Better estimates of soil carbon from geographical data: a revised global approach

s.duarteguardia@gmail.com; peri.pablo@inta.gob.ar; wulf.amelung@uni-bonn.de; douglas.sheil@nmbu.no; shawn.laffan@unsw.edu.au; nils.borchard@gmail.com; michael.bird@jcu.edu.au; wouter.dieleman@gmail.com; david.pepper@canberra.edu.au; brian.zutta@gmail.com; jobbagy@gmail.com; lsilva7@uoregon.edu; s.bonser@unsw.edu.au; bgonzalo@agro.uba.ar; pineiro@agro.uba.ar; martinez.luna.mariajose@gmail.com; annette.cowie@dpi.nsw.gov.au; brenton.ladd@gmail.com

SANDRA DUARTE-GUARDIA[a]*, PABLO L. PERI[b], WULF AMELUNG[c], DOUGLAS SHEIL[d], SHAWN W. LAFFAN[e], NILS BORCHARD[f,g,h,i], MICHAEL I. BIRD[j], WOUTER DIELEMAN[j], DAVID A. PEPPER[k], BRIAN ZUTTA[l], ESTEBAN JOBBAGY[m], LUCAS C.R. SILVA[n], STEPHEN P. BONSER[o], GONZALO BERHONGARAY[p], GERVASIO PIÑEIRO[q], MARIA-JOSE MARTINEZ[r], ANNETTE L. COWIE[s,t], BRENTON LADD[o,r]

[a]Universidad Nacional de la Patagonia Austral (UNPA) (9400) Rio Gallegos, Santa Cruz, Argentina.
[b]INTA EEA, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)-cc332 (9400) Rio Gallegos, Santa Cruz, Argentina.
[c]Institute of Crop Science and Resource Conservation (INRES), Soil Science and Soil Ecology, University of Bonn, D-53115 Bonn, Germany.
[d]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003 NO-1432 Ås, Norway. and Center for International Forestry Research (CIFOR), