Reforestation can sequester two petagrams of carbon in US topsoils in a century

Lucas E. Nave,a,b,1 Grant M. Domke,c Kathryn L. Hofmeister,a,d Umakanth Mishra,d Charles H. Perry,e Brian F. Walters,f and Christopher W. Swanstonf

aBiological Station, University of Michigan, Pellston, MI 49769; bDepartment of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109; cNorthern Research Station, US Department of Agriculture-Forest Service, St. Paul, MN 55018; dDepartment of Natural Resources, Cornell University, Ithaca, NY 14850; eEnvironmental Science Division, Argonne National Laboratory, Argonne, IL 60439; and fNorthern Research Station, US Department of Agriculture-Forest Service, Houghton, MI 49931

Edited by William H. Schlesinger, Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, NC, and approved January 26, 2018 (received for review November 13, 2017)

Soils are Earth’s largest terrestrial carbon (C) pool, and their responsiveness to land use and management make them appealing targets for strategies to enhance C sequestration. Numerous studies have identified practices that increase soil C, but their inferences are often based on limited data extrapolated over large areas. Here, we combine 15,000 observations from two national-level databases with remote sensing information to address the impacts of reforestation on the sequestration of C in topsoils (uppermost mineral soil horizons). We quantify C stocks in cultivated, reforested, and natural forest topsoils; rates of C accumulation in reforesting topsoils; and their contribution to the US forest C sink. Our results indicate that reforestation increases topsoil C storage, and that reforestation lands, currently occupying >500,000 km² in the United States, will sequester a cumulative 1.3–2.1 Pg C within a century (13–21 Tg C y⁻¹). Annually, these C gains constitute 10% of the US forest sector C sink and offset 1% of all US greenhouse gas emissions.

In the midst of growing urgency to enhance terrestrial C sequestration, mitigate climate change, and sustain soil-based ecosystem services, there is widespread agreement on the need to understand how land use practices impact soil C. Indeed, for years, scientists have estimated the C cycle impacts of land use and soil management strategies, often calling for new or improved practices to be implemented on increased land areas. These forward-looking studies have defined targets (1, 2), developed networks and methodologies (3, 4), and advanced soil and C cycle science along the way. However, they have rarely used large observational datasets to directly address how land use affects soil C (5). Far more often, these big-picture studies employ expert judgment, literature review, or sophisticated model-data fusion methods (6–9). Against this range of approaches, there is a recognized need for empirical analyses of large datasets to develop observational constraints on modeled and/or extrapolated results (10–13). Without a doubt, observational approaches have their own limitations, and large databases incorporate natural variation and associated uncertainty. Nonetheless, the magnitude of the problem and the nature of science in the collaboration age—which applies complementary methods to pressing questions—establish a strong role for observations in addressing large-scale questions of soil C sequestration. Here, we demonstrate an approach that synthesizes multiple data types and sources to address a problem at the nexus of soils, land use, and the C cycle.

Soils to 1 m depth hold 74% of all terrestrial C (14–16), with North America and the conterminous US (CONUS) representing 35% and 4% of all soil C, respectively (17, 18). Because adding even a few percent to this large C stock translates into a globally relevant increase, it is essential to quantify the increases achievable through changes in land use and management. Reforestation of marginal croplands and active reforestation (replanting) on understocked forestlands are two promising strategies for increasing terrestrial C sequestration (19–21). Reviews of site-level studies (22, 23) suggest reforestation and other land use and management changes increase soil C by 0.1–0.4 Mg of C ha⁻¹ y⁻¹, with potential increases of 50–100 Tg y⁻¹ estimated for US and European agricultural lands (1, 24). However, it may be more important to develop a framework for constraining these estimates with observations than to precisely quantify the soil C increases associated with specific land use or management changes. To that end, we offer an approach for synthesis—and a set of empirical data resources for collective use—while also directly quantifying C sequestration in the topsoils (uppermost mineral soil horizons) of lands that are currently undergoing reforestation in the United States.

