Environmental Research Letters

LETTER

Estimating aboveground live understory vegetation carbon in the United States

Kristofer D Johnson[1,5] , Grant M Domke[2], Matthew B Russell[3], Brian Walters[2], John Hom[1], Alicia Peduzzi[1], Richard Birdsey[1], Katelyn Dolan[4] and Wenli Huang[4]

[1] US Forest Service, Northern Research Station, Newtown Square, PA, United States of America
[2] US Forest Service, Northern Research Station, St Paul, MN, United States of America
[3] Department of Forest Resources, University of Minnesota, St. Paul, MN, United States of America
[4] Department of Geographical Sciences, University of Maryland, College Park, MD, United States of America
[5] Author to whom any correspondence should be addressed.

E-mail: kdj888@hotmail.com

Abstract

Despite the key role that understory vegetation plays in ecosystems and the terrestrial carbon cycle, it is often overlooked and has few quantitative measurements, especially at national scales. To understand the contribution of understory carbon to the United States (US) carbon budget, we developed an approach that relies on field measurements of understory vegetation cover and height on US Department of Agriculture Forest Service, Forest Inventory and Analysis (FIA) subplots. Allometric models were developed to estimate aboveground understory carbon. A spatial model based on stand characteristics and remotely sensed data was also applied to estimate understory carbon on all FIA plots. We found that most understory carbon was comprised of woody shrub species (64%), followed by nonwoody forbs and graminoid species (35%) and seedlings (1%). The largest estimates were found in temperate or warm humid locations such as the Pacific Northwest and southeastern US, thus following the same broad trend as aboveground tree biomass. The average understory aboveground carbon density was estimated to be 0.977 Mg ha^{-1}, for a total estimate of 272 Tg carbon across all managed forest land in the US (approximately 2% of the total aboveground live tree carbon pool). This estimate is more than twice as low as previous FIA modeled estimates that did not rely on understory measurements, suggesting that this pool may currently be overestimated in US National Greenhouse Gas reporting.

Introduction

Understory vegetation (UVEG) plays a key role in ecosystem function and the terrestrial carbon cycle (Hou *et al* 2015, Saitoh *et al* 2014). The composition of shrubs, forbs, graminoids and seedlings interact with and influence plant diversity, forest productivity and nutrient cycling (Yarie 1980, Mallik 2003, Gilliam 2007, Moore *et al* 2007), and wildlife diversity and abundance (Pardini *et al* 2005, Russell *et al* 2017). Despite the importance of UVEG, relatively few measurements of UVEG characteristics are available compared to the overstory. The majority of studies that quantify UVEG attributes are limited to specific forest types and/or are focused on a particular plant functional group such

as shrubs or herbs. Additionally, UVEG abundance is highly variable as it can be dramatically influenced by wildlife grazing, weather, and fire disturbance (Hart and Chen 2006, Nilsson and Wardle 2005). Therefore, quantifying UVEG abundance and composition remains problematic, especially at regional or national scales (Suchar and Crookston 2010, Russell *et al* 2014).

UVEG is also a component of the terrestrial carbon budget, albeit a relatively minor one, and currently reported as 4.7% of total aboveground biomass in the United States (US) (Smith *et al* 2013). The UVEG component refers to all live understory shrubs, forbs, graminoids, and seedlings. It does not include dead components of the understory such as woody debris and litter layer, which contains a much larger

proportion of forest carbon (Hudak *et al* 2012). In the US, the Department of Agriculture Forest Service's Forest Inventory and Analysis Program (FIA) is responsible for compiling estimates of forest carbon stocks and stock changes, including UVEG carbon stocks (UVEGC), for national and international reporting instruments. UVEG-related measurements, such as percent cover and height of shrubs, forbs, graminoids, and seedlings have been collected in many FIA plots since 2001. However, the current national estimates do not depend on these plot measurements, rather they are calculated by models based on live tree density and forest type (Birdsey 1996, Smith *et al* 2013). To date, there has never been a comprehensive investigation of the appropriateness of the existing approach to quantifying national stocks of UVEGC. Additionally, the reason existing UVEG characteristics have not been applied for quantifying UVEGC may be due to a lack of understory allometric models. In general, very few of such models have been published (Chojnacky and Milton 2008).

Remotely sensed data may be useful proxies of factors that control the spatial distribution of UVEG (Martinuzzi *et al* 2009, Wing *et al* 2012). Relationships between environmental variables and UVEG are well-documented at stand to regional scales (Suchar and Crookston 2010, Tuanmu *et al* 2010). Some of the major factors influencing UVEG include light, water, nutrients, natural disturbances, and management practices (Royo and Carson 2006, Barbier *et al* 2008). Ideally, data from sensors that directly measure tree canopy density (LiDAR) as a proxy for light availability, could be used to correlate to UVEGC. However, in the absence of this information, it is possible that broad patterns of reflectance data (AVHRR/MODIS) as a proxy for productivity may also correlate with UVEG spatial patterns.

The main purpose of this study is to quantify UVEGC in the conterminous US, and more specifically to: 1) estimate the UVEGC pool at the national level using stand and geospatial data, and 2) compare the results to previous US estimates. We develop allometric models that can make use of UVEG measurements collected on FIA plots. We also validate our national approach with independent UVEGC estimates. The mixed probabilistic and modeled approach applied here at a large scale provides information for UVEGC comparisons in other countries and lays out a method that other countries may find useful in reporting their own UVEGC estimates.

Methods

Digital photo series database
We use data from the Digital Photo Series database to develop a suite of allometric models to estimate UVEGC for dominant forest types (Ottmar *et al* 2004, Wright *et al* 2010). The Digital Photo Series Database

is a collection of UVEG data from field sites located in major forest types throughout the US (Wright *et al* 2010). Although the data are principally used for fire fuel modeling and monitoring, the current study takes advantage of plots that were clipped and weighed to measure UVEG biomass, including shrubs, nonwoody material (e.g. forbs and graminoids), and seedlings. Additionally, percent cover, height measurements, and seedling densities were collected at most plots allowing for the development of allometric models for predicting UVEG biomass. The biomass value was multiplied by 0.5 to estimate carbon in each respective UVEG pool. The Digital Photo Series Database is maintained by the USDA Forest Service Fire and Environmental Research Applications Team in Seattle, Washington (http://depts.washington.edu/nwfire/dps/).

