	Corine Land Cover Classification, 2018	EEA
DEM	Digital elevation model, SRTM	SRTM, NASA
Sentinel1 variables (B1S2, B2S2, B3S2, B1S1min, B1S1max, B1S1media, B2S1min, B2S1max, B2S1media, B3S1min, B3S1max, B3S1media)	Sentinel-1 satellite imagery, 2019, 3-Bands (VV, VH,VV/VH), Min-Max-Median values per each 3-bands	Google Earth Engine calculation
Sentinel-2 variables (B1S2, B2S2, B3S2)	Sentinel-2 satellite imagery, composite image of 2019, (RGB (B4-B3-B2)), Median values per each 3-bands	Google Earth Engine calculation

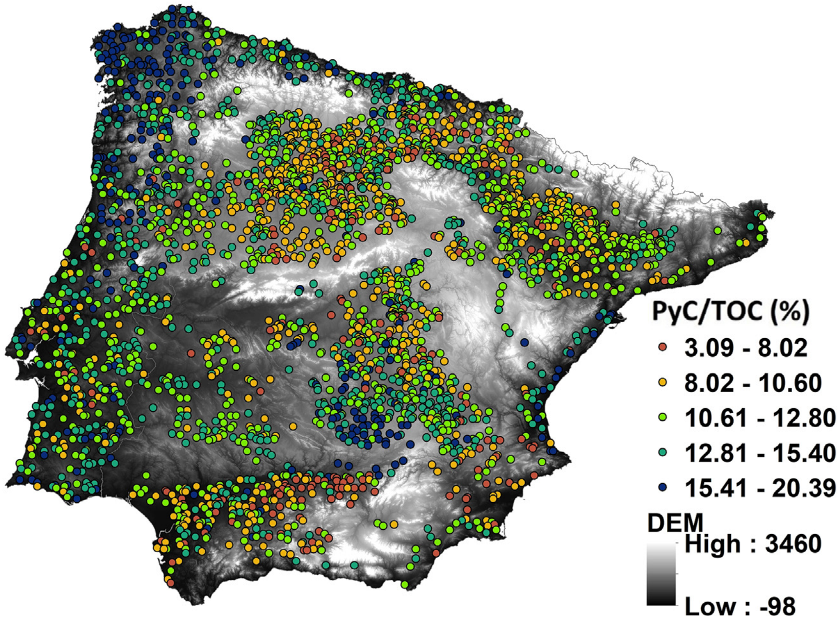


Fig. 2. Geographic distribution of pyrogenic carbon relative to total soil organic carbon (%) in Iberian LUCAS samples (topsoils) overlaid on a digital elevation model.**). The location of the samples has been obtained according to the information provided by the LUCAS database (European Soil Data Centre; ESDAC)

3. Results & discussion

3.1. Chemometrically-predicted topsoil pyrogenic carbon contents

A significant cross-validation plot was obtained (Fig. 1a) despite the limited number of samples (42 soil samples) used to generate the model. The validation of the model used samples not used to develop the model and provided significant results (Fig. 1b), and the regression result between the real and predicted values of PyC was $R^2 = 0.72$. The results of the PyC content predicted by PLS regression from the IR spectra of the soils are shown in Fig. 2. The PyC/TOC (%) values predicted for the 2961 samples ranged from 3.09% to 20.39%. In general, samples from the northwest of the Iberian Peninsula showed a higher relative and absolute abundance of PyC. The following sections will discuss the possible causes of this trend.

The median PyC/TOC (%) value for the dataset was 11.9%: first quartile at 10.1%, third quartile at 13.7%. This is in agreement with the slightly higher median of 12.3% (range from 0% to 50%) that Reisser et al. (2016) found for a smaller global dataset (569 cases) of measured PyC values (with a variety of methods) extracted from the scientific literature and similar to the 11% found for Amazon Basin forest soils (Koele et al., 2017). The range of the current study (17.3%) is less than half that the 50% reported by Reisser et al. (2016), which may be caused in part by the six different methods for measuring PyC that they included, and the known heterogeneity of values for the same sample (Hammes et al., 2007). In addition, the larger range could also be due to the greater variability of soils and climate types reported by Reisser et al. (2016). Skjemstad et al. (2002) reported that PyC/TOC ranged 10–35% for US soils based on a NMR spectroscopy methodology. Along the same lime, Sanderman et al. (2021) calculated on average 24% of PyC/TOC on the Great Plains of the US by mid-IR spectroscopy. However, the latter study only included non-forest and non-shrub areas, making it difficult to compare with our study. In addition, Wang et al. (2018) reported PyC/TOC from 7% to 37% in two locations of south-eastern Australia by using a M-IR–PLSR model similar to this study. A much lower average PyC content was reported for French forest topsoils (4.4% PyC/TOC) by Soucémarianadin et al. (2019), which they attributed to the ongoing fire suppression over Europe that started in the 18th century (Pyne, 1997).Supplementary Fig. 1 shows the estimated PyC topsoil stocks for the Iberian Peninsula, which varied from 61 to 3575 g m^{-2}, with a median value of 384 g m^{-2}. This result overlaps with the ranges reported by Soucémarianadin et al. (2014), i.e. over 200 g PyC m^{-2} for mineral soils and 200–1200 g PyC m^{-2} for forest floor soils located throughout the province of Quebec (Canada).

However, our estimates have to be considered as preliminary results because of the multiple uncertainties (see Section 3.7.1 Limitations). Nevertheless, the preliminary assessment of the total topsoil stock of PyC in Portugal and Spain mainland territories are approximately 6×10^{10} kg and 2.9×10^{11} kg, respectively, which correspond to about 667 kg km^{-2} and 587 kg km^{-2} of PyC, respectively.

3.2. Random Forest model results

The geospatial distribution of topsoil PyC relative to TOC and topsoil stocks, predicted by the RF model across the Iberian Peninsula, are provided in Fig. 3. Fig. 4 shows the importance of the auxiliary variables used in RF regression modelling (Table 1). The results of the model showed that the R^2 of PyC to TOC (%) was 0.41, root mean square error was 2.18; and bias was 0.064. The most important variables to

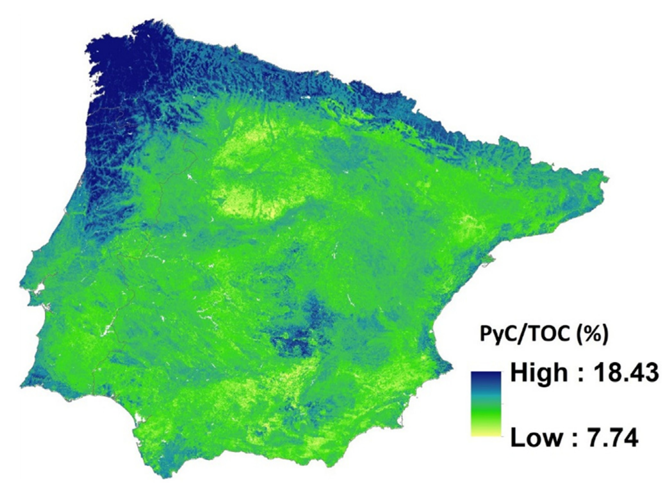


Fig. 3. Random forest (RF) digital soil map of topsoil pyrogenic carbon relative to total organic carbon content in the Iberian Peninsula.

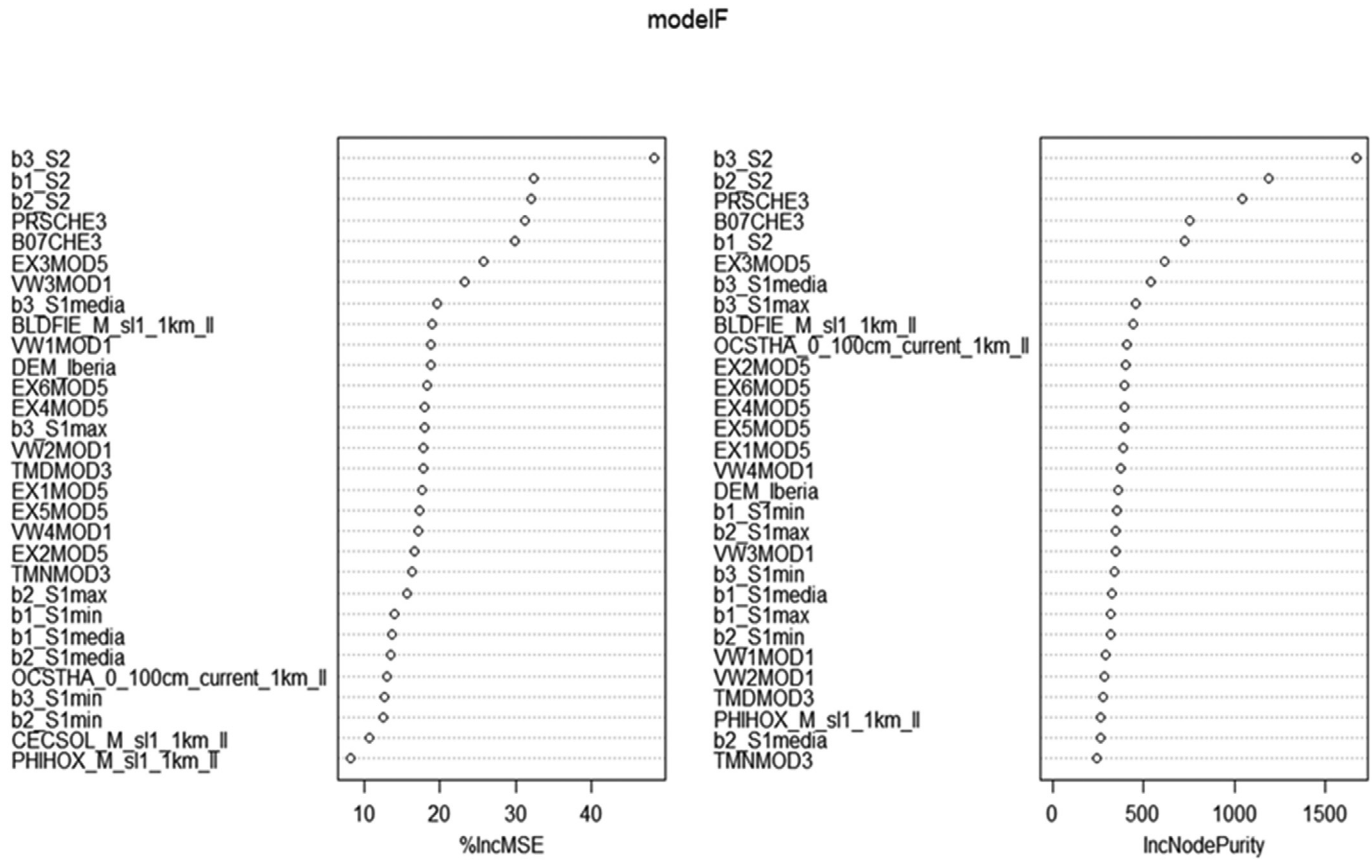


Fig. 4. Variable importance plot relative to TOC: %IncMSE: Mean Decrease Accuracy (percent increase in the mean squared error of each tree after permuting a variable); IncNodePurity: Mean Decrease Gini (total increase in node purities from splitting on the variable).**). The detailed meaning of the codes is given in Table 1.

shape the model were Sentinel-2 bands and precipitation for both of the independent predicted variables. The least important variable in the model was soil pH for both of the variables. The northwest of the Iberian Peninsula has the most acidic soils and an average rainfall of more than 1000 mm per year, in some cases more than 2000 mm yr^{-1}. On the other hand, in the south-western (Mediterranean) strip, average rainfall does not exceed 400 mm per year, and the soils are more alkaline. As usual, there is a close relationship between rainfall and soil pH. In addition, eucalypt plantations, and to a lesser extent pine plantations, are concentrated in the northwest, and are recurrently affected by forest fires. We hypothesize that the abundance of precipitation and human action are probably masking the relationship between soil pH and PyC content in the Iberian Peninsula.

The averages and the standard deviations of the variables per CORINE land cover (CLC) classes are shown in Table S1. The averages for those CLC classes that cover >1% of the Iberian Peninsula are in Fig. 5. The highest PyC/TOC (%) averages were found in CLCs associated with moors and heathlands > forests (deciduous, mixed, coniferous) > grasslands (pastures and natural grasslands) > arable (permanently irrigated, not irrigated, agroforestry). CLCs associated with permanent crops (vineyards, olive groves, fruit and berry trees) showed variable PyC contents across the Iberian Peninsula.

RF is one of the most used algorithms since it is flexible, has a high predictive performance, low correlation, small bias and variance, reliable error estimates and provision of information on the relative importance of predictors (Breiman, 2001). It constructs a large number of uncorrelated decision trees based on averaging random selection of predictor variables and decision trees have proven to be very successful in solving classification problems of statistical learning (Carvajal et al., 2018). In our study, the RF model explained 44.6% of the PyC stock variation estimated from the spectral data of the LUCAS soil sample dataset, compared to 33% for the model of Reisser et al. (2016). This higher percentage may be due to the greater number of variables used in the current study. In the previous estimation of Reisser et al. (2016), five variables were used to generate the model, while in the current study, 33 variables have been used to estimate the PyC content. Furthermore, this difference may also be due to methodological differences between our study and Reisser et al. (2016), as is detailed in Section 4.3.

3.3. PyC by land use

The range in PyC content as a proportion of total SOC content per land use, was from 11.3% (olive groves) to 13.7% (Moors and heathland), see Fig. 5. The total range in mean PyC/TOC (%) for the CLC categories was 3.5%. This relatively small difference between land uses may be caused in part by a relatively high historic-baseline PyC content of Iberian soils, i.e. the background PyC from natural wildfires before humans converted the land to the various non-natural categories. Another cause may be an upper limit of topsoil PyC caused by erosion of the low density PyC particles on the soil surface (Abney and Berhe, 2018).

