Science of the Total Environment 654 (2019) 94–106

Contents lists available at ScienceDirect

Science of the Total Environment

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

Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States

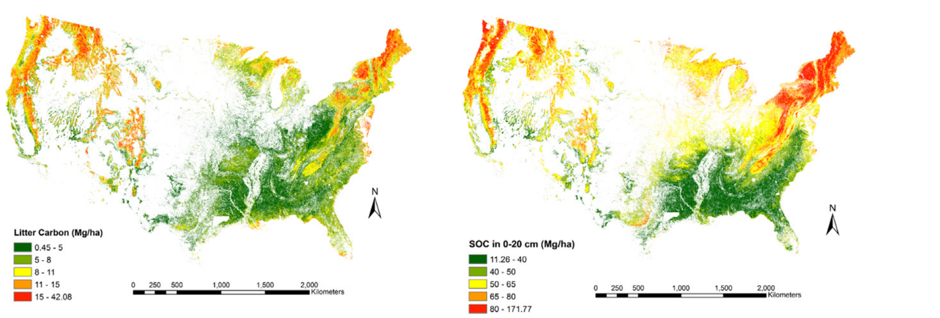
Baijing Cao [a], Grant M. Domke [a,b,*], Matthew B. Russell [a], Brian F. Walters [b]

[a] *Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA*
[b] *USDA Forest Service, Northern Research Station, St. Paul, MN 55108, USA*

HIGHLIGHTS

- Spatial patterns found in the estimated litter and soil carbon stocks in forests
- Including Normalized Difference Vegetation Index facilitated the model predictions.
- Forest disturbances caused statistically significant differences in litter carbon.
- Estimates of litter and soil carbon stocks were 2.07 Pg and 14.68 Pg, respectively.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 August 2018
Received in revised form 25 October 2018
Accepted 27 October 2018
Available online 29 October 2018

Editor: Elena PAOLETTI

Keywords:
National Forest Inventory
Machine learning
NDVI
Google Earth Engine

ABSTRACT

Forest ecosystems contribute substantially to carbon (C) storage. The dynamics of litter decomposition, translocation and stabilization into soil layers are essential processes in the functioning of forest ecosystems, as these processes control the cycling of soil organic matter and the accumulation and release of C to the atmosphere. Therefore, the spatial distribution of litter and soil C stocks are important in greenhouse gas estimation and reporting and inform land management decisions, policy, and climate change mitigation strategies. Here we explored the effects of spatial aggregation of climatic, biotic, topographic and soil variables on national estimates of litter and soil C stocks and characterized the spatial distribution of litter and soil C stocks in the conterminous United States (CONUS). Litter and soil variables were measured on permanent sample plots ($n = 3303$) from the National Forest Inventory (NFI) within the United States from 2000 to 2011. These data were used with vegetation phenology data estimated from LANDSAT imagery (30 m) and raster data describing environmental variables for the entire CONUS to predict litter and soil C stocks. The total estimated litter C stock was 2.07 ± 0.97 Pg with an average density of 10.45 ± 2.38 Mg ha^{-1}, and the soil C stock at 0–20 cm depth was 14.68 ± 3.50 Pg with an average density of 62.68 ± 8.98 Mg ha^{-1}. This study extends NFI data from points to pixels providing spatially explicit and continuous predictions of litter and soil C stocks on forest land in the CONUS. The approaches described illustrate the utility of harmonizing field measurements with remotely sensed data to facilitate modeling and prediction across spatial scales in support of inventory, monitoring, and reporting activities, particularly in countries with ready access to remotely sensed data but with limited observations of litter and soil variables.

1. Introduction

Forests cover about 42 million km^2 of the world's land surface (Bonan, 2008), and 50 to 90% of the total annual carbon (C) flux of

* Corresponding author at: USDA Forest Service, Northern Research Station, St. Paul, MN 55108, USA.
E-mail address: gmdomke@fs.fed.us (G.M. Domke).

https://doi.org/10.1016/j.scitotenv.2018.10.359
0048-9697/© 2018 Published by Elsevier B.V.

terrestrial ecosystems occurs at the interface between forests and the atmosphere (Beer et al., 2010). In the United States (US), litter C stocks account for between 5 and 7% (Domke et al., 2016; US Environmental Protection Agency (US EPA), 2018) of total forest ecosystem C stocks which is consistent with the contribution of the litter C pool globally (Pan et al., 2011). The interaction of C from forest litter to soil organic matter (SOM) is key to understanding how soil C stocks and microbial decomposition will respond to climate change and whether soil C sinks can be enhanced.

Litter decomposition alters the organic matter chemistry over time with different litter species and tissue types degrading at different rates (Lemma et al., 2007; Moore et al., 2011; Prescott, 2005; Prescott et al., 2004). That said, it is not clear how forest litter quantity and source (i.e., roots versus aboveground litter) are linked to C retention in soils. Thus, the soil C sequestration potential through management remains speculative (Dungait et al., 2012). To detect the temporal change of forest litter and soil C, we must understand current C storage in the soil, from the incorporation and transformation of litter inputs to its ultimate mineralization. Recently, Nave et al. (2018) indicated that reforestation increases topsoil C storage, and that reforesting lands may add as much as 1.3–2.1 Pg C within a century in the continental US (CONUS).

Soil C, including mineral and organic soils, hereafter referred to as SOC is the largest stock in the global terrestrial C pool (Amundson, 2001) and accounts for approximately 56% of the total forest ecosystem C on the managed land in the CONUS and southeast and southcentral coastal AK (Domke et al., 2017). Therefore, even small changes in SOC content can have tremendous impact on national and global C stocks and stock changes.

Digital soil mapping applies geostatistical tools to determine the quantitative relationships between soil properties and environmental variables (McBratney et al., 2003). Considering the importance of SOC in the global C cycle and its potential contribution in national C budgets (US Environmental Protection Agency (US EPA), 2018), many studies have quantified SOC stocks or its changes at national scales: in the conterminous United States (CONUS) (Guo et al., 2006a); India (Bhattacharyya et al., 2000); France (Martin et al., 2010); Australia (Bui et al., 2009), Laos (Phachomphon et al., 2010); China (Li et al., 2007); and Ireland (Zhang et al., 2011). These studies have used a range of digital soil mapping techniques by which point estimates of SOC are extended to national or continental scales.

Soil and ecological processes and spatial patterns are related across different scales, and their spatial and temporal distributions are mutually influenced as well. There are many studies that have shown relationships between environmental factors (e.g., climate and topographic factors) and litter and soil properties at continental (Domke et al., 2017, 2016), regional (McKenzie and Ryan, 1999; Xiong et al., 2014) and local scale (Odeh et al., 1995). Some studies have been successful in including secondary information such as land use, soil type, lithology, topography and other environmental factors in predicting SOC at regional scales (Schulp and Verburg, 2009; Simbahan et al., 2006).

Large-scale datasets of soil information, based on documented procedures and standards, are necessary for SOC assessments at continental scales (Batjes, 2009). Harmonized legacy soil data that arise from traditional soil surveys can also facilitate new digital soil mapping activity (Hartemink et al., 2008). The present analysis is the first known study to use digital mapping techniques to predict litter and SOC stocks across the CONUS using litter and soil data sampled at the same location and harmonized with auxiliary environmental data for prediction.

Machine learning methods offer opportunities to predict and characterize the spatial patterns of litter and SOC and the relationships between C stocks and environmental covariates allowing insights into pedogenic processes (Domke et al., 2017, 2016). This approach has been implemented in soil-landscape modeling (Bui et al., 2009; Hengl et al., 2017; Vasques et al., 2012; Xiong et al., 2014), with the advantage that it allows for both the prediction of litter and SOC and new insights into biogeochemical processes.

To better understand the distribution of litter and SOC patterns at a continental scale, it is necessary to identify the underlying ecological processes responsible for these patterns. The objectives of this study were to 1) determine and predict the spatial distribution of litter and SOC stocks using environmental data available across the CONUS; 2) investigate the relationships between soil order, climate, vegetation, topography, parent material, and litter and SOC stocks; 3) predict the distribution of litter and SOC stocks using environmental variables and assess the uncertainties associated with the litter and SOC predictions; 4) evaluate the impact of forest disturbance on litter and SOC stocks, and 5) analyze spatial patterns of the litter and SOC ratio. Such quantification of litter and SOC stocks will contribute to national and global efforts to estimate C storage in forest ecosystems and guide long-term management activities. Our study is the first to investigate both litter and soil C stocks at the same site in the CONUS.

2. Materials and methods

2.1. Study area

The study area in this analysis represents forest land in the CONUS, which includes 48 of the US states on the continent of North America. The CONUS occupies a combined area of over 8 million km^2, which represents 1.58% of the total surface area of the Earth (including water and land), with latitudes ranging from 24°31′ N to 49°23′ N and longitude from 66°53′ W to 124°50′ W. The climate of the CONUS is highly diverse and variable. Annual average temperature (1981–2010) ranges from <4 °C in the higher mountain areas and along the northern border with Canada to >21 °C in the desert southwest, south Texas, and south Florida (Kunkel et al., 2013). Average annual precipitation also varies from <381 mm annually in much of the western US to >1016 mm in much of the eastern US (Kunkel et al., 2013). Ten of the twelve soil orders can be found in forest soils in the CONUS: Mollisols, Entisols, Alfisols, Inceptisols, Ultisols, Aridisols, Vertisols, Spodosols, Histosols, Andisols.

Forest land accounts for 3.10 million km^2, or 38% of the CONUS land area (Oswalt et al., 2014). The CONUS has a varied topography, with elevations ranging from around 86 m below sea level (Death Valley, California) to 4421 m above sea level (Mt. Whitney, California).

2.2. Litter and soil organic carbon data

Litter and soil data were obtained from National Forest Inventory (NFI) plots maintained and measured by the Forest Inventory and Analysis (FIA) Program within the US Department of Agriculture, Forest Service. NFI plots are systematically distributed approximately every 2428 ha across the CONUS. Each plot which contains at least one forest land condition (determined by reserved status, owner group, forest type, stand size class, regeneration status or tree density) is comprised of a series of smaller plots (i.e., subplots) where tree- and site-level attributes – such as diameter at breast height (DBH) and tree height – are measured at regular temporal intervals (USDA Forest Service, 2011). Litter and soil samples are collected along with other non-standing tree ecosystem attributes (e.g., downed dead wood) on every 16th base intensity NFI plot distributed approximately every 38,848 ha (USDA Forest Service, 2011). Litter material is sampled adjacent to three of four subplots at each plot and the entire litter thickness (i.e., duff and LFH horizons) is measured to the nearest 0.25 cm at points in each cardinal direction within the sampling frame, which is a circle of 30.28 cm in diameter, to the point where mineral soil begins. The entire litter layer within the confines of the sampling frame is removed for lab analysis (Domke et al., 2016). In this study, litter C stock does not include any fine woody debris between 0.635 and 7.366 cm diameter. Soil attributes are measured by collecting cores from two depths: 0 to 10 cm and 10 to 20 cm beneath the litter sample adjacent to subplot 2 (i.e., one sample per FIA plot; Domke et al., 2017). The texture of each soil layer is estimated in the field, while physical and chemical properties are determined in the laboratory. The FIA program analyzes soils for total C concentration by dry combustion (O'Neill et al., 2005). The laboratory results of both litter and soil samples are managed as part of the Soils Lab Table (SOILS_LAB) in the publicly-available FIA DataMart (USDA Forest Service, 2014).

In this study, there were 3303 coincident litter and soil samples (0–20 cm) collected on NFI plots from 2000 to 2011. Based on the protocols for the data collected in this study, we characterized the SOC in the topsoil (0–20 cm) layer. Litter and SOC estimates used in this study were compiled following methods in Domke et al. (2016, 2017).

2.3. Environmental variables

We assembled our environmental variables based on the SCORPAN conceptual model (McBratney et al., 2003), which allows incorporation of existing soil forming information as covariates via the soil factors. Forty categorical and continuous environmental variables represented each soil forming factor in the SCORPAN model. They were compiled from various data sources and resolutions with ArcGIS 10.5 (Environmental Systems Research Institute, ESRI Inc., Redlands, CA) (Table 1). Soil (S) factors included soil order, soil moisture and hydrologic soil groups (Ross et al., 2018) variables; climate (C) factors included mean and maximum precipitation and temperature; organism (O) factors were represented by NDVI variables, LANDFIRE vegetation variables, and Major Land Resources Areas (MLRA); relief (R) factors contained elevation, slope and aspect; parent material (P) was presented as Gamma ray variables and surficial geology; and latitude and longitude were used as the coordinates (n). The aboveground biomass (AGB) was obtained from the NFI database and methods and equations for its estimation were produced by the USDA Forest Service (Woodall et al., 2011). For high-resolution remote sensing data, we used Google Earth Engine to extract Landsat 7 monthly NDVI data (30 m) in the growing season (June to September) from 2000 to 2011. The maximum of NDVI values from the growing season was selected as a predictor of litter C and SOC stocks.

Forest disturbances are events that cause change in the structure and composition of a forest ecosystem, beyond the growth and death of individual organisms. We also used two sources of forest disturbance data to evaluate the impact of disturbances on forest litter C and SOC stocks: 1) NFI data, including disturbance types (e.g., animal, human, fire, weather, insect, disease) and 2) LANDFIRE (2010) forest disturbance product for 2000 to 2010 compiled using remotely sensed information. For NFI data, a disturbance is defined as one that occurred since the last measurement or within the last 5 years for new plots. The area affected by the disturbance must be at least 0.40 ha in size. A significant level of disturbance (mortality or damage to 25% of the trees in the condition) is required (USDA Forest Service, 2014). For both datasets, only disturbances which occurred prior to litter and soil sampling on the NFI plots were included. In our nonparametric analyses, we used the Kruskal-Wallis test with Dunn's multiple comparison to indicate whether litter and SOC stocks differed significantly on disturbed and non-disturbed plots.

2.4. Modeling approaches

Three supervised machine learning methods, random forest (RF), quantile regression forest (QRF) and k-nearest neighbor (kNN) were chosen to model the distribution of litter C and SOC stocks. They are

Table 1
Covariates used to model the relationships with 3303 litter and soil carbon samples across the CONUS. Variables in bold were predictors in the partial models.

SCORPAN model factor	Variable[a]	Variable abbreviation	Data type[b]	Source[c]	Original resolution	Year
Soil	**Available water**	**AvaiWater**	**Con.**	**PRISM**	**800 m**	**1981–2010**
	Soil order	**order**	**Cat.**	**STATSGO**	**1:250,000**	**1997**
	Soil Moisture Annual	SM2010	Con.	ESA	1 km	2010
	Soil Moisture July	SM201007	Con.	ESA	1 km	2010
	Hydrologic soil groups	HYDROGRU	Cat.	USGS	250 m	2018
Climate	**Mean annual precipitation**	**ppt**	**Con.**	**PRISM**	**800 m**	**1981–2010**
	Mean annual maximum temperature	**tmax**	**Con.**	**PRISM**	**800 m**	**1981–2010**
	Ratio of precipitation to potential evapotranspiration	GMI	Con.		250 m	N/A
Organism	AGB	above	Con.	NFI	N/A	2000–2010
	Forest type group	fortypgrp	Cat.	NFI	250 m	N/A
	NDVI (monthly, maximum and sum)	**NDVI**	**Con.**	**Landsat 7**	**30 m**	**2000–2010**
	Canopy bulk density	Lfcbd	Con.	LANDFIRE	30 m	2010
	Canopy base height	Lfcbh	Con.	LANDFIRE	30 m	2010
	Canopy cover	**Lfcc**	**Con.**	**LANDFIRE**	**30 m**	**2010**
	Canopy height	Lfch	Con.	LANDFIRE	30 m	2010
	Existing vegetation cover	Lfevc	Con.	LANDFIRE	30 m	2010
	Existing vegetation height	Lfevh	Con.	LANDFIRE	30 m	2010
	Existing vegetation type	Lfevt	Cat.	LANDFIRE	30 m	2010
	Land Resource Regions	**LRR**	**Cat**	**USDA**	**N/A**	**N/A**
Relief	**Elevation**	**Elev**	**Con**	**NED**	**30 m**	**N/A**
	Aspect	Aspect	Con	NED	30 m	N/A
	Slope	Slope	Con	NED	30 m	N/A
Parent material	Surficial geological	surfgeo	Cat.	USGS	1:250,000	N/A
	Gamma ray Potassium concentrations	**GR_K**	**Con.**	**USGS**	**2 km**	**1999–2005**
	Gamma ray Thorium concentrations	**GR_Th**	**Con.**	**USGS**	**2 km**	**1999–2005**
	Gamma ray absorbed dose rate	GR_exp	Con.	USGS	2 km	1999–2005
	Gamma ray Uranium concentrations	GR_U	Con.	USGS	2 km	1999–2005
	Gamma ray Bouguer gravity anomaly	GR_bouguer	Con.	USGS	2 km	1999–2005
	Gamma ray Magnetic anomaly	GR_origmrg	Con.	USGS	2 km	1999–2005
	Gamma ray Magnetic anomaly	GR_CM	Con.	USGS	2 km	1999–2005
	Gamma ray Magnetic anomaly	GR_hp500	Con.	USGS	2 km	1999–2005
	Gamma ray isostatic residual gravity anomaly	GR_isograv	Con.	USGS	2 km	1999–2005
N	Latitude, Longitude[d]	Lat, Lon	Con.	NFI	N/A	N/A

[a] Abbreviations: AGB, aboveground biomass; NDVI, Normalized Difference Vegetation Index.
[b] Abbreviations: Cat., categorical; Con., continuous.
[c] Abbreviations: NED, national elevation dataset; NFI, National Forest Inventory.
[d] Latitude and longitude are the projected coordinates in NAD_1983_Albers where litter samples were collected.

all non-parametric methods and accommodate both continuous and categorical predictors. The first two models allow estimating the relative importance of the predictor variables based on the performance of the model if the data for each predictor were permuted randomly (Prasad et al., 2006). Each of these methods can deal with complex non-linear relationships between litter C and SOC stocks and environmental variables. These approaches have been used successfully to predict litter C and SOC stocks in various climatic regions (Fuchs et al., 2009; Grimm et al., 2008; Rudiyanto et al., 2016; Suchenwirth et al., 2014). Each type of machine learning approach has specific and different tuning parameters to control how the relationship between input predictors and response is determined. These parameters must be optimized to generate the best model between covariates and target properties (Brungard et al., 2015).

Random forest (RF) is a machine learning method consisting of an ensemble of randomized classification and regression trees (CART) (Breiman, 2001). The RF algorithm grows different trees by randomly and repeatedly selecting predictor variables and training cases to develop a random population of trees. In regression it is the average of the individual tree predictions. The RF algorithm can be very efficient, especially when the number of descriptors is very large (Svetnik et al., 2003). In this study, RF was used to predict litter and SOC stocks. Two options were presented in the final models: number of trees to grow (ntree) and number of variables randomly sampled as candidates at each split (mtry). We used these two parameters to prune our models.

Quantile regression forest (QRF) is a non-parametric technique used to estimate the conditional quantiles of multidimensional predictor variables. Therefore, QRF is able to predict more accurate results for the conditional distribution of the response variable (Meinshausen, 2017). In this application, the conditional quantiles were computed for $\alpha =$ 0.10, 0.50, 0.90 at each node of the grid from the estimated conditional distribution.

The k-nearest neighbor (kNN) approach is also a non-parametric approach that has been used since the early 1970's in statistical applications (Duda and Hart, 1973; Franco-Lopez et al., 2001). The basic concept of kNN is that in the calibration dataset, it finds a group of k samples that are nearest in parameter space to unknown samples (e.g., based on distance functions). The k plays an important role in the model performance of the kNN, since it determines the tuning parameter of kNN (Qian et al., 2014). The parameter k is generated using a bootstrap procedure. When kNN is used for regression, the predictions of unknown samples are based on the mean of the response variables. In this study, we examined k values from 1 to 25 to identify the optimal k value for all training sample sets.

2.5. Model validation

The three model approaches were assessed using independent validation. The whole dataset (n = 3303) was split 70 (n = 2312)/30 (n = 991) into calibration and validation sets. The 70% calibration samples were randomly selected from the whole dataset. The Kolmogorov-Smirnov test on the SOC distribution of the two sets (i.e., calibration and validation sets) was applied to ensure they have the similar distributions. The error metrics used to compare models were the coefficient of determination (R^2, Eq. (1)), root mean squared error (RMSE, Eq. (2)),

$$R^2 = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y}_i)^2} \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2/n} \tag{2}$$

residual prediction deviation (RPD, Eq. (3)),

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSE}\sqrt{n/(n-1)}} \tag{3}$$

and ratio of performance to inter-quartile range (RPIQ, Eq. (4)) (Bellon-Maurel et al., 2010),

$$\mathrm{RPIQ} = \frac{\mathrm{IQ}}{\mathrm{RMSE}} \tag{4}$$

where, $\hat{y}_i$ is the ith mean of posterior prediction distribution, y_i is the ith observation, $\overline{y}_i$ is the mean of y_i, n is the number of predicted or observed values with $i = 1,2, \ldots, n$, SD is the standard deviation, and IQ is the interquartile range, which is the difference between the third and first quartiles (IQ = Q3 − Q1). Also, the predictions from the validation dataset were compared to the observed values using two one-sided tests (Wellek, 2003).

In this study, all analyses were conducted with R 3.4.3 (R Development Core Team, 2017). The 'randomForest' (Liaw and Wiener, 2002), 'quantregForest' (Meinshausen, 2017), and 'caret' (Kuhn, 2018) packages were used for litter and SOC stock predictions and the 'equivalence' package (Robinson, 2016) was used for two one-sided tests. We used continuous and categorical variables to generate the predicted litter C and SOC stock (0-20 cm) maps, produced by the 'ModelMap' package (Freeman and Frescino, 2009). All variable layers were projected and resampled to the same extent and resolution (1 km) to create spatial characterizations of litter C and SOC predictions for the CONUS. The details of the modeling procedure can be found in Fig. S1.

3. Results

3.1. Litter and soil carbon predictions

Estimated mean litter C varied from 0 to 144.62 Mg ha^{-1}, while the estimated SOC varied from 0.26 to 524.83 Mg ha^{-1} in the 0–20 cm profile (Fig. 1). The litter C and SOC data were both positively skewed (Table 2), with most estimates < 50 Mg ha^{-1} and 100 Mg ha^{-1}, respectively. After log transformation of litter C and SOC, the mean and median values were similar to each other, approximating a normal distribution. The range of the calibration data encompassed that of the validation sample, and its frequency distribution was slightly more positively skewed and more variable than that of the validation sample. Log transformation of litter C, soil C, NDVI and AGB reduced their variability.

3.2. Correlations

The strongest positive correlation was found between the maximum NDVI of growing season and AGB ($r = 0.46$, $p < 0.01$, Fig. 2a). Both maximum NDVI and AGB were significantly and positively correlated with log-transformed litter C stock (Fig. 2b and c). Moreover, the correlation between litter C and the maximum NDVI was still slightly higher than that between SOC and the maximum NDVI (Fig. 2b and d). Therefore, it is reliable to use NDVI data to replace AGB data in the modeling stage, since AGB data is point based.

3.3. Model performance and variable importance

The results from the optimal models for each non-parametric approach are summarized in Table 3. Generally, the models representing each of the three non-parametric approaches performed better for SOC than litter C stocks. Although the predictive power of the model is variable depending on model type and C pool, for SOC the QRF model, RPD was 1.22 for validation, and this was close to the value of 1.40 that is considered fair for models and predictions which may be used

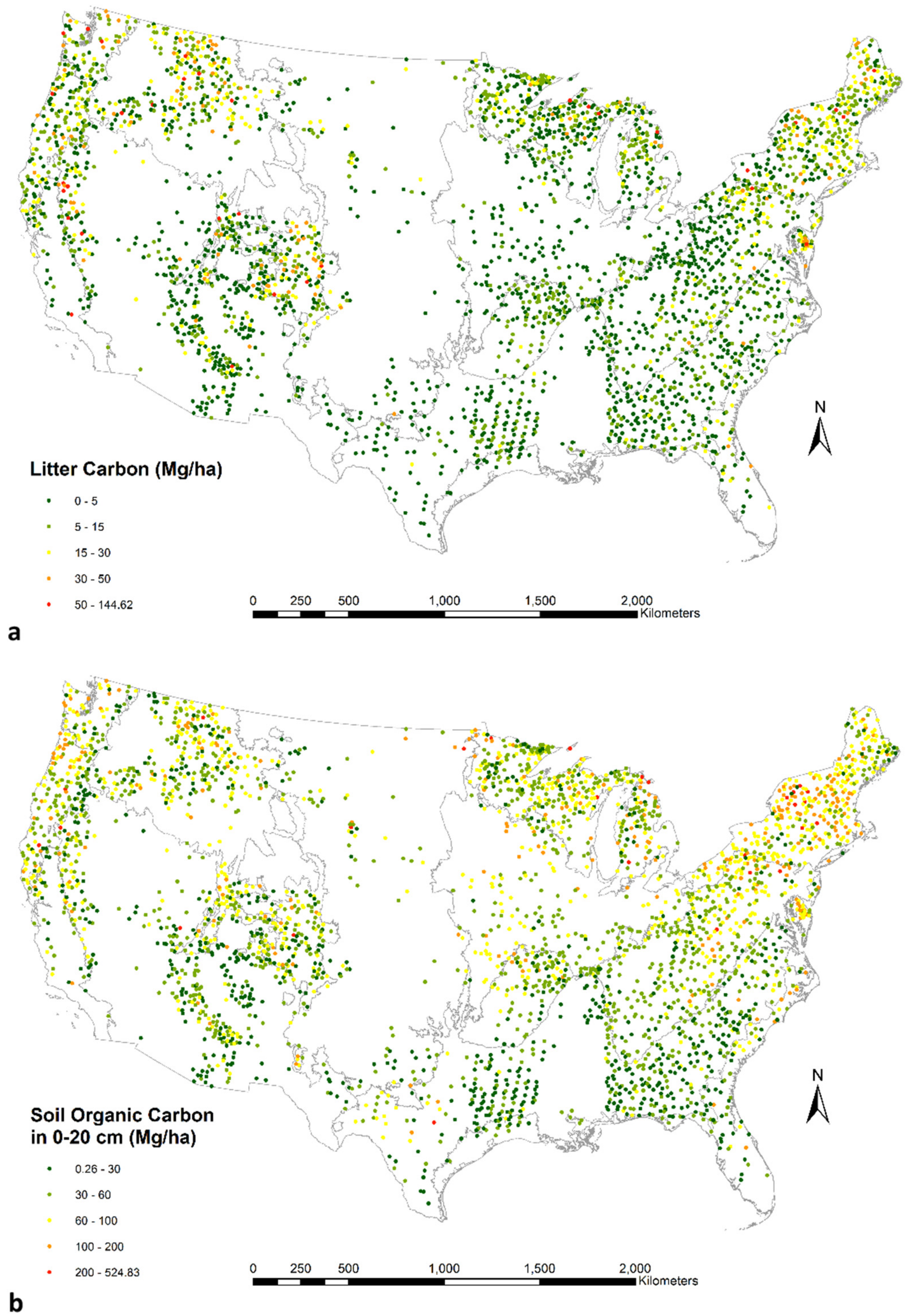


Fig. 1. a) Litter carbon samples (n = 3303); b) soil carbon samples (n = 3303) in the CONUS.

for assessment and correlation (Viscarra Rossel et al., 2006). Average prediction performance of the validation data were expressed by the RPIQ index, with values of 1.01 (litter C) and 1.39 (SOC). The best coefficient of determination (R^2) was 0.20 and 0.35 for litter (RF model) and SOC (QRF model), with an RMSE of 9.23 Mg ha^{-1} and 28.15 Mg ha^{-1} respectively. The RPD values (1.11 of litter C and 1.22

Table 2
Summary statistics of litter and SOC stocks (Mg ha^{-1}) acquired from NFI data collected across the conterminous US (Domke et al., 2016, 2017).

Variable	Mean	SD	Median	Minimum	Maximum	Range	Skewness	Kurtosis	SE
Litter C	8.11	10.75	4.41	0	144.62	144.62	3.48	21.11	0.19
Soil C	56.87	38.46	48.74	0.26	524.83	524.57	3.05	20.29	0.68

Abbreviations: SD, standard deviation; SE, standard error.

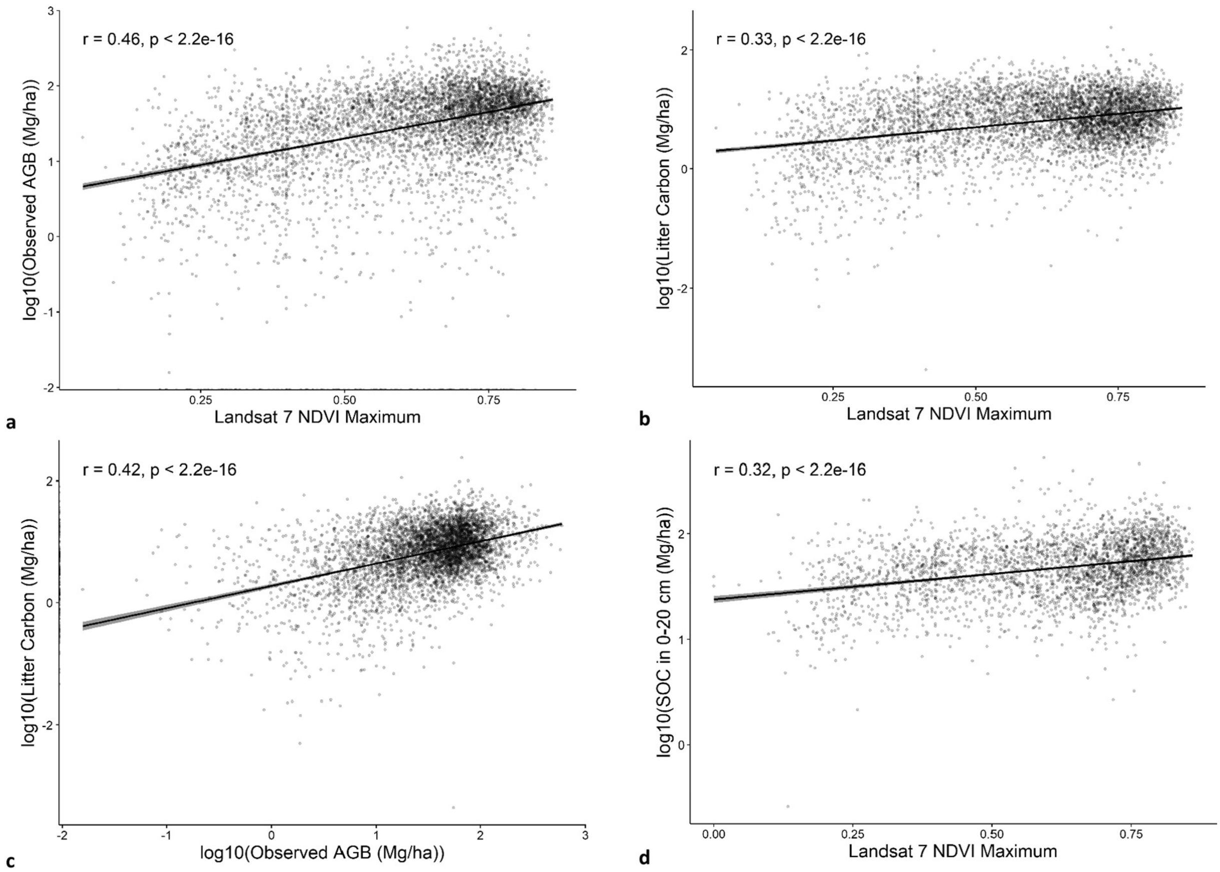


Fig. 2. Correlations with litter carbon stocks and other variables: a) NDVI and log-transformed aboveground biomass (AGB); b) NDVI and log-transformed litter carbon stock; c) log-transformed litter carbon stock and AGB; d) log-transformed soil carbon stocks (0–20 cm) and NDVI.

of SOC) indicated fair models that could be improved with additional data and co-variates (Chang et al., 2001). Despite the variability in the validation plots (Fig. 3a), the litter C model showed slight overestimation and the model fitted the lower estimates better (i.e., for litter C values < 20 Mg ha^{-1}). A slight overestimation of SOC for lower estimates (SOC < 50 Mg ha^{-1}) and underestimation for estimates higher than 150 Mg ha^{-1} was apparent (Fig. 3b). By only using the best correlated predictors to litter and SOC (i.e., 10 covariates, bold in Table 1), the independent validation of the QRF SOC model showed 29% of explained data variability, while RF litter C model showed 15% of explained data variability (Table 4). Variable importance in this study was determined by IncNodePurity, which is used to measure quality of a split for every variable (node) of a tree by means of the Gini Index. It revealed different dominating environmental features for litter C and SOC models (Fig. 4). Variable importance among predictors showed similar patterns for litter C and SOC. Climate moisture index (GMI) proved to be the most important predictor for litter C, while it was the third strongest for SOC. Spatial location (latitude and longitude) played a dominant role in both models, while elevation ranked high for both models. NDVI also presented relatively strong prediction power when compared to other top ranked variables. Soil moisture variables (Available water and soil moisture) ranked in the middle of variable importance. For both models, the categorical predictor forest type group was the least important variable. Gamma ray variables appeared to be important to litter C, especially Potassium (GR_K) and Thorium (GR_Th) concentrations. The land resource regions variable did not rank as important in both litter and soil C models.

The optimal models for litter C and SOC were further validated using Two-One-Sided Tests of equivalence (Wellek, 2003). The RF approach provided the best summary statistics and the model predictions were statistically equivalent with litter C estimates in the validation dataset (Table 4). The QRF approach provided the best fit to the training dataset and the predictions were statistically equivalent to the SOC estimates in the validation dataset.

Table 3
Summary statistics of random forest (RF), quantile regression forest (QRF), and k nearest neighbor (kNN) models using all variables to predict litter carbon (C) and soil organic carbon (SOC) stocks (Mg ha^{-1}) across the conterminous US.

Method	Dataset	R^2	RMSE (Mg ha^{-1})	RPD	RPIQ	Bias
RF	Litter C	0.20	9.23	1.11	1.01	0.36
	SOC	0.33	28.39	1.21	1.37	3.18
QRF	Litter C	0.16	10.90	1.05	0.89	−2.91
	SOC	0.35	28.15	1.22	1.39	−3.17
kNN	Litter C	0.11	10.89	1.06	0.79	−0.38
	SOC	0.28	29.32	1.17	1.33	0.74

Abbreviations: RMSE, root mean square error; RPD, residual predict; RPIQ, ratio of performance to inter-quantile range.

3.4. Spatial characterization of litter and soil carbon

The distribution of litter C and SOC stocks (Mg ha^{-1}) (Figs. 5a and 6a) and their SD (Figs. 5b and 6b) and coefficient of variation (CV) (Fig. S4) were characterized at a resolution of 1 km × 1 km. The total litter C and SOC stocks in 0–20 cm for the CONUS was 2.07 ± 0.97 (total ± SD) Pg C with an average density of 10.45 ± 2.38 Mg ha^{-1}, and 14.68 ± 3.50 (total ± SD) Pg C with an average density of 62.68 ± 8.98 Mg ha^{-1}, respectively, for a surface of about 3.10 million km^2. The total surface represented by the grid is 24% of the actual CONUS (8,080,464.3 km^2). The highest stocks shown on the predicted maps were forecast by the models in three areas: northeastern US, northern Wisconsin and northwest coastal forest. For litter C, SD varied from 0.95 to 33.80 Mg ha^{-1},

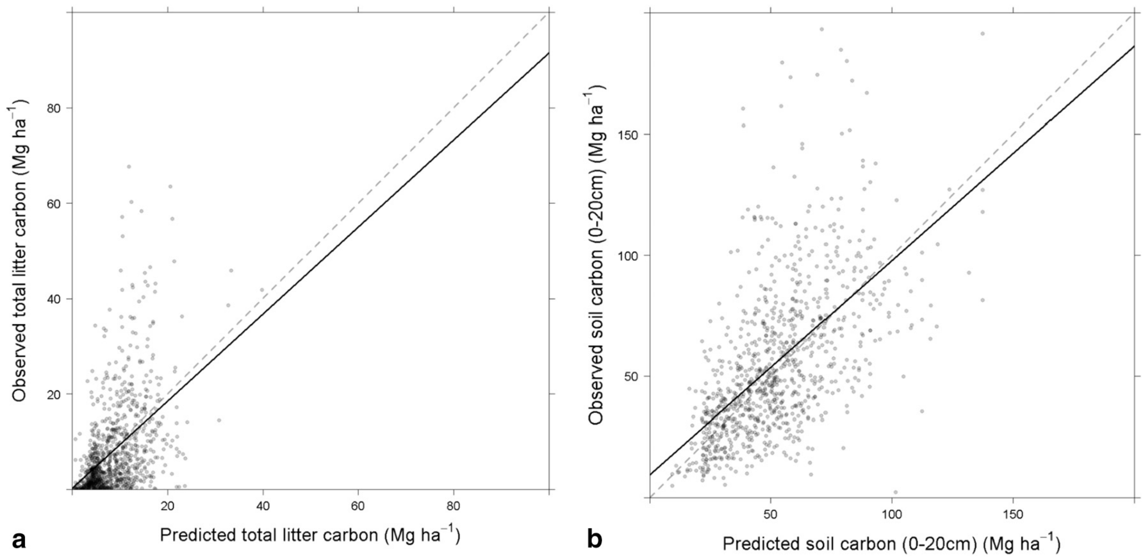


Fig. 3. Observed and predicted values of the validation dataset from random forest models for a) Litter C; quantile regression forest model for b) soil organic carbon (SOC) in 0–20 cm. Black line is the regression line and the dash line is the 1:1 line.

while coefficient of variation ranged from 0.33% to 4.28%; for SOC, SD ranged from 7.60 to 148.75 Mg ha^{-1}, while coefficient of variation ranged from 0.20% to 1.70%.

3.5. Forest disturbance impact on forest carbon stocks

Litter C and SOC stocks were summarized by forest disturbance types in Fig. S2 based on FIA data. The fire disturbance group presented the second highest SOC stocks and the third lowest in litter C stocks. The disease group of forest disturbance presented the highest litter C, but the lowest soil C. The Kruskal-Wallis test results suggested that the difference of litter C and SOC stocks between the disturbed and non-disturbed plots were not statistically significant. However, litter C stocks were significantly different between disturbance groups ($p = 0.0091$), while the disease group had the highest litter C and the animal group had the lowest. Also, the forest disturbance (LANDFIRE) showed significant impact on litter C stocks (dataset from Domke et al. (2016), $n = 5236$, $p = 0.0056$). For both litter and soil, the mean and median values of C stocks in non-disturbance groups tended to be slightly higher but not statistically significantly different than that of the disturbance group.

3.6. Litter and soil carbon ratio distribution

A map characterizing the ratio of litter C to SOC showed that 75% of sample sites had a litter and soil C ratio lower than 0.21, and only 1.5% (51 sample sites) with higher litter C than SOC (ratio > 1, red points in Fig. S3). There were some spatial patterns of regions with high ratios, including areas located in the Sierra Nevada Mountains, northern and southern Rocky Mountains, Lake States, and southeast coast flatwoods regions.

Table 4
Summary statistics of best models using partial covariates (variables that are bold in Table 1) to predict litter carbon (C) and soil organic carbon (SOC) stocks (Mg ha^{-1}) across the conterminous US. TOST is two-one-sided test results where NE = not equivalent and E = equivalent where the absolute value of the mean of the differences is ±25% of the standard deviation.

Dataset	Method	R^2	RMSE (Mg ha^{-1})	RPD	RPIQ	Bias	$C_{NFI} - C_{pred}$		
							Mean	SE	TOST
Litter C	RF	0.15	10.22	1.09	0.78	−0.51	0.17	0.20	E
SOC	QRF	0.29	29.14	1.18	1.34	−4.42	−1.56	0.45	E

Abbreviations: RMSE, root mean square error; RPD, residual predict; RPIQ, ratio of performance to inter-quartile range; SE, standard error of the mean difference; C_{NFI}, carbon measured in National Forest Inventory; C_{pred}, carbon predicted with method.

4. Discussion

4.1. Carbon stock estimates

Globally, it was reported that the current litter C stock in the world's forests was 43 ± 3 Pg C (5% of total forest C) (Pan et al., 2011). Our prediction of litter C in the CONUS accounts for approximately5%of the world's litter C stocks. Sanderman et al. (2017) reported that the global forest SOC in 0–30 cm profile was 223.4 Pg (148.93 Pg of SOC in 0–20 cm after depth-weighted average). Thus, our analysis suggests that forest SOC of 0–20 cm in CONUS accounts for approximately 10% of the world's.

Nationally, our results indicate that litter C and SOC in 0–20 cm account for 4.83% and 34.22% of total US forest C stocks (40.19 Pg) (US EPA, 2015), respectively. Our results for both litter C and SOC stocks are higher than those of China and Mexico. Our results are consistent with Domke et al.'s (2016, 2017) findings in terms of total C stocks which is not surprising given the same NFI data were used. However, our study extended the work of Domke et al. (2016, 2017) from site-specific to continuous predictions across the CONUS. The predicted litter and soil C maps (Figs. 5 and 6) contain much greater spatial detail than the point-based estimates previously published (Domke et al., 2016, 2017). This is important for future forest C modeling and site-specific management activities. The mean of all Rapid Carbon Assessment (RaCA) site SOC stocks to 100 cm was reported as 345.4 Mg ha^{-1}, with a median of 183.2 Mg ha^{-1} (Wills et al., 2014) in the CONUS. These were much higher than our SOC density and might be because RaCA included samples with extreme high values from agricultural farmland and wetland. In a study of China's forests, it was estimated that litter C stock (not including woody debris) in the period of 2004–2008 was 0.50 ± 0.024 Pg, with an average density of 5.95 ± 0.35 Mg C ha^{-1} (Zhu et al., 2017). Also, although forest area size (3.10 million km^2) in our study is about two times that of China (1.53 million km^2), litter C in the CONUS forest (2.07 ± 0.97 Pg) is four times of that in China's, and showed much higher average density of 10.45 ± 2.38 Mg ha^{-1}. Compared with China's 189 sample sites, this study contained higher sample density and covered more diverse

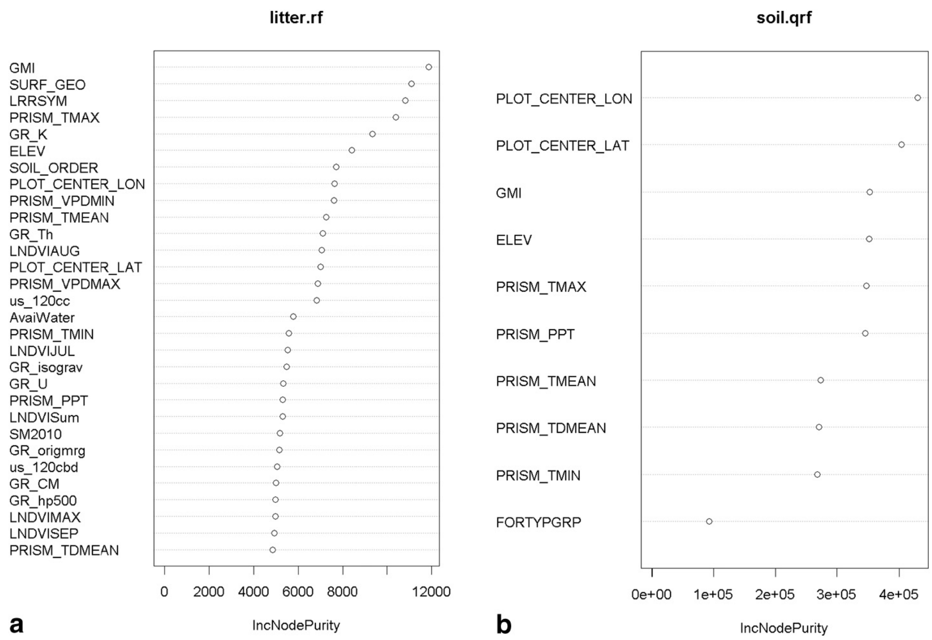


Fig. 4. Variable importance from random forest model of litter and quantile regression forest model of soil carbon predictions.

forest types. Considering the areal extent, our results are in the same magnitude as these previous predictions from China and Mexico.

Our estimates based on available NFI data provided a baseline for litter and soil C stocks in the CONUS. The presented stocks reflect the total C of forest litter and soil for the period 2000 to 2011 and can be used as a benchmark for future comparison with new data or enhance modeling methodologies. This background information is valuable as a reference to evaluate spatio-temporal litter and soil C changes in the CONUS.

4.2. Sample variation and density

Geographically speaking, the high variation presented in those areas with high C density, such as northeastern area, west coastal area, lake state area and rocky mountain area. Not surprisingly, the SD was much smaller in litter C prediction than those in SOC prediction, likely because the range of SOC was more than two times that of litter C. Interestingly, the spatial distribution of litter C SD also exhibited strong spatial patterns (Fig. 5b), with clusters containing the highest level of variation in heavily forested regions in the northeastern US. When the results showed low levels of litter and soil C density, the spatial distribution of SD was more homogeneous across landscapes (e.g., in the southeastern US).

In this study, we demonstrated relationships and predictions using remote sensing image-based data (NDVI, LANDFIRE and Gamma Ray), allowing for continental scale analysis of forest C stocks where limited field data are available. In terms of sample density, Somarathna et al. (2017) suggested a minimum of 15 samples of soil C per square kilometer to reach models' maximum predictive capability for a hill area in Australia, and Vaysse and Lagacherie (2017) suggested sample density as 1/13.5 km^2 for a regional study in France. Comparing to those studies, our sample density is much lower (about 1/1000 km^2) and our results of soil C prediction are better as indicated by the R^2. This could be a reflection of the greater resolution (30 m) of remote sensing data that we applied, which enabled the expansion of litter and soil C stocks from plot-based observations to a pixel-based continuous layer across the CONUS.

4.3. Relationships between carbon stock and environmental variables

The prediction method, such as RF and QRF using R software, was capable of exploring the relationship of litter and soil C to their environment predictors. Based on the validation indices, the models showed a higher performance (i.e., higher R^2, lower RMSE, higher RPD and RPIQ) in calibration data (70%) than in validation data (30%).

The positive correlation between the SOC and NDVI was also found in a tropical forest in India, which demonstrated the fact that NDVI can be considered to be an effective spectral vegetation index to estimate SOC (Kumar et al., 2016). There was also a study in the western US which reported that NDVI is one of the most important variables of affecting soil pyrogenic and particulate C (two forms of SOC) stocks (Ahmed et al., 2017). Our findings are in agreement with studies that indicate organic C inputs to the soil, in addition to the loss due to erosion and mineralization, are the important factors explaining variation of SOC stocks.

The distribution of litter C presented a similar spatial pattern as with temperature that warmed from north to south. This is because temperature and precipitation decrease the C density of litter (Zhu et al., 2017). A microcosm experiment in a rain forest in China found that a rise of 10 °C in temperature significantly decreased the total mass of litter for the primary forest (He et al., 2009). However, the northeastern coastal area of Florida (Everglades) showed high litter C stocks, and this might be because of the high production and low decomposition rates in mangrove ecosystems (Liu et al., 2017). Also, soil C stocks presented a spatial pattern due to the temperature. Forests in the northeast and west coastal areas of the US accumulated higher SOC than forests in the southeast, due to the longer C turnover time in high-latitude compared to low-latitude zones (Hakkenberg et al., 2008; Wang et al., 2018).

Moisture-related factors (e.g., precipitation, climate moisture index) control SOC in Florida (Xiong et al., 2014), and this also applied to the CONUS scale. Factors relating to moisture were the dominant variables in the models highlighting the importance of the interaction between climate and soil variables in controlling SOC stocks. Guo et al. (2006b) studied climatic effects on SOC in the CONUS, and also found that the effect of temperature on SOC is complicated or confounded by other SCORPAN factors. For example, it was corroborated that soil C mineralization responds quite differently to climate change by parent material and litter type in a temperate forest (Rasmussen et al., 2008).

Interestingly, forest type group ranked low in importance for litter C, but ranked high in importance for soil C. One contributing factor may be the mobility of leaf and woody material, which was suggested by a study in Western Australia that observed roughly 40–60% of the existing

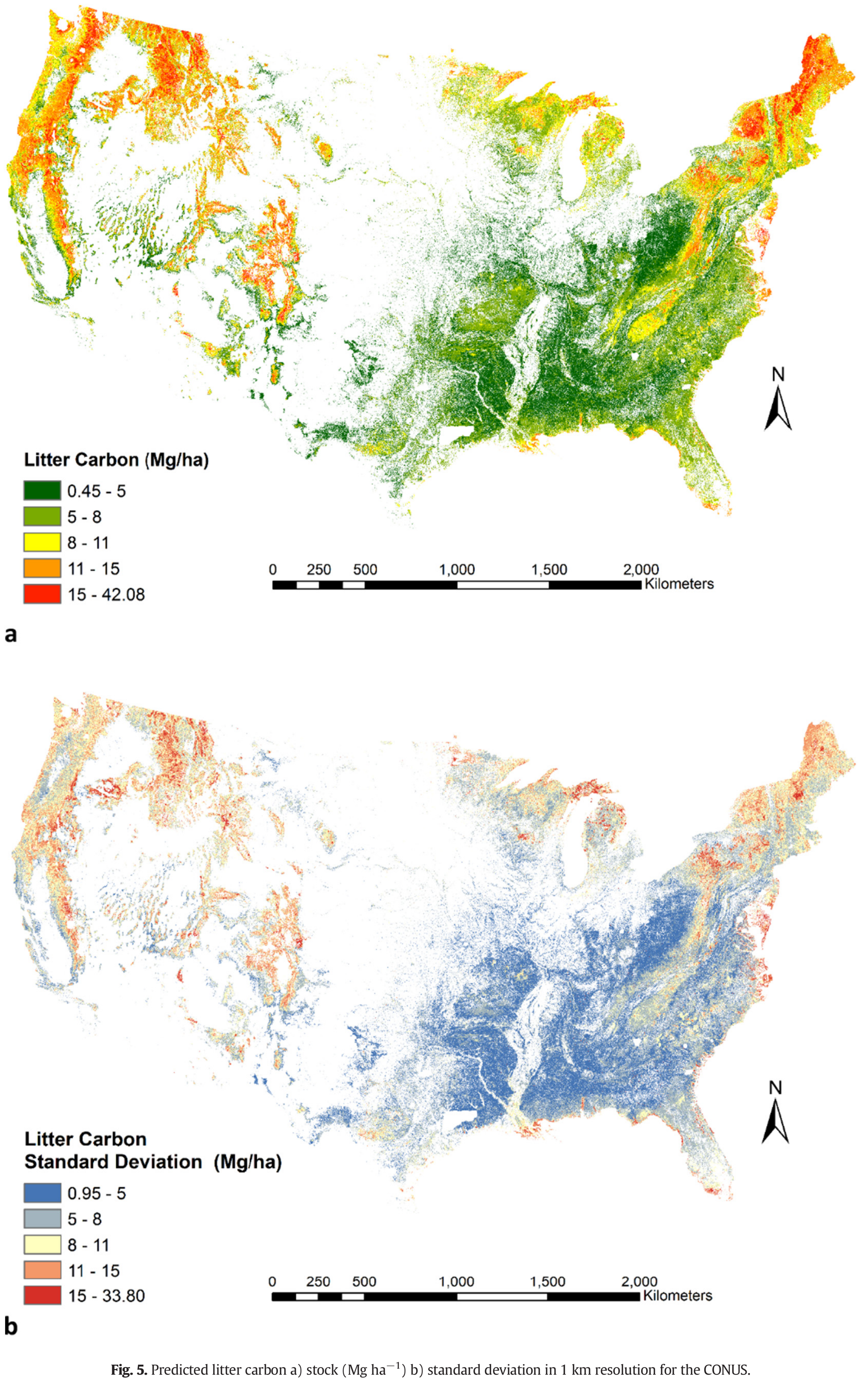


Fig. 5. Predicted litter carbon a) stock (Mg ha^{-1}) b) standard deviation in 1 km resolution for the CONUS.

litter was removed by wind or flood annually from all the experiment sites (Kumada et al., 2009). This also explained the complexity and uncertainty on litter C predictions and why adding a climate variable such as wind speed could possibly improve the model. Another model improvement could come from insights through litter decomposition studies. A study in British Columbia tested 14 overstory tree species

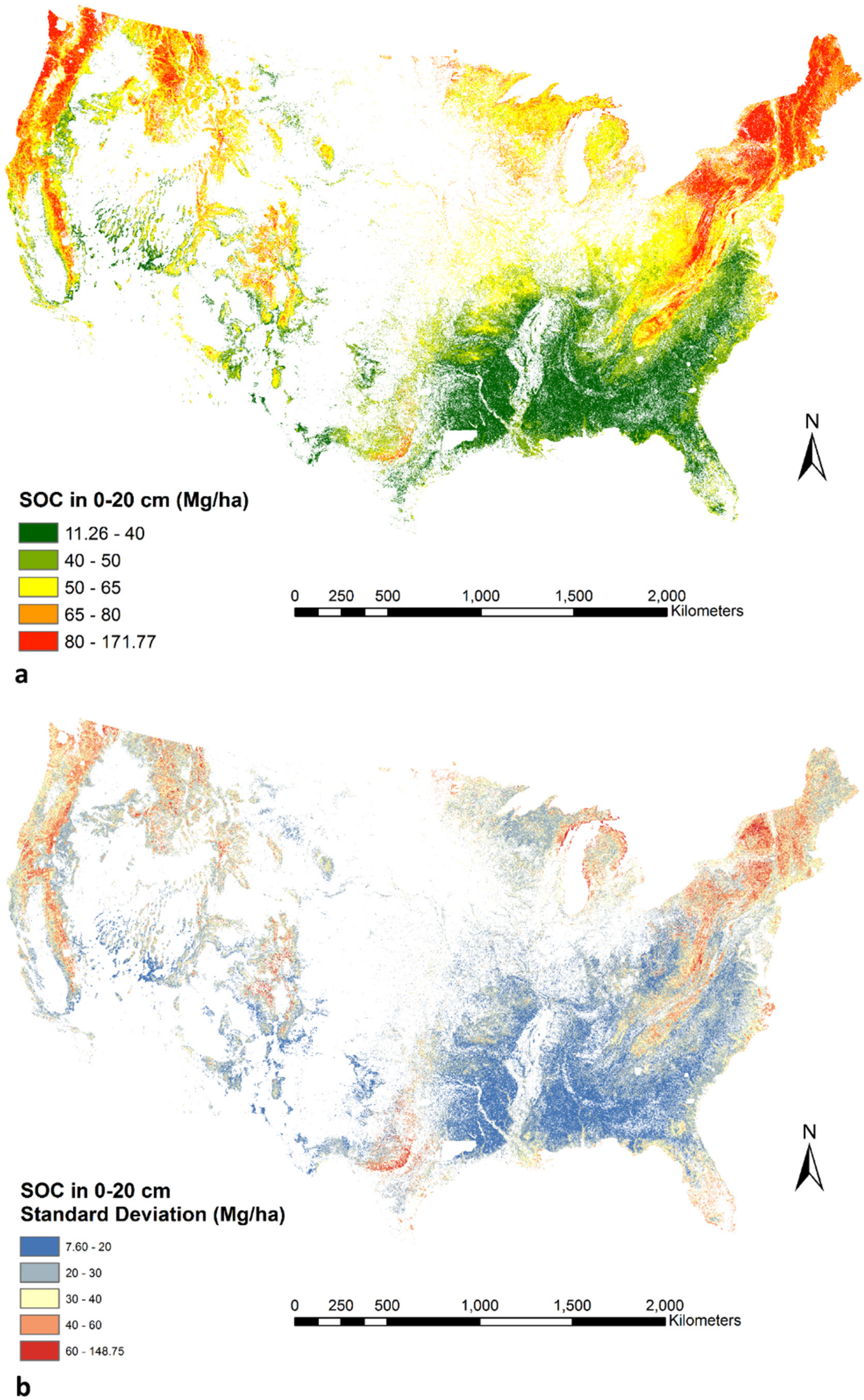


Fig. 6. Predicted soil organic carbon in 0–20 cm a) stock (Mg ha^{-1}) b) standard deviation in 1 km resolution for the CONUS.

(broadleaf, needleleaf, or mixedwood) and found that long-term litter decay rates may not differ much among forests of different tree species composition (Prescott et al., 2004).

4.4. Forest disturbance and litter soil carbon ratio

We evaluated the influence of disturbance on litter C and SOC stocks using two different data products: FIA and LANDFIRE. The FIA data are more accurate with site-specific forest disturbance types, while the LANDFIRE data are applicable to large scale disturbances, since they were obtained from remote sensing images and created with a 500 m buffer. The disease group of forest disturbance presented the highest litter C, but the lowest SOC, which suggested these sites had high potential for a site-specific disturbance to transfer C from live to dead pools (McGarvey et al., 2015). The mean difference between the two litter and soil estimates and the observed values account for only 2.1% and 2.7% of mean estimated litter C and SOC, respectively. Statistical equivalence between NFI estimates and model predictions for litter C and SOC provides evidence that employing remotely sensed information can enable the expansion of forest C pool estimates across a range of spatial scales for which limited field observations often exist. For longer-term forecasts, these data provide information on forest dynamics linked to past changes in climate, land-use, and disturbances. These studies can often offer landscape-scale reconstructions, or in some situations, can be used to infer stand-level histories.

Litter C to SOC ratios larger than 1 appeared in either mountainous areas or lake and coastal flatwoods areas. This is because soil layers and depth to bedrock or other impervious surfaces were likely much shallower in mountainous areas and slower decomposition of leaves resulted in denser litter layers in flatwoods (Rayamajhi et al., 2010). In the US Lake States, it is possible that wetlands and streams with more run-off and less water storage may have lower average water tables and the potential for more aerobic conditions which can reduce SOC storage (Barksdale et al., 2014). A litter dynamic study (Upton et al., 2018) suggested that litter inputs strongly impact litter chemistry of the topsoil, however, ultimately the total C storage at their study sites is controlled by how variation in environmental conditions (e.g., all the SCORPAN environmental variables included in our study) govern long term decay processes rather than litter inputs linked to specific vegetation types.

5. Conclusions

In this study we characterized the spatial distributions of litter C and SOC stocks in the CONUS using environmental variables from the NFI and obtained from remotely sensed data products. Six main conclusions can be drawn from this study. 1) Spatially continuous and explicit estimates of litter C stocks on forest land in the CONUS were 2.07 $\pm$ 0.97 Pg and SOC stocks at 0–20 cm depth was 14.68 $\pm$ 3.50 Pg, 2) our estimates are comparable to other national estimates for the US and our results suggest that litter C and SOC in 0–20 cm account for 4.83% and 34.22% of total forest C stocks nationally, 3) the high-resolution litter C and SOC stock predictions can be used to facilitate forest C spatial modeling and provide insight to site-specific management activities at different scales, 4) the role of forest disturbances on litter C and SOC stocks in forest ecosystems is complex, 5) our results suggested that the RF and QRF prediction models performed better than kNN models although results across the three methods were similar, and 6) despite the variability in litter and soil observations across a large geographic scale, all modeling approaches performed better for soil compared to litter layers and the spatial pattern of association between litter, SOC, and environmental covariates observed from the RF and QRF models may reflect spatial patterns in litter decomposition, soil chemistry, and plant and microbial communities. Furthermore, these methods can be expanded to data-limited areas by the application of remote sensing techniques. These conclusions illustrate the application of harmonizing field measurements with remotely sensed data to improve modeling and estimation across spatial scales in support of inventory, monitoring, and reporting activities, particularly in countries with ready access to remotely sensed data but with limited field observations of litter and soil variables.

Acknowledgements

This study was funded by USDA Forest Service, Forest Inventory and Analysis Program. The authors would like to thank Ronald McRoberts for early discussion on possible stratification techniques.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.10.359.

References

Ahmed, Z.U., Woodbury, P.B., 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. J. Geophys. Res. Biogeosci. 122, 354–369. https://doi.org/10.1002/2016JG003488.

Amundson, R., 2001. The carbon budget in soils. Annu. Rev. Earth Planet. Sci. 29, 535–562. https://doi.org/10.1146/annurev.earth.29.1.535.

Barksdale, W.F., Anderson, C.J., Kalin, L., 2014. The influence of watershed run-off on the hydrology, forest floor litter and soil carbon of headwater wetlands: run-off effects on hydrology, leaf litter and soils of headwater wetlands. Ecohydrology 7, 803–814. https://doi.org/10.1002/eco.1404.

Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use Manag. 25, 124–127. https://doi.org/10.1111/j.1475-2743.2009.00202.x.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838. https://doi.org/10.1126/science.1184984.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends Anal. Chem. 29, 1073–1081. https://doi.org/10.1016/j.trac.2010.05.006.

Bhattacharyya, T., Pal, D.K., Mandal, C., Velayutham, M., et al., 2000. Organic carbon stock in Indian soils and their geographical distribution. Curr. Sci. 79, 655–660.

Bonan, G.B., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320, 1444–1449. https://doi.org/10.1126/science.1155121.

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

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019.

Bui, E., Henderson, B., Viergever, K., 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. Glob. Biogeochem. Cycles 23, GB4033. https://doi.org/10.1029/2009GB003506.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480. https://doi.org/10.2136/sssaj2001.652480x.

Domke, G.M., Perry, C.H., Walters, B.F., Woodall, C.W., Russell, M.B., Smith, J.E., 2016. Estimating litter carbon stocks on forest land in the United States. Sci. Total Environ. 557–558, 469–478. https://doi.org/10.1016/j.scitotenv.2016.03.090.

Domke, G.M., Perry, C.H., Walters, B.F., Nave, L.E., Woodall, C.W., Swanston, C.W., 2017. Toward inventory-based estimates of soil organic carbon in forests of the United States. Ecol. Appl. 27, 1223–1235. https://doi.org/10.1002/eap.1516.

Duda, R.O., Hart, P.E., 1973. Pattern Classification and Scene Analysis. 1 edition. Wiley, New York.

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S., Whitmore, A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Glob. Change. Biol. 18, 1781–1796. https://doi.org/10.1111/j.1365-2486.2012.02665.x.

Franco-Lopez, H., Ek, A.R., Bauer, M.E., 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sens. Environ. 77, 251–274. https://doi.org/10.1016/S0034-4257(01)00209-7.

Freeman, E., Frescino, T., 2009. ModelMap: Modeling and Map Production Using Random Forest and Stochastic Gradient Boosting. USDA Forest Service, Rocky Mountain Research Station, 507 25th Street, Ogden, UT, USA.

Fuchs, H., Magdon, P., Kleinn, C., Flessa, H., 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. Remote Sens. Environ. 113, 518–531. https://doi.org/10.1016/j.rse.2008.07.017.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — digital soil mapping using Random Forests analysis. Geoderma 146, 102–113. https://doi.org/10.1016/j.geoderma.2008.05.008.