Data Synthesis

We used several large data sources as starting points in a series of filtering, gap-filling, validation, and analysis steps, described in Methods and detailed in Supporting Information. Because these independent data sources have different applications and align with different questions, we must first summarize them briefly. The first source was the International Soil Carbon Network (ISCN) Gen3 Database (25), containing data for >433,000 individual soil layers worldwide. Individual layers make up soil profiles; profiles are one to many per site, and sites are georeferenced. The ISCN Database contains data contributed by individuals, networks, and government agencies. Geospatial data are the second data type we used for this analysis; we extracted point attributes from two geospatial datasets to overlay these upon the geographical coordinates of ISCN sites. These secondary geospatial (overlay) data include the following: (i) land cover attributes from all four versions of the National Land Cover Dataset (NLCD; refs. 26–29), a LANDSAT-derived, 30-m

Significance

Forestland in the United States is a carbon (C) sink, offsetting ~10% of annual greenhouse gas emissions and mitigating climate change. Most of the C in forests is held in soils, and the capacity of forest soils to sequester C makes them a major component of the US forest C sink. Where reforestation is presently occurring, either through deliberate replanting after forestland is disturbed (e.g., burned), or where previously nonforested lands (e.g., cultivated) are converting to forestland, topsoils are accumulating C. However, these C accumulation rates are poorly constrained; quantifying them with empirical data is critical to accurately represent the role of reforestation in the US C budget and forecast the longevity of the US forest C sink.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

1To whom correspondence should be addressed. Email: lukeneve@umich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719685115/-/DCSupplemental.

Published online February 26, 2018.

http://www.pnas.org/content/115/11/2776

2776–2781 | PNAS | March 13, 2018 | vol. 115 | no. 11
resolution data product; (ii) estimates of forest stand age from a North American canopy age map (30), a 250-m product developed from ground-based observations and remote sensing information. We refer to ISCN sites with their remotely sensed overlay data as “ISCN-NLCD sites,” which we use to compare cultivated, reforesting (previously cultivated), and natural forest (i.e., never cultivated) lands. The third major data source was the US Department of Agriculture (USDA)-Forest Service (FS). Specifically, we acquired plot-level C data and forest area estimates from the National Forest Inventory (NFI) program’s Forest Inventory and Analysis (FIA) Database (31) and ecoregional classifications from the FS Geodatabase Clearinghouse (32). We present results from the NFI plot network, which only samples forestland, to address reforestation (i.e., replanting) on continuously forested lands.

Land Use, Land Use Change, and Soil Properties at a National Level

At present, there is a spatially nonrandom pattern of land use change operating in CONUS, not detectable from land cover alone, which is having a major impact on terrestrial C sequestration at a national level. While remotely sensed land cover products such as the NLCD effectively recognize land uses with distinctive land cover signals (e.g., cultivation), remote sensing products may not reliably distinguish intergrading land cover types, such as forest and shrub/scrub (26, 27, 33). Furthermore, these products cannot identify lands that began transitioning to a different use before the remote sensing era. However, soil observations can reveal historic land use transitions, indicate its spatial patterns and drivers, and assess its impacts on C sequestration (ref. 34 and Fig. 1).

Evidence of preferential land use and broad-scale land use change in CONUS can be found in several observations of topsoil properties in cultivated, reforesting, and natural forest ISCN-NLCD sites. First, cultivated lands have higher topsoil clay contents, lower sand, and stone contents than natural forest sites with never-cultivated soils (Fig. 2). These differences indicate a widespread tendency for preferential cultivation of finer-textured soils, while forests are allowed to persist on stonier and coarser-textured soils. In turn, lands that are presently reforesting have topsoil clay and stone contents that are intermediate between cultivated soils and never-cultivated natural forest soils, indicating nonrandom abandonment of cultivation on these marginal soils. Within this same wide-ranging dataset, significant differences in median soil C stocks between the three land uses suggest that reforestation increases soil C storage (Fig. 3). The detectability and magnitude of these differences depends on the depth of reporting (see ISCN Dataset Preparation for details and full discussion). For topsoils (the top 10 cm of the uppermost mineral horizon), the sample size of the nationwide dataset is sufficiently large to detect a 5% difference in median C stocks between cultivated and reforesting topsoils as statistically significant (20 vs. 21 Mg of C·ha⁻¹, respectively). Detecting a small difference in soil C stocks is difficult due to spatial variability (35, 36), and in meta-analyses of management impacts on soil C, we have found the uppermost portions of the soil to be the most responsive (34, 37, 38). In this case, detecting such a subtle change may only be possible because the depth we consider is the most superficial portion of the topsoil, which is in direct contact with physical disturbances and detritus inputs (e.g., crop residues, forest litter). Within this same volume of soil, the median C stock for natural forests (37 Mg of C·ha⁻¹) is much higher than for cultivated or reforesting soils. However, if C stocks are computed to 30 cm depth (typical of international greenhouse gas inventory and reporting programs), an apparent 2% relative difference between reforesting and cultivated soils (medians of 47 and 48 Mg of C·ha⁻¹, respectively) is not statistically significant, while both are significantly less than natural forest soils (60 Mg of C·ha⁻¹). In this case, the differences between natural forest soils and the other two land uses are smaller, possibly indicating that past cultivation mixed surface C downward (39, 40). If this is true, we overestimate the potential for C gain by comparing natural forest to cultivated (or reforesting) topsoils. Conversely, the preferential cultivation of soils with higher clay, lower sand, and stone contents (Fig. 2) could mean that soils used for cultivation have inherently higher productivity or greater capacity for C storage below 10 cm (e.g., through illuviation with clay minerals), and that these fundamental differences in soil properties are responsible for patterns of C storage in deeper horizons. Ultimately, because our analysis is not intended to provide C estimates for international reporting, but rather so much as a quantitative assessment of reforestation impacts on soil C in a highly responsive surface layer, we maintain the focus on topsoils for the rest of this paper.