Despite the relatively small range, some broad patterns may be identified (Fig. 5). For example, PyC content as a proportion of SOC followed the sequence forests > grasslands/pastures > arable. This trend is consistent with the low PyC content measured by Rovira et al. (2009) for Spanish shrubland topsoils on old agricultural fields diversely affected by fires (8.6%). Less PyC in cropland than grassland samples was also found on the Great Plains of the United States by Sanderman et al., (2021). In this study, this result was also expected based on the coincidence of large biomass and frequent large burnt areas for eucalypt and pine plantations in the northwest of the Iberian Peninsula (Vázquez de la Cueva, 2012). This suggests a strong "char flux" to the soil, during

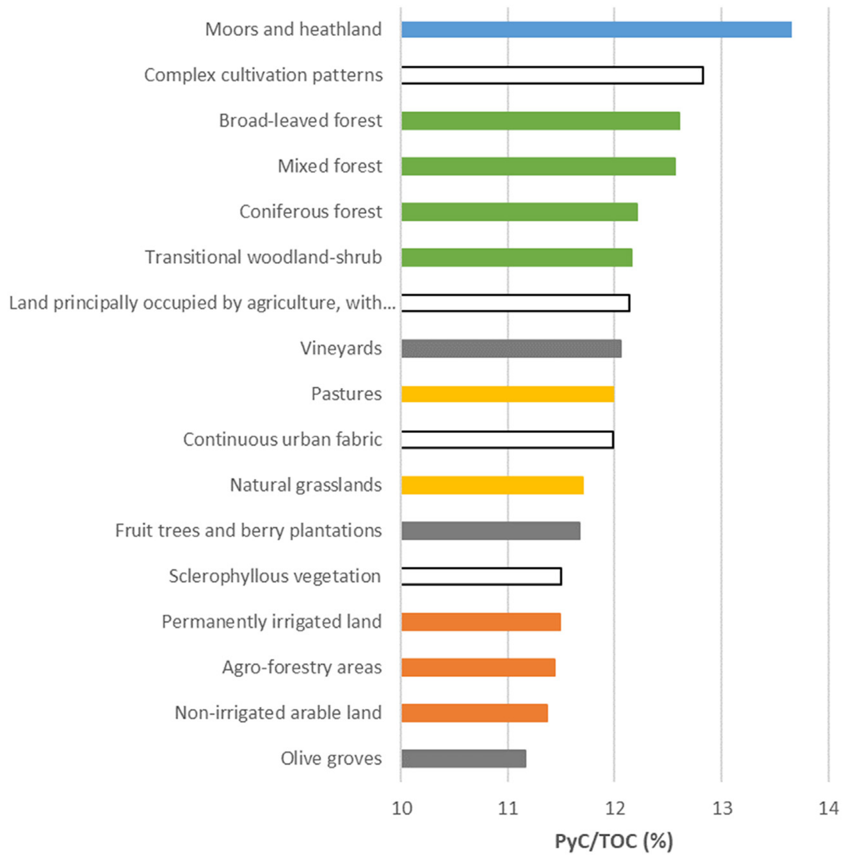


Fig. 5. Mean topsoil pyrogenic carbon to total organic carbon ratio (%) values for each Corine Land Cover category with >1% coverage in the Iberian Peninsula. Forest categories are depicted in green, grasslands in yellow, arable in orange, permanent crops in grey, moors and heathlands in blue, and others in white. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the last decades. At the very top of the table is a CLC category related to cold and moist organic soils (moors and heathland). This may have been caused by frequent traditional burning to improve grazing quality by shepherds, which is now also promoted as a conservation technique (Prichard et al., 2017). Beside this, the inhibited decomposition of PyC in this areas should be also considered, due to reduced microbial activity of those ecosystems. One possible explanation for olive groves exhibiting the lowest PyC/TOC (%) of any CLC category, may be the relatively recent large-scale extension of olive orchards in the Iberian Peninsula, where the soil management includes inversion of the soil profile, i.e. the topsoil is buried and the subsoil is at the surface. This soil management increases soil erosion rates and may have inverted the natural concentration gradient in PyC that exists down the soil gradient and buried PyC that was in the topsoil below the sampling depth of the LUCAS sampling (LUCAS, 2009). Cotrufo et al. (2016) showed that PyC redistributes after wildfire in patterns that are consistent with erosion and deposition of low-density sediments. Lopez-Martin et al. (2018) reported a fast loss of PyOM from topsoils in SW Spain taken 4 weeks and 7 years after a severe fire, which was explained by erosion, transport and microbial decomposition. A clear example of how soil management and fire affect PyC dynamics in soil can be seen in Matosziuk et al. (2019), who reported an increase in PyC content (BPCA method) in soils following prescribed fires in the fall compared to unburned controls, 8.42 g BPCA/kg C, but no difference in mean PyC concentration of the mineral soil between the spring burns and the unburned controls in ponderosa pine stands in the southern Blue Mountains (Oregon, USA). Selvalakshmi et al. (2018) showed that PyC content in soil slightly increased with the number of prescribed fires. Nevertheless, charcoal stocks decreased sharply with increasing stand age after each slash-and-burn event. In fact, over 25% of the charcoal stock was lost during the period from 12- to 21-years after the first slash-and-burn, and 85% was lost in the 97-year old stand. The repetition on the future of the LUCAS sampling campaign as well as subsoils sampling (not part of the 2009 LUCAS sampling) are required to monitor relevant changes in PyC.

The olive plantations developed over the last decades in the Iberian Peninsula are mostly intensive mechanized monoculture exploitations in which the land between each tree is kept free of vegetation, is often plowed, and suffers from high rates of soil erosion. The intensive management of the olive plantations has increased, especially the rate of laminar and trench erosion, leading to high average soil loss rates (Inventario Nacional de Erosion, 2021). The erosion data for the olive-provinces par excellence in Spain, such as Jaen (the largest olive oil producing region in the world), where olive trees represent more than 65% of the surface area of the province, show that more than 50% of these lands have high or very high soil losses, i.e. greater than 50 t $ha^{-1}y^{-1}$, compared to an average rate of soil loss for Spain of 4 t $ha^{-1}y^{-1}$ (*Inventario Nacional de Erosión*), and a sustainable erosion rate of 1.4 t $ha^{-1}y^{-1}$ (Verheijen et al., 2009). On the contrary, PyC data from Wang et al. (2018) suggested that erosion may not preferentially transport PyC over other soil organic carbon forms. However, that study was performed on only two sites never used for crop production and had not been plowed. Soil erosion has been identified as the dominant mechanism for the fate, transport and redistribution of PyC in the areas affected by wildfire (Wang et al., 2018). As such, PyC from ancient wildfires may have eroded with those soils, enriching nearby sediments and areas further down the catchment, but becoming depleted locally. Forest lands show much lower erosion rates, especially in dense Mediterranean hardwood formations, and these have the highest PyC proportions, which supports this hypothesized mechanism although the evidence is circumstantial.

Compared to the study by Reisser et al. (2016), the mean PyC/TOC (%) estimation for grasslands was similar between the two studies: 12.2% (current study) and 12.1, respectively (Table 2). However, the CLC Moors & Heathland and Forest is higher while the CLC Agriculture is lower in the current study. We hypothesize that forests may have averaged higher in the current study because of the short fire interval and large burnt areas in the pine and eucalypt plantations of north-western Iberian Peninsula as reported by Vázquez de la Cueva, (2012), relative to global forests. Why agriculture would average a lower PyC mean in the Iberian Peninsula than in the rest of the world is perhaps more surprising considering that before the conversion of the land from natural to agricultural, wildfires are likely to have been more frequent in the Iberian Peninsula than in most other places of Earth's surface. Potential causal issues are the same as discussed in Section 3.5.3.

Reisser et al. (2016) found climate effects on PyC abundance to be minimal, and the connection between PyC abundance and fire regimes to be weak. However, the methodological differences between the current study and that of Reisser et al. (2016) – global vs Iberian Peninsula, higher proportion of samples taken in colder/continental climates vs all samples taken in the Iberian Peninsula - likely explain the differences in the model outputs. In addition, the availability of detailed wildfire history maps for the Iberian Peninsula, combined with nearly 3000 samples, also allowed for a more robust comparison in the current study.

