

# Correction

## SUSTAINABILITY SCIENCE

Correction for “Soil carbon debt of 12,000 years of human land use,” by Jonathan Sanderman, Tomislav Hengl, and Gregory J. Fiske, which was first published August 21, 2017; 10.1073/pnas.1706103114 (*Proc Natl Acad Sci USA* 114:9575–9580).

The authors would like to note the following: “We regret that two small errors were found in the code used to produce the findings of this study.

“1. The model calculates organic carbon density (OCD) at standard depths. These OCD values then must be multiplied by layer thickness to calculate organic carbon stocks (OCS). For the 30–100 cm layer, the OCD value was multiplied by only 60 cm instead of 70 cm, resulting in an underestimation of OCS when summed to 100 cm and 200 cm.

“2. There was a mismatch in the land mask used in producing the AD 2010 map and the historic no land-use map (NoLU), which results in a narrow band of erroneous soil carbon debt values along the northern Eurasian coastline. (This can actually be seen by zooming in on Fig. S5.) This error affected the global and International Geosphere-Biosphere Programme (IGBP) land cover class based total carbon stocks. The country-level calculations were not impacted.

“Both errors have been rectified. The net result is that the recalculated OCS to 100 cm and 200 cm depth are greater by ~5%, but because this error affected both the current and historic OCS, the net result on the pixel-level soil carbon debt values was negligible. Given that error #1 affected all pixels proportionally, there is no noticeable visual change to the patterns of OCS loss and the interpretation of the findings has not changed. However, because of the land mask issue (error #2), the soil carbon debt for the global and specific IGBP land cover categories are now lower than previously reported. The global total is now 116 Pg C compared with the published value of 133 Pg C. Table S3 in the Supporting Information has been updated to reflect the revised soil carbon stock and loss values.

“We would like to apologize for any inconvenience these errors have caused.

“Data availability: Updated code and new spatial layers can be found at <https://github.com/whrc/Soil-Carbon-Debt>.”

The *SI Appendix* has been updated online to include the corrected Table S3.

Published under the [PNAS license](#).

[www.pnas.org/cgi/doi/10.1073/pnas.1800925115](http://www.pnas.org/cgi/doi/10.1073/pnas.1800925115)

# Soil carbon debt of 12,000 years of human land use

Jonathan Sanderma<sup>a,1,2</sup>, Tomislav Hengl<sup>b,1</sup>, and Gregory J. Fiske<sup>a</sup>

<sup>a</sup>Woods Hole Research Center, Falmouth MA 02540; and <sup>b</sup>The International Soil Resource and Information Center – World Soil Information, 6708 PB Wageningen, The Netherlands

Edited by William H. Schlesinger, Cary Institute of Ecosystem Studies, Millbrook, NY, and approved July 14, 2017 (received for review April 12, 2017)

Human appropriation of land for agriculture has greatly altered the terrestrial carbon balance, creating a large but uncertain carbon debt in soils. Estimating the size and spatial distribution of soil organic carbon (SOC) loss due to land use and land cover change has been difficult but is a critical step in understanding whether SOC sequestration can be an effective climate mitigation strategy. In this study, a machine learning-based model was fitted using a global compilation of SOC data and the History Database of the Global Environment (HYDE) land use data in combination with climatic, landform and lithology covariates. Model results compared favorably with a global compilation of paired plot studies. Projection of this model onto a world without agriculture indicated a global carbon debt due to agriculture of 133 Pg C for the top 2 m of soil, with the rate of loss increasing dramatically in the past 200 years. The HYDE classes “grazing” and “cropland” contributed nearly equally to the loss of SOC. There were higher percent SOC losses on cropland but since more than twice as much land is grazed, slightly higher total losses were found from grazing land. Important spatial patterns of SOC loss were found: Hotspots of SOC loss coincided with some major cropping regions as well as semiarid grazing regions, while other major agricultural zones showed small losses and even net gains in SOC. This analysis has demonstrated that there are identifiable regions which can be targeted for SOC restoration efforts.

agriculture | soil organic matter | climate change | soil degradation

The incredible rise of human civilizations and the continuing sustainability of current and future human societies are inextricably linked to soils and the wide array of services soils provide (1–3). Human population and economic growth has led to an exponential rise in use of soil resources. Roughly 50 million km<sup>2</sup> of soils are currently being managed to some degree by humans for food, fiber, and livestock production (4), leading to the declaration that we live on a “used planet” (5). The consequences of human domination of soil resources are far ranging (6, 7): accelerated erosion, desertification, salinization, acidification, compaction, biodiversity loss, nutrient depletion, and loss of soil organic matter (SOM).

Of these soil threats, loss of SOM has received the most attention, due to the critical role SOM plays in the contemporary carbon cycle (8, 9) and as a key component of sustaining food production (10, 11). Despite the intense research interest in SOM and soil organic carbon (SOC) as the dominant component of SOM, there remain many unknowns (12) that impede progress in implementing sound land management strategies to rebuild SOC stocks (13).

Conversion of native soil to agricultural uses typically leads to a decline in SOC levels (14–16). The rate and extent of decline in SOC stocks should vary greatly across the globe, due to differences in soil properties, climate, type of land-use conversion, and, importantly, the specific management implementation of a given form of land use. Loss of SOC under agricultural land use is not universal; modest gains are seen when soil of naturally low fertility is improved and the previous constraint (e.g., moisture, fertility, and hardpan) on plant growth is alleviated (17–19). However, for the vast majority of land, SOC loss is more

common. In fact, in a metaanalysis of the available literature (*SI Appendix*), we found median SOC loss values of 26% for the upper 30 cm and 16% for the top 100 cm of soil, but ranges of –36 to 78% and –25 to 61%, respectively, have been reported for these two depth increments (*SI Appendix, Fig. S2*). Scaling these limited point measurements to calculate a cumulative SOC loss for the world’s agricultural land has been difficult, with estimates ranging from 40 Pg C to over 500 Pg C (20). Recent estimates from dynamic global vegetation models run with actual land use versus with potential natural vegetation have put this figure at 30 Pg C to 62 Pg C for the industrial post-1850 period (21, 22).

A credible estimate of the global total and spatial distribution of SOC loss is a critical step in understanding the potential for soil carbon sequestration to be an effective climate abatement strategy. To quantify the cumulative impact of human land use on changes in SOC at the global scale, we have developed a machine learning-based data-driven statistical model, which is based on a global compilation ( $N = 158,147$ ) of soil profile observations and samples collated at The International Soil Resource and Information Center (ISRIC) – World Soil Information over the last decade (23). We predicted current SOC stocks as a function of climatic, topographic, geologic, and land-use characteristics. By using the History Database of the Global Environment (HYDE) v3.2 spatially explicit database of historic pattern of human land use (4), we projected the predictive SOC model back in time, holding climate constant, to estimate historic SOC levels, with the difference between historic and current SOC levels representing the amount of SOC lost due to human land use.

## Significance

Land use and land cover change has resulted in substantial losses of carbon from soils globally, but credible estimates of how much soil carbon has been lost have been difficult to generate. Using a data-driven statistical model and the History Database of the Global Environment v3.2 historic land-use dataset, we estimated that agricultural land uses have resulted in the loss of 133 Pg C from the soil. Importantly, our maps indicate hotspots of soil carbon loss, often associated with major cropping regions and degraded grazing lands, suggesting that there are identifiable regions that should be targets for soil carbon restoration efforts.

Author contributions: J.S. and T.H. designed research; J.S., T.H., and G.J.F. performed research; J.S., T.H., and G.J.F. contributed new reagents/analytic tools; J.S., T.H., and G.J.F. analyzed data; and J.S. and T.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Meta-analysis data have been deposited on Harvard University’s website, <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QQQM8V>.

Code and model outputs have been archived on GitHub, <https://github.com/whrc/Soil-Carbon-Debt>.

<sup>1</sup>J.S. and T.H. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [jsanderma@whrc.org](mailto:jsanderma@whrc.org).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706103114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706103114/-DCSupplemental).

## Results and Discussion

**Model Performance and Predictors.** Models explaining SOC density distribution as a function of depth, climate, relief, lithology, land cover, and land use were evaluated using fivefold cross-validation with repeated fitting. For the 0- to 2-m depth interval,  $R$ -square was 54% with a root mean square error (RMSE) of 12 kg C·m<sup>-3</sup> based on random forest model, and 45% with a RMSE of 14 kg C·m<sup>-3</sup> based on the gradient boosting model. Correlation plots indicated that the ensemble model for the 0- to 2-m depth interval was unbiased (no overestimation/underestimation) over the whole span of values. Model errors were lowest where sample density was high, and were greatest in regions with sparse data and with high SOC values (*SI Appendix, Fig. S4*).

While climate and topographic attributes were the most important variables in both models (*SI Appendix, Fig. S3*), multiple land-use variables have been shown to be also important for explaining the distribution of current SOC. Individual correlation plots for the land-use variables generally decline with increasing intensity of land use within a pixel (*SI Appendix, Fig. S3*). The relatively high importance of the HYDE land use classes “grazing” and “cropland” in our SOC models validates our conceptual approach to estimating historic SOC stocks in the absence of human land use.

**Current SOC Stocks.** Projecting our data-driven statistical model across the globe for the year 2010 suggested that global SOC stocks were 863 Pg C, 1,824 Pg C, and 3,012 Pg C in the upper 0.3 m, 1 m, and 2 m of soil, respectively (*SI Appendix, Fig. S5A and Table S3*). On a per hectare basis, land classified as cropland [International Geosphere-Biosphere Programme (IGBP) class 12 (24)] contained an average of 62 [31 to 96, 95% confidence intervals (CI)], 127 (61 to 200, 95% CI), and 198 (96 to 315, 95% CI) Mg C·ha<sup>-1</sup> for the same depth increments. The global total SOC stocks and depth distribution are consistent with most recent estimates of this quantity using a number of approaches (25, 26). This agreement is particularly encouraging because we excluded remote sensing data products characterizing current vegetation, which has been found to be particularly important in explaining SOC distribution (27), because of the requirement to be able to project this model back in time.