FIA data
The FIA program employs a multi-phase inventory, with each phase contributing to the subsequent phase. First, current aerial photography (e.g. National Agriculture Imagery Program, USDA Farm Services Agency (2008)) is used in a prefield process to determine the land use (e.g. forest or cropland) at all sampling points (i.e. plot locations). Next, each sample point is assigned to a stratum using imagery or thematic products (e.g. National Land Cover Database, Homer *et al* 2012). A stratum is a defined geographic area (e.g. state or estimation unit) that includes plots with similar attributes; in many regions, strata are defined by predicted percent canopy cover. Permanent ground plots are distributed approximately every 2428 ha across the 48 conterminous states. Each permanent ground plot comprises a series of smaller fixed-radius (7.32 m) plots (i.e. subplots) spaced 36.6 m apart in a triangular arrangement with one subplot in the center. Tree- and site-level attributes—such as diameter at breast height (dbh) and tree height—are measured at regular temporal intervals on plots that have at least one forested condition defined in the prefield process (USDA Forest Service 2016). On a subset of the base intensity plots distributed approximately every 38 848 ha additional forest ecosystem attributes are measured including UVEG characteristics. The UVEG, or 'Vegplot' measurements include species type and percent cover of shrub and nonwoody species for height layers of 0–0.61 m, 0.62–1.83 m, 1.84–4.88 m, and >4.88 m (Schulz *et al* 2009). However, no actual heights were collected in the Vegplots. Only a subset of intensive plots had Vegplot measurements (24 392 subplots, or 5% of the total), and those that did were mainly concentrated in the South Central US and West Coast. The years sampled of the Vegplot data spans from 2001 to 2011, with the majority sampled between 2007 and 2010.

To fill the spatial gaps in Vegplot data we combined them with similar 'Microplot' data from the FIA program collected throughout the US This dataset not only contains percent cover of shrub and herbaceous

species, but also their heights (Woodall and Monleon 2008). Similar to Vegplot data, the measurements were not collected for all intensive plots, rather a subset (45 700 subplots, or 15% of the total). Importantly, the area of Microplots is much smaller ($13.5\,\mathrm{m^{-2}}$) than the Vegplots, which cover the whole subplot ($168\,\mathrm{m^{-2}}$). The Microplots were sampled from 2002 to 2013 with the majority sampled between 2005 and 2010.

Microplot and Vegplot data that were collected in the same location and in the same year were used to develop models for harmonizing the two datasets. Scatterplots of percent cover observed on the Microplots and Vegplots indicated a consistent bias in the Microplot measurements. Microplot percent cover tended to be lower than Vegplot percent cover for small values, but higher for larger values. This may be due to the differences in perceptions of field crews in determining shrub and nonwoody covers over the smaller Microplots versus the larger Vegplots. In any case, a polynomial model was used to adjust the Microplot percent cover to be more similar to the Vegplot. Additionally, due to the fact that there are no height measurements recorded on the Vegplots (only height classes), height was predicted from percent cover from models developed from the Microplot data (figure 1). The new allometric models developed from the Digital Photo Series database were then used to estimate UVEGC on each FIA subplots where measurements of UVEG existed.

Information on seedlings (<2.54 cm or 1 inch diameter) is collected on all FIA plots, not only intensive plots. Actual seedling data are not used in the current calculation and reporting of UVEGC, though it is considered part of this pool (Smith *et al* 2013). In the current approach, we used counts of seedlings collected in the Microplots to calculate seedling density, which could then be related to seedling carbon density based on models developed from the Digital Photo Series Database (figure 1).

FIA stand metrics at the subplot level were also calculated and used as variables for subsequent random forest modeling. These included: forest type, stand age (years), slope (%), mean and maximum tree height (ft), basal area ($\mathrm{m^2\,ha^{-1}}$), tree C ($\mathrm{Mg\,ha^{-1}}$), mean diameter (cm), trees per hectare, and latitude and longitude. Although the variables are in imperial units, all results were converted to common metric units. The majority of stand characteristics were collected from 2005 to 2010, similar to the UVEG measurements, and they most closely approximate the year 2007.

Spatial data
Several spatial datasets were used in the random forest modeling of UVEGC. Data from the NOAA Advanced Very High Resolution Radiometer (AVHRR) sensor provide phenological metrics that indicate plant distribution and relative greenness at 1 km resolution. The current study uses a 5 year average (2006–2010) of start of season normalized difference vegetation index (NDVI) (SOSN), end of season NDVI (EOSN), duration (DUR), maximum NDVI (MAXN), and time integrated NDVI (TOTND). The 2006–2010 period was used because it coincides with the period that most of the UVEG data was collected by the FIA program (i.e. from 2005–2010). These data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at USGS/EROS, Sioux Falls, South Dakota (http://lpdaac.usgs.gov). Mean annual temperature and annual precipitation averaged over the same period was obtained from DAYMET (Thornton *et al* 2016). Elevation was obtained from the Shuttle Radar Topography Mission (Jarvis *et al* 2008). All datasets were aligned and converted to the same projection.

Random forest modeling
We used the subset of FIA plots having UVEG measurements in combination with geospatial data to develop a random forest model of UVEGC in the US. Random forest modeling is a machine learning approach that predicts observations using both categorical and continuous data (Liaw and Wiener 2002), and has been applied to predict litter layer carbon on FIA plots (Domke *et al* 2016). In this study, random forest modeling was applied to predict total UVEGC at the national scale (shrub, nonwoody, and seedling) using: (1) field collected variables, (2) geospatial variables, and (3) both field and geospatial variables. The best model approach was determined by comparing the different groups of variables against random hold-out observations of 15% of the total dataset. The model was then applied to all Phase 3 subplots across the country (130 547 plots, or 438 089 subplots), thus preserving a spatially unbiased and expanded estimate. Variable importance was assessed by percent increase in MSE of each variable after 100 permutations (%incMSE). The R package randomForest was used for all analyses (Liaw and Wiener 2002, R Development Core Team 2014).