Guo, Y., Amundson, R., Gong, P., Yu, Q., 2006a. Quantity and spatial variability of soil carbon in the conterminous United States. Soil Sci. Soc. Am. J. 70, 590. https://doi.org/10.2136/sssaj2005.0162.
Guo, Y., Gong, P., Amundson, R., Yu, Q., 2006b. Analysis of factors controlling soil carbon in the conterminous United States. Soil Sci. Soc. Am. J. 70, 601. https://doi.org/10.2136/sssaj2005.0163.
Hakkenberg, R., Churkina, G., Rodeghiero, M., Boerner, A., Steinhof, A., Cescatti, A., 2008. Temperature sensitivity of the turnover times of soil organic matter in forests. Ecol. Appl. 18, 119–131. https://doi.org/10.1890/06-1034.1.
Hartemink, A., McBratney, A., Mendonça-Santos, M., Mendonça-Santos, M., 2008. Digital Soil Mapping With Limited Data. Springer, Dordrecht.
He, X., Zhang, P., Lin, Y., Li, A., Tian, X., Zhang, Q.-H., 2009. Responses of litter decomposition to temperature along a chronosequence of tropical montane rainforest in a microcosm experiment. Ecol. Res. 24, 781–789. https://doi.org/10.1007/s11284-008-0549-2.
Hengl, T., de Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748.
Kuhn, M., 2018. caret: Classification and Regression Training. R Package Version 6.0-80.
Kumada, S., Kawanishi, T., Hayashi, Y., Hamano, H., Kawarasaki, S., Aikawa, S., Takahashi, N., Egashira, Y., Tanouchi, H., Kojima, T., Kinnear, A., Yamada, K., 2009. Effects of different mobilities of leaf and woody litters on litter carbon dynamics in arid ecosystems in Western Australia. Ecol. Model. 220, 2792–2801. https://doi.org/10.1016/j.ecolmodel.2009.07.009.
Kumar, P., Pandey, P.C., Singh, B.K., Katiyar, S., Mandal, V.P., Rani, M., Tomar, V., Patairiya, S., 2016. Estimation of accumulated soil organic carbon stock in tropical forest using geospatial strategy. Egypt. J. Remote Sens. Space. Sci. 19, 109–123. https://doi.org/10.1016/j.ejrs.2015.12.003.
Kunkel, K.E., Stevens, L.E., Stevens, S.E., Sun, L., Janssen, E., Wuebbles, D., Dobson, J.G., 2013. Regional Climate Trends and Scenarios for the U.S. National Climate Assessment. Part 9. Climate of the Contiguous United States (No. NOAA Technical Report NESDIS 142-9). National Oceanic and Atmospheric Administration (NOAA).
LANDFIRE, 2010. LANDFIRE 1.2.0. U.S. Department of the Interior, Geological Surveyhttp://landfire.cr.usgs.gov/viewer/, Accessed date: January 2016.
Lemma, B., Nilsson, I., Kleja, D.B., Olsson, M., Knicker, H., 2007. Decomposition and substrate quality of leaf litters and fine roots from three exotic plantations and a native forest in the southwestern highlands of Ethiopia. Soil Biol. Biochem. 39, 2317–2328. https://doi.org/10.1016/j.soilbio.2007.03.032.
Li, Z.P., Han, F.X., Su, Y., Zhang, T.L., Sun, B., Monts, D.L., Plodinec, M.J., 2007. Assessment of soil organic and carbonate carbon storage in China. Geoderma 138, 119–126. https://doi.org/10.1016/j.geoderma.2006.11.007.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
Liu, X., Xiong, Y., Liao, B., 2017. Relative contributions of leaf litter and fine roots to soil organic matter accumulation in mangrove forests. Plant Soil 421, 493–503. https://doi.org/10.1007/s11104-017-3477-5.
Martin, M.P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., Arrouays, D., 2010. Spatial distribution of soil organic carbon stocks in France. Biogeosci. Discuss. 7, 8409–8443.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.
McGarvey, J.C., Thompson, J.R., Epstein, H.E., Shugart, H.H., 2015. Carbon storage in old-growth forests of the Mid-Atlantic: toward better understanding the eastern forest carbon sink. Ecology 96, 311–317.
McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94. https://doi.org/10.1016/S0016-7061(98)00137-2.
Meinshausen, N., 2017. Quantile regression forests. J. Mach. Learn. 17.
Moore, T.R., Trofymow, J.A., Prescott, C.E., Titus, B.D., 2011. Nature and nurture in the dynamics of C, N and P during litter decomposition in Canadian forests. Plant Soil 339, 163–175. https://doi.org/10.1007/s11104-010-0563-3.
Nave, L.E., Domke, G.M., Hofmeister, K.L., Mishra, U., Perry, C.H., Walters, B.F., Swanston, C.W., 2018. Reforestation can sequester two petagrams of carbon in US topsoils in a century. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1719685115 (201719685).
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226. https://doi.org/10.1016/0016-7061(95)00007-B.
O'Neill, K.P., Amacher, M.C., Perry, C.H., 2005. Soils as an Indicator of Forest Health: A Guide to the Collection, Analysis, and Interpretation of Soil Indicator Data in the Forest Inventory and Analysis Program. https://doi.org/10.2737/NC-GTR-258.
Oswalt, S.N., Brad Smith, W., Miles, D.P., Pugh, A.S., 2014. Forest Resources of the United States, 2012: a technical document supporting the Forest Service 2010 update of the RPA Assessment. Gen Tech Rep WO-91. US Dep. Agric. For. Serv. Wash. Off., Wash. DC https://doi.org/10.2737/WO-GTR-91 (218 P 91).
Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Science https://doi.org/10.1126/science.1201609.
Phachomphon, K., Dlamini, P., Chaplot, V., 2010. Estimating carbon stocks at a regional level using soil information and easily accessible auxiliary variables. Geoderma 155, 372–380. https://doi.org/10.1016/j.geoderma.2009.12.020.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199. https://doi.org/10.1007/s10021-005-0054-1.
Prescott, C.E., 2005. Do rates of litter decomposition tell us anything we really need to know? For. Ecol. Manag. 220, 66–74. https://doi.org/10.1016/j.foreco.2005.08.005.
Prescott, C.E., Vesterdal, L., Preston, C.M., Simard, S.W., 2004. Influence of initial chemistry on decomposition of foliar litter in contrasting forest types in British Columbia. Can. J. For. Res. 34, 1714–1729. https://doi.org/10.1139/x04-040.
Qian, Y., Zhou, W., Yan, J., Li, W., Han, L., 2014. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sens. 7, 153–168. https://doi.org/10.3390/rs70100153.
R Development Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org/.
Rasmussen, C., Southard, R.J., Horwath, W.R., 2008. Litter type and soil minerals control temperate forest soil carbon response to climate change. Glob. Chang. Biol. 14, 2064–2080. https://doi.org/10.1111/j.1365-2486.2008.01639.x.
Rayamajhi, M.B., Pratt, P.D., Center, T.D., Van, T.K., 2010. Exotic tree leaf litter accumulation and mass loss dynamics compared with two sympatric native species in South Florida, USA. Eur. J. For. Res. 129, 1155–1168. https://doi.org/10.1007/s10342-010-0404-1.
Robinson, A., 2016. equivalence: Provides Tests and Graphics for Assessing Tests of Equivalence. R Package Version 0.7.2.
Ross, C.W., Prihodko, L., Anchang, J., Kumar, S., Ji, W., Hanan, N.P., 2018. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. Sci. Data 5, 180091.
Rudiyanto, Minasny, B., Setiawan, B.I., Arif, C., Saptomo, S.K., Chadirin, Y., 2016. Digital mapping for cost-effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. Geoderma 272, 20–31. https://doi.org/10.1016/j.geoderma.2016.02.026.
Sanderman, J., Hengl, T., Fiske, G.J., 2017. Soil carbon debt of 12,000 years of human land use. Proc. Natl. Acad. Sci. U. S. A. 114, 9575–9580. https://doi.org/10.1073/pnas.1706103114.
Schulp, C.J.E., Verburg, P.H., 2009. Effect of land use history and site factors on spatial variation of soil organic carbon across a physiographic region. Agric. Ecosyst. Environ. 133, 86–97. https://doi.org/10.1016/j.agee.2009.05.005.
Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J., Haddix, M.L., 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. Geoderma 132, 471–489. https://doi.org/10.1016/j.geoderma.2005.07.001.
Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. Soil Sci. Soc. Am. J. 81, 1413. https://doi.org/10.2136/sssaj2016.11.0376.
Suchenwirth, L., Stümer, W., Schmidt, T., Förster, M., Kleinschmit, B., 2014. Large-scale mapping of carbon stocks in riparian forests with self-organizing maps and the k-nearest-neighbor algorithm. Forests 5, 1635–1652. https://doi.org/10.3390/f5071635.
Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Model. 43, 1947–1958. https://doi.org/10.1021/ci034160g.
Upton, A., Vane, C.H., Girkin, N., Turner, B.L., Sjögersten, S., 2018. Does litter input determine carbon storage and peat organic chemistry in tropical peatlands? Geoderma 326, 76–87. https://doi.org/10.1016/j.geoderma.2018.03.030.
US Environmental Protection Agency (US EPA), 2015. Forest sections of the land use, land use change, and forestry chapter, and annex. In: US Environmental Protection Agency (Ed.), Inventory of US Greenhouse Gas Emissions and Sinks: 1990–2013 http://www.epa.gov/climatechange/ghgemissions/usinventoryreport.html, Accessed date: 4 March 2016.
US Environmental Protection Agency (US EPA), 2018. Forest sections of land use, land-use change, and forestry chapter, and annex. In: Environmental Protection Agency, US (Ed.), Inventory of US Greenhouse Gas Emission and Sinks: 1990-2016 https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2016, Accessed date: 12 April 2018.
USDA Forest Service, 2011. Phase 3 field guide — soil measurements and sampling (version 5.1). http://www.fia.fs.fed.us/library/field-guides-methods-proc/docs/2012/field_guide_p3_5-1_sec22_10_2011.pdf, Accessed date: 28 October 2015.
USDA Forest Service, 2014. The Forest Inventory and Analysis Database: database description and user guide for phase 2 (version 6.0.1). http://www.fia.fs.fed.us/library/database-documentation/current/ver6.0/FIADB%20User%20Guide%20P2_6-0-1_final.pdf, Accessed date: 28 October 2015.
Vasques, G.M., Grunwald, S., Myers, D.B., 2012. Associations between soil carbon and ecological landscape variables at escalating spatial scales in Florida, USA. Landsc. Ecol. 27, 355–367. https://doi.org/10.1007/s10980-011-9702-3.
Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64. https://doi.org/10.1016/j.geoderma.2016.12.017.
Viscarra Rossel, R.A., McGlynn, R.N., McBratney, A.B., 2006. Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. Geoderma 137, 70–82. https://doi.org/10.1016/j.geoderma.2006.07.004.
Wang, J., Sun, J., Xia, J., He, N., Li, M., Niu, S., 2018. Soil and vegetation carbon turnover times from tropical to boreal forests. Funct. Ecol. 32, 71–82. https://doi.org/10.1111/1365-2435.12914.
Wellek, S., 2003. Testing Statistical Hypotheses of Equivalence. CRC Press URL. https://www.crcpress.com/Testing-Statistical-Hypotheses-of-Equivalence-and-Noninferiority-Second/Wellek/p/book/9781439808184, Accessed date: 13 July 2018 (WWW Document).
Wills, S.A., Loecke, T., Sequeira, C., Teachman, G., Grunwald, S., West, L.T., 2014. Overview of the U.S. Rapid Carbon Assessment Project: sampling design, initial summary and uncertainty estimates. Soil Carbon. Springer International Publishing.
Woodall, C.W., Heath, L., Domke, G., Nichols, M., 2011. Methods and Equations for Estimating Aboveground Volume, Biomass, and Carbon for Trees in the U.S. Forest Inventory, 2010.

Xiong, X., Grunwald, S., Myers, D.B., Kim, J., Harris, W.G., Comerford, N.B., 2014. Holistic environmental soil-landscape modeling of soil organic carbon. Environ. Model. Softw. 57, 202–215. https://doi.org/10.1016/j.envsoft.2014.03.004.
Zhang, C., Tang, Y., Xu, X., Kiely, G., 2011. Towards spatial geochemical modelling: use of geographically weighted regression for mapping soil organic carbon contents in Ireland. Appl. Geochem. 26, 1239–1248. https://doi.org/10.1016/j.apgeochem.2011.04.014.
Zhu, J., Hu, H., Tao, S., Chi, X., Li, P., Jiang, L., Ji, C., Zhu, J., Tang, Z., Pan, Y., Birdsey, R.A., He, X., Fang, J., 2017. Carbon stocks and changes of dead organic matter in China's forests. Nat. Commun. 8, 151. https://doi.org/10.1038/s41467-017-00207-1.