

RESEARCH ARTICLE

10.1002/2016JG003488

Key Points:

- About 16% of the area throughout four Western USA states is moderately to highly vulnerable to soil organic carbon mineralization in surface soil
- Much variation in the recalcitrant pyrogenic soil carbon fraction in surface soil could be predicted based on vegetation, climate, and terrain variables

Supporting Information:

- Supporting Information S1

Correspondence to:

Z. U. Ahmed,
z.ahmed@cgiar.org

Citation:

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

Received 16 MAY 2016

Accepted 25 JAN 2017

Accepted article online 26 JAN 2017

Published online 18 FEB 2017

Assessing soil carbon vulnerability in the Western USA by geospatial modeling of pyrogenic and particulate carbon stocks

Zia U. Ahmed^{1,2} , Peter B. Woodbury², Jonathan Sanderman^{3,4} , Bruce Hawke³ , Verena Jaus³, Dawit Solomon² , and Johannes Lehmann^{2,5} 

¹International Maize and Wheat Improvement Center (CIMMYT), Dhaka, Bangladesh, ²Crop and Soil Sciences Section, School of Integrative Plant Science, Cornell University, Ithaca, New York, USA, ³CSIRO Land and Water, Glen Osmond, South Australia, Australia, ⁴Woods Hole Research Center, Falmouth, Massachusetts, USA, ⁵Atkinson Center for a Sustainable Future, Cornell University, Ithaca, New York, USA

Abstract To predict how land management practices and climate change will affect soil carbon cycling, improved understanding of factors controlling soil organic carbon fractions at large spatial scales is needed. We analyzed total soil organic (SOC) as well as pyrogenic (PyC), particulate (POC), and other soil organic carbon (OOC) fractions in surface layers from 650 stratified-sampling locations throughout Colorado, Kansas, New Mexico, and Wyoming. PyC varied from 0.29 to 18.0 mg C g⁻¹ soil with a mean of 4.05 mg C g⁻¹ soil. The mean PyC was 34.6% of the SOC and ranged from 11.8 to 96.6%. Both POC and PyC were highest in forests and canyon bottoms. In the best random forest regression model, normalized vegetation index (NDVI), mean annual precipitation (MAP), mean annual temperature (MAT), and elevation were ranked as the top four important variables determining PyC and POC variability. Random forests regression kriging (RFK) with environmental covariables improved predictions over ordinary kriging by 20 and 7% for PyC and POC, respectively. Based on RFK, 8% of the study area was dominated (≥50% of SOC) by PyC and less than 1% was dominated by POC. Furthermore, based on spatial analysis of the ratio of POC to PyC, we estimated that about 16% of the study area is medium to highly vulnerable to SOC mineralization in surface soil. These are the first results to characterize PyC and POC stocks geospatially using stratified sampling scheme at the scale of 1,000,000 km², and the methods are scalable to other regions.

1. Introduction

Soils contain two to three times as much carbon as does the atmosphere [Scharlemann *et al.*, 2014; Schmidt *et al.*, 2011; Trumbore, 2009]. Thus, small changes in soil organic carbon (SOC) pools can have large impacts on concentrations of CO₂ in the atmosphere. Changes in SOC are driven by land management practices including land use change, as well as by environmental factors including ongoing climatic change [Schmidt *et al.*, 2011]. However, different fractions of SOC vary tremendously in their longevity in soil, with particulate organic carbon (POC) being highly vulnerable to rapid turnover of between months and a few years and pyrogenic carbon (PyC) often having very high persistence of decades to thousands of years [Schmidt *et al.*, 2011]. Because of the long residence time of PyC in soils, and the influence PyC has on soil properties, fire can be considered as a soil-forming factor [Certini, 2014].

The ability to predict how SOC will respond to ongoing climatic change and land management practices depends on using Earth systems models at continental to global scales and land management models at local, regional, and national scales [Tonitto *et al.*, 2016]. Models of soil organic matter dynamics such as CENTURY and RothC represent soil organic matter (SOM) in conceptual pools with different turnover times [Parton *et al.*, 1987; Jenkinson and Rayner, 1977]. Much research effort has focused on linking these conceptual pools with measurable pools of SOC, but since stable SOC pools such as PyC can be very long lived, accurate parameterization of such pools is critical for model prediction accuracy [Falloon *et al.*, 2000; Lehmann *et al.*, 2008]. Numerous investigators have quantified spatial patterns of total SOC [Minasny *et al.*, 2013]. However, spatially explicit information on different forms of SOC, such as PyC, and thus vulnerability to mineralization, is lacking, particularly at scales as large as 1,000,000 km² and for ecosystems besides boreal forests [e.g., Soucérianadin *et al.*, 2014] and limited data for native grasslands [e.g., Glaser and Amelung, 2003; Rodionov *et al.*, 2010].

Limited spatial assessments have shown that the distribution of PyC in soil is highly variable. It may depend on climatic factors [Glaser and Amelung, 2003; Jauss et al., 2015], on vegetation [Lehmann et al., 2008], on moisture levels [Kane et al., 2007], or on other factors such as erosion [Rumpel et al., 2006]. However, no large-scale spatial analyses have determined which factors are most important in controlling PyC distribution in landscapes. For these reasons, we selected a fire-prone region with a wide range in elevation, climate, and ecosystem types including forests, grasslands, shrublands, and arable lands comprising the states of Kansas, Colorado, New Mexico, and Wyoming, USA (1,051,029 km²). Very few PyC data are available for this region. In the northern and eastern portions of the region, PyC at five grassland sites was analyzed [Glaser and Amelung, 2003], but no geospatial analysis would be possible with so few samples. Moreover, the topic of PyC in soils in forests of the Rocky Mountain West has been reviewed [DeLuca and Aplet, 2008], but this review presented only a qualitative description of the mechanisms of PyC sequestration because data on PyC contents of soils were not available.

Our overall goal was to develop a map of SOC vulnerability by geospatial modeling of relatively persistent (PyC) and easily mineralizable (POC) fractions of SOC. Our specific objectives were to (1) quantify spatially explicit stocks of PyC, POC, and other soil organic carbon (OOC) fractions and (2) quantify the role of climate, fire regime, terrain variables, land cover, and vegetation drivers in controlling the spatial distribution of PyC, POC, and OOC throughout the 1,000,000 km² region.

2. Materials and Methods

2.1. Study Region

We used 650 soil samples from Colorado, Kansas, New Mexico, and Wyoming. These samples were collected by the United States Geological Survey (USGS) as a part of the USGS Geochemical Landscapes Project [Smith et al., 2011]. The total area of the study region is approximately 1,051,029 km² and was chosen for its highly variable topography, vegetation zones, climate, and fire regimes. It encompasses the Rocky Mountains in the west, grassland prairie of the Great Plains in the center, and Alluvial River and Osage Plains in the east. The mean altitude above sea level is 662 m, ranging from < 300 m in the east (Alluvial River Plains and Osage Plains in KS) to > 4000 m in the west (Rocky Mountains) (Figure 1a). The climate in this region varies greatly. Mean annual temperatures range from less than -5°C in the high elevations to greater than 15°C in the southern and southeastern portions (Figure 2a and Table S1 in the supporting information). While some southern portions of the region typically have 70 to 100 days each year over 32°C, the northern portions typically have only 10 to 20 days above 32°C [Global Climate Change Impacts in the United States, 2009]. Mean annual precipitation ranges from less than 200 mm in the western Great Plains that are in the rain shadow of the Rocky Mountains to more than 2000 mm in the east (Figure 2b and Table S1). These diverse environmental conditions are reflected by variation in land cover. The largest area is under grasslands (31%) in the High Plains and shrublands (31%) in semiarid regions of the Great Plains followed by cultivated land (15%) in plains and forest (15%) in high-altitude portions of the Rocky Mountains (Figure 2f).

2.2. Soil Sampling and Preparation

The soil samples were collected throughout the A horizon with sampling densities of 1 sample per ~1600 km² [Smith et al., 2011]. The sampling strategy was a generalized random tessellation stratified design [Smith et al., 2011]. If a target site was inaccessible for any reason, an alternative site was chosen as close as possible to the original site, matching the landscape and soil characteristics of the original site. Sites were also not sampled near roads (50 to 200 m depending on road size) or buildings (100 m) or downwind of major industrial sites (5 km). Samples were air dried and sieved through a 2 mm stainless steel screen and then crushed to <150 μm in a ceramic mill.

2.3. Soil Analysis

The concentration of SOC fractions was predicted by means of mid-infrared (MIR) spectroscopy and partial least squares regression (PLSR) analysis using methods based on algorithms by Haaland and Thomas [1988] (PLSplus/IQ™, Thermo-Electron GRAMS™) described previously [Janik et al., 2007]. The acquired MIR spectra were truncated to 6000–1030 cm⁻¹, baseline-corrected utilizing a simple baseline offset, and