Model validation

Estimates of UVEGC at the national scale were validated by searching the literature for studies that reported actual field measurements of UVEG over landscapes and regions. Relatively few data sources are available on understory biomass stocks to allow for a robust and spatially unbiased independent validation of either previous FIA estimates or the estimates of this study. Nonetheless, some studies have included data on clipped and weighed UVEG biomass for multiple plots at a site. For example, the Database for Landscape-scale Carbon Monitoring Sites contains eight sites with UVEGC sampled systematically over a 9 km^2 area (Cole *et al* 2013). Other studies have developed allometric models from clipped and weighed individuals and applied them over larger areas (e.g.
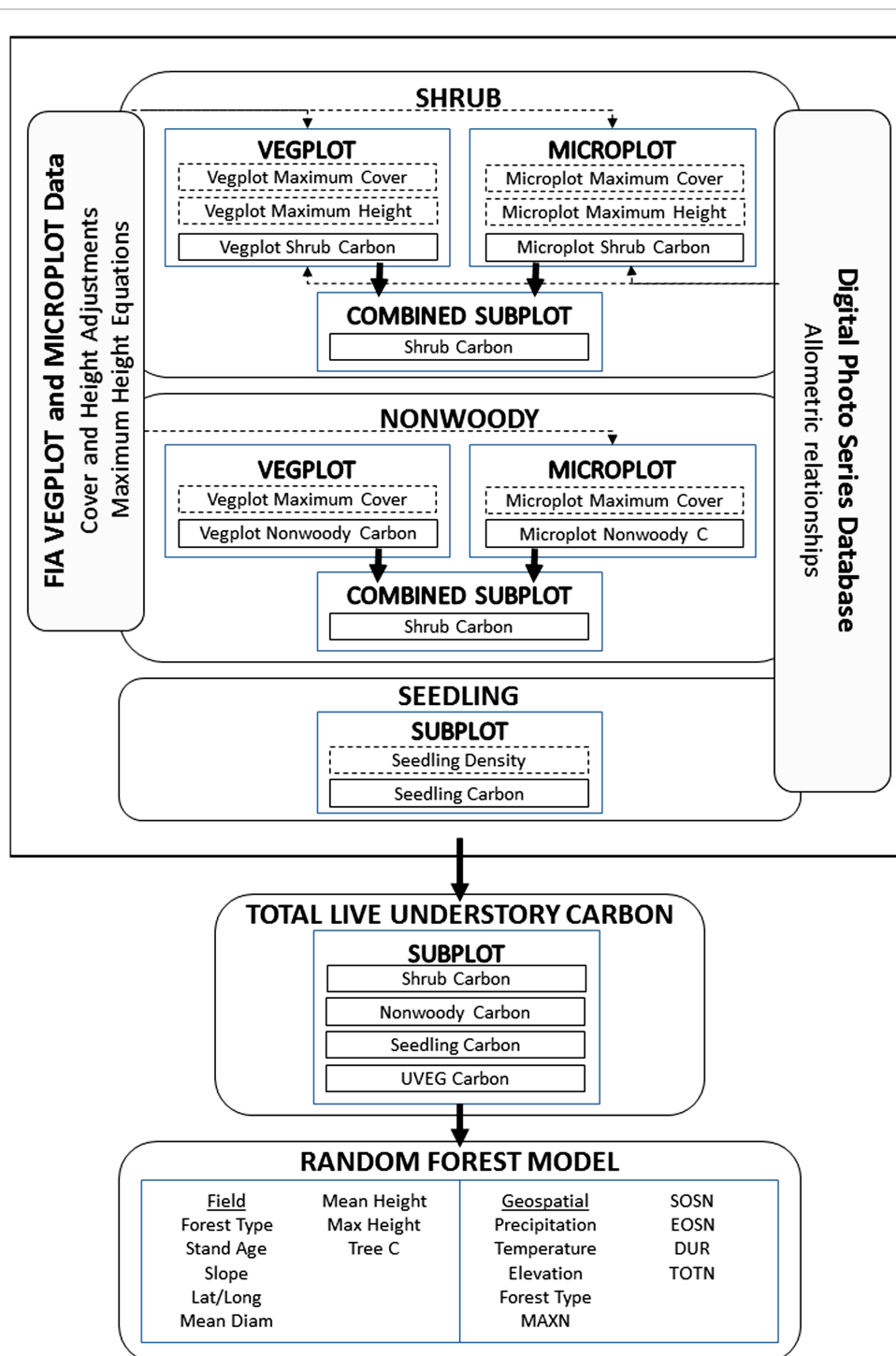
Figure 1. Flow diagram for the estimation of UVEGC (i.e. aboveground shrub, nonwoody, and seedling carbon) on FIA plots. Dotted line boxes represent the measurements used with allometric models to estimate specific UVEGC pools (solid line boxes). Shaded boxes to the side indicate the data sources and their roles in the UVEGC estimation.

Smithwick *et al* 2002). These datasets were used to validate our best random forest model and current FIA estimates. Therefore it was a separate validation activity than the hold-out dataset used above for random forest model validation, and in this case emphasized comparing the performance of the previous FIA model or the new random forest model. Overall, we compared UVEGC estimates from 14 sites, accounting for 10 of the 32 forest types, to both UVEGC estimates (table 1). Validation sites were only included in our analysis if an

adequate UVEG sample was available, ranging from 7 to 96 observations. Basic *p*-value and R^2 statistics were performed to examine correlations.

Results

Allometric models
Of the several linear and non-linear models that were explored to fit shrub height and shrub cover to

Table 1. Validation sites chosen for comparing random forest and current FIA estimates of UVEGC.

Forest type	n	Location	Reference
Fir/spruce/mountain hemlock	96	Niwot Ridge and Fraser Experimental Forest, Colorado	(Bradford *et al* 2008, Cole *et al* 2013)
Fir/spruce/mountain hemlock	48	Glacier Lakes Ecosystem Experiments Site, Wyoming	(Bradford *et al* 2008, Cole *et al* 2013)
Aspen/birch	63	Marcell Experimental Forest, Minnesota	(Bradford *et al* 2009, Cole *et al* 2013)
Maple/beech/birch	17	Marcell Experimental Forest, Minnesota	(Bradford *et al* 2009, Cole *et al* 2013)
Spruce/fir	9	Marcell Experimental Forest, Minnesota	(Bradford *et al* 2009, Cole *et al* 2013)
Longleaf/slash pine	16	Northern Florida	(Cronan *et al* 2015)
Maple/beech/birch	10	Quebec, Canada (similar to northeastern US)	(Tremblay and Larocque 2001)
Lodgepole pine	16	Yellowstone National Park, Wyoming	(Smithwick *et al* 2005)
Loblolly/shortleaf pine	25	Silas Little and Cedar Bridge, New Jersey	(Clark *et al* 2010, Cole *et al* 2013)
Loblolly/shortleaf pine	19	Parker Tract, North Carolina	(Cole *et al* 2013)
Douglas-fir	14	H J Andrews, Oregon	(Smithwick *et al* 2002)
Hemlock/sitka spruce	7	Olympic Peninsula, Washington	(Smithwick *et al* 2002)
Fir/spruce/mountain hemlock	9	Mt. Rainier National Park, Washington	(Smithwick *et al* 2002)
Oak/hickory	12	Butler County, Missouri	(Knapp *et al* 2015)

Table 2. Allometric models for shrub, nonwoody, and seedling biomass.