**Historic SOC Stocks.** Reprojection of our SOC model to a no land-use (NoLU) condition with all other variables held constant resulted in global SOC stocks of 899 Pg C, 1,899 Pg C, and 3144 Pg C in the upper 0.3 m, 1 m, and 2 m of soil, respectively (*SI Appendix, Fig. S5B and Table S3*), suggesting that human-driven land-use decisions have resulted in substantial reductions in global SOC levels. In the absence of accurate SOC data from past millennia, we have attempted to assess the accuracy of these historic projections by comparing the modeled NoLU SOC stocks to SOC measurements taken in remnant patches of native vegetation which were compiled from the literature (*SI Appendix, Fig. S6*). Given the limitations of this comparison between point measurements and 10-km model output, model results for NoLU SOC compared favorably to the measured literature values ( $R$ -square = 0.33 to 0.34, with RMSE values of 17 Mg C·ha<sup>-1</sup> and 37 Mg C·ha<sup>-1</sup> for 0.3- and 1-m depths, respectively).

**SOC Loss Due to Land Use.** Subtracting current (2010) SOC stocks from historic (NoLU) SOC stocks, we found that 37 Pg C, 75 Pg C, and 133 Pg C have been lost due to land-use change in the upper 0.3 m, 1 m, and 2 m of soil (Fig. 1B and *SI Appendix, Table S3*). The mean absolute loss due to land use to 2 m across all pixels with some degree of land use was 17.7 Mg C·ha<sup>-1</sup> (−22.7 to 54.0, 95% CI). As a percentage of initial SOC stocks

(*SI Appendix, Fig. S7*), this represents a mean loss of 8.1% (−3.9 to 21.8%, 95% CI). While SOC was lost throughout the soil profile, consistent with findings from experimental studies (14), there was an exponential decline in loss with increasing depth (*SI Appendix, Fig. S8*).

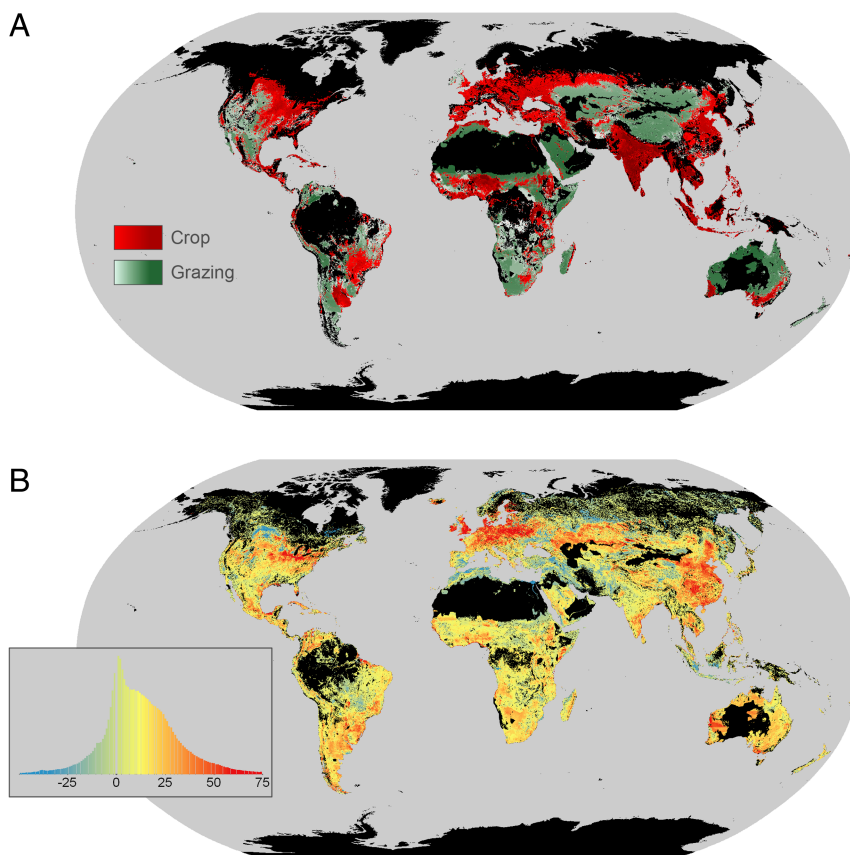
Comparison of model results with the native remnant database suggests that modeled SOC loss due to land use was likely a conservative estimate. First, comparing measured SOC in native patches to predicted SOC stocks with NoLU, we found a negative bias of −10 Mg C·ha<sup>-1</sup> (*SI Appendix, Fig. S6*). Second, it was found that there was 13.8 to 28.0% (95% CI,  $n = 62$ ) less SOC to 1 m under agricultural land compared with paired remnant native vegetation plots (see *SI Appendix, Fig. S2* and extended discussion); however, modeled loss (NoLU − 2010) for the same sites was only 7.7 to 12.8% (95% CI,  $n = 62$ ) of the NoLU SOC to 1 m.