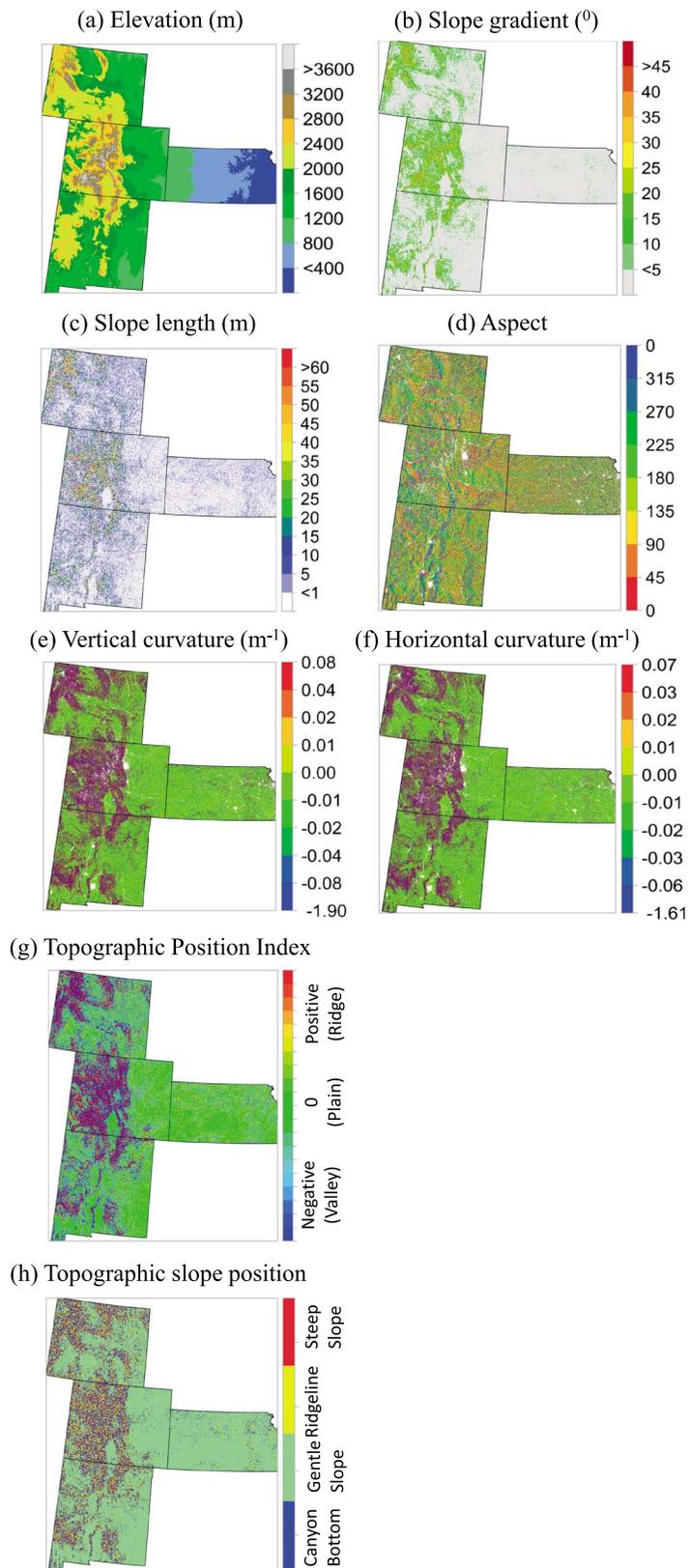


Figure 1. Spatial variation in topographic variables including (a) elevation, (b) slope gradient, (c) slope length, (d) aspect, (e) vertical curvature, (f) horizontal curvature, (g) topographic position index (2000 m \times 2000 m), and (h) topographic position in relation to slope gradient.

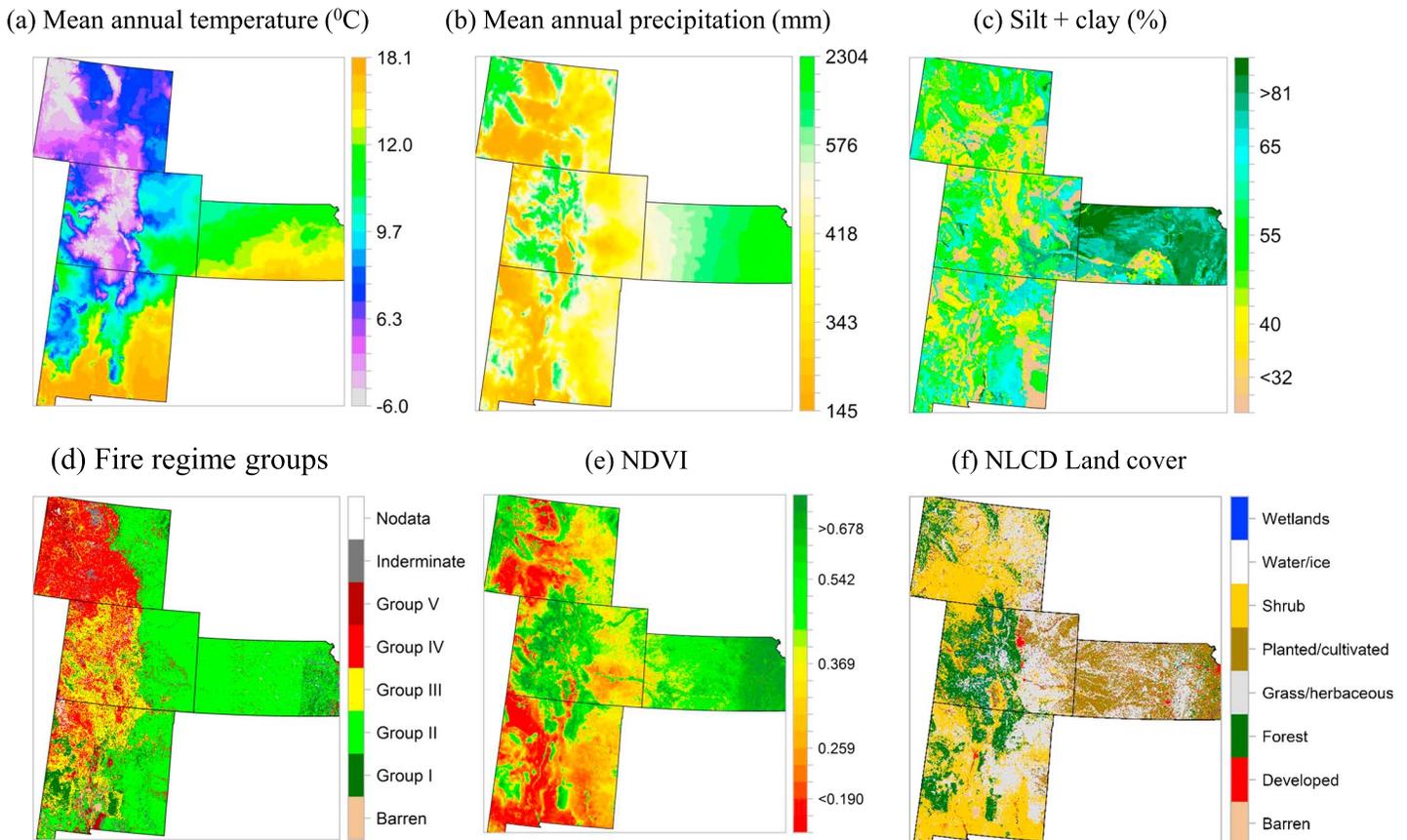


Figure 2. Spatial variation in the independent variables used in the model including (a) mean annual temperature, (b) mean annual precipitation; (c) soil silt + clay, (d) fire regime group; (e) mean normalized difference vegetation index (NDVI) for the month of June and July from 2000 to 2011; and (f) land cover in nine classes as reclassified from the NLCD (see text for details).

mean centered, all using the Unscrambler 10.2 software (CAMO Software AS, Oslo, Norway). Following this spectral preprocessing, predictions for total organic carbon (SOC), particulate organic carbon (POC), other organic carbon (OOC), and pyrogenic carbon (PyC) were made in Unscrambler, using MIR PLSR calibration models from a large Soil Carbon Research Project (SCaRP) conducted in Australia [Baldock *et al.*, 2013b]. The content of soil OC and its fractions in a subset of these calibration data ($n=312$) were measured using an automated Dumas combustion carbon analyzer. The POC fraction was measured based on the amount of OC retained on a $53\ \mu\text{m}$ sieve and PyC was measured using ultraviolet photooxidation and HF treatment followed by solid-state ^{13}C nuclear magnetic resonance (NMR) spectroscopy [Janik *et al.*, 2007]. The OOC fraction was calculated as the soil OC minus POC and PyC fractions. Calibration models for POC, OOC and PyC were produced from these 312 samples. The details of the methodology for determining these fractions and the resultant data used to produce MIR PLSR calibration models for each of the fractions are described in Baldock *et al.* [2013a]. This fractionation analysis was an improvement on methods used to produce previous MIR soil organic carbon fractionation calibration models [Janik *et al.*, 2007] in that it was an automated approach and that it allowed for allocation of PyC to both the $53\text{--}2000\ \mu\text{m}$ and the $<53\ \mu\text{m}$ size fractions. Prediction data generated from the Unscrambler consisted of predicted values with associated deviation or uncertainty values which can be used to calculate an estimate of a 95% confidence interval about the predicted value. In addition, a sample inlier distance and Hotellings T^2 distance can be used to determine how well represented the predicted samples were in the calibration sample set based on their spectral characteristics. This enabled the identification of outlier samples which could be excluded from further analysis. The recovery of SOC within the fractions was calculated as a percentage (equation (1)) and an absolute deviation (equation (2)) where PyC, POC, OOC, and SOC are all expressed in units of mg C g^{-1} soil.

$$\text{Recovery}(\%) = \frac{(\text{PyC} + \text{POC} + \text{OOC})}{\text{SOC}} \times 100 \quad (1)$$

$$\text{Absolute deviation} = (\text{PyC} + \text{POC} + \text{OOC}) - \text{SOC} \quad (2)$$

It should be noted that expressing the recovery of SOC within the fractions as a percentage of SOC can be misleading when SOC contents are low. Under such conditions small measurement errors or sample variability can translate into a large deviation from 100% recovery. To create a more robust data set for further analysis, only samples for which the absolute deviation between the sum of OC in all fractions and the analyzed SOC was within $\pm 5 \text{ mg C g}^{-1}$ soil and percent PyC and POC is less than 100 were selected; these comprised 73% of the total (473/650). We used this subset of 473 samples for further statistical analysis and mapping of PyC and POC.

As mentioned above, the calibration data set includes only Australian soils; and therefore, it is important to evaluate its efficacy for predictions for U.S. soils. We therefore validated our MIR predicted values for total C and SOC using independent data set from the United States Geological Survey [Smith *et al.*, 2014]. Total C (TC) was determined by the use of an automated carbon analyzer at 1370°C to oxidize C to carbon dioxide (CO₂), and the CO₂ gas was measured by a solid state infrared detector. The concentration of SOC was calculated by subtracting the amount of inorganic C (carbonate) from TC concentration. We found a very strong correlation between MIR predicted soil C and measured soil C ($R^2 = 0.91$ for TC, and $R^2 = 0.92$ for SOC; $n = 47$). The residual prediction deviation (RPD = StDEV/root-mean-square error (RMSE)) for TC and SOC was > 2 (2.14 for TC and 2.10 for SOC) which indicates successful prediction [Sinnavee *et al.*, 2001].

2.4. Statistical Analyses

Before statistical analysis, the data set ($n = 473$) was randomly split into 384 calibration data, which were used for training the random forest (RF) regression model and 89 validation data (Figures 3c and 3d) which were used for evaluating the spatial prediction models. Statistical analyses including bivariate correlation, analysis of variance, and RF regression analysis were performed using R statistical software [R Core Team, 2016].

Random forests, based on the assemblage of multiple iterations of decision trees, have become a major data analysis tool that performs well in comparison to single iteration classification and regression tree analysis [Heidema *et al.*, 2006]. Each tree is made by bootstrapping of the original data set which allows for robust error estimation with the remaining test set, the so-called Out-Of-Bag (OOB) sample. The excluded OOB samples are predicted from the bootstrap samples and by combining the OOB predictions from all trees. The RF algorithm can outperform linear regression, and unlike linear regression, RF has no requirements considering the form of the probability density function of the target variable [Hengl *et al.*, 2015; Kuhn and Johnson, 2013]. One major disadvantage of RF is that it is difficult to interpret the relationships between the response and predictor variables. However, RF allows estimation of the importance of variables as measured by the mean decrease in prediction accuracy before and after permuting OOB variables. The difference between the two are then averaged over all trees and normalized by the standard deviation of the differences.

First, we trained the random forests regression model with 384 training data set with bootstrapped 5000 trees and five splits in each tree and then the RF model with the lowest error component and with a subset of optimum numbers of predictors was selected using RF model selection approach [Evans and Cushman, 2009]. "RandomForests" [Liaw and Wiener, 2002] and "rfUtilities" [Evans and Murphy, 2016] packages in the R statistical computing environment [R Core Team, 2016] were used for all statistical analysis.

2.5. Spatial Mapping of POC and PyC

Both ordinary (OK) and random forest regression kriging (RFK) [Hengl *et al.*, 2015] were used to generate prediction maps of POC and PyC concentration from the calibration data set ($n = 384$) at a resolution of $1 \text{ km} \times 1 \text{ km}$ throughout forest, planted/cultivated land, grassland/herbaceous land, and shrubland of the study area. The RFK method combines the RF predicted values with simple kriging (SK) of the regression residuals [Hengl *et al.*, 2015]. The experimental variogram of residuals of the RF regression model was first computed and modeled, and then SK was applied to the residuals to estimate the spatial prediction of the residuals. The RF regression predicted results, and the kriged residuals were added to estimate the interpolated target variable. Because the distribution of SOC fractions was positively skewed (skewness > 1 ; Figures 3a and 3b), we reduced the effect of outliers on variogram properties through the use of robust variograms [Cressie and Hawkins, 1980]. The RFK method was evaluated with the validation data sets ($n = 89$) in

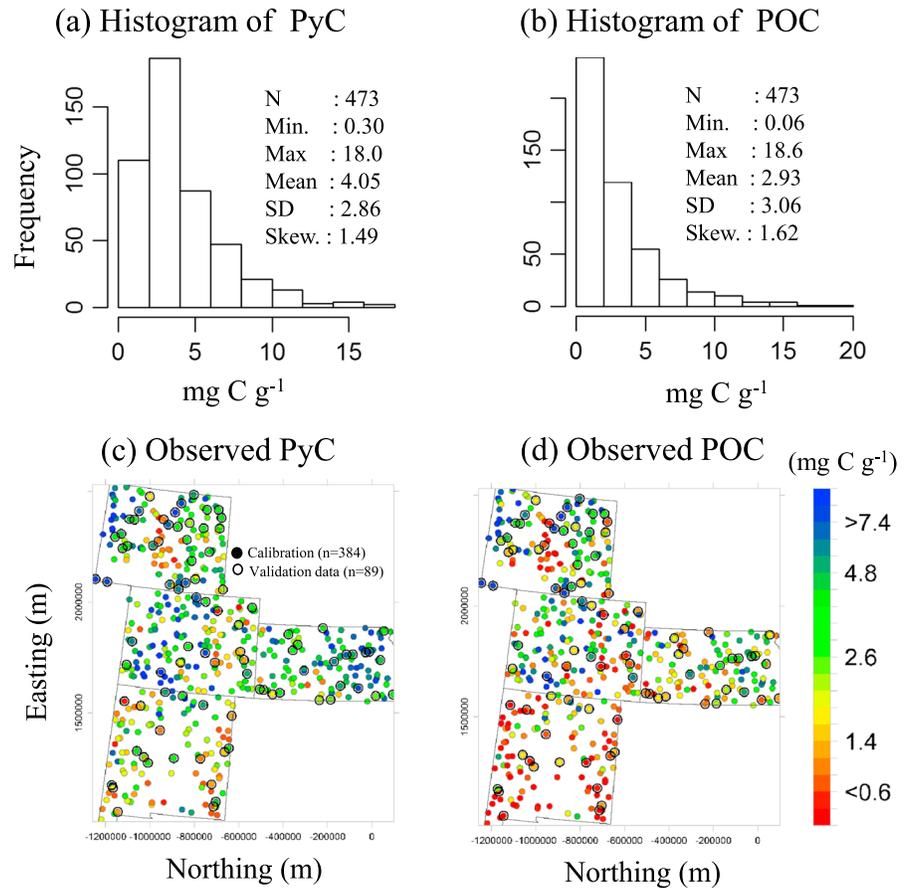


Figure 3. Frequency distributions and spatial distributions of observed pyrogenic carbon (PyC) and particulate organic carbon (POC) of 473 sampling points.

order to obtain an independent error estimate. We used OK as a benchmark method. Diagnostic measures of map quality were the mean absolute error (MAE) and root-mean-square error (RMSE).

$$MAE = \frac{1}{n} \sum_{a=q}^n |z(u_a) - z(u_a)^*| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{a=1}^n [z(u_a) - z(u_a)^*]^2} \tag{4}$$

where n is the number of observation points and $z(u)$ and $z(u)^*$ are the observed and predicted values at location a . The mean MAE shows if the prediction is biased whereas the RMSE measures the accuracy of the predictions.

We also modeled SOC and OOC (Table S3 and Figure S3) to calculate the percentage of SOC contributed by each fraction for the prediction grid. Finally, we developed an index map of vulnerability to SOC mineralization by defining the grids with POC:PyC values ≤ 0.5 as very low, 0.5–1.0 as low, 1.0–1.5 as medium, 1.5–2.0 as high, and ≥ 2.0 as very high. To estimate the area vulnerable to SOC mineralization under different land use and slope position, we overlaid the vulnerability index, National Land Cover Database (NLCD) land use and slope position maps.

3. Environmental Data

We assembled a comprehensive set of spatial environmental data to characterize the region and used these data to predict SOC, PyC, POC, and OOC throughout the study region. These data included terrain, climate, soil, fire, and land cover as described below.

Table 1. Pearson Correlation Coefficient Values (r) Between Environmental Variables and Pyrogenic Carbon (PyC) and Particulate Organic Carbon (POC) for 473 Soil Samples

Variables	r Value	
	PyC	POC
Elevation (ELEV)	0.181***	0.232***
Slope gradient (G)	0.297***	0.256***
Slope length (SL)	0.190***	0.187***
Slope aspect (A)	0.051 ^a	0.029 ^a
Topographic position index (TPI)		
TPI ₂₅₀	−0.042 ^a	−0.0001 ^a
TPI ₅₀₀	−0.123***	−0.113***
TPI ₁₀₀₀	−0.162***	−0.170***
TPI ₁₅₀₀	−0.210***	−0.204***
TPI ₂₀₀₀	−0.225**	−0.213**
Vertical curvature (C_v)	−0.067 ^a	0.049 ^a
Horizontal curvature (C_h)	−0.007 ^a	0.065 ^a
Temperature (MAT)	−0.359***	−0.417***
Precipitation (MAP)	0.491***	0.312**
NDVI	0.607***	0.443***
K-factor (K_f)	−0.102**	−0.131**
Silt + Clay	0.220***	0.083 ^a

^aNot significant.

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

(TPI₂₅₀, TPI₅₀₀, TPI₁₀₀₀, TPI₁₅₀₀, and TPI₂₀₀₀). Finally, we developed a topographic slope position (TSP) map by defining the cells with TPI values ≤ -8 m as canyon bottoms, cells with TPI values ≥ 8 m as ridgelines, and cells with TPI values between -8 and $+8$ m as gentle slopes (slope $< 6^\circ$) or steep slopes (slope $\geq 6^\circ$) (Figure 1h) as described by Dickson and Beier [2007].

3.2. Climate Data

Thirty year (1971–2000) mean annual air temperature (MAT) (Figure 2a) and mean annual precipitation (MAP) (Figure 2b) were obtained from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) climate mapping system (Daly et al. [2001]; accessed at <http://prism.oregonstate.edu/>). PRISM uses point measurements and a digital elevation model (DEM) to generate gridded estimates of annual, monthly, and event-based climatic parameters.

3.3. Soil Data

Soil data were derived from the Digital General Soil Map of the United States (STATSGO2; map scale 1:250,000), an inventory of soils and nonsoil areas that occur in a repeatable pattern on the landscape [Natural Resources Conservation Service (NRCS), United States Department of Agriculture, 2012]. We used the component weighted mean of silt + clay content for surface soil (Figure 2c) and the soil erodibility factor (K factor).

3.4. Fire Regime Groups Data

Fire Regime Groups (FRG) data were obtained from the LANDFIRE (LF) project also known as the Landscape Fire and Resource Management Planning Tools Project (accessed at <http://www.landfire.gov/datatool.php>). The FRG data layer characterizes the estimated historical fire regimes within landscapes based on interactions between vegetation dynamics, fire spread, fire effects, and spatial context. Fire regimes are classified into five groups (Figure 2d): Fire Regime Group (FRG) I ≤ 35 year fire return interval, low and mixed severity; FRG II ≤ 35 year fire return interval, replacement severity; FRG III 35–200 year fire return interval, low and mixed severity; FRG IV 35–200 year fire return interval, replacement severity; FRG V > 200 year fire return interval and replacement severity; and Indeterminate (unknown) fire regime characteristics.

3.1. Terrain Data

For digital terrain modeling (Figure 1), we used 30 m spatial resolution digital elevation model (DEM) data obtained from a USGS database [Multi-Resolution Land Characteristics Consortium, 2007]. The 30 m DEM was resampled to 90 m resolution for further analysis. From DEM data (Figure 1a), we calculated slope gradient (G) (Figure 1b), slope length (SL) (Figure 1c), aspect (A) (Figure 1d), vertical curvature (C_v) (Figure 1e), horizontal curvature (C_h) (Figure 1f), topographic position index (TPI) (Figure 1g), and slope position (TSP) (Figure 1h). We used DEM Surface Tools [Jenness, 2013] for ArcGIS 10 [Environmental Systems Research Institute (ESRI), 2011] to calculate G, C_v , C_h , and TSP. Topographic position indices (TPI) were calculated at each cell of the DEM by calculating the difference between the elevation of the cell and the mean elevation calculated for all cells of a moving rectangular window centered on the cell of interest. Five topographic positions were calculated for grid cells ranging from 250 to 2000 m

Table 2. Summary Statistics of Percentage of Particulate Organic (POC), Pyrogenic (PyC), and Other Organic Carbon (OOC) in Soil Organic C (N = 473)

	POC (%)	PyC (%)	OOC (%)
Mean	20.5	34.6	47.5
Median	19.2	31.9	46.9
SD	12.3	11.16	12.2
Max	87.1	96.6	85.5
Min	0.1	11.8	0.01
25%	13.9	27.7	41.5
75%	25.8	37.2	51.7

3.5. Normalized Difference Vegetation Index

The normalized difference vegetation index (NDVI) data for the months of June and July of the year 2000 to 2011, derived from the Moderate-Resolution Imaging Spectroradiometer Bands 1 (red) and 2 (near infrared), were obtained from the NASA Land Processes Distributed Active Archive Center, USGS/Earth Resources Observation and Science Center, Sioux Falls, South Dakota (accessed at https://lpdaac.usgs.gov/get_data). The study region is covered by six granules of the MOD13Q1 composites. Data download was performed with an R script [R Core Team, 2016], and postprocessing was performed with ArcPython [ESRI,

2011]. We used the R-package “raster” [Hijmans and van Etten, 2016] to calculate mean NDVI for the time period as discussed previously (Figure 2e).

3.6. National Land Cover

We used the National Land Cover Database 2006 (NLCD), which was developed using an unsupervised classification of Landsat Enhanced Thematic Mapper+ (ETM+) circa 2006 satellite data. The NLCD includes 16 land cover classes applied consistently across the conterminous United States at a spatial resolution of 30 m [Fry et al., 2011]. The results were reclassified into nine major classes (Figure 2f): (1) open water; (2) ice/snow; (2) developed (all developed lands); (4) barren; (5) forest (deciduous, evergreen, and mixed forest); (6) shrubland; (7) grass/herbaceous; (8) planted/cultivated (crops, pasture, and hay); and (9) wetlands (woody and herbaceous wetlands).

4. Results

4.1. Descriptive Statistics and Spatial Distribution POC and PyC

The distributions of POC and PyC in the 473 selected soil samples were positively skewed (skewness > 1.5) and showed statistically significant differences ($p < 0.05$) from the normal distribution based on the Kolmogorov-Smirnov (KS) normality test (Figure 3). The POC fraction made up the smallest component of the SOC (mean of 20.3% and median of 19.2%; Table 2). Concentrations of POC ranged from 0.06 to 18.6 mg g⁻¹ soil with a mean of 2.9 mg C g⁻¹ soil and standard deviation of 3.1 mg C g⁻¹ soil (Figure 3b). The PyC fraction ranged from 11.8 to 96.6% with a mean of 34.6% and median of 31.9% of the SOC (Table 2). The concentrations of PyC ranged from 0.3 to 18.0 mg g⁻¹ soil with a mean of 4.0 mg C g⁻¹ soil and a standard deviation of 2.9 mg C g⁻¹ soil (Figure 3a). There was no sample with a lower percentage of PyC/SOC than 11.8%, which was much higher than the minimum for other SOC fractions (Table 2). However, the fraction of SOC comprised of PyC was more variable than for other fractions (Table 2). The

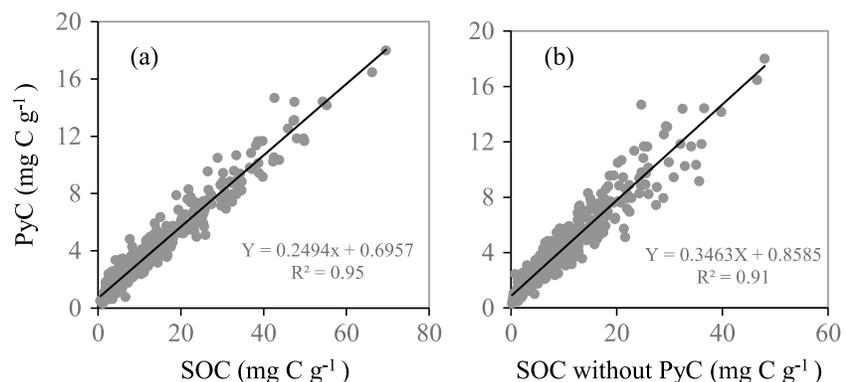


Figure 4. Relationship (a) between pyrogenic soil organic carbon (PyC) and soil organic carbon (SOC) and (b) between PyC and non-PyC SOC throughout four Western USA states (WY, CO, NM, and KS).

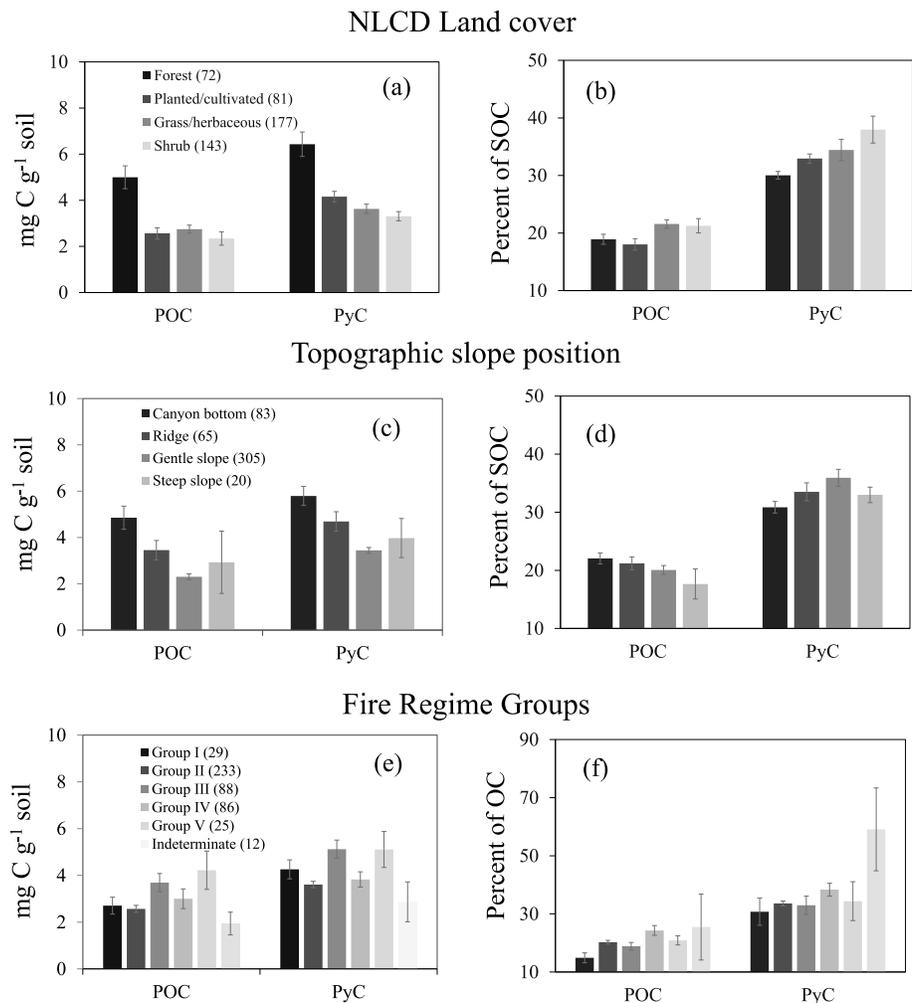


Figure 5. (a, c, and e) Mean and (b, d, f) percentage of total organic carbon contributed by particulate organic carbon (POC) and pyrogenic carbon (PyC) in relation to land cover, topographic positions, and fire regime groups.

PyC fraction was highly correlated with both the total SOC ($r^2 = 0.95$; Figure 4a) and with the SOC minus the PyC fraction ($r^2 = 0.91$; Figure 4b).

Maps of observed concentrations of PyC and POC show clear spatial distribution patterns throughout the study region (Figures 3c and 3d). High values (75th percentile) of PyC and POC were observed mostly in the high altitude forested area of the Rocky Mountains and in the Alluvial River and Osage Plains in the eastern part of Kansas. The mean PyC and POC were higher in forests than other land cover classes (Figure 5a). However, the percentages of SOC contributed by the POC and PyC fractions were higher in shrubland and herbaceous land (Figure 5b). In addition, the PyC and POC concentrations were lower in gentle slope position (slope $< 6^\circ$) than in other topographic positions (Figure 5c), and percentage of POC and PyC in the SOC were lower in steep slope position (Figure 5d). For fire regime groups, the concentration of both POC and PyC were greatest for FRG-III and FRG-V (Figure 5e). However, the percentage of PyC in the SOC was higher in areas with FRG-V. This FRG has > 200 year fire return interval with replacement severity and $> 75\%$ mean top kill within a typical fire perimeter for a given vegetation type (Figure 5f).

4.2. Factors Affecting Soil PyC and POC

We found that PyC and POC stocks were positively correlated with elevation, slope gradient (G), and slope length (SL) and negatively correlated with topographic position indexes (TPI) greater than $250\text{ m} \times 250\text{ m}$ (Table 1). For both fractions, the highest correlation coefficients (r) with TPI were obtained with the largest

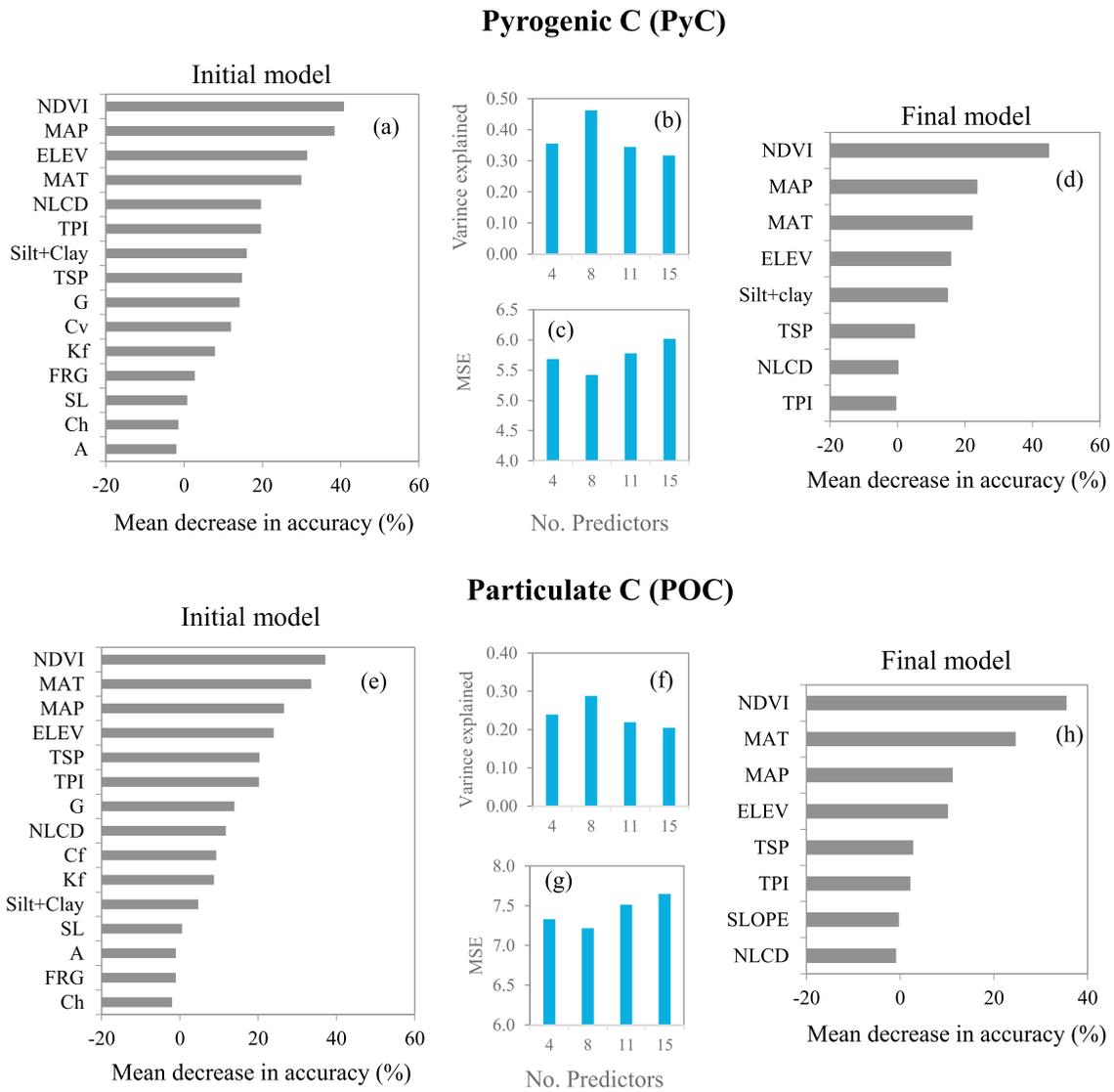


Figure 6. (a–d) Importance plots for pyrogenic (PyC) and (e–h) particulate C derived from initial (Figures 6a and 6e) and final (Figures 6d and 6h) random forest regression models. Bar plots in Figures 6b, 6c, 6f, and 6g show the percentage of the explained variance and mean square error (MSE) of RF models with different numbers of predictors. (ELEV = elevation; G = slope gradient; SL = slope length; A = aspect; TPI = topographic position index; TSP = topographic slope position; C_v = vertical curvature; C_h = horizontal curvature; MAT = mean annual temperature; MAP = mean annual precipitation; NDVI = normalized vegetation index; NLCD = national land cover data; FRG = fire regime group; and K_f = K factor).

distance analyzed (2000 m; TPI₂₀₀₀). PyC and POC contents were positively and strongly correlated with both NDVI and monthly mean precipitation (MAP), and negatively correlated with temperature (MAT). Both PyC and POC showed weak but significant positive correlation with the soil erodibility factor (K_f). Only PyC showed a weak but statistically significant positive relationship with finer soil texture (silt + clay).

The RF regression analysis using all variables listed in Table 1 explained 32 and 20% variability of PyC and POC, respectively. The NDVI, MAP, MAT, and elevation (ELEV) were identified as the four most important variables while FRG, SL, C_h, and A were the four least important variables for predicting PyC (Figure 6a) and POC (Figure 6e). When we applied RF model selection approach, we found that model with eight predictors had the lower mean square error (MSE) (Figures 6c and 6g) and explained more the variability both PyC and POC than the model with all predictors (Figures 6b and 6f). In the best random forests regression model, NDVI, MAP, MAT, ELEV, soil texture, TSP, NLCD, and TPI were the most important variables of PyC and the model explained 46% of the variation in PyC (Figure 6d). The most important

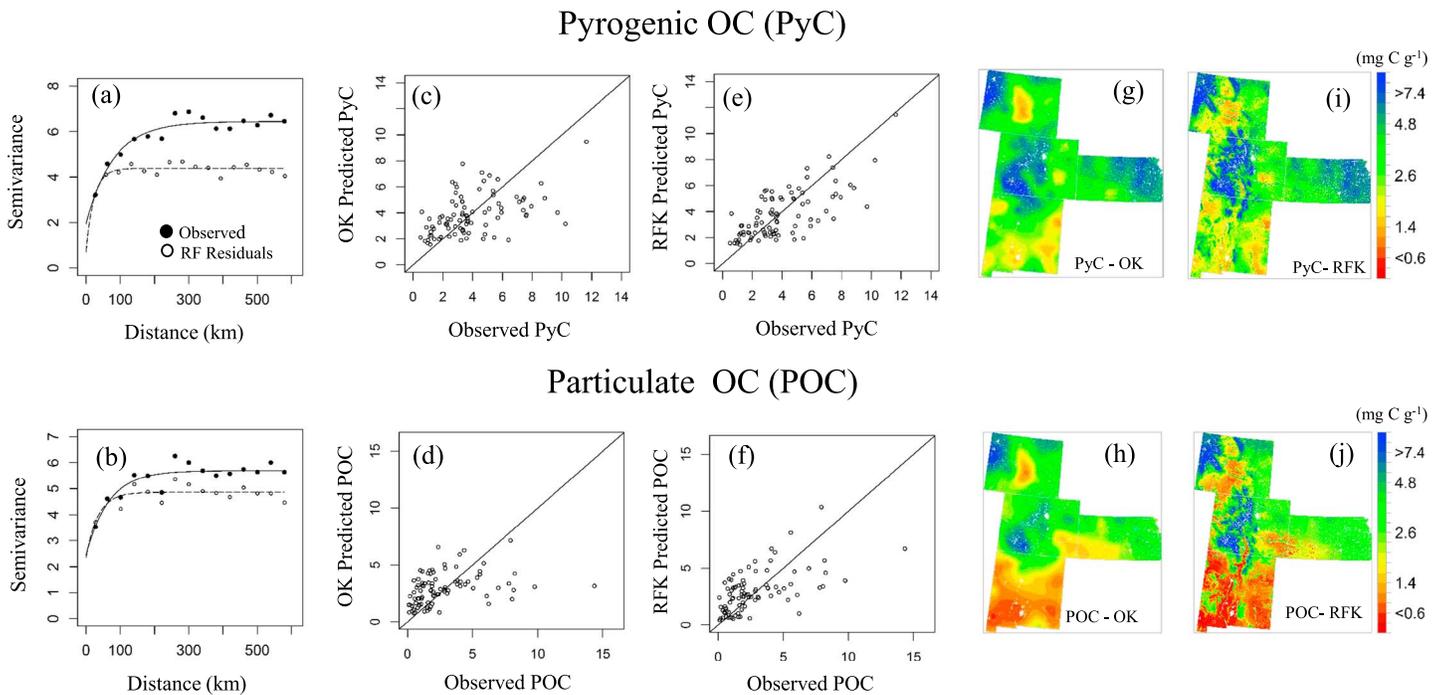


Figure 7. Variograms with fitted models, 1:1 plots of observed versus predicted values and ordinary (OK) and random forest kriging (RFK) predicted maps of (a, c, e, g, and i) pyrogenic carbon (PyC) and (b, d, f, h, and j) particulate organic carbon (POC).

variables for prediction POC were in rank order: NDVI, MAT, MAP, ELEV, G, TSP, NLCD, and TPI and explaining 28% of the variation (Figure 6h).

4.3. Geospatial Analysis and Mapping of PyC and POC

Figures 7a and 7b show the isotropic variograms with fitted model (solid lines) of PyC and POC (closed circles) and the residuals of their RF regression models with selected environmental covariables. The variograms of these fractions and their residuals of RF regression models showed a clear spatial dependence with bounded sills and were fitted well with exponential models. The variance of PyC and POC increases steadily with increasing lag distance and approaches its sill asymptotically within a range of 79.7 and 65.4 km and autocorrelation may be extended to the effective range of 239.1 and 196.2 km, respectively (Table 3). However, the residual variogram of PyC showed much shorter range (about 56.8 km shorter than original) than that of POC (30.5 km shorter range). These results indicate that environmental covariables better explained local variation of PyC than POC. The scatter plots of predicted vs observed values show that RFK predicted values (Figures 6e and 6f) are more closer to the 1:1 line than that of OK predicted values (Figures 6c and 6d) and RFK had lower MAE and RMSE values (Table 3). The relative improvement of RFK over OK was 20% for PyC and 7% for POC.

Maps produced by RFK (Figures 7i and 7j) at a resolution of 1 km × 1 km show a similar global pattern and hot spot locations of both PyC and POC as found using OK (Figures 7g and 7h). However, the RFK maps were less smooth than the OK maps and thus provide more detail regarding local variability in SOC concentration reflecting the effects of environmental covariables. In the high-altitude forested area of the Rocky Mountains concentrations of both PyC and POC were consistently higher than other locations within the regions studied. In large portions of the area in the Alluvial River and Osage Plains in eastern Kansas, the concentration of PyC was higher than POC. From the maps of percentage of SOC contributed by POC and PyC fractions (Figure 8), we calculated that PyC was the dominant fraction of SOC (≥50% contribution) for 8% of the study area. A large portion of this area was in shrubland with gentle slope under FRG-V. In contrast, POC dominated less than 1% of the study area. Because POC has a faster turnover time than other SOC fractions, this small area would likely be highly vulnerable to mineralization of SOC. Furthermore, from the map of the ratio of POC to PyC (Figure 9a), we estimated about 16% of the

Table 3. Properties of Fitted Variogram Models and Validation Results^a

Variables	Variogram Properties					Validation Results		
	Model	Co	C + Co	A	A _e	MAE	RMSE	RI
Organic Carbon						Ordinary Kriging		
PyC	Exp	1.97	6.44	79.7	239.1	1.64	2.10	
POC	Exp	2.44	5.68	65.4	196.2	1.72	2.24	
Residuals						Random Forest Kriging		
PyC	Exp	0.73	4.39	22.9	68.7	1.32	1.67	20.3
POC	Exp	2.36	4.85	34.9	104.7	1.55	2.09	7.0

^aPyC = pyrogenic OC; POC = particulate OC; Exp = exponential; Co = nugget; C + Co = sill; A = range (km); A_e = effective range (km), range × 3 for exponential model; MAE = mean absolute error; RMSE = root mean square error; and RI = relative improvement (%) of RMSE of random forest regression kriging over ordinary kriging.

study area is medium to highly vulnerable to SOC mineralization. Most of this area is under grassland and shrubland with gentle slopes of less than 6%.

5. Discussion

The mean percentage of PyC to SOC is much higher than found previously in central and Western North America, using a range of different methods for quantification. For example, PyC also quantified by NMR excluding the fraction >53 μm, for five long-term agricultural experiments in Minnesota, Illinois, Texas, and Oregon consisting of three Mollisols, Vertisol, and Alfisol, was on average 24% of SOC (10–35%) [Skjemstad *et al.*, 2002]. Across a climosequence of North American prairies, PyC determined by benzene-polycarboxylic acid (BPCA) contributed a mean of only 9% of SOC across 18 sites, with a very similar value of 95 mg g⁻¹ of SOC for the five samples that fell within our study region [Glaser and Amelung, 2003]. This same study found a mean of 2.41 g kg⁻¹ of PyC for these five samples, while we found a regional mean of 4.0 g kg⁻¹. Similarly, across five Ponderosa pine sites with different fire histories within each site in western Montana and northern Idaho, PyC quantified using chemical oxidation with H₂O₂/dilute HNO₃ comprised only 11% of SOC [Kurth *et al.*, 2006; DeLuca and Sala, 2006]. Across a soil catena covering both forest and agricultural lands in the Pacific Northwest of the USA, PyC determined by the same method as in our study ranged from 8 to 27%, with a mean of 16% of SOC [Jauss *et al.*, 2015]. In comparison, we found that throughout the study region PyC had a mean of 34.6% of SOC (Table 3). Reisser *et al.* [2016] recently estimated that global PyC (*n* = 560) represent on average 13.7% of the SOC, ranging from 0 to 60%. Because PyC on average has a longer turn-over time compared to other forms of SOC, such as POC, the high percentage of SOC found to be PyC throughout this large region of 1,051,029 km² has important implications for the vulnerability of SOC to mineralization with ongoing climate change and land management.

Many different methods have been, and are still, used to quantify PyC in soils [Bird, 2015], including the literature cited above. The NMR-MIR prediction method adopted in our study captures the widest range of PyC among other chemical or spectroscopic methods [Hammes *et al.*, 2007]; especially with the recent modification to include PyC fragments >53 μm (see section 2). This method is expected to generate higher estimates than other PyC quantification approaches cited here, possibly about double the estimates compared to BPCA or oxidative techniques [Hammes *et al.*, 2007; Bird, 2015]. These methods estimated by MIR have been demonstrated to be reliable when properly calibrated [Janik *et al.*, 2007; Baldock *et al.*, 2012], have been recommended recently [Reeves, 2012], and are feasible for such large numbers of samples and are therefore particularly valuable for regional and landscape studies [e.g., Lehmann *et al.*, 2008; Jauss *et al.*, 2015].

Of all SOC fractions, we found that only PyC showed a weak but statistically significant positive relationship with finer soil texture (silt + clay), and this variable ranked sixth as variable of the importance RF regression model. The only previously published data on PyC from this region did not find such a correlation [Glaser and Amelung, 2003]. However, that study sampled only 18 sites compared to 473 sites in the current study, and our large sample size likely allowed us to detect a significant but weak association. This result may indicate that long-term stabilization of PyC depends in part on associated with soil minerals, as shown by mineralization studies [Bruun *et al.*, 2014], but it might also suggest that other factors are more important in controlling PyC contents. The amount of aboveground biomass is the most important factor throughout

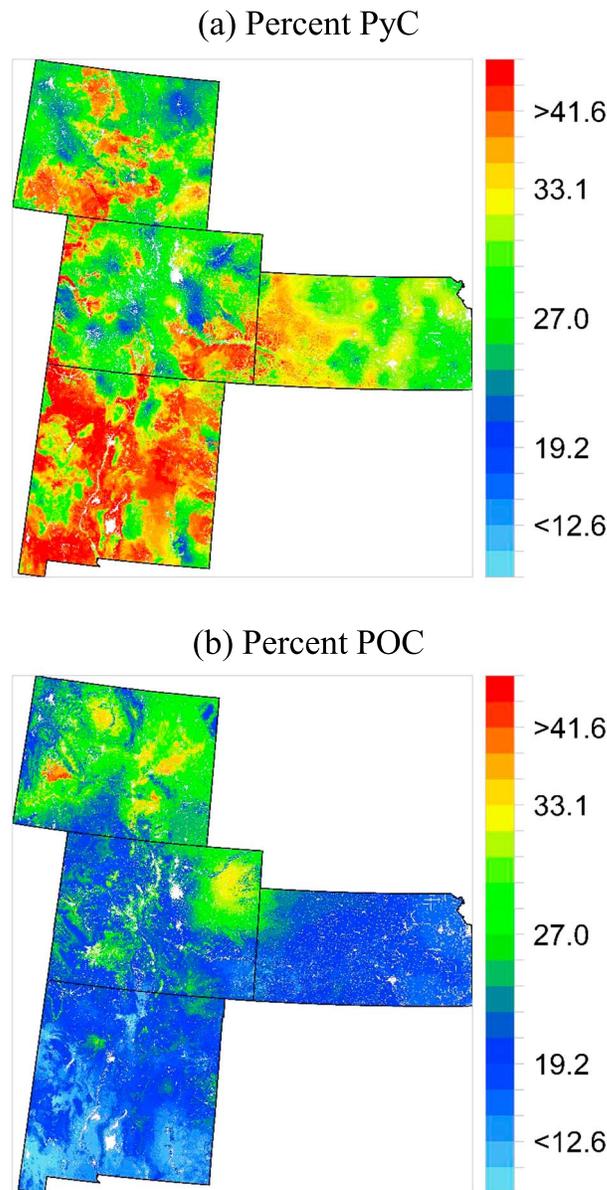


Figure 8. Map of the percentage of soil total organic carbon contributed by (a) pyrogenic carbon (PyC) and (b) particulate carbon (POC).

landscape position to accumulate at lower positions. This confirmed the results from the present studies that both movement and accumulation are important in determining landscape-scale distribution of PyC.

The most important predictor of both PyC and POC was NDVI. The strong influence of NDVI can also be seen in the maps developed with RFK (Figures 7i and 7j): the spatial pattern is most similar to that of NDVI (Figure 2e). The strong relationships between both NDVI and forest cover and PyC and other SOC fractions likely indicates that organic carbon (OC) inputs to the soil rather than loss due to erosion and mineralization are the dominant factor explaining variation of all SOC fractions. Furthermore, the finding of the same ranking for both PyC and other OOC indicates that all OC fractions respond similarly to variation in vegetation, climate, and landscape position. *Jauss et al.* [2015] found a moderate relationship ($r^2 = 0.44$), between PyC and POC + OOC for topsoils of 11 sites in Oregon, indicating a weak but relevant landscape-scale connection between the fractions using the same PyC methodology as used in our study. In our study, we found a much stronger relationship between PyC and non-PyC SOC ($r^2 = 0.91$) in the topsoil (Figure 4b). In the Great Plains, *Glaser and Amelung* [2003] found a strong correlation ($r^2 = 0.89$) between PyC and SOC, which we found to be

this large region as indicated by the very strong association of both PyC and OOC with NDVI based on the results of (1) correlation analysis, (2) the RF regression model, and (3) the association with forest vegetation which has high aboveground biomass.

In a previously published model of the fate of PyC, erosion (runoff) was identified by difference as the dominant flux of PyC, accounting for 76% of the loss over 100 years [*Foereid et al.*, 2011]. We found that although soil erodibility (as indicated by the k factor) was significantly associated with PyC concentration, it was not an important predictor of PyC or other SOC fractions. However, TPI and TSP were relatively important. These results suggest that landscape position may be more important than soil properties in predicting the impact of erosion and deposition on SOC fractions, at least for ecosystems dominated by perennial vegetation. Furthermore, the finding that TPI at the largest distance tested (2000 m) was a more significant predictor than TPI at shorter distances indicates that landscape processes are important at scales of multiple kilometers across the wide range of topography, climate, and ecosystems present in the study region. However, it is also likely that more intensive local sampling could provide additional information about local-scale factors that affect SOC fractions such as PyC [*Kane et al.*, 2007]. *Guerena* [2015] found significant changes in PyC distributions in soils of watersheds with a size of 10 ha over decadal timescales after fire clearing, changing from greater PyC stocks at higher land-

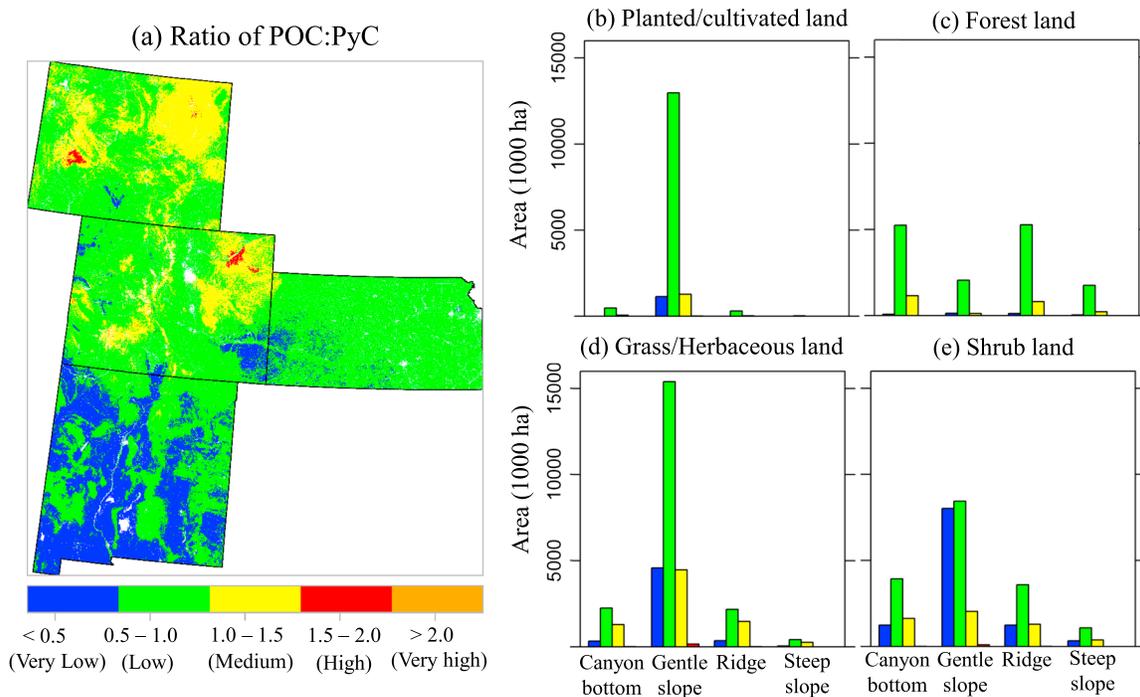


Figure 9. (a) Ratio of particulate organic carbon (POC) to pyrogenic carbon (PyC) showing the vulnerability of soil C degradation and calculated area under five vulnerability classes in relation to topographic position and (b) planted/cultivated, (c) forest, (d) grass/herbaceous and (e) shrub land.

even stronger for our data set ($r^2 = 0.95$). These results support the finding that PyC responds similarly to variation in vegetation, climate, and landscape position as does the non-PyC portion of SOC.

The very weak association of PyC with FRG may reflect the complexity of fire effects on PyC, acting both as the source of PyC but potentially also as a sink if PyC is combusted in a subsequent fire [Ohlson and Tryterud, 2000; Czimczik et al., 2005; DeLuca and Aplet, 2008]. However, it is not clear whether fire always leads to large reburning of PyC [Santin et al., 2013], even if it is placed on the soil surface [Saiz et al., 2014]. Also, the effects of fire can be extremely variable over short distances, so the relatively coarse-scale geospatial data on fire regime groups may not reflect local conditions where the soil samples were taken. However, even with collocated sampling of fire interval and PyC, previous investigations in Montana [Kurth et al., 2006] and Sweden [Zackrisson et al., 1996] also did not find significant associations between fire return interval and PyC in forested ecosystems. Fire characteristics could only explain PyC content, if site-specific information was available [Reisser et al., 2016]. Taken together, our results and those from the literature suggest that FRG may not be an important predictor of PyC contents in surface soils. However, it should be noted that we addressed only the lateral movement of PyC in this study. The vertical movement of PyC therefore needs to be addressed for better understanding the effect of FRG on PyC.

The relatively high percentage of persistent PyC along with the relatively low percentage found as easily mineralizable POC has important implications for the vulnerability of SOC to mineralization under ongoing climate change and land management practices. The PyC has been found to be virtually unchanged over several decades of cultivation [Skjemstad et al., 2004]. However, there was substantial variation in the percentage of PyC and POC fractions in SOC and therefore in the vulnerability index derived from them. Only a tiny fraction of the entire region (<1%) was found to be highly vulnerable to SOC mineralization with POC as the dominant fraction of SOC (>50%). A larger area of 16% of the region had medium to high vulnerability, concentrated in northeast Wyoming, southwest Wyoming, and northeast Colorado. The southernmost portion of the region, including most of New Mexico, was found to have a comparatively very low vulnerability to SOC mineralization (Figure 9). Cultivated and forest lands had mostly low vulnerability with some very low vulnerability to SOC mineralization, whereas grasslands and shrublands had nearly not only all of the medium vulnerability area but also substantial areas of very low vulnerability (Figures 8b–8f). These results suggest that

the most highly managed lands generally have low vulnerability to further SOC mineralization, while uncultivated grasslands and shrublands have much more variability in vulnerability, ranging from medium to very low. The absolute values of such vulnerability indices greatly depend on the methods used to isolate PyC and POC and have to be evaluated in this context, whereas the changes between locations likely allow more robust interpretation.

Medium- and high-vulnerability locations were concentrated in just three contiguous portions of the region (Figure 9a). These regions are different from the spatial patterns of each of the environmental variables (Figures 1 and 2). This result demonstrates that combining environmental data with analysis of SOC fractions using geostatistical analysis can improve understanding of the vulnerability of SOC to loss due to ongoing climatic changes as well as opportunities to increase SOC sequestration with spatially targeted management practices.

6. Conclusions

We found that the mean PyC fraction in A horizon comprised approximately one third of the SOC on average but ranged widely from 11.8 to 96.6%. Much variation in the recalcitrant pyrogenic soil carbon fraction could be predicted based on vegetation, climate, and terrain variables. This result is important because PyC is often refractory and thus less susceptible to mineralization compared to other soil carbon fractions. Using the ratio of POC to PyC as an index of the vulnerability of soil carbon to mineralization, we found that 16% of the study area throughout four Western USA states is moderately to highly vulnerable to soil organic carbon mineralization.

Further application of methods such as ours will improve the understanding of spatial patterns of SOC and its fractions. Improved maps of SOC fractions, including PyC are needed for improved parameterization of OC cycling models [Falloon *et al.*, 2000; Lehmann *et al.*, 2008; Tonitto *et al.*, 2016]. The combination of spatial analyses with OC cycling models can improve understanding of the vulnerability of SOC to mineralization due to ongoing climatic changes as well as opportunities to increase SOC sequestration with spatially targeted management practices.

Acknowledgments

Funding through NASA-USDA award 2008-35615-18961 and NIFA-USDA award 2014-67003-22069 are gratefully acknowledged. Many thanks to L.R. Spouncer for help with MIR analyses and data processing and David Smith for providing the samples. Both data and R-code are available from the authors upon request (z.ahmed@cgiar.org).

References

- Baldock, J., B. Hawke, J. Sanderman, L. Macdonald, E. Schmidt, S. Szarvas, A. Puccini, and J. McGowan (2012), Soil carbon research program: Project 2 Developing a cost effective soil carbon analytical capability, A collaborative project supported by the Climate Change Reduction Program of the Australian Department of Agriculture, Fisheries and Forestry and the Grains Research and Development Corporation.
- Baldock, J. A., J. Sanderman, L. M. Macdonald, A. Massis, B. Hawke, S. Szarvas, and J. McGowan (2013a), Quantifying the allocation of soil organic carbon to biologically significant fractions, *Soil Res.*, *51*, 561–576.
- Baldock, J. A., B. Hawke, J. Sanderman, and L. M. Macdonald (2013b), Predicting contents of soil carbon and its component fractions from diffuse reflectance mid-infrared spectra, *Soil Res.*, *51*, 577–595.
- Bird, M. (2015), Test procedures for biochar analysis in soils, in *Biochar for Environmental Management: Science, Technology, and Implementation*, edited by J. Lehmann and S. Joseph, pp. 679–716, Earthscan, London.
- Bruun, S., S. Clauson-Kaas, L. Bobulska, and I. K. Thomsen (2014), Carbon dioxide emissions from biochar in soil: Role of clay, microorganisms and carbonates, *Eur. J. Soil Sci.*, *65*, 52–59.
- Certini, G. (2014), Fire as a soil-forming factor, *Ambio*, *43*, 191–195.
- Cressie, N., and D. M. Hawkins (1980), Robust estimation of the variogram: 1, *J. Int. Assoc. Math. Geol.*, *12*, 115–125.
- Czimczik, C. I., M. W. I. Schmidt, and E. D. Schulze (2005), Effects of increasing fire frequency on black carbon and organic matter in Podzols of Siberian Scots pine forests, *Eur. J. Soil Sci.*, *56*, 417–428.
- Daly, C., G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson, and P. Pasteris (2001), High quality spatial climate data sets for the United States and beyond, *Trans. ASAE*, *43*, 1957–1962.
- DeLuca, T. H., and G. H. Aplet (2008), Charcoal and carbon storage in forest soils of the Rocky Mountain West, *Front. Ecol. Environ.*, *6*, 18–24.
- DeLuca, T. H., and A. Sala (2006), Frequent fire alters nitrogen transformations in ponderosa pine stands of the inland northwest, *Ecology*, *87*, 2511–2522.
- Dickson, B. G., and P. Beier (2007), Quantifying the influence of topographic position on cougar (*Puma concolor*) movement in southern California, USA, *J. Zoology*, *271*, 270–277.
- Environmental Systems Research Institute (ESRI) (2011), *ArcGIS Desktop: Release 10*, Environ. Syst. Res. Inst., Redlands, Calif.
- Evans, J. S., and S. A. Cushman (2009), Gradient modeling of conifer species using random forest, *Landscape Ecol.*, *24*, 673–683.
- Evans, J. S., and M. A. Murphy (2016), rfUtilities. R package version 2.0-0. [Available at <http://CRAN.R-project.org/package=rfUtilities>.]
- Falloon, P., P. Smith, K. Coleman, and S. Marshall (2000), How important is inert organic matter for predictive soil carbon modelling using the Rothamsted carbon model? *Soil Biol. Biochem.*, *32*, 433–436.
- Foeroid, B., J. Lehmann, and J. Major (2011), Modeling black carbon degradation and movement in soil, *Plant Soil*, *345*, 223–236.
- Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold, and J. Wickham (2011), Completion of the 2006 national land cover database for the conterminous United States, *Photogramm. Eng. Remote Sens.*, *77*, 858–864.
- Glaser, B., and W. Amelung (2003), Pyrogenic carbon in native grassland soils along a climosequence in North America, *Global Biogeochem. Cycles*, *17*(2), 1064, doi:10.1029/2002GB002019.

- Global Climate Change Impacts in the United States (2009), *Global Climate Change Impacts in the United States*, edited by T. R. Karl, J. M. Melillo, and T. C. Peterson, Cambridge Univ., U.S. Global Change Res. Program, New York.
- Guerena, D. T. (2015), Anthropogenic and natural pyrogenic carbon in tropical environments: From rhizosphere to landscape, PhD Dissertation, Cornell Univ.
- Haaland, D. M., and V. T. Thomas (1988), Partial least-squares methods for spectral analyses: 1. Relation to other quantitative calibration methods and the extraction of qualitative information, *Anal. Chem.*, *60*, 1193–1202.
- Hammes, K., et al. (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, *Global Biogeochem. Cycles*, *21*, GB3016, doi:10.1029/2006GB002914.
- Heidema, A. G., J. M. A. Boer, N. Nagelkerke, E. C. M. Mariman, D. L. van der A, and E. J. M. Feskens (2006), The challenge for genetic epidemiologists: How to analyze large numbers of SNPs in relation to complex diseases, *BMC Genet.*, *7*, 23, doi:10.1186/1471-2156-7-23.
- Hengl T., et al. (2015), Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions, *PLoS One*, *10*(6), e0125814, doi:10.1371/journal.pone.0125814.
- Hijmans, R. J., and J. van Etten (2016), Geographic analysis and modeling with raster data. R package version 2.0-12. [Available at <http://CRAN.R-project.org/package=raster>.]
- Janik, L. J., J. O. Skjemstad, K. D. Shepherd, and L. R. Spouncer (2007), The prediction of soil carbon fractions using mid-infrared-partial least square analysis, *Austr. J. Soil Res.*, *45*, 73–81.
- Jauss, V., M. Johnson, E. Krull, M. Daub, and J. Lehmann (2015), Pyrogenic carbon controls across a soil catena in the Pacific Northwest, *Catena*, *12*, 53–59.
- Jenkinson, D. S., and J. H. Rayner (1977), The turnover of soil organic matter in some of the Rothamsted classical experiments, *Soil Sci.*, *123*, 298–305.
- Jenness, J. (2013), *DEM Surface Tools for ArcGIS 10*, Jenness Enterprises, Flagstaff, Ariz.
- Kane, E. S., E. S. Kasischke, D. W. Valentine, M. R. Turetsky, and A. D. McGuire (2007), Topographic influences on wildfire consumption of soil organic carbon in interior Alaska: Implications for black carbon accumulation, *J. Geophys. Res.*, *112*, G03017, doi:10.1029/2007JG000458.
- Kuhn, M., and K. Johnson (2013), *Applied Predictive Modeling*, Springer, London.
- Kurth, V. J., M. D. MacKenzie, and T. H. DeLuca (2006), Estimating charcoal content in forest mineral soils, *Geoderma*, *137*, 135–139.
- Lehmann, J., J. Skjemstad, S. Sohi, J. Carter, M. Barson, P. Falloon, K. Coleman, P. Woodbury, and E. Krull (2008), Australian climate-carbon cycle feedback reduced by soil black carbon, *Nat. Geosci.*, *1*, 832–835.
- Liaw, A., and M. Wiener (2002), Classification and regression by randomForest. R News 2(3), pp. 18–22. [Available at <http://CRAN.R-project.org/doc/Rnews>.] Verified: August 18.
- Minasny, B., A. B. McBratney, B. P. Malone, and I. Wheeler (2013), Digital mapping of soil carbon, *Adv. Agron.*, *118*, 1–47.
- Multi-Resolution Land Characteristics Consortium (2007), *2001 National Land Cover Data (NLCD 2006)*, USEPA, Multi-Resolution Land Characteristics Consortium, Washington, D. C.
- Natural Resources Conservation Service (NRCS), United States Department of Agriculture (2012), General soil map (STATSGO). [Available at <http://soildatamart.nrcs.usda.gov>, Accessed 08/12/2012.]
- Ohlson, M., and E. Tryterud (2000), Interpretation of the charcoal record in forest soils: Forest fires and their production and deposition of macroscopic charcoal, *Holocene*, *10*, 519–525.
- Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima (1987), Analysis of factors controlling soil organic-matter levels in Great-Plains grasslands, *Soil Sci. Soc. Am. J.*, *51*, 1173–1179.
- R Core Team (2016), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available at <http://www.R-project.org>.]
- Reeves, J. B. (2012), Mid-infrared spectroscopy of biochars and spectral similarities to coal and kerogens: What are the implications? *Appl. Spectrosc.*, *66*, 689–695.
- Reisser, M., R. S. Purves, M. W. I. Schmidt, and S. Abiven (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.
- Rodionov, A., W. Amelung, N. Peinemann, L. Haumaier, X. D. Zhang, M. Kleber, B. Glaser, I. Urusevskaya, and W. Zech (2010), Black carbon in grassland ecosystems of the world, *Global Biogeochem. Cycles*, *24*, GB3013, doi:10.1029/2009GB003669.
- Rumpel, C., V. Chaplot, O. Planchon, J. Bernadou, C. Valentin, and A. Mariotti (2006), Preferential erosion of black carbon on steep slopes with slash and burn agriculture, *Catena*, *65*, 30–40.
- Saiz, G., I. Goodrick, C. M. Wurster, M. Zimmermann, P. N. Nelson, and M. I. Bird (2014), Charcoal re-combustion efficiency in tropical savannas, *Geoderma*, *219*(220), 40–45.
- Santin, C., S. H. Doerr, C. Preston, and R. Bryant (2013), Consumption of residual pyrogenic carbon by wildfire, *Int. J. Wildland Fire*, *22*, 1072–1077.
- Scharlemann, J. P. W., E. V. J. Tanner, R. Hiederer, and V. Kapos (2014), Global soil carbon: Understanding and managing the largest terrestrial carbon pool, *Carbon Manage.*, *5*, 81–91.
- Schmidt, M. W. I., M. S. Torn, and S. Abiven (2011), Persistence of soil organic matter as an ecosystem property, *Nature*, *478*, 49–56.
- Sinnaeve, G., J. L. Herman, V. Baeten, P. Dardenne, M. Frankinet (2001), Performances of an on board diode array NIR instrument for the analysis of fresh grass, Journée Thématique AFMEX. Appareils Embarqués de Mesure de la Biomasse, Rennes, France.
- Skjemstad, J. O. D., C. Reicosky, A. R. Wilts, and J. A. McGowan (2002), Charcoal carbon in US agricultural soils, *Soil Sci. Soc. Am. J.*, *66*, 1249–1255.
- Skjemstad, J. O., L. R. Spouncer, B. Cowie, and R. S. Swift (2004), Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools, *Austr. J. Soil Res.*, *42*, 79–88.
- Smith, D. B., W. F. Cannon, and L. G. Woodruff (2011), A national-scale geochemical and mineralogical survey of soils of the conterminous United States, *Appl. Geochem.*, *26*, S250–S255.
- Smith, D. B., W. F. Cannon, L. G. Woodruff, F. Solano, K. J. Ellefsen (2014), Geochemical and mineralogical maps for soils of the conterminous United States, U.S. Geological Survey Open-File Report 2014-1082 (386 pp). [Available at <http://pubs.usgs.gov/of/2014/1082/>.]
- Soucémarianadin, L. N., S. A. Quideau, and M. D. MacKenzie (2014), Pyrogenic carbon stocks and storage mechanisms in podzolic soils of fire-affected Quebec black spruce forests, *Geoderma*, *217–218*, 118–128.
- Tonitto, C., N. Gurwick, and P. B. Woodbury (2016), Quantifying greenhouse gas emissions from agricultural and forest landscapes for policy development and verification, in *Synthesis and Modeling of Greenhouse Gas Fluxes and Carbon Changes in Agricultural and Forest Systems to Guide Mitigation and Adaptation*, edited by S. L. Del Grosso, L. R. Ahuja, and W. J. Parton, Am. Soc. of Agron., Crop Sci. Soc. of Am., Soil Sci. Soc. of Am., Madison, Wis.
- Trumbore, S. (2009), Radiocarbon and soil carbon dynamics, *Ann. Rev. Earth Planet. Sci.*, *37*, 47–66.
- Zackrisson, O., M. C. Nilsson, and D. A. Wardle (1996), Key ecological function of charcoal from wildfire in the boreal forest, *Oikos*, *77*, 10–19.