Forest type	b1	b2	R2	RMSE	n
*Shrub biomass, Mg ha^{-1} = (b1 * (MaxCover * MaxHT))/(b2 + (MaxCover * MaxHT))*					
AK Aspen/birch group	9457705.1	650148.5	0.26	2.4	15
AK Fir/spruce/mountain hemlock group		use PWW Douglas-fir group			
Loblolly/shortleaf pine group	39.0	1.8	0.38	2.3	37
NE Maple/beech/birch group	12839404.7	2049530.3	0.88	0.0	9
NE Spruce/fir group	2688.8	683.8	0.32	0.0	10
NLS White/red/jack pine group	7.7	0.2	0.75	0.3	21
Pinyon/juniper group	59.2	1.1	0.65	1.5	37
Ponderosa pine group	413718721.3	11202103.1	0.73	5.6	23
PWW Douglas-fir group	11.4	0.6	0.68	0.4	19
RMN, RMS Aspen/birch group	84.3	2.3	0.52	0.7	13
RMN, RMS Lodgepole pine group	19.4	0.6	0.90	0.2	13
RMN, RMS Woodland hardwoods group	6.5	0.1	0.34	0.8	9
SC, SE Longleaf/slash pine group	13.7	0.2	0.65	1.0	11
SC, SE Oak/hickory group	3300.4	683.8	0.99	0.00	11
Western oak group	14.1	0.4	0.65	0.7	20
*Nonwoody biomass, Mg ha^{-1} = (b1 * Cover)/(b2 + Cover)*					
All UVEG forest types	2646.4	90.2	0.61	0.7	318
*Seedling biomass, Mg ha^{-1} = b1 + b2 * SeedDense*					
AK Aspen/birch group		use RMN, RMS Woodland hardwoods group			
AK Fir/spruce/mountain hemlock group	0	6.51 E–05	0.27	0.8	26
Loblolly/shortleaf pine group	0	5.74 E–05	0.87	0.6	16
NE Maple/beech/birch group	0	1.05 E–06	0.95	0.0	9
NE Spruce/fir group	0	8.10 E–06	0.94	0.2	10
NLS White/red/jack pine group		use NE Spruce/fir group			
Pinyon/juniper group	0	1.02 E–04	0.49	0.1	14
Ponderosa pine group		use Pinyon/juniper group			
PWW_Douglas-fir group		use AK Fir/spruce/mountain hemlock group			
RMN, RMS Aspen/birch group		use RMN, RMS Woodland hardwoods group			
RMN, RMS Lodgepole pine group		use Pinyon/juniper group			
RMN, RMS Woodland hardwoods group	0	4.19 E–05	0.53	1.0	9
SC, SE Longleaf/slash pine group		use Loblolly/shortleaf pine group			
SC, SE Oak/hickory group		use NE Maple/beech/birch group			
Western oak group		use RMN, RMS Woodland hardwoods group			

AK–Alaska, NE–Northeast, NLS–Northern Lake States, PWW–Pacific Northwest Westside, RMN–Rocky Mountain North, RMS–Rocky Mountain South, SC–South Central, SE–Southeastern.
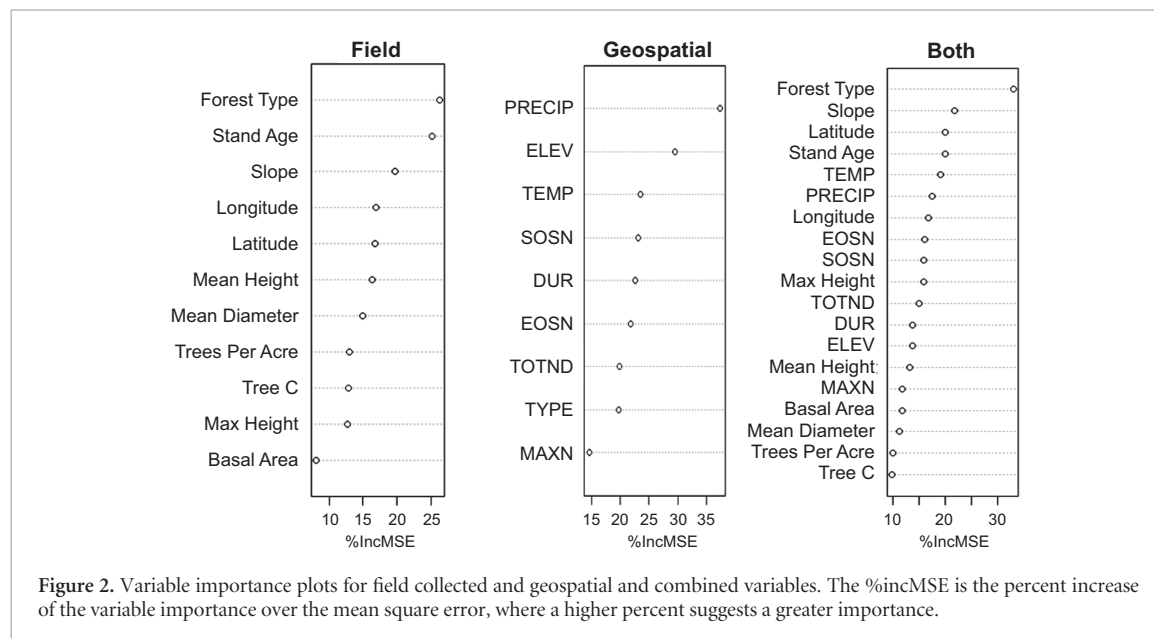
MaxCover and Cover units are %; MaxHT units are meters; SeedDense units are stems per hectare.

aboveground shrub biomass, the Michaelis Menten form performed the most consistently (table 2, supplementary table S1, available at stacks.iop.org/ERL/12/125010/mmedia). Semi-arid forest types (ponderosa, pinyon/juniper, and aspen/birch) increased more in shrub biomass with shrub cover and height compared to other forest types, suggesting a higher density. Other conifer types (douglas fir and white/red/jack pine) had the lowest densities. For the nonwoody component, only one Michaelis Menten model using nonwoody cover was used for all forest types. Finally,

for the seedling component, a linear model using seedling density was used for six forest types (table 2, supplementary table S1). The allometric relationships developed for the shrub, nonwoody, and seedling components allowed for the estimation of total UVEGC for 15 of the 32 forest types and 76% of the intensive plot observations. For the remaining forest types not directly represented by allometric models, close substitutions were made (e.g. douglas fir was used for redwood). Each of the allometric equations are recorded both in metric (table 2) and imperial units (supplementary table S1).

Table 3. Random forest results for national scale analysis. The regression estimate (i.e. slope, indicating bias), R2, and RMSE of each random forest model against a hold-out dataset of 2805 observations. Units are in Mg ha^{-1}.

RF model	Estimate	RSquare	RMSE
Field	1.13	0.34	0.76
Remote sensing	0.81	0.34	0.77
Both	1.06	0.40	0.73

Figure 2. Variable importance plots for field collected and geospatial and combined variables. The %incMSE is the percent increase of the variable importance over the mean square error, where a higher percent suggests a greater importance.

UVEGC in the conterminous US

Random forest models based on field collected and geospatial variables each explained 34% of the variation in UVEGC of the hold-out validation dataset, and each had similar RMSEs (table 3). However, the field collected variables tended to slightly underestimate UVEGC while the remotely sensed variables tended to slightly overestimate UVEGC (table 3). When all variables were combined, 40% of UVEGC was explained and with less bias than either of the two sets of variables alone. The most important variable of the field collected variables was forest type followed by stand age and slope. The most important geospatial variables were precipitation and elevation (figure 2). When both sets of variables were combined, the most important variables were field collected variables, such as forest type, slope, and stand age, followed by temperature, and to a lesser degree precipitation, and NDVI variables.

Perhaps not surprisingly, there were similar trends across environmental and stand age gradients at the national scale to those that would be expected at the stand level. For example, UVEGC increased with increasing temperature and precipitation, and decreased with increasing elevation and stand age (p-value < 0.0001 in all cases). In contrast, variables such as trees per hectare, basal area, and tree carbon were much less important in the random forest model.