This analysis clearly demonstrates that, while, on average, agricultural land use leads to SOC loss, there are important spatial patterns and contrasts (Fig. 1) suggesting that simple accounting methods for SOC change [e.g., Intergovernmental Panel on Climate Change default emission factors (28)] will misrepresent SOC change in many regions of the world. The majority of data used to generate published emission factors come from North America and Europe (29). In our own metaanalysis (*SI Appendix*), 82 of the 140 paired comparisons were from these two regions. Model results from the agricultural heartland of the United States and much of Europe showed large losses (Fig. 1B), primarily from cropping, that are consistent with the average documented changes in SOC stocks from field investigations (14, 15).

The largest per pixel losses were found to coincide with cropping regions (Fig. 1B); however, grazing, especially in arid and semiarid regions, with its larger spatial extents (Fig. 1A), was responsible for at least half of the total SOC loss (*SI Appendix, Table S3*). The grassland and savanna IGBP land classification categories collectively lost more SOC than the cropland and crop/natural vegetation mosaic categories (48 Pg C vs. 35 Pg C). In particular, the rangelands of Argentina, southern Africa, and Australia stand out as hotspots of SOC loss when viewed as a percent of historic SOC (*SI Appendix, Fig. S7*).

While land use is the underlying anthropogenic driver of SOC loss, the degree to which land use results in SOC loss is at least partially dependent upon the degree to which the soil resource has been exploited (2). Using the Global Land Degradation Information System biophysical status of land index, a quantitative expression of a given land area's ability to provide four categories of ecosystem services (biomass, soil, water, and biodiversity) (30), we found a strong correlation between land degradation and SOC loss (*SI Appendix, Fig. S9*). Results from *SI Appendix, Fig. S9*, coupled with the finding that grazing was the single most important land-use variable in the model (*SI Appendix, Fig. S3*), suggest that grazing of relatively unmanaged rangelands may be a stronger driver of SOC loss than previously acknowledged.

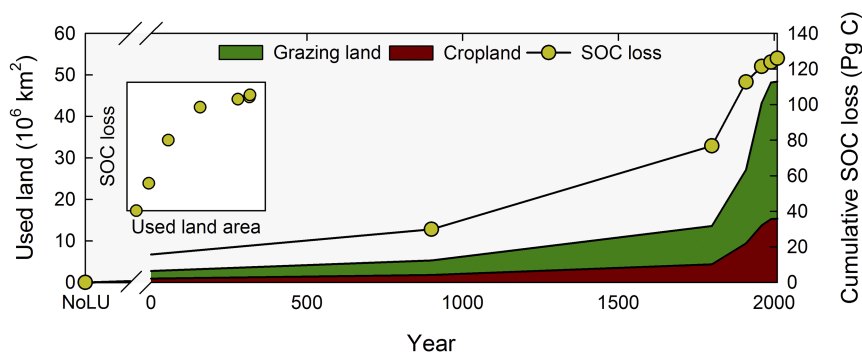
Agricultural land uses do not always result in large losses of SOC. For example, the 2-million-km<sup>2</sup> seasonally dry moist savanna region in Brazil known as “the Cerrado” was thought to have soils too poor to support intensive agriculture but, over the past few decades, has been transformed into one of the most productive agricultural regions of the world, with ~750,000 km<sup>2</sup> of crops and 800,000 km<sup>2</sup> of pasture, through liming, fertilization, and weed control (31). Because of the naturally infertile soil state in the Cerrado, agricultural expansion has resulted in little loss in SOC (Fig. 1B), and there are large areas that have actually accumulated modest amounts of SOC after the advent of agriculture, which is consistent with results from field studies (17, 32, 33).



**Fig. 1.** Global distribution of cropping and grazing in 2010 from (A) HYDE v3.2 and (B) modeled SOC change in the top 2 m. In A, color gradients indicate proportion of grid cell occupied by given land use. In B, legend is presented as histogram of SOC loss ( $\text{Mg C}\cdot\text{ha}^{-1}$ ), with positive values indicating loss and negative values depicting net gains in SOC.

**Historic SOC Loss Trend.** By using the HYDE v3.2 land-use dataset, we predicted the temporal evolution of SOC stocks due to changes in land use alone (Fig. 2). Globally, SOC loss follows the exponential rise in used land, but not in a linear fashion (Fig. 2, *Inset*). There were low annual rates of SOC loss ( $<0.05$  Pg C $\cdot$ y $^{-1}$ ) until AD 1800, followed by a century of losses  $>0.3$  Pg C $\cdot$ y $^{-1}$ , with a slight moderation of this rate during the last century ( $0.13$  Pg C $\cdot$ y $^{-1}$ ). Comparison of this historic trend among the 10 countries with the greatest losses reveals some interesting contrasts (*SI Appendix, Fig. S10*), with old-world countries (e.g., India, Kazakhstan, and Germany) showing large losses before AD 1000 but new-world countries (e.g., Brazil and Argentina) only losing significant amounts of SOC in the past century.