We have to bear in mind that the present study is using a single methodology for the calculation of the PyC content of topsoils, which carries the risk of over- or under-estimating the PyC contents, as we only will "see" a part of the PyC continuum (Masiello, 2004). Nevertheless, it has been demonstrated that the range of the PyC continuum detected by ^{13}C NMR-based methods is broader than for example either of the methods benzene polycarboxylic acid (BPCA) biomarker and hydrogen pyrolysis technique (Reisser et al., 2016; Schmidt et al., 2001). We suggest that future metadata studies on soil PyC contents, as with TOC data, should use PyC data from comparable methodologies.

3.4. Drivers of PyC contents

As shown in Section 3.2, the main drivers of the spatial distribution of PyC content are vegetation and climate, as Sentinel-2 bands (B1S2, B2S2, B3S2), average precipitation, temperature and May–June Enhanced Vegetation Index (EVI) best explained our model. Our results reveal that PyC concentration parameters were mostly related with quantity of biomass and soil moisture and temperature. The quantity of biomass and vegetation type play an important role in wildfire periods (Vázquez de la Cueva et al., 2006), which is well known to increase levels of PyC in soil (Santín et al., 2015). In the same way, the moisture and the temperature are also important predictors. In this respect, the northwest region of the Iberian Peninsula, with the greatest precipitation rates, concentrates eucalyptus and pine plantations, whose high primary productivity and biomass concentration drive the largest area affected by fires every year. The characteristic weather conditions in summer (dry and hot) facilitate the appearance of fires on the Iberian Peninsula, with a large number of fires in the dry season (Vázquez de la Cueva, 2012). The drivers identified by our models confirm wildfires in the Iberian Peninsula as important producers of PyC that can be measured in the topsoil. This agrees with expectations and so increases confidence in the results obtained from this model. Another indicator of primary production – i.e. the normalized difference vegetation index – has also been found to be the strongest predictor of PyC stocks, in the Wyoming, Colorado, Kansas and New Mexico region (Ahmed et al., 2017), as. Topsoil pH was not found to be an important driver, contrary to what other previous studies have shown (e.g. Reisser et al., 2016; Braadbaart et al., 2009). In fact, archaeologists usually use the pH as a parameter to identify sites where charcoal remains may be found. However, this contrasting result may be explained in part by the net primary production (NPP) in the Iberian Peninsula being more strongly limited by water availability than by topsoil pH. Actually, the region with the lowest topsoil pH is the northwest, which has the highest NPP thanks to tree species that grow well under acidic conditions and the highest precipitation of the Iberian Peninsula. In addition, the spatial overlap between precipitation and topsoil pH gradients may have reduced the importance of soil pH as a driver of topsoil PyC in the model.

Table 2
Comparison of PyC by land use class.

Land use	Current study	Reisser et al. (2016)
Agriculture	11.7%	16.0%
Grasslands	12.2%	12.1%
Forests	12.8%	9.7%
Moors and heathland	13.5%	12.3%

3.5. Validation by primary productivity & wildfire history

The flux of PyC from vegetation to topsoils during and after wildfires, over the time period relevant to validate our chemometrically-estimated topsoil PyC contents, is largely a function of the fuel load in the standing biomass and the wildfire history of a given location. Therefore, we compared our results against data on primary productivity (including key species) and burnt area integrated over the three decades before the LUCAS sampling in 2009.

3.5.1. Portugal

The distribution of PyC/TOC (%) (Fig. 3) shows a clear north-south contrast, with high topsoil PyC contents in the north and low in the south, roughly separated by the river Tagus. Two notable exceptions to this general trend are the high PyC topsoil contents in western Algarve, and the low PyC topsoil contents in the SE part of the Castelo Branco district. This pattern is closely aligned with forest types. The high PyC topsoil contents overlap with the eucalypt and pine forest plantations north of the Tagus and in western Algarve (Fig. 1 in: Gouveia et al., 2010). The low PyC topsoil contents overlaps with the mainly cork oak (*Quercus suber*) forest (montado) and arable landscape south of the Tagus and in the SE of the Castelo Branco district. The fire susceptibility (Fig. 19 in. Verde and Zêzere, 2010) and forest cover are closely related, thereby providing a potential PyC flux mechanism that could explain the topsoil PyC distribution observed in our chemometrically-modelled values (Fig. 3). Measured burnt area data, for the three decades preceding the 2009 LUCAS sampling (Fig. S2), also reflects the observed patterns in our PyC map (Fig. 3). The districts south of the river Tagus, i.e. Evora, Setubal, Beja and Faro have cumulative burnt areas smaller combine than those for districts north of the river Tagus.

3.5.2. Spain

The north-western part of Spain (Galicia, Asturias and Cantabria regions) exhibited very high contents of PyC relative to TOC. Vazquez de la Cueva et al. (2012) showed that these regions, which have a much greater forested area than the average for the whole country, suffer the highest number of fires and burned area per municipality. Official data (Ministry of Agriculture, Food and Environment of Spain) reported that of the total area burned, the most affected were, in order; *Pinus halepensis* (22.0%), *Pinus pinaster* (21.3%) and *Eucalyptus globulus* (12.3%). In contrast, the Guadalquivir (SW) and the Ebro (NE) valleys show low density of wildfires, coinciding with low PyC/TOC rates (%; Fig. 3).

In the south-eastern region, the "*Sierra de Cazorla, Segura y las Villas*" national park, a UNESCO Biosphere Reserve since 1983, hosts the largest continuous wooded area (pine forests) in Spain. Over the last 50 years it has been affected by numerous wildfires, some of them >5000 ha, such as the one that affected the western slope of the Tranco reservoir in 2005. In this area the topsoil PyC estimates are very high in relation to the surrounding areas. Another area of similar behavior than "*Sierra de*

Cazorla" is the *Alcornocales* National Park, located in the extreme south of Spain, on the eastern slope of the Strait of Gibraltar, showing a high incidence of wildfires concurrent with PyC to TOC rates.

3.5.3. Iberian Peninsula

The only other published soil PyC/TOC (%) map that included the Iberian Peninsula (Reisser et al., 2016) shows a near opposite pattern compared to the map of this current study, i.e. low PyC contents in the NW compared to the rest of the Iberian Peninsula. There are several methodological differences between the two studies that may have contributed to this discrepancy. First, and perhaps most important, is the roughly factor 1000 greater resolution of sites in the Iberian Peninsula that are used in the current study. Reisser et al. (2016) used 569 sites globally, with only three sites for the Iberian Peninsula (2961 in the current study). Considering that the Iberian Peninsula is a very diverse area in terms of wildfire activity, climate, vegetation and soil type, the spatial resolution of the current study seems more appropriate to map PyC contents in soils. Second, Reisser et al., (2016) used data from six different PyC quantification techniques for the PyC model, including thermal, optical and spectroscopic methods, while the current study used a single method to quantify PyC and the samples were taken on a systematic grid with the same methodology, sample pre-treatment, storage and analysis (Tóth et al., 2013; Orgiazzi et al., 2018). Hammes et al. (2007) demonstrated in an interlaboratory comparison study that the direct comparison of PyC data in soils from such different techniques yielded "widely different BC contents for the environmental matrices".

3.6. Validation by anthropogenic sources: power plants

An additional potential flux of PyC to soils is the dry/wet deposition of PyC aerosols emitted by anthropogenic activity, e.g. coal power plants, traffic, domestic fires. Coal and biomass power plants generate a high quantity of aerosols by combustion (Li et al., 2019; Nzihou and Stanmore, 2015). In 2009, the Iberian Peninsula had 23 coal power plants (two in continental Portugal and 21 in continental Spain (Fig. 6; CarbonBrief, 2021; Portuguese Institute for the Conservation of Nature and Forests (ICNF, I.P.), 2021, Resourcewatch, 2021 and Endcoal, 2021). Some plants have been in production for more than 40 years, with a production capacity of >1000 MW. In addition, there are power plants that use biomass to generate energy, 15 in Portugal (central and north of Portugal), and over 60 in Spain. The energy production of these plants is roughly 10% of the coal power plants, not exceeding the capacity of 150 MW (Resourcewatch, 2021).

3.6.1. Portugal

The southern coal power plant is located in a coastal area, without nearby forest but with measured soil PyC/TOC of 12.49–15.19%, i.e. above the median for the Iberian Peninsula. It is possible that the power station has contributed to the PyC contents of local topsoils, but further research is required to confirm this. No evidence of contributions to topsoil PyC can be observed for the other coal power plant. However, it is surrounded by forests with frequent wildfires, and so PyC/TOC ranged from 15.41 to 20.39%. As such, any contribution may have been masked by the greater contribution of forest fires. Regarding the biomass power plants, six are located in coastal areas where wildfires are not common. Downwind from these power plants, there are forests where frequent wildfires occur within 10 km, again hindering observation of evidence of contributions to topsoil PyC. This confounding issue is also present in the location of other biomass power plants where soil PyC/TOC range from 12.80 to 20.39%, and hence the source of PyC cannot be contributed to the power plants.

3.6.2. Spain

The north of Spain contains an area with a high density of coal energy: 13 coal power plants (Fig. 6), which represent 56% of all coal power plants in the Iberian Peninsula. This area overlaps with a high frequency of wildfires (Vázquez de la Cueva et al., 2006; Vázquez de la Cueva, 2012) and a high topsoil PyC/TOC content according to our

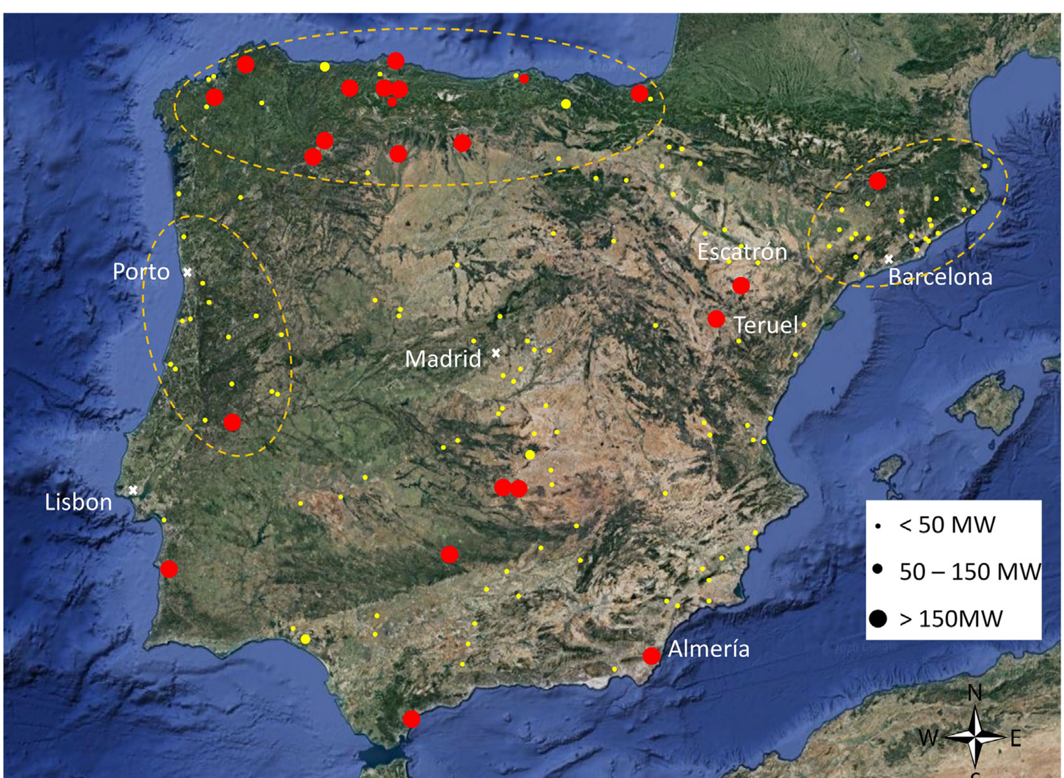


Fig. 6. Location of power plants operating in the Iberia peninsula in 2009. Coal power plants are represented as red circles. Yellows circles represent the power plants that use biomass. Dashed lines mark the areas with high density of power plants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model (10.61–20.39%). The area close to Barcelona has one coal power plant and 23 biomass power plants, with soils there also exhibiting high PyC/TOC values ranging from 8.02 to 15.40%. This suggests evidence for potential contribution to topsoil PyC from power plants as these are located in areas where wildfires are very rare. In Fig. 6, there are three big coal power plants which are located in areas practically unaffected by wildfires (Escatrón and Teruel in the NE of Spain, and Almería in the SE of Spain) and these areas show higher PyC/TOC contents 10.61–20.39% for Escatrón and Teruel, and 10.61–15.40% for Almería, possibly showing the contribution of this industry to the PyC content in the soil, although more research is needed to confirm this hypothesis. This certainly would require the contribution of different types of analysis to distinguish qualitative differences in PyC composition and sources. For example, molecular composition and isotopic studies would be necessary.

3.7. Implications of the results

The results presented here show that it is possible to model the expected proportions of PyC in total SOC by applying the approaches described above.

By applying the techniques comprised in this paper, combined with further ^{13}C NMR spectroscopy of soil samples to develop a more robust chemometric model, it will be possible to develop an accurate global scale database/map. Global quantification of individual soil C pools, including SOC, PyC and inorganic carbon will allow more effective prediction of carbon flows between global compartments (soil/ atm/ ocean and sediment), as well as facilitating MRV of large-scale application of geo-engineering approaches such as biochar use.

3.7.1. Limitations

The first limitation of the procedure developed in this work is the need to use ^{13}C NMR spectroscopy to determine the aryl-C content and validate the model. This methodology demands a lot of analysis time and cost per sample, required by the ^{13}C NMR spectroscopy, which makes the number of samples used to build the model still low. There is a need to expand the number of samples and include other soil types to build a more robust model, but funding would be required. Thus, the chemometric method can be improved. In this study, the number of samples to generate the prediction model by PLS were limited to 42 soil samples. Using more soil samples for the chemometric method, with a wider range of PyC content to develop the prediction model, would improve the precision in the prediction of PyC and consequently the method could be applied to higher scale maps with more confidence.

Another important aspect focuses on the use of IR in the study of the soil. The mineral matrix in the soil is an important fraction that can mask information about organic bands that are usually less intense. In this work, we try to reduce this effect in the model, discarding samples with low TOC as well as high carbonate content, but completely avoiding this effect is impossible. Laboratory treatment of the samples (acid demineralisation/carbonate removal) would improve the quality of the IR spectra but adds other sources of uncertainty. We must also consider that the IR spectrum provides an overview of the functional groups present on the SOM, making it difficult to evaluate the effect of possible interferents that may contribute to these structures. The effect of the presence of water and particle size are always present and the affect the quality of the IR spectra. Higher accuracy of spectral model performance has been reported for “fine-ground” samples (Wijewardane et al., 2021). The 42 samples used to generate the model were carefully milled and dried to minimize the artifacts, but we have to consider that the LUCAS samples were only crushed and sieved over 2 mm (Ward et al., 2020). Further improvements in PyC estimates could be achieved in future surveys if the soil sample material is milled to <250 μm prior analysis (as described in ISO11464:2006).

In this study, the final digital soil maps showed that RF was able to explain substantial part of the spatial variation in the PyC by using the available covariates, although some of the variation remains unexplained because of the relatively poor density of observations or sampling design. In general, model performances and prediction accuracy were successful but limited. The map accuracy is partly determined by the number and spatial locations of the measurements used to calibrate the machine learning model (Wadoux et al., 2019). While some of the error is a result of sampling, some error can be caused by limitations of the predictive variables, for example, technical uncertainties and limited accuracies of the modelled environmental layers (soil depth, DEM, etc.), or the sampling sites measurements. Based on these results, the spatial distribution of topsoil PyC is highly variable due to small scale input variations and structural variability of PyC, which also limits the assessment performance.

This study estimates PyC as a proportion of total SOC. Estimates of PyC topsoil stocks are presented in the Supplement. There are three main sources of uncertainty in the PyC stock data: i) it uses a PTF for BD, not measured data; ii) it assumes a single rock density to convert gravimetric to volumetric stone content, iii) the LUCAS database uses a 20 cm standard topsoil depth (some soils are known to have topsoils <20 cm so this will have overestimated the PyC stocks for these soils, which will also have affected the RF model). These multiple sources of uncertainty may limit the accuracy of PyC stock estimates. These uncertainties could be addressed in future LUCAS – or other – sampling campaigns by inclusion of BD measurements in the field (ISO 11272:2017), a measurement of volumetric stone content, e.g. by the immersion method, and a record of actual topsoil depth, respectively. Another required aspect is to better understand dynamics in deeper soil layers; the present work studies only the topsoil (< 20 cm) but is well known that the carbon content in deeper layers is important too (Batjes, 2016; Jobbágy and Jackson, 2000).

4. Conclusions

The spatial distribution of topsoil PyC measured by a chemometric analysis of about 3000 LUCAS topsoil samples, correlates with known drivers of PyC production (biomass, wildfires, power plants). PyC content of topsoils varied by land use: forest > pastures/grasslands > arable soils. The PLS regression using the mid-IR spectra (400–1800 cm^{-1}) to predict the PyC followed by modelling to predict geospatial distribution can be a tool for surveying and monitoring PyC distribution and contents in different areas, which may be useful for future soil management strategies and policy validation. PyC represents a significant pool of terrestrial carbon that has not been effectively included in soil carbon modelling and mapping. This study confirmed that forests accumulate more PyC in the topsoil than agricultural soils and that recurrent wildfires have a notable influence on topsoil PyC contents. Increased understanding of this topic will facilitate the effectiveness of future global carbon modelling.

CRediT authorship contribution statement

F.G.A. Verheijen: Conceptualization, Methodology, Investigation, Visualization, Data analysis, Writing - original draft. **M.A. Jiménez**: Methodology, Investigation, Visualization, Data analysis, Writing - original draft. **J.M. de la Rosa**: Methodology, Investigation, Data analysis, Writing - original draft. **E. Aksoy**: Visualization, Data curation. **S. Jeffery**: Methodology, Investigation, Data analysis, Writing - review & editing. **B.R.F. Oliveira**: Data analysis, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

José M. De la Rosa thanks the former Spanish Ministry of Economy and Competitiveness (MINECO) for his "Ramón y Cajal" post-doctoral contract (RYC2014-16338). We gratefully acknowledge the Portuguese Foundation for Science and Technology (FCT) for the Assistant Researcher funding of F.G.A. Verheijen (CEECIND/02509/2018). Thanks are also due for the financial support to CESAM (UID/AMB/50017/2019), to FCT/MCTES through national funds, and the cofunding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020. The LUCAS topsoil dataset used in this work was made available by the European Commission through the European Soil Data Centre managed by the Joint Research Centre (JRC), http://esdac.jrc.ec.europa.eu/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.148170.

References

4 PER 1000, last access 14th December 2020, https://www.4p1000.org/.

Abney, R.B., Berhe, A.A., 2018. Pyrogenic carbon erosion: implications for stock and persistence of pyrogenic carbon in soil. Front. Earth Sci. 6, 26. https://doi.org/10.3389/feart.2018.00026.

Ahmed, Z., Woodbury, P., Sanderman, J., Hawke, B., Jauss, V., Solomon, D., Lehmann, J., 2017. Assessing soil carbon vulnerability in the Western USA by geospatial modeling of pyrogenic and particulate carbon stocks: soil carbon vulnerability in Western USA. J. Geophys. Res. Biogeosci. 122, 354–369. https://doi.org/10.1002/2016JG003488.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723. https://doi.org/10.1007/978-1-4612-1694-0_16.

Batjes, N.H., 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma 269, 61–68. https://doi.org/10.1016/j.geoderma.2016.01.034.

Bird, M.I., Wynn, J.G., Saiz, G., Wurster, C.M., McBeath, A., 2015. The pyrogenic carbon cycle. Annu. Rev. Earth Planet. Sci. 43, 273–298.

Braadbaart, F., Poole, I., van Brussel, A.A., 2009. Preservation potential of charcoal in alkaline environments: an experimental approach and implications for the archaeological record. J. Archaeol. Sci. 36, 1672–1679. https://doi.org/10.1016/j.jas.2009.03.006.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1023/A:1018054314350.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software https://doi.org/10.1201/9781315139470.

CarbonBrief, last access 17th January 2021. https://www.carbonbrief.org/mapped-worlds-coal-power-plants.

Carvajal, G., Maucec, M., Cullick, S., 2018. Chapter four - components of artificial intelligence and data analytics. Intell. Digit. Oil Gas Fields, 101–148 https://doi.org/10.1016/B978-0-12-804642-5.00004-9.

Cotrufo, M.F., Boot, C., Kampf, S., Nelson, P., Brogan, D.J., Covino, T., Haddix, M.L., MacDonald, L.H., Rathburn, S., Ryan-Bukett, S., Schmeer, S., Hall, E., 2016. Redistribution of pyrogenic carbon from hillslopes to stream corridors following a large montane wildfire. Glob. Biogeochem. Cycles 30, 1348–1355. https://doi.org/10.1002/2016GB005467.

Dangal, S.R.S., Sanderman, J., 2020. Is standardization necessary for sharing of a large mid-infrared spectral library? Sensors 20, 6729. https://doi.org/10.3390/s20236729.

De la Rosa, J.M., Sánchez García, L., de Andrés, J.R., González-Vila, F.J., González-Pérez, J.A., Knicker, H., 2011. Contribution of black carbon in recent sediments of the Gulf of Cadiz: applicability of different quantification methodologies. Quat. Int. 243, 264–272. https://doi.org/10.1016/j.quaint.2011.01.034.

De la Rosa, J.M., Rosado, M., Paneque, M., Miller, A.Z., Knicker, H., 2018. Effects of aging under field conditions on biochar structure and composition: implications for biochar stability in soils. Sci. Total Environ. 613-614, 969–976. doi: https://doi.org/10.1016/j.scitotenv.2017.09.124.

De la Rosa, J.M., Jiménez-González, M.A., Jiménez-Morillo, N.T., Knicker, H., Almendros, G., 2019. Quantitative forecasting black (pyrogenic) carbon in soils by chemometric analysis of infrared spectra. J. Environ. Manag. 251, 109567. https://doi.org/10.1016/j.jenvman.2019.109567.

Díaz-Uriarte, R., de Andrés, S.A., 2006. Gene selection and classification of microarray data using random forest. Bmc Bioinforma. 7. https://doi.org/10.1186/1471-2105-7-3.