When the random forest model using all variables was applied to all intensive FIA plots, the lowest UVEGC was generally associated with high latitude and arid regions while higher UVEGC was generally associated with forests with greater precipitation and warmer temperatures (figure 3(*a*)). For example, the two forest types that contained the most UVEGC were loblolly/shortleaf pine (17%) and oak/hickory (16%). More specifically, of the major forest types, the highest UVEGC density estimates were found in the Pacific Northwest alder/maple (1.9 Mg ha^{-1}) and Southeast loblolly/shortleaf pine (1.9 Mg ha^{-1}) groups, while the lowest densities were found in the Northeast spruce/fir (0.52 Mg ha^{-1}) and South Central Woodland hardwoods (0.52 Mg ha^{-1}) groups (supplementary table S2). These patterns were broadly similar to patterns of the current FIA estimates (figure 3(*b*)). However, the random forest estimates were almost always lower—as much as 100% in some areas such as Texas (figure 3(*c*); supplementary table S2). For all forest types, the majority of UVEGC was stored in shrubs (64%), followed by nonwoody (35%) and seedling (1%).

The UVEGC estimates from the random forest model for the conterminous US ranged from 0.047–7.605 MgC ha^{-1}. The mean was 0.977 ± 0.0008 MgC ha^{-1} (mean $\pm$ standard error), or 1.7% of the total aboveground live tree carbon reported by Smith *et al* (2013) (56.5 MgC ha^{-1}). The previous UVEGC estimate reported by Smith *et al* (2013) was 2.8 MgC ha^{-1} or 4.7% of the total live aboveground carbon. Therefore, the previous estimates were nearly three times higher than the random forest estimate. When means were summarized by state and forest type and multiplied by their forest areas, the sum total UVEGC for the conterminous US was 272 ± 0.21 Tg (total
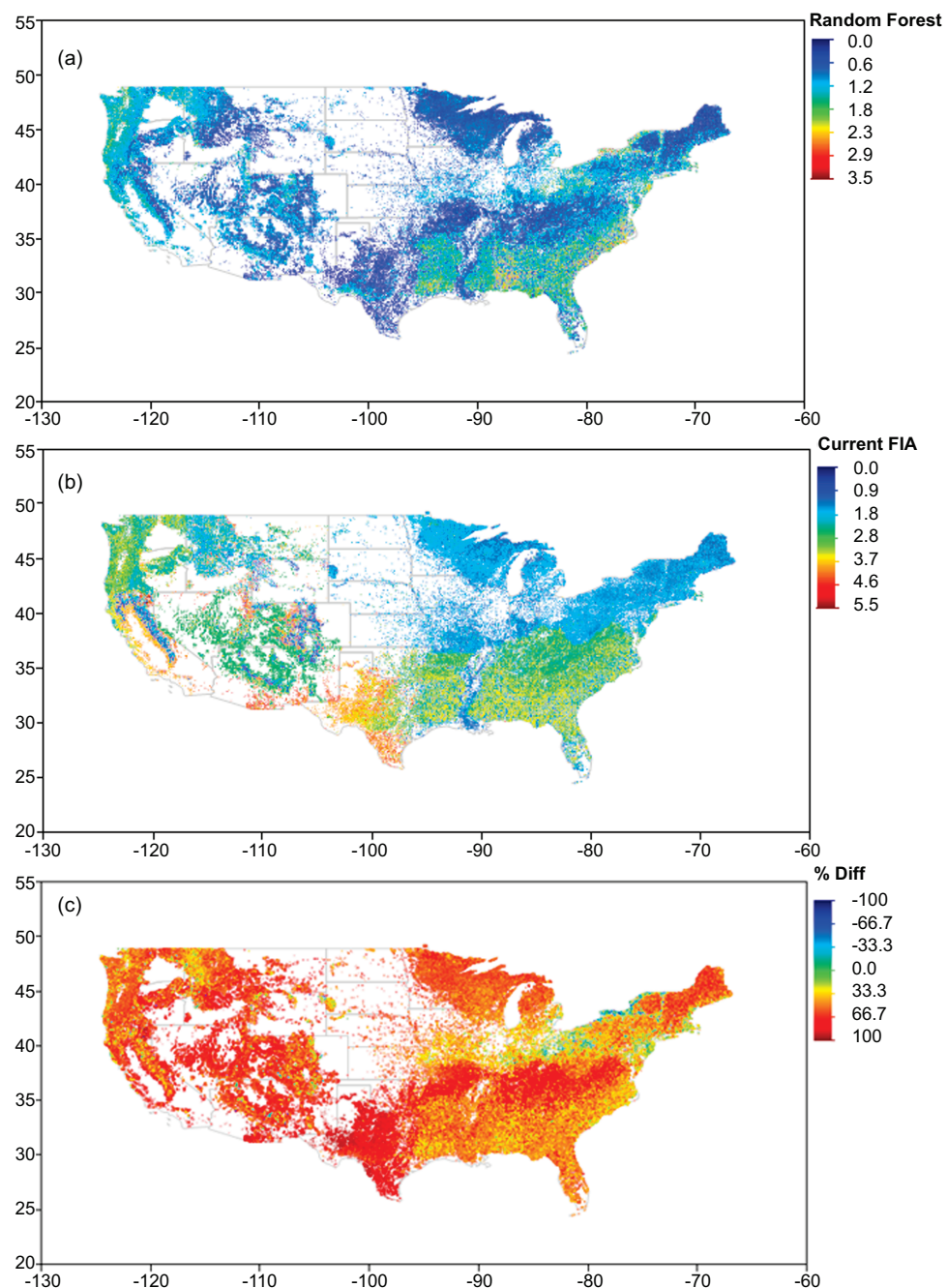
Figure 3. Map showing the P3 plot locations and spatial patterns of UVEGC, Mg ha^{-1} of (*a*) current FIA, (*b*) predicted by random forest models of both field collected and remotely sensed data, (*c*) the difference between the FIA and random forest estimates calculated as ((Current FIA—Random Forest)/Current FIA) ∗ 100.

stock ± standard error). The cumulative distribution plots for both estimates reveal that, in addition to being consistently higher, the current FIA estimate minimum was 0.5 Mg ha^{-1} (figure 4). In contrast, about 10% of the observations of the new estimate were below this threshold, making the overall average even lower. Both estimates were weakly, but significantly, positively correlated ($R2 = 0.12$; *p*-value < 0.0001).