**Limitations of This Study.** While great care has been taken to ensure that the input data were of the highest quality possible (23, 27), there remain several limitations in the underlying datasets and therefore predicted SOC change. First, the training dataset used to build spatial predictions models was not ideal for testing the hypotheses. ISRIC's soil profile dataset is a compilation of national inventories from a large number of nations with different reasons for undertaking soil surveys and different methods of laboratory analysis. Data were collected over a 50-y period, which is likely smoothing out some of the SOC loss in the model. In addition, the mismatch in scale between a soil pedon (0.5 m on a side) and the pixel size of the HYDE v3.2 land-use data (10 km) can create situations where the dominant



**Fig. 2.** Historic reconstruction of loss in SOC relative to 10,000 BC (assumed NoLU). Temporal evolution of cropland and grazing land is given in stacked area plots. (*Inset*) Biplot of SOC loss (Pg C) v. total used land area ( $10^6 \text{ km}^2$ ) for each predicted time interval.



soil properties of a pixel are not represented by the particular soil pedons that were sampled within that pixel (34). Relatedly, regions with low sampling density may be overly influenced by a few data points that may not be representative of that region as a whole. These last two issues are the major drivers of the spatial distribution of model error (*SI Appendix*, Fig. S4).

The HYDE dataset itself also presents a few limitations. First, these land-use data are a combination of national and subnational level statistics and remotely sensed land-use change (4), which, in some cases, can create artificial changes in SOC along political boundaries. This became particularly apparent when we used the reconstructed land-use histories in prior centuries to estimate SOC stocks for those time periods (Fig. 2). Second, only very coarse land-use categories are represented, so management-specific practices which can influence SOC levels (13), such as tillage practices, rotations, and cover crops, are not represented. Relatedly, HYDE does not contain direct information on forest or wetland loss, both known drivers of SOC loss. Third, in our model formation, there is no indication of the duration of a given land use. SOC stocks, while often declining most rapidly in the first decade after land-use change (35), often take many decades to over a century to reach a new steady state (36, 37). Finally, the HYDE dataset describes the extent of land use but not the intensity. This limitation may be particularly important for the grazing category, as SOC levels have been shown to decline with increasing grazing pressure (38), although this effect appears to be dependent upon grass species composition, with C3 grasses showing large declines and C4 grasses showing small gains in SOC with increased grazing pressure (39). It is very likely that, taking these limitations together, our estimate of soil carbon debt covers only a smaller fraction of the actual debt due to human influence.

**Implications.** This analysis indicates that the majority of the used portions of planet Earth have lost SOC, resulting in a cumulative loss of ~133 Pg C due to agricultural land use. These SOC losses are on par with estimates of carbon lost from living vegetation primarily due to deforestation (40) and are nearly 100 Pg C higher than earlier estimates of land use and land use change-driven losses of SOC (41). Importantly, as Fig. 1 demonstrates, there are hotspots of SOC loss, associated with extensive cropping regions but also with highly degraded grazing land (*SI Appendix*, Fig. S9), suggesting that there are identifiable regions which should be targets for SOC restoration efforts.

The potential to recover lost SOC may be more limited than is often assumed. The amount of SOC that has been lost historically can be thought of as the carbon sink potential of the soil (42). Our analysis has found that this sink potential is ~133 Pg C (*SI Appendix*, Table S3). A widely repeated figure is that, with adoption of best management practices, two thirds of lost SOC can be recovered (42). If the two-thirds figure is accurate, then SOC sequestration has the potential to offset 88 Pg C (322 Pg CO<sub>2</sub>) of emissions. However, bottom-up estimates of the maximum biophysical potential for carbon sequestration on cropping and grazing land range from 0.4 Pg C·y<sup>-1</sup> to 1.4 Pg C·y<sup>-1</sup> (20, 43). Assuming SOC reaches a new steady state in 20 y (35, 44), this calculation suggests that 8 Pg C to 28 Pg C can be recaptured. Even the range of 8 Pg C to 28 Pg C is likely overly ambitious given the various social, economic, and technical constraints on universal adoption of best management practices (45), suggesting that the amount of the carbon sink that can be filled is on the order of, at best, 10 to 30% globally and may well be <10%.

## Conclusions

Our data-driven statistical analysis confirms that agricultural land use is a significant driver of SOC levels. Importantly, we have generated estimates for the global cumulative loss of SOC

which potentially represent a maximum estimate of the SOC sink capacity, and have demonstrated that there are hotspots of SOC loss which are closely associated with land that has been identified as highly degraded. This analysis also demonstrated that not all land use is associated with large losses in SOC, particularly in regions with naturally infertile soils. These results provide a basis for national and international policies to target SOC restoration efforts but also suggest that more effort needs to be put into collecting, integrating, and using legacy soil profile data, especially historic data 50+ y old, so that even more reliable models of SOC dynamics can be produced.

## Materials and Methods

SOC varies in complex but mostly predictable ways across the landscape. It can be best modeled as a function of climate, potential vegetation, topographic relief, soil parent material, and time (46, 47). Numerous spatially explicit data layers now exist that cover most of these state factors of soil formation. This is the foundation of the current state of the art in predictive soil mapping (27, 48),

$$\text{SOC} = f(\text{CL}, \text{R}, \text{L}, \text{LC}, \text{LU}), \quad [1]$$

where CL represents climate, R = relief, L = lithology, LC = land cover, and LU = land use. Our interest here is in applying such soil × environment + land-use models to historic land use while holding all other factors constant. Once we have estimated a model for SOC (*m*) that relates SOC to environmental + land-use data (using current data), we can use it to predict global distribution of both SOC also for historical land cover/land-use data,