Land Use Has a Strong, Direct Impact on Topsoil C Stocks with a Consistent Pattern Across Geographic Areas

Placing ISCN-NLCD sites into an ecoregional framework reveals that in 9 of 10 ecoregional divisions, natural forests have significantly higher mean topsoil C stocks than cultivated topsoils (Fig. 4). The sole exception is the tropical/subtropical desert division, where data are too limited to detect significant differences among any land uses. In 5 of 10 divisions, reforesting lands have significantly greater mean topsoil C storage than cultivated lands, and in no division are there significant land use differences that deviate from the overall national trend. Similarly, in most divisions, mean topsoil C stocks in reforesting soils are much closer to cultivated than natural forest levels, suggesting considerable remaining C sequestration potential. Because land use interacts with significant regional variation in topsoil C storage (P < 0.001 for main effects and interactions), and spatial variation in soil C stocks is important in its own right (36, 41), regionalized estimates of topsoil C storage for different land uses may allow more nuanced projections and predictions of land use impacts on soil C sequestration at a national scale.

Reforestation and C Sequestration at the CONUS Level

At the national level, our ISCN-NLCD and NFI synthesis datasets provide quantitative estimates for topsoil C sequestration on two types of reforesting land, collectively occurring on >500,000 km² (Table 1; see Scaling: Projection and Prediction of Topsoil C Stocks for detailed methods). Because these datasets constrain the independent contributions of cultivated land reforestation (ISCN-NLCD) and forestland reforestation (NFI), the C gains associated with these land use and management decisions are additive. Over the course of a century, topsoils on reforesting cultivated lands will accumulate 0.8–1.6 Pg C, depending on which forest area estimate is applied to the ISCN-NLCD dataset, while reforestation
on continuous forestland will contribute an additional 0.5 Pg C. Collectively, this total C gain (1.3–2.1 Pg C) is on lands that are already reforesting, i.e., it is not a theoretical projection based on increased replanting efforts or large-scale crop-to-forest land use changes, as is often suggested in arguments for enhancing soil C sequestration. However, realizing these topsoil C gains over the course of a century requires previously cultivated lands now undergoing reforestation to remain in forest land use, as recultivation could cause larger soil C losses than the C gains achievable by substituting reforestation on a similar area of cultivated land. However, in light of relatively low areal percentages for both types of reforestation, modest increases in area would have a multiplicative impact on nationwide C sequestration. For example, replanting could be implemented on more than 7% of current forestland (Table 1), or additional marginal lands could be transferred from cultivation to forest, increasing the contribution of reforesting cultivated lands to the national forest land base. In either case, a small increase in reforestation area as a percentage of the total forest land area, over a long time period, would create significant C cycle impacts that are readily quantifiable using the empirical models we present here.