Independent site validation

The Random Forest UVEGC estimates compared well to the independent estimates in nine of the ten forest types, but poorly for two of the three fir/spruce/mountain hemlock forest sites located in Colorado and Wyoming (figure 5). The UVEGC measurements at these sites were 2.5–4 times higher than the modeled value. However, the same forest type in Washington was comparable. Neither this study's, nor the previous study's estimate, were significantly correlated ($p < 0.05$) with all the validation sites. However, when the fir/spruce/mountain hemlock forest types were removed, both this study's and the previous estimate were positively correlated ($R2 = 0.72$ and 0.56, respectively, $p < 0.01$). The current FIA model estimate was always biased towards higher values (figures 4 and 5).
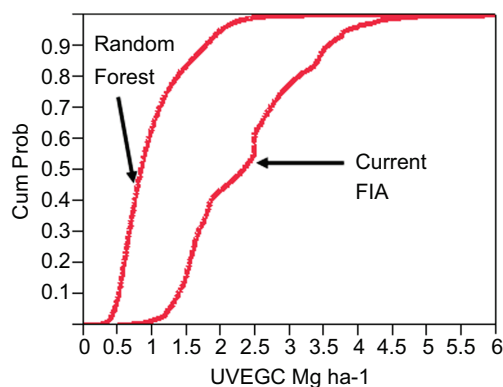
Figure 4. The cumulative distribution plots of UVEGC predictions of this study compared to previous FIA estimates.

Discussion

A major challenge for estimating UVEGC at the national level is that UVEGC measurements are only available for a limited number of forest types and are often measured using different methods. To address this challenge, we developed UVEGC allometric models from the Digital Photo Series Database, which provides a wealth of consistent measurements of shrub and nonwoody covers and heights, and seedling mass, for a variety of forest types across the country. Comparisons with independent validation data suggest that using UVEG measurements on FIA subplots appears to be an improvement over the previous method. Still, some forest types were not represented by our allometric models, and substantial uncertainty exists in the allometric models of certain forest types (table 2). Additional UVEG observations would provide greater understanding about the accuracy of the allometric models of this study.

Expanding UVEGC across all forest conditions was somewhat successful using the random forest models and resulted in an R2 of 0.40. Considering the high variability of UVEGC this is not surprising and is consistent with results from Suchar and Crookston (2010) who used a similar suite of variables to predict UVEGC in the Northwestern US. Their models yielded adj-R2's that ranged from 0.05 − 0.76. Perhaps more important for large scale estimates is that the model is not overly biased. It appears that combining both field and remote sensing variables for UVEGC prediction improves model predictions both in terms of the variation explained and the bias.

Based on these findings, the live aboveground UVEGC estimate currently reported by FIA for national greenhouse gas accounting may be high relative to field observations. This probably has to do with the relatively simple models that were first developed to describe this pool which were based on even fewer resources than are currently available. It also appears that the previous model did not capture as well conditions with very small amounts of UVEGC. In either case, however, it appears that UVEGC in the

conterminous US is still a minor component of the terrestrial ecosystem carbon budget, comprising between 2% and 5% of the total aboveground carbon. Furthermore, the live UVEGC component is much lower than the dead down woody debris component (six times lower than the national down woody debris estimate reported by Smith *et al* 2013). We also note that our assumed carbon content of 50%, which we used because of the lack of data about this ratio across all forest types, of the biomass may be too high. For example, Jain *et al* (2009) found that for forest types in the Rocky Mountains the carbon content was between 41 and 47% of the biomass. Assuming a value of 41% would further lower our estimate and underscore that the current FIA estimate may be overestimated (i.e. only 1.4% of aboveground biomass).

It was challenging to find estimates of forests outside the United States for comparing to our UVEGC, especially given the variety of definitions of what is the understory. Still, some comparisons are informative. For example, Fang *et al* (2007) estimated UVEGC in temperate forests in China that was lower than similar forests in the Northern Lakes States (0.14 vs. 0.8 MgC ha^{-1} in birch forests; 0.03 vs. 0.7 MgC ha^{-1} in oak forests). However, they did not count the nonwoody portion, which was about a third of the understory biomass in our estimate. Similarly, in a study by Ordóñez *et al* (2008) estimates of a fir forest in Mexico were somewhat lower than the Pacific Southwest fir forest of our study (0.39 vs. 0.67 MgC ha^{-1}). Finally, in a Himalayan maple forest the UVEGC estimate from Garkoti (2008) was higher than our estimate of maple/beech/birch forests in the Northeastern US (1.45 v. 0.75 MgC ha^{-1}).

Our total estimate of 272 Tg carbon in the live aboveground UVEG pool deserves some disclaimers. The small standard error reported is an artifact of the random forest model since it tends to predict mean understory densities and not extreme values. In other words, we did not apply model-assisted estimators of variance that more accurately reflect the true population variance, but future studies should address this. Furthermore, although our validation exercise
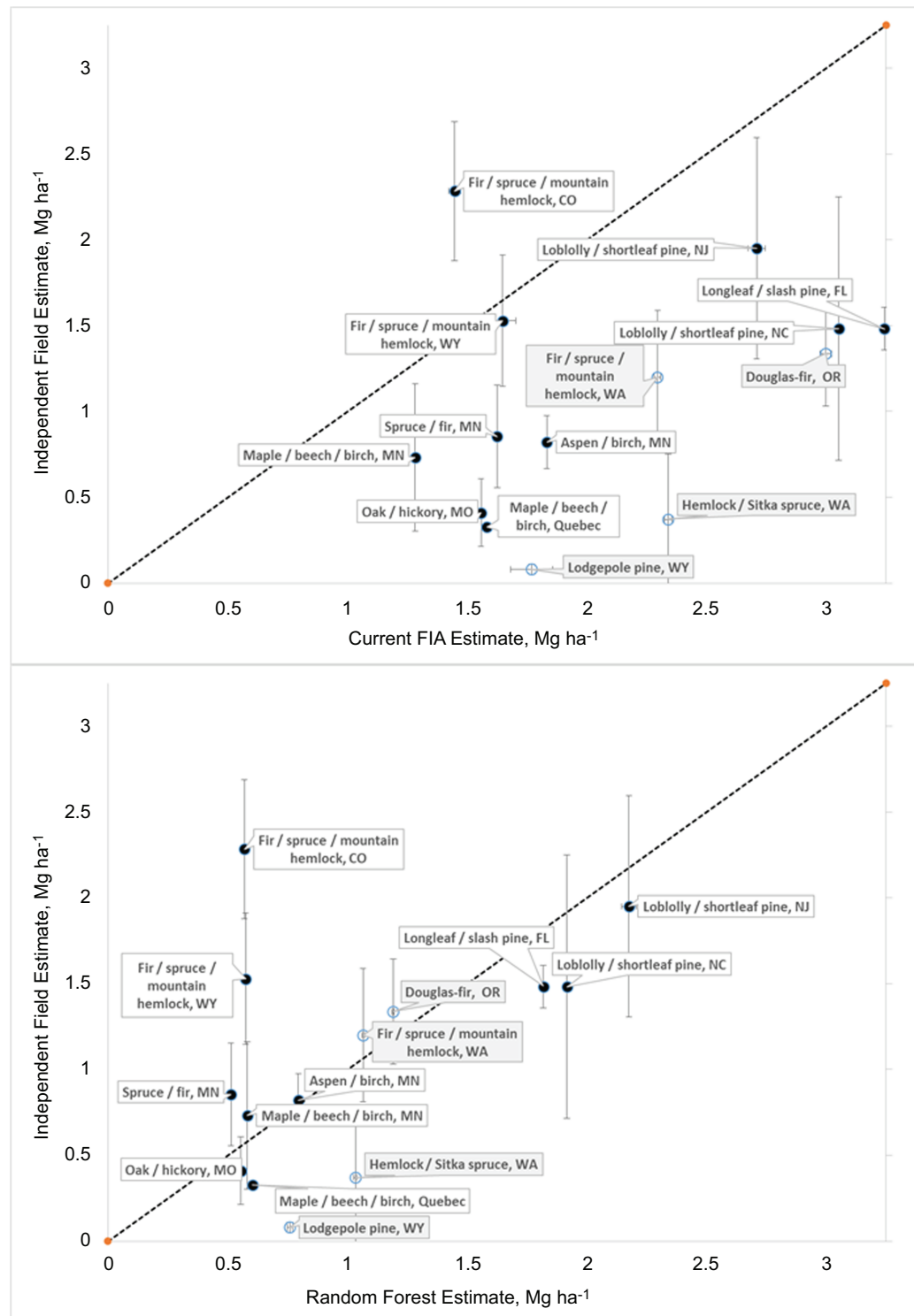
Figure 5. Validation sites for UVEGC estimates in MgC ha^{-1}. Top figure is the comparison with current FIA data and the bottom figure with results from the random forest model. The dotted black line is the 1:1 line. Closed symbols indicate sites with destructive harvest samples where as open symbols indicate where allometric models were applied to a larger population.