$$\hat{\text{SOC}}(y) = E[m|\text{CL}, \text{R}, \text{L}, \text{LC}, \text{LU}(y)], \quad [2]$$

where *m* is the statistical model used that relates SOC to environmental conditions and LU(*y*) are the historical land-use images for year *y*. By using the HYDE v3.2 dataset (4), we have a spatially explicit representation of land use [LU(*y*)] over the past 12,000 y, so that we can determine the difference in SOC between the current and historical land use,

$$\Delta\text{SOC} = \hat{\text{SOC}}(y_2) - \hat{\text{SOC}}(y_1). \quad [3]$$

In this work, we have used the year 2010 to represent current conditions, and, to estimate the cumulative impact of human land use on SOC, we have replaced current land use with NoLU, which corresponds also to the preagricultural times, for example, 10,000 BC.

**Soil Profile Data.** ISRIC ([www.isric.org](http://www.isric.org)) curates the largest repository of spatially explicit soil profile observations and samples (*N* = 158,147) covering all climatic zones and biomes. These data come from both public and private collections, and ISRIC has gone to great lengths to try to harmonize differences among different collections (23). Given that multiple depths are sampled in each profile, there are over 850,000 measurements of SOC content (in grams per kilogram), bulk density (in kilograms per cubic meter), and coarse fragments (27). Based on these, we derive SOC density (in kilograms C per cubic meter) for each horizon using

$$\text{SOCD} = \frac{\text{OC}}{1000} \cdot \text{BD} \cdot \frac{100 - \text{CF}}{100}, \quad [4]$$

where OC is SOC mass fraction in permilles, BD is soil bulk density in kilograms per cubic meter, and CF is volumetric fraction of coarse fragments (>2 mm) in percent. Unfortunately, BD is not available for many of the soil horizons and needs to be determined either using global SoilGrids predictions (27) or, for organic soils (C > 8%), using a simple pedotransfer rule relating BD to C, for example, from Köchy et al. (49),

$$\text{BD} = [1.38 - 0.31 \cdot \log(\text{C}/100)] \cdot 1000. \quad [5]$$

After all missing values for BD have been filled, we can determine SOC density (SOCD) for every soil horizon.

**Spatial Prediction Model for Organic Carbon Density.** We overlay the training points and environmental covariates and fit spatial prediction models following the formula

$$\text{SOCD}(xyd) = d + X_1(xy) + X_2(xy) + \dots + X_p(xy), \quad [6]$$

where *d* is the depth of observation and *X<sub>p</sub>*(*xy*) are covariates. We model spatial distribution of SOCD in three dimensions (soil depth used as a predictor) using all soil horizons layers at different depths. The derived spatial prediction model is then used to predict SOCD at standard depths 0 cm, 30 cm, 100 cm, and 200 cm, so that the SOC stock for 0 m to 2 m can be

derived as a cumulative sum of the four layers. Aggregated SOCD can be converted to stock by using

$$\text{SOCs} [\text{kg m}^{-2}] = \text{SOCD} [\text{kg m}^{-3}] \cdot h [\text{m}], \quad [7]$$

where  $h$  is the layer thickness in meters.

In addition to spatial prediction of OCD values, we also modeled associated spatial prediction uncertainty using the following procedure: (i) Run fivefold cross-validation using the machine learning framework and derive prediction residuals; (ii) overlay prediction residuals (absolute values) and covariates and fit a new spatial prediction model using the same machine learning framework; and (iii) apply this error model over the whole area of interest to produce maps of absolute errors.

**Environmental Covariates.** For modeling purposes, we use a large stack of spatially explicit covariate raster data layers. These covariate layers have been compiled from numerous sources representing the major state factors mentioned above, including the following: (i) The HYDE 3.2 Historic land-use dataset (<http://ftp.pbl.nl/hyde/hyde3.2/>) (4) contains the distribution of main agricultural systems from 10,000 BC (prehistoric NoLU condition) to present time. Each raster layer represents the area (square kilometers) of each pixel occupied by a given land-use category, with the 10 categories being as follows: total cropping, total grazing, pasture (improved grazing land), rangeland (unimproved grazing land), total rainfed cropping, and total irrigated cropping with further subdivisions for rice and nonrice cropping systems for both rainfed and irrigated cropping. (ii) The Climate Research Unit TS2.1 climatic surfaces for period 1960 to 1990 ([www.ipcc-data.org/observ/clim/](http://www.ipcc-data.org/observ/clim/)) (50) include precipitation, ground frost frequency, near-surface air temperature (daily minimum, daily maximum, and mean), water vapor pressure, wet day frequency, cloud area fraction, near-surface air temperature, and diurnal range. (iii) Current and historic forest extent was estimated from the United Nations Environment Programme World Conservation Monitoring Centre Generalised Original and Current Forest cover map ([www.unep-wcmc.org/resources-and-data/generalised-original-and-current-forest](http://www.unep-wcmc.org/resources-and-data/generalised-original-and-current-forest)). (iv) Topographic parameters were derived from Global EarthEnv-Digital Elevation Model 90 (51) including slope, curvature, topographic index, topographic openness, valley depth, and multiresolution

valley bottom index. Topographic properties were derived using the System for Automated Geoscientific Analyses Geographic Information System (52) at finer resolution (250 m) and then resampled to 10-km resolution. (v) Landform and geologic substrate were determined from the United States Geologic Survey Global Ecophysiology landform classification and lithological map ([rmgsc.cr.usgs.gov/outgoing/ecosystems/Global/](http://rmgsc.cr.usgs.gov/outgoing/ecosystems/Global/)) (53).

If not already available at 10-km resolution, spatial data layers were resampled to 10-km resolution using the Geospatial Data Abstraction Library software (54).

**Statistical Modeling and Prediction.** The statistical modeling was accomplished using machine learning techniques implemented in R environment for statistical computing (55). We used an ensemble prediction of two algorithms: (i) random forest as implemented in the package ranger (56) and (ii) gradient boosting as implemented in the package xgboost (57). For model fitting, we used all soil profiles, then used this model to predict SOCD at all grid nodes and all depths for current and historic land use. During the prediction, we derived ensemble estimates as a weighted average between the two models, with weights estimated based on overall model performance/accuracy. In most cases, random forest and gradient boosting resulted in almost equal prediction accuracy; hence the weights are almost equal.

All computing was run on ISRIC High Performance Computing servers with 48 cores and 256 GB RAM. Total computing time required to produce all outputs from scratch is about 18 h of optimized computing (or about 1,000 central processing unit hours). All derived maps (as geotiff files) and R code used in this analysis can be found at Woods Hole Research Center's github repository (<https://github.com/whrc/Soil-Carbon-Debt>). Metaanalysis data can be found at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QQQM8V>.

**ACKNOWLEDGMENTS.** We thank Emily Cheney for compiling the remnant native vegetation soil carbon database and Rebecca McCulley for contributing data to the remnant native vegetation soil database. Funding was provided by The Nature Conservancy and the Doris Duke Charitable Foundation. ISRIC – World Soil Information is a nonprofit foundation primarily funded by The Netherlands government.

- Koch A, et al. (2013) Soil security: Solving the global soil crisis. *Glob Pol* 4:434–441.
- Amundson R, et al. (2015) Soil and human security in the 21st century. *Science* 348:1261071.
- Montanarella L, et al. (2016) World's soils are under threat. *SOIL* 2:79–82.
- Klein Goldewijk K, Beusen A, Van Dreht G, De Vos M (2011) The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Glob Ecol Biogeogr* 20:73–86.
- Ellis EC, Fuller DQ, Kaplan JO, Lutters WG (2013) Dating the anthropocene: Towards an empirical global history of human transformation of the terrestrial biosphere. *Elem Sci Anthropocene* 1:000018.
- Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils (2015) *Status of the World's Soil Resources (SWSR) — Main Report* (Food Agric Org, Rome).
- Keesstra S, et al. (2016) Forum paper: The significance of soils and soil science towards realization of the UN sustainable development goals (SDGs). *Soil Discuss* 2016: 1–28.
- Schlesinger W (2005) *Biogeochemistry*, Treatise on Geochemistry (Elsevier, New York), Vol 8.
- Crowther T, et al. (2016) Quantifying global soil carbon losses in response to warming. *Nature* 540:104–108.
- Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature* 418:671–677.
- Oldfield EE, Wood SA, Palm CA, Bradford MA (2015) How much SOM is needed for sustainable agriculture? *Front Ecol Environ* 13:527–527.
- Stockmann U, et al. (2013) The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agric Ecosyst Environ* 164:80–99.
- Paustian K, et al. (2016) Climate-Smart soils. *Nature* 532:49–57.
- Guo LB, Gifford R (2002) Soil carbon stocks and land use change: A meta analysis. *Glob Change Biol* 8:345–360.
- Don A, Schumacher J, Freibauer A (2011) Impact of tropical land-use change on soil organic carbon stocks—A meta-analysis. *Glob Change Biol* 17:1658–1670.
- Wei X, Shao M, Gale W, Li L (2014) Global pattern of soil carbon losses due to the conversion of forests to agricultural land. *Sci Rep* 4:4062.
- Neill C, et al. (1997) Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern Brazilian Amazon. *Ecol Appl* 7:1216–1225.
- Hoyle F, D'Antuono M, Overheu T, Murphy D (2014) Capacity for increasing soil organic carbon stocks in dryland agricultural systems. *Soil Res* 51:657–667.
- Churchman GJ, Noble A, Bailey G, Chittleborough D, Harper R (2014) *Clay addition and redistribution to enhance carbon sequestration in soils*, Soil Carbon (Springer, New York), pp 327–335.
- Lal R (2001) World cropland soils as a source or sink for atmospheric carbon. *Adv Agron* 71:145–191.
- Pugh T, et al. (2015) Simulated carbon emissions from land-use change are substantially enhanced by accounting for agricultural management. *Environ Res Lett* 10:124008.
- Smith P, et al. (2016) Global change pressures on soils from land use and management. *Glob Change Biol* 22:1008–1028.
- Batjes NH, et al. (2017) Wosis: Providing standardised soil profile data for the world. *Earth Syst Sci Data* 9:1–14.
- Friedl MA, et al. (2010) Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ* 114:168–182.
- Scharlemann JP, Tanner EV, Hiederer R, Kapos V (2014) Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manage* 5:81–91.
- Batjes N (2016) Harmonized soil property values for broad-scale modelling (wise30sec) with estimates of global soil carbon stocks. *Geoderma* 269:61–68.
- Hengl T, et al. (2017) Soilgrids250m: Global gridded soil information based on machine learning. *PLoS one* 12:e0169748.
- Intergovernmental Panel on Climate Change (2006) *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, Prepared by the National Greenhouse Gas Inventories Programme, eds Eggleston HS, Buendia L, Miwa K, Ngara T, Tanabe K (Institute for Global Environmental Strategies, Hayama, Japan).
- Ogle SM, Breidt FJ, Paustian K (2005) Agricultural management impacts on soil organic carbon storage under moist and dry climatic conditions of temperate and tropical regions. *Biogeochemistry* 72:87–121.
- Nachtergaele F, et al. (2010) *Global Land Degradation Information System (GLADIS)*. An Information Database for Land Degradation Assessment at Global Level Land Degradation Assessment in Drylands (Food Agric Org, Rome), Tech Rep 17 Beta Vers.
- Spera SA, Galford GL, Coe MT, Macedo MN, Mustard JF (2016) Land-use change affects water recycling in Brazil's last agricultural frontier. *Glob Change Biol* 22:3405–3413.
- Conant RT, Paustian K, Elliott ET (2001) Grassland management and conversion into grassland: Effects on soil carbon. *Ecol Appl* 11:343–355.
- Fujisaki K, et al. (2015) From forest to cropland and pasture systems: A critical review of soil organic carbon stocks changes in Amazonia. *Glob Change Biol* 21:2773–2786.
- Pachepsky Y, Hill RL (2017) Scale and scaling in soils. *Geoderma* 287:4–30.
- West TO, et al. (2004) Carbon management response curves: Estimates of temporal soil carbon dynamics. *Environ Manage* 33:507–518.
- Sanderman J, Baldock JA (2010) Accounting for soil carbon sequestration in national inventories: A soil scientist's perspective. *Environ Res Lett* 5:034003.
- Senthilkumar S, Basso B, Kravchenko A, Robertson G (2009) Contemporary evidence of soil carbon loss in the US corn belt. *Soil Sci Soc Am J* 73:2078–2086.
- Han G, et al. (2008) Effect of grazing intensity on carbon and nitrogen in soil and vegetation in a meadow steppe in inner Mongolia. *Agric Ecosyst Environ* 125:21–32.
- McSherry ME, Ritchie ME (2013) Effects of grazing on grassland soil carbon: A global review. *Glob Change Biol* 19:1347–1357.