Our estimates of the realized and potential topsoil C gains on cultivated lands that are undergoing reforestation help to constrain C sequestration under this type of land use change at multiple levels. Most importantly, at the CONUS level, the C gains that have been realized thus far for this type of reforestation are only about 10% of their potential, highlighting the substantial C sink capacity of this land use transition if these lands are allowed to continue returning toward a natural forest condition. However, because our results are not based upon a longitudinally sampled set of sites undergoing land use change, but rather on a space-for-time substitution, it is important to recognize that individual sites may show no change in topsoil C, or changes over time that differ from the estimates we calculate based on statistical differences (i.e., between medians or means). In this regard, our broader inferences are based on statistical differences and are only as robust as the assumptions that underlie them. Regardless of whether all prerequisites of space-for-time substitution are met, the metrics that we have termed realized and potential topsoil C gains (Table 1) can be used to prioritizing ecoregions for C management, because they indicate significant statistical differences in soil C between land uses. For example, some ecoregions have less potential overall (dry domain or Mediterranean division), while some have much remaining potential (humid temperate domain or hot continental division). Our empirical models can be used similarly; ecoregions with particularly low (subtropical) or high (warm continental) topsoil C stocks are indicated by corresponding differences in their y intercepts, while differing C accumulation rates are suggested by differences in model slopes. Overall, our modeled rates of topsoil C accumulation range from 0.11 to 0.34 Mg of C·ha⁻¹·y⁻¹, confirming the rates reported in expert reviews and cross-site studies while providing a more detailed view that can help move C management in an increasingly regionalized direction.

To estimate how long C accumulation will continue during reforestation on cultivated lands, we need some way to constrain the age of the forests that are currently growing at ISCN-NLCD.
sites. Unfortunately, while the forest canopy age dataset proved sufficient for broadly attributing ages and estimating rates of topsoil C accumulation across ISCN-NLCD sites, it may not be applicable at finer scales or to more specific questions due to its coarse resolution (250 m) and intended application to forestland with homogenous physical and age structure. In lieu of relying exclusively on that data product for questions it cannot confidently address, literature reviews are useful for approximating relationships between age and soil C during forest recovery. One global analysis that highlighted the potential of cultivation-to-forest land use change suggested a 6% increase in soil C (relative to cultivation) during the 10- to 30-y period, with 19% increases over longer-term (>30 y) reforestation (42). In an analysis focused on the United States, we developed look-up tables for several domains and divisions. At CONUS level, empirical models and 100 y C gains are presented for reforestation occurring on two previous land uses (34). On soils previously used for agriculture, the first detectable change in soil C storage (+5%) occurred during the period 35–45 y since forest establishment, with gains reaching 18% during the 115- to 125-y period. In the context of this relevant literature, the 5% difference in median C storage that we find between cultivated and reforesting topsoils suggests that, on average, previously cultivated reforesting lands in CONUS have been growing forests for roughly 20–40 y, meaning that many more decades are required to realize their full potential for topsoil C sequestration.

**Key Uncertainties: Deep Soil C Dynamics During Forest Development**

Here, we must highlight two sources of uncertainty due to their potential effects on inferences drawn from our empirical models of forest age and topsoil C storage. First, if deep soils (below 30 cm) can undergo significant C stock changes over decadal time periods, then the topsoil C gains we report here may not be representative of the whole profile. Some literature reports subsoil C losses during postagricultural reforestation (43–45), and if these are generalizable across the soils in our nationwide dataset, the topsoil C gains that we report may offset by deep soil C losses. Second, our reliance on stand age as a predictor variable in empirical models of topsoil C storage introduces some uncertainty into our predictions. Not only is the very concept of “stand age” questionable, particularly in older uneven-aged forests, but methods of measuring stand age (or extracting it from maps) may introduce uncertainty into our predictions. While we suggest that there is no likely directional bias in our empirical predictions, it is important to acknowledge that mixed-age forests are extensive in the United States and attributing ages for them can be challenging (46). Ultimately, future analyses similar to our empirical data may be able to refine our predictions for topsoils, or supplement them for subsols, but for the sake of this work, we refer the reader to ISCN Dataset Preparation and Empirical Models: Supporting Results and Critical Appraisal of Stand Age Data for fuller consideration of these caveats.

**Reforesting Topsoils Play a Major Role in the US Forest C Sink**

The forest sector is a critical component of the US greenhouse gas budget, offsetting ~11% of all US greenhouse gas emissions with a

---

Table 1. Reforestation areas and impacts on C sequestration are unevenly distributed across the CONUS

<table>
<thead>
<tr>
<th>Scenario/Domain/Division</th>
<th>Forest, km²</th>
<th>Reforestation, km²</th>
<th>%</th>
<th>Model</th>
<th>100 y C, Pg</th>
<th>Realized C, Tg</th>
<th>Potential C, Tg</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONUS (ISCN)*</td>
<td>3,715,005</td>
<td>631,551</td>
<td>17</td>
<td>C = 26 + 0.249 × age</td>
<td>1.6</td>
<td>104</td>
<td>1,015</td>
</tr>
<tr>
<td>CONUS (ISCN)†</td>
<td>1,970,143</td>
<td>334,924</td>
<td>17</td>
<td>C = 26 + 0.249 × age</td>
<td>0.8</td>
<td>55</td>
<td>538</td>
</tr>
<tr>
<td>CONUS (NFI)</td>
<td>2,773,628</td>
<td>197,754</td>
<td>7</td>
<td>C = 19 + 0.233 × age</td>
<td>0.5</td>
<td>12</td>
<td>47</td>
</tr>
<tr>
<td>Dry</td>
<td>457,312</td>
<td>32,012</td>
<td>7</td>
<td>C = 25 + 0.237 × age</td>
<td>0.1</td>
<td>12</td>
<td>47</td>
</tr>
<tr>
<td>Humid Temperate</td>
<td>2,316,316</td>
<td>463,263</td>
<td>20</td>
<td>C = 27 + 0.215 × age</td>
<td>1.0</td>
<td>774</td>
<td></td>
</tr>
<tr>
<td>Total across domains</td>
<td>2,773,628</td>
<td>495,275</td>
<td></td>
<td></td>
<td>1.1</td>
<td>12</td>
<td>821</td>
</tr>
<tr>
<td>Hot Continental</td>
<td>437,829</td>
<td>137,811</td>
<td>22</td>
<td>C = 28 + 0.155 × age</td>
<td>0.2</td>
<td>28</td>
<td>331</td>
</tr>
<tr>
<td>Marine</td>
<td>137,680</td>
<td>3,129</td>
<td>2</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediterranean</td>
<td>146,957</td>
<td>1,062</td>
<td>1</td>
<td>NS</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Prairie</td>
<td>111,512</td>
<td>59,241</td>
<td>34</td>
<td>C = 25 + 0.110 × age</td>
<td>0.1</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Subtropical</td>
<td>677,857</td>
<td>406,312</td>
<td>36</td>
<td>C = 19 + 0.289 × age</td>
<td>0.8</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td>Temperate Desert</td>
<td>127,858</td>
<td>9,465</td>
<td>5</td>
<td>C = 39 + 0.118 × age</td>
<td>0.0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Temperate Steppe</td>
<td>395,083</td>
<td>102,721</td>
<td>16</td>
<td>NS</td>
<td></td>
<td></td>
<td>216</td>
</tr>
<tr>
<td>Trop./Subtrop. Desert</td>
<td>65,918</td>
<td>1,156</td>
<td>2</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trop./Subtrop. Steppe</td>
<td>274,420</td>
<td>37,183</td>
<td>10</td>
<td>NS</td>
<td></td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>Warm Continental</td>
<td>398,515</td>
<td>74,722</td>
<td>15</td>
<td>C = 37 + 0.339 × age</td>
<td>0.3</td>
<td>67</td>
<td>374</td>
</tr>
<tr>
<td>Total across divisions</td>
<td>2,773,628</td>
<td>822,803</td>
<td></td>
<td></td>
<td>1.4</td>
<td>217</td>
<td>1,438</td>
</tr>
</tbody>
</table>

Table shows forestland area (km²), reforestation area (as km² and as a percentage of total forestland), and C gains during reforestation for CONUS overall, and subdivided by ecoregions (domains and divisions). At CONUS level, empirical models and 100 y C gains are presented for reforestation occurring on two types of land: cultivation-to-forest (using the ISCN-NLCD dataset) and forestland reforestation (replanting, using NFI data). For reforestation on previously cultivated lands (ISCN-NLCD sites), realized topsoil C gains (Tg relative to cultivated baseline) are presented for spatial units where site-level mean C storage (Mg C ha⁻¹) is significantly greater for reforested than cultivated sites; potential topsoil C gains are presented for units where C storage is significantly greater for natural forest than cultivated C sites (see Fig. 3 for land use differences in C storage by ecoregion).