accounted for most of the vegetation types found in the US, the UVEGC of some forest types were not available for comparison (e.g. ponderosa pine, pinyon-juniper, redwood). Our analysis is only for forest conditions in the conterminous US We did not report UVEG stocks in Alaska, although we do provide allometric models for two forests types found in Southeastern Alaska. Therefore, it remains unknown how much

shrub biomass associated with Alaskan boreal forest and tundra vegetation contributes to the US carbon budget. Additionally, UVEGC stocks occur in non-forest areas such as urban landscapes and semi-arid regions, but they are unaccounted for here (Keeling and Phillips 2007). Finally, the allometric models and stocks reported are only for the aboveground portion of UVEG. Smith *et al* (2013) assumes that the

belowground portion is 11% of the aboveground, so this factor could be applied to our estimates. However, some have argued that the belowground contribution to understory biomass could be substantially greater, especially when the fine root portion is included (Gonzalez *et al* 2013). Despite these limitations the new estimate appears to give a more conservative estimation of this carbon pool than previous estimates.

Conclusions

The live UVEGC is an important part of forest ecosystem carbon stocks and is recognized as a component of the aboveground and belowground biomass pools in Intergovernmental Panel on Climate Change Good Practice Guidance (IPCC 2006). This study presents a parsimonious approach for predicting UVEGC which leverages observations from intensive FIA plots, the Digital Photo Series database, and other remotely sensed and climatic geospatial data. Three primary conclusions may be drawn from this study: (1) the live aboveground UVEGC pool is a small component of the terrestrial carbon budget in the US, (2) the new approach based on actual measurements of UVEG characteristics resulted in a lower UVEGC estimate than previously reported, and (3) new UVEG cover, height, and weight measurements in under-represented sites in this study will improve national level representation of this highly variable pool. These conclusions, in addition to the general approach outlined in this study, may be useful for other countries as they consider how they will account for this pool in their carbon reporting.

Acknowledgments

ORCID iDs

Kristofer D Johnson ⓘ https://orcid.org/0000-0002-4015-6910
Katelyn Dolan ⓘ https://orcid.org/0000-0002-1119-2277

References

Barbier S, Gosselin F and Balandier P 2008 Influence of tree species on understory vegetation diversity and mechanisms involved—a critical review for temperate and boreal forests *For. Ecol. Manage.* **254** 1–15

Birdsey R A 1996 Carbon storage for major forest types and regions in the conterminous United States *For. Glob. Change* **2** 1–25

Bradford J B, Birdsey R A, Joyce L A and Ryan M G 2008 Tree age, disturbance history, and carbon stocks and fluxes in subalpine Rocky Mountain forests *Glob. Change Biol.* **14** 2882–97

Bradford J, Weishampel P, Smith M-L, Kolka R, Birdsey R A, Ollinger S V and Ryan M G 2009 Detrital carbon pools in temperate forests: magnitude and potential for landscape-scale assessment *Can. J. For. Res.* **39** 802–13

Chojnacky D C and Milton M 2008 Measuring carbon shrubs *Field Measurements for Forest Carbon Monitoring: A Landscape-Scale Approach* ed M Coeli (New York: Springer Science + Business Media) ch 5

Clark K L, Skowronski N and Hom J 2010 Invasive insects impact forest carbon dynamics *Glob. Change Biol.* **16** 88–101

Cole J A *et al* 2013 Database for landscape-scale carbon monitoring sites *General Technical Report NRS-119* (Newtown Square, PA: US Department of Agriculture, Forest Service, North Central Station) p 12

Cronan J B, Wright C S and Petrova M 2015 Effects of dormant and growing season burning on surface fuels and potential fire behavior in northern Florida longleaf pine (*Pinus palustris*) flatwoods *Forest Ecol. Manage.* **354** 318–33

Domke G M, Perry C H, Walters B F, Woodall C W, Russell M B and Smith J E 2016 Estimating litter carbon stocks on forest land in the United States *Sci. Total Environ.* **557–558** 469–78

Fang J, Liu G, Zhu B, Wang X and Liu S 2007 Carbon budgets of three temperate forest ecosystems in Dongling Mt., Beijing, China *Sci. China* D **50** 92–101

Garkoti S C 2008 Estimates of biomass and primary productivity in a high-altitude maple forest of the west central Himalayas *Ecol. Res.* **23** 41–9

Gilliam F S 2007 The ecological significance of the herbaceous layer in temperate forest ecosystems *Bioscience* **57** 845

Gonzalez M *et al* 2013 Contribution of understory species to total ecosystem aboveground and belowground biomass in temperate *Pinus pinaster* Ait. forests *For. Ecol. Manage.* **289** 38–47

Hart S A and Chen H Y H 2006 Understory vegetation dynamics of North American boreal forests *CRC. Crit. Rev. Plant Sci.* **25** 381–97

Homer C H, Fry J A and Barnes C A 2012 The national land cover database (No. 2012–3020) US Geological Survey

Hou L, Xi W and Zhang S 2015 Effect of understory on a natural secondary forest ecosystem carbon budget *Russ. J. Ecol.* **46** 51–8

Hudak A T, Strand E K, Vierling L A, Byrne J C, Eitel J U H, Martinuzzi S and Falkowski M J 2012 Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys *Remote Sens. Environ.* **123** 25–40