40. Houghton RA, et al. (2012) Carbon emissions from land use and land-cover change. *Biogeosciences* 9:5125–5142.
41. Houghton R (1999) The annual net flux of carbon to the atmosphere from changes in land use 1850–1990. *Tellus B* 51:298–313.
42. Lal R (2004) Soil carbon sequestration impacts on global climate change and food security. *Science* 304:1623–1627.
43. Smith P, et al. (2008) Greenhouse gas mitigation in agriculture. *Philos Trans R Soc Lond B Biol Sci* 363:789–813.
44. Eggleston H, Buendia L, Miwa K, Ngara T, Tanabe K (2006) *IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme* (Intergov Panel Clim Change, Hayama, Japan).
45. Smith P, et al. (2005) Carbon sequestration potential in European croplands has been overestimated. *Glob Change Biol* 11:2153–2163.
46. Jenny H (1994) *Factors of Soil Formation: A System of Quantitative Pedology* (Courier Corp, North Chelmsford, MA).
47. McBratney AB, Santos MM, Minasny B (2003) On digital soil mapping. *Geoderma* 117:3–52.
48. Scull P, Franklin J, Chadwick O, McArthur D (2003) Predictive soil mapping: A review. *Prog Phys Geogr* 27:171–197.
49. Köchy M, Don A, van der Molen MK, Freibauer A (2015) Global distribution of soil organic carbon-part 2: Certainty of changes related to land use and climate. *Soil* 1:367–380.
50. Harris I, Jones P, Osborn T, Lister D (2014) Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int J Climatol* 34:623–642.
51. Robinson N, Regetz J, Guralnick RP (2014) Earthenv-dem90: A nearly-global, void-free, multi-scale smoothed, 90m digital elevation model from fused ASTER and SRTM data. *ISPRS J Photogramm Remote Sens* 87:57–67.
52. Conrad O, et al. (2015) System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci Model Dev* 8:1991–2007.
53. Sayre R, et al. (2014) *A New Map of Global Ecological Land Units—An Ecophysio-graphic Stratification Approach* (US Geol Surv, Washington, DC).
54. Mitchell T, GDAL Developers (2014) *Geospatial Power Tools: GDAL Raster & Vector Commands* (Locate Press, Chugiak, AK).
55. R Development Core Team (2009) *R: A Language and Environment for Statistical Computing* (R Found Stat Comput, Vienna).
56. Wright MN, Ziegler A (2016) Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 77:18.
57. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. arXiv:1603.02754.