*Areas include lands classified by NLCD as shrub/scrub.
†Areas exclude lands classified by NLCD as shrub/scrub.
mean net removal of 651 Tg CO2 equivalents (176 Tg of C) annually during the past 5 y (47). Based on estimates from our empirical models, annual topsoil C gains on reforesting lands (1.3–2.1 Pg of C over 100 y = 13–21 Tg of C) represent 7–12% of the entire forest sector greenhouse sink, effectively offsetting 0.8–1.3% of all US greenhouse gas emissions. Given that our results suggest this C sink will persist for decades, C sequestration in reforesting topsoils can be a significant long-range solution to the problem of the declining sink strength of the US forest sector. For over 20 y, strategic planning analyses and reports have warned that the C sink strength of US forests is declining, and is expected to become neutral beyond the mid-21st century (48–51). An additional 100–200 Tg of C yr−1 has been identified as an achievable target that would stabilize the long-range trajectory of the forest C sink (52). In light of these far-reaching assessments, our analysis is significant for two reasons. First, there are no published empirical estimates for the role of reforesting topsoils in US forest C sequestration; researchers overwhelmingly focus on more readily quantified (but admittedly complex) aboveground components. Second, even at their current extent, topsoils that are now undergoing reforesting represent a >10% contribution toward the additional C sequestration needed to stabilize the C sink trajectory beyond the mid-21st century. Given that the majority of their potential C gains are yet to be realized, and even slight increases in reforestation area will have a multiplicative impact on long-term CONUS-level C sequestration, topsoils in reforesting lands are indeed a natural climate solution (1).

Conclusion

Through a synthesis that integrates 15,000 soil observations with remote sensing and geospatial information, we estimate topsoil C gains on lands undergoing two types of reforestation in CONUS. By quantifying the impact of replanting on topsoil C accumulation in continuous forestland, revealing differences in the underlying soil properties of cultivated, reforesting, and natural forest lands nationwide, and placing our results in an ecoregional framework, we reveal the physical basis for patterns of land use change on marginal soils, and provide regionalized insight into rates of topsoil C accumulation and long-term C sequestration potentials. Even if continued topsoil C accumulation rates are slower than we estimate, or whole-profile C stocks follow a different trajectory, we demonstrate the use of complementary empirical data sources suggesting that topsoil C stocks have already increased in CONUS due to ongoing reforestation. Currently, reforesting topsoils are accumulating 13–21 Tg of C yr−1, with the potential to accumulate hundreds more Tg of C within a century. Collectively, the cumulative potential for topsoil C sequestration on these lands may exceed 2 Pg of C—double the global land use change emission (53)—highlighting the significance of continued forest recovery on these lands, as well as the magnitude of additional C accumulation that could be achieved with modest area increases in reforestation.

Methods

ISCN Dataset. We used the ISCN map-based data tool to download a data set of 319,316 soil layers from 52,178 profiles in CONUS. The dataset included geographic coordinates, descriptive, physical, chemical, and metadata for all soils, profiles, and layers contained in the download, which contained data from multiple sources (25, 54–58). Next, we proceeded through a series of filtering, gap-filling, and validation steps, described in greater detail in ISCN Dataset Preparation. The present analysis focuses on results for 12,617 topsoil layers (A horizons) across the CONUS, representing those layers that met all criteria and were appropriate to address the questions of interest.

Overlay Data: Land Cover and Canopy Age. We extracted land cover and canopy age for each ISCN site, assigning each location to its closest (in time) NLCD product. We assumed that land cover attributes for soils sampled between January 1, 1989, and December 31, 1998, were reasonably represented by the NLCD1992 product; soils from 1997 to 2001 were represented by NLCD2001; soils from 2002 to 2006 by NLCD2006; soils from 2007 to most recent (2014) by NLCD2011. After matching each ISCN site to the appropriate NLCD classification, we performed a series of land cover aggregation and validation steps described in ISCN Overlay Data: Aggregating and Validating Land Cover Classification.

For our second overlay dataset, we extracted canopy ages (year 2006) from a 250-m resolution North American forest canopy age map produced by combining plot-level and multiple remote sensing data sources. We performed this extraction only for a subset of the ISCN-NLCD sites; specifically, those matched to NLCD2001 and NLCD2006, to obtain forest canopy ages only for ISCN-NLCD sites that were closely temporally aligned with the input data and estimated canopy age attributes of the data product. To overlay an age value on each site, we extracted the age for the pixel containing the ISCN-NLCD geographic coordinates using ArcGIS (ESRI), which was also our approach for obtaining other overlay data.