IPCC 2006 *IPCC Guidelines for National Greenhouse Gas Inventories* ed H S Eggleston, L Buendia, K Miwa, T Ngara and K Tanabe (www.ipcc-nggip.iges.or.jp/public/2006gl/)

Jain T B, Graham R T and Adams D 2009 Carbon concentrations and carbon pool distributions in dry, moist, and cold mid-aged forests of the rocky mountains *Integrated Management of Carbon Sequestration and Biomass Utilization Opportunities in a Changing Climate: Proc. of the 2009 National Silviculture Workshop* ed T B Jain, R T Graham and J Sandquist (Boise, ID) pp 39–59 (www.treesearch.fs.fed.us/pubs/37311)

Jarvis A *et al* 2008 Hole-filled SRTM for the globe version 4 *CGIAR-CSI SRTM 90 m Database* (http://srtm.csi.cgiar.org)

Keeling H C and Phillips O L 2007 The global relationship between forest productivity and biomass *Glob. Ecol. Biogeogr.* **16** 618–31

Knapp B O, Stephan K and Hubbart J A 2015 Structure and composition of an oak-hickory forest after over 60 years of repeated prescribed burning in Missouri, USA *For. Ecol. Manage.* **344** 95–109

Liaw A and Wiener M 2002 Classification and regression by random Forest *R News* vol **2/3** (http://cogns. northwestern.edu/cbmg/LiawAndWiener2002.pdf)

Mallik A U 2003 Conifer regeneration problems in boreal and temperate forests with ericaceous understory: role of disturbance, seedbed limitation, and keytsone species change *CRC. Crit. Rev. Plant Sci.* **22** 341–66

Martinuzzi S, Vierling L A, Gould W A, Falkowski M J, Evans J S, Hudak A T and Vierling K T 2009 Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability *Remote Sens. Environ.* **113** 2533–46

Moore P T, Van Miegroet H and Nicholas N S 2007 Relative role of understory and overstory in carbon and nitrogen cycling in a southern Appalachian spruce–fir forest *Can. J. For. Res.* **37** 2689–2700

Nilsson M-C and Wardle D A 2005 Understory vegetation as a forest ecosystem driver: evidence from the northern Swedish boreal forest *Front. Ecol. Environ.* **3** 421–8

Ordóñez J A B, de Jong B H J, García-Oliva F, Aviña F L, Pérez J V, Guerrero G, Martínez R and Masera O 2008 Carbon content in vegetation, litter, and soil under 10 different land-use and land-cover classes in the central highlands of Michoacan, Mexico *Forest Ecol. Manage.* **255** 2074–84

Ottmar R D, Vihnanek R E, Wright C S and Olson D L 2004 Stereo photo series for quantifying natural fuels. Vol 7: Oregon white oak, California deciduous oak, and mixed-conifer with shrub types in the western United States *Report* PMS 839 Boise, ID National Wildfire Coordinating Group, National Interagency Fire Center

Pardini R, de Souza S M, Braga-Neto R and Metzger J P 2005 The role of forest structure, fragment size and corridors in maintaining small mammal abundance and diversity in an Atlantic forest landscape *Biol. Conserv.* **124** 253–66

R Development Core Team 2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (www.r-project.org)

Royo A A and Carson W P 2006 On the formation of dense understory layers in forests worldwide: consequences and implications for forest dynamics, biodiversity, and succession *Can. J. For. Res.* **36** 1345–62

Russell M B, Woodall C W, Potter K M, Walters B F, Domke G M and Oswalt C M 2017 Interactions between white-tailed deer density and the composition of forest understories in the northern United States *For. Ecol. Manage.* **384** 26–33

Russell M, D'Amato A and Schulz B 2014 Quantifying understorey vegetation in the US Lake States: a proposed framework to inform regional forest carbon stocks *Forestry* **87** 629 –38

Saitoh T M, Nagai S, Yoshino J, Kondo H, Tamagawa I and Muraoka H 2014 Effects of canopy phenology on deciduous overstory and evergreen understory carbon budgets in a cool-temperate forest ecosystem under ongoing climate change *Ecol. Res.* **30** 267–77

Schulz B K, Bechtold W A and Zarnoch S J 2009 Sampling and estimation procedures for the vegetation diversity and structure indicator *US Forest Service Pacific Northwest Research Station General Technical Report* PNW-GTR 1–53

Smith J E, Heath L S and Hoover C M 2013 Carbon factors and models for forest carbon estimates for the 2005–2011 national greenhouse gas inventories of the United States *For. Ecol. Manage.* **307** 7–19

Smithwick E A H, Harmon M E, Remillard S M, Acker S A and Franklin J F 2002 Potential upper bounds of carbon stores in forests of the Pacific Northwest *Ecol. Appl.* **12** 1303–17

Smithwick E A H, Turner M G, Metzger K L and Balser T C 2005 Variation in NH_4^+ mineralization and microbial communities with stand age in lodgepole pine (*Pinus contorta*) forests, Yellowstone National Park (USA) *Soil Biol. Biochem.* **37** 1546–59

Suchar V A and Crookston N L 2010 Understory cover and biomass indices predictions for forest ecosystems of the Northwestern United States *Ecol. Indic.* **10** 602–9

Thornton M M, Thornton P E, Wei Y, Mayer B W, Cook R B and Vose R 2016 *Daymet: Annual Climate Summaries on a 1 km Grid for North America* (Oak Ridge, TN) (http://daac.ornl.gov/DAYMET/guides/Daymet_V3_Annual_Climatology.html)

Tremblay N O and Larocque G R 2001 Seasonal dynamics of understory vegetation in four Eastern Canadian forest types *Int. J. Plant Sci.* **162** 271–86

Tuanmu M-N, Viña A, Bearer S, Xu W, Ouyang Z, Zhang H and Liu J 2010 Mapping understory vegetation using phenological characteristics derived from remotely sensed data *Remote Sens. Environ.* **114** 1833–44

USDA Farm Services Agency 2008 *Qualitative and quantitative synopsis on NAIP usage from 2004–2008* (www.fsa.usda.gov/Assets/USDA-FSA-Public/usdafiles/APFO/naip_usage.pdf)

USDA Forest Service 2016 *The Forest Inventory and Analysis Database: Database Description and User Guide Version 6.1.1 for Phase 2* (www.fia.fs.fed.us/library/database-documentation/current/ver611/FIADB_User_Guide_P2_6-1-1_final.pdf)

Wing B M, Ritchie M W, Boston K, Cohen W B, Gitelman A and Olsen M J 2012 Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest *Remote Sens. Environ.* **124** 730–41

Woodall C W and Monleon V J 2008 Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA Program *US Forest Service General Technical Report* NRS-22

Wright C C, Eagle P C and Olson D L 2010 A high-quality fuels database of photos and information *Fire Manag. Today* **70** 27–31

Yarie J 1980 THe role of understory vegetation in the nutrient cycle of forested ecosystems in the mountain hemlock biogeoclimatic zone *Ecology* **61** 1498–514