Endcoal, last access 17th January 2021, https://endcoal.org/tracker/.

Fernández-Getino, A.P., Hernández, Z., Piedra Buena, A., Almendros, G., 2013. Exploratory analysis of the structural variability of forest soil humic acids based on multivariate processing of infrared spectral data. Eur. J. Soil Sci. 64, 66–79. https://doi.org/10.1111/ejss.12016.

Forbes, M.S., Raison, R.J., Skjemstad, J.O., 2006. Formation, transformation and transport of black carbon (charcoal) in terrestrial and aquatic ecosystems. Sci. Total Environ. 370, 190–206. https://doi.org/10.1016/j.scitotenv.2006.06.007.

Goldberg, E.D., 1985. Black Carbon in the Environment. vol. 198pp. John Wiley & Sons, New York, NY.

Gouveia, C., DaCamara, C.C., Trigo, R.M., 2010. Post-fire vegetation recovery in Portugal based on spot/vegetation data. Nat. Hazards Earth Syst. Sci. 10, 673–684. https://doi.org/10.5194/nhess-10-673-2010.

Hammes, K., Schmidt, M.W.I., Smernik, R.J., Currie, L.A., Ball, W.P., Nguyen, T.H., Louchouarn, P., Houel, S., Gustafsson, O., Elmquist, M., Cornelissen, G., Skjemstad, J.O., Masiello, C.A., Song, J., Peng, P., Mitra, S., Dunn, J.C., Hatcher, P.G., Hockaday, W.C., Smith, D.M., Hartkopf-Fröder, C., Böhmer, A., Lüer, B., Huebert, B.J., Amelung, W., Brodowski, S., Huang, L., Zhang, W., Gschwend, P.M., Flores-Cervantes, D.X., Largeau, C., Rouzaud, J.-N., Rumpel, C., Guggenberger, G., Kaiser, K., Rodionov, A., Gonzalez-Vila, F.J., Gonzalez-Perez, J.A., De la Rosa, J.M., Manning, D.A.C., López-Capél, E., Ding, L., 2007. Comparison of quantification methods to measure fire-derived (black/elemental) carbon in soils and sediments using reference materials from soil, water, sediment and the atmosphere. Glob. Biogeochem. Cycles 21, GB3016. https://doi.org/10.1029/2006GB002914.

Inventario Nacional de Erosion, last access 13th January 2021, https://www.mapa.gob.es/es/desarrollo-rural/temas/politica-forestal/inventario-cartografia/inventario-nacional-erosion-suelos/.

ISO11464:2006, 2006. Soil Quality — Pretreatment of Samples for Physico-chemical Analysis. International Organization for Standardization, Genève, Switzerland https://www.iso.org/standard/37718.html.

Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.

Knicker, H., Müller, P., Hilscher, A., 2007. How useful is chemical oxidation with dichromate for the determination of "Black Carbon" in fire-affected soils? Geoderma 142, 178–196. https://doi.org/10.1016/j.geoderma.2007.08.010.

Köchy, M., Hiederer, R., Freibauer, A., 2015. Global distribution of soil organic carbon–Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. Soil 1, 351–365. https://doi.org/10.5194/soil-1-351-2015.

Koele, N., Bird, M., Haig, J., Marimon-Junior, B.-H., Schwantes, B., Phillips, O.L., Oliveira, E., Quesada, C.A., Feldpausch, T., 2017. Amazon Basin forest pyrogenic carbon stocks: first estimate of deep storage. Geoderma 306, 237–243. https://doi.org/10.1016/j.geoderma.2017.07.029.

Laird, A.D., 2008. The charcoal vision: A win-win-win scenario for simultaneously producing bioenergy, permanently sequestering carbon, while improving soil and water quality. Agron. J. 100, 178–181. https://doi.org/10.2134/agronj2007.0161.

Lehmann, J., Joseph, S., 2009. Biochar systems. In: Lehmann, C.J., Joseph, S. (Eds.), Biochar for Environmental Management: Science and Technology. Earthscan, London.

Lehmann, J., Skjemstad, J., Sohi, S., Carter, J., Barson, M., Falloon, P., Coleman, K., Woodbury, P., Krull, E., 2008. Australian climate-carbon cycle feedback reduced by soil black carbon. Nat. Geosci. 1 (12), 832–835. https://doi.org/10.1038/ngeo358.

Li, Z., Wang, Y., Lu, Y., Biswas, P., 2019. Investigation of aerosol and gas emissions from a coal-fired power plant under various operating conditions. J. Air Waste Manage. Assoc. 69, 34–46. https://doi.org/10.1080/10962247.2018.1503981.

Lopez-Martin, M., Gonzalez-Vila, F.J., Knicker, H., 2018. Distribution of black carbon and black nitrogen in physical soil fractions from soils seven years after an intense forest fire and their role as C sink. Sci. Total Environ. 637, 1187–1196. https://doi.org/10.1016/j.scitotenv.2018.05.084.

LUCAS, 2009. TOPSOIL data. European Soil Data Centre (ESDAC). European Commission, Joint Research Centre esdac.jrc.ec.europa.eu.

Masiello, C.A., 2004. New directions in black carbon organic geochemistry. Mar. Chem. 92, 201–213. https://doi.org/10.1016/j.marchem.2004.06.043.

Masiello, C.A., Druffel, E.R.M., 1998. Black carbon in deep-sea sediments. Science 280 (5371), 1911–1913. https://doi.org/10.1126/science.280.5371.1911.

Matosziuk, L.M., Alleau, Y., Kerns, B., Bailey, J., Johnson, M.G., Hatten, J.A., 2019. Effects of season and interval of prescribed burns on pyrogenic carbon in ponderosa pine stands in the southern Blue Mountains, Oregon, USA. Geoderma 348, 1–11. https://doi.org/10.1016/j.geoderma.2019.04.009.

Matthews, J.A., 2008. Carbon negative biofuels. En. Pol. 36, 940–945. https://doi.org/10.1016/j.enpol.2007.11.029.

Ministry of Agriculture, Food and Environment of Spain, last access 23th December 2020 https://www.mapa.gob.es/es/desarrollo-rural/estadisticas/Incendios_default.aspx.

Nocita, M., Stevens, A., van Wesemael, B., et al., 2015. Soil spectroscopy: an opportunity to be seized. Glob. Chang. Biol. 21, 10–11. https://doi.org/10.1111/gcb.12632.

Nzihou, A., Stanmore, B.R., 2015. The formation of aerosols during the co-combustion of coal and biomass. Waste Biomass Valor. 6, 947–957. https://doi.org/10.1007/s12649-015-9390-3.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS soil, the largest expandable soil dataset for Europe: A review. Eur. J. Soil Sci. 69 (1), 140–153. https://doi.org/10.1111/ejss.12499.

Portuguese Institute for the Conservation of Nature and Forests (ICNF, I.P.), last access 4 January 2021. http://www2.icnf.pt/portal/florestas/dfci/inc/estat-sgif#balanco-anual,%20last%20visited%20on%2004%20January%202021.

Prichard, S.J., Stevens-Rumann, C.S., Hessburg, P.F., 2017. Tamm review: shifting global fire regimes: lessons from reburns and research needs. For. Ecol. Manag. 396, 217–233. https://doi.org/10.1016/j.foreco.2017.03.035.

Pyne, S.J., 1997. Vestal Fire: An Environmental History, Told through Fire, of Europe and Europe's Encounter with the World. Weyerhaeuser Environmental Books. University of Washington Press, Seattle, WA, USA.

Reisser, M., Purves, R.S., Schmidt, M.W., Abiven, S., 2016. Pyrogenic carbon in soils: a literature-based inventory and a global estimation of its content in soil organic carbon and stocks. Front. Earth Sci. 4, 80. https://doi.org/10.3389/feart.2016.00080.

Resourcewatch, last access 15th January 2021, https://resourcewatch.org/.

Rovira, P., Rovira, Duguy, Vallejo R, B., 2009. Black carbon in wildfire-affected shrubland Mediterranean soils. J. Plant Nutr. Soil Sci. 172, 43–52. https://doi.org/10.1002/jpln.200700216.

Sanderman, J., Baldock, J.A., Dangal, S.R.S., Ludwig, S., Potter, S., Rivard, C., Savage, K., 2021. Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. Biogeochemistry https://doi.org/10.1007/s10533-021-00755-1.

Santín, C., Doerr, S.H., Preston, C., González-Rodríguez, G., 2015. Pyrogenic organic matter production from wildfires: a missing sink in the global carbon cycle. Glob. Chang. Biol. 21, 1621–1633. https://doi.org/10.1111/gcb.12800.

Santín, C., Doerr, S.H., Kane, E.S., Masiello, C.A., Ohlson, M., Rosa, J.M., Preston, C.M., Dittmar, T., 2016. Towards a global assessment of pyrogenic carbon from vegetation fires. Glob. Chang. Biol. 22, 76–91. https://doi.org/10.1111/gcb.12985.

Schlesinger, W.H., 1997. Biogeochemistry: An Analysis of Global Change. 2nd edn. Academic Press, San Diego.

Schmidt, M.W.I., Noack, A.G., 2000. Black carbon in soils and sediments: analysis, distribution, implications, and current challenges. Glob. Biogeochem. Cycles 14, 777–794. https://doi.org/10.1029/1999GB001208.

Schmidt, M.W.I., Skjemstad, J.O., Czimczik, C.I., Glaser, B., Prentice, K.M., Gelinas, Y., Kuhlbusch, T.A., 2001. Comparative analysis of black carbon in soils. Glob. Biogeochem. Cycles 15, 163–167. https://doi.org/10.1029/2000GB001284.

Selvalakshmi, S., de la Rosa, J.M., Zhijun, H., Guo, F., Ma, X., 2018. Effects of ageing and successive slash-and-burn practice on the chemical composition of charcoal and yields of stable carbon. CATENA 162, 141–147. https://doi.org/10.1016/j.catena.2017.11.028.

Simeone, G.R., Benesch, M., Glaser, G., 2018. Degradation products of polycondensed aromatic moieties (black carbon or pyrogenic carbon) in soil: methodological improvements and comparison to contemporary black carbon concentrations. J. Plant Nutr. Soil Sci. 181, 714–720. https://doi.org/10.1002/jpln.201800079.

Simpson, M., Hatcher, P.G., 2004. Determination of Black Carbon in natural organic matter by chemical oxidation and solid-state ^{13}C nuclear magnetic resonance spectroscopy. Org. Geochem. 35, 923–935. https://doi.org/10.1016/j.orggeochem.2004.04.004.

Skjemstad, J.O., Reicosky, D.C., Wilts, A.R., McGowan, J.A., 2002. Charcoal carbon in US agricultural soils. Soil Soil Sci. Soc. Am. J. 66, 1249–1255. https://doi.org/10.2136/sssaj2002.1249.

Sohi, S., Krull, E., Lopez-Capel, E., Bol, R., 2010. A review of biochar and its use and function in soil. Adv. Agron. 105, 47–82. https://doi.org/10.1016/S0065-2113(10)05002-9.

Soucémarianadin, L., Quideau, S.A., MacKenzie, M.D., 2014. Pyrogenic carbon stocks and storage mechanisms in podzolic soils of fire-affected Quebec black spruce forests. Geoderma 217–218, 118–128. https://doi.org/10.1016/j.geoderma.2013.11.010.

Soucémarianadin, L., Reisser, M., Cécillon, L., Barréa, P., Nicolas, M., Abiven, S., 2019. Pyrogenic carbon content and dynamics in top and subsoil of French forests. Soil Biol. Biochem. 133, 12–15. https://doi.org/10.1016/j.soilbio.2019.02.013.

Suman, D.O., Kuhlbusch, T.A.J., Lim, B., 1997. Marine sediments: a reservoir for black carbon and their use as spatial and temporal records of combustion. In: Clark, J.S., Cachier, H., Goldammer, J.G., Stocks, B.J. (Eds.), Sediment Records of Biomass Burning and Global Change. Springer, Berlin, Heidelberg, pp. 271–293.

Tóth, G., Jones, A., Montanarella, L., 2013. LUCAS Topsoil Survey. Methodology, Data and Results. JRC Technical Reports. Luxembourg. Publications Office of the European Union, EUR26102 – Scientific and Technical Research series – ISSN 1831-9424 (online). https://doi.org/10.2788/97922 (ISBN 978-92-79-32542-7).

Vázquez de la Cueva, A., 2012. Relative contributions of Forest vegetation, land cover, topography and climate in explaining fire regime patterns (1974–2005) in Peninsular Spain. ISRN Forestry 2012, 479491. https://doi.org/10.5402/2012/479491.

Vázquez de la Cueva, A., García del Barrio, J.M., Quero, M.O., Palomares, O.S., 2006. Recent fire regime in peninsular Spain in relation to forest potential productivity and population density. Int. J. Wildland Fire 15, 397–405. https://doi.org/10.1071/WF05071.

Verde, J.C., Zêzere, J.L., 2010. Assessment and validation of wildfire susceptibility and hazard in Portugal. Nat. Hazards Earth Syst. Sci. 10, 485–497. https://doi.org/10.5194/nhess-10-485-2010.

Verheijen, F.G., Jones, R.J., Rickson, R.J., Smith, C.J., 2009. Tolerable versus actual soil erosion rates in Europe. Earth Sci. Rev. 94 (1–4), 23–38. https://doi.org/10.1016/j.earscirev.2009.02.003.

Viscarra Rossel, R.P., 2008. ParLeS: software for chemometric analysis of spectroscopic data. Chemom. Intell. Lab. Syst. 90, 72–83. https://doi.org/10.1016/j.chemolab.2007.06.006.

Wadoux, A.M.J.-C., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest. Geoderma 355, 113913. https://doi.org/10.1016/j.geoderma.2019.113913.

Wang, X., Sanderman, J., Yoo, K., 2018. Climate-dependent topographic effects on pyrogenic soil carbon in Southeastern Australia. Geoderma 322, 121–130. https://doi.org/10.1016/j.geoderma.2018.02.025.

Ward, K.J., Chabrillat, S., Brell, M., Castaldi, F., Spengler, D., Foerster, S., 2020. Mapping soil organic carbon for airborne and simulated EnMAP imagery using the LUCAS soil database and a local PLSR. Remote Sens. 12 (20), 3451.

Wijewardane, N.K., Ge, Y., Sanderman, J., Ferguson, R., 2021. Fine grinding is needed to maintain the high accuracy of mid-infrared diffuse reflectance spectroscopy for soil property estimation. Soil Sci. Soc. Am. J. 85 (2), 263–272.

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports vol. 106. FAO, Rome, 1–203.

Zong, W., Zhang, J., Jiang, Y., 2020. Chapter 15 - life-oriented household energy consumption research. Transp. Energy Res. Elsevier, 373–391. https://doi.org/10.1016/B978-0-12-815965-1.00015-6.