FS Data: Plots, Land Areas, and Ecoregions. Data contained in the FIA Database are taken as part of the NFI, an equal-probability sample of forestland across CONUS. There is one FPI plot on approximately every 2,400 ha across the United States, placed randomly within a hexagonal grid. We queried the FIA Database for plots with topsoil C measurements (Mg of Chao−1 to 10 cm), and filtered query returns to use topsoil C stocks only for single-condition plots, i.e., those not subdivided into different forest conditions (e.g., stand ages). We made this exclusion to ensure straightforward data interpretation and avoid potential edge effects. As additional constraints, we utilized only the most recent observation of each plot (FPI plots are periodically remeasured), and only plots observed since 2000, to ensure NFI data were concurrent with the observations from the ISCN-NLCD dataset. Altogether, our NFI soil C dataset held 2,383 observations. To test relationships between soil and bio-mass C, we also assembled a dataset of plot-level aboveground and belowground C stocks for the same plots that provided soils information. We obtained forestland areas for ecoregions of CONUS (described in Ecoregional Framework) using standard FIA estimation procedures (59).

Data Analysis. We used nonparametric and parametric tests, in a phased approach, to address our questions of interest. We used nonparametric analyses in the first phase with the goal of showing data distributions, describing patterns, and testing for significant differences in what is essentially a national soil census. Nonparametric tests are suited for this portion of our analysis because skewed distributions that challenge parametric criteria are common in large datasets, and reflect the reality of the populations of interest (e.g., a small number of topsoils have very large C stocks). We used nonparametric tests rather than eliminate outliers from our populations of interest, retain them and incur skewness, or attempt to normalize the distributions with transformations.

In our nonparametric analyses, we used Kruskal–Wallis tests with Dunn’s multiple comparisons to indicate whether topsoil textures, thicknesses, and C stocks differed significantly between land uses. Observations included in these tests were ISCN-NLCD sites with cultivated, reforesting, or natural forest land uses; we assigned sites into these land use groups using a combination of remote sensing (NLCD) and soil profile information (ISCN). We review our approach for assigning land use (Fig. 1) here because it contextualizes our results. Because cultivation is reliably indicated by a distinctive land cover, ISCN-NLCD sites with cultivated land cover represent cultivation as a land use. However, ISCN-NLCD sites attributed as forest are separable into two land uses: natural forests (soils never cultivated) and forests that have become established on previously cultivated lands, with an Ap horizon serving as the distinguishing feature. An Ap horizon (i.e., plow layer) is readily recognized in a soil pit by its consistent (often deep) thickness and clear abrupt boundary over underlying horizons, and may persist for centuries following agricultural abandonment (60, 61). In contrast, A horizons in forest soils are usually thinner and have wavy, irregular, or otherwise variable thicknesses and boundaries.

For the second phase of our data analysis, we used parametric statistics to estimate mean values, SEs, and significant differences for topsoils from various land use by ecoregion groups. In these tests, we did not include topsoils with C stocks separated from the grand mean by >2σ (3% of the 12,617 total observations). Also, we diagnosed transformations of the right-skewed distribution of topsoil C stocks, selecting the square-root transformation which provided the greatest mitigation of nonnormality. Subsequently, we ran ANOVA with Fishers Least Significant Difference multiple comparisons to indicate whether land uses had significantly different topsoil C stocks across the entire CONUS, as well as within individual ecoregions. We back-transformed the means of ecoregion by land use for scaling exercises described in Scaling: Projection and Prediction of Topsoil C Stocks.

Nave et al.
In all statistical tests comprising our data analysis, conducted using SigmaPlot (SYSTAT), we set $P < 0.05$ as the threshold for accepting test results as significant.

ACKNOWLEDGMENTS. We thank Kailey Marcinckowski of the Northern Institute of Applied Climate Science for figure illustration and two anonymous reviewers for their careful reading and thoughtful feedback. This work was facilitated by the ISCN and benefited from extensive data contribution from the USDA-Natural Resources Conservation Service (National Cooperative Soil Survey) and the US Geological Survey. Financial support was provided by the USDA-Forest Service, Northern Research Station Agreements 13-CR112306-077 and 16-CR112306-071 and National Science Foundation Award 13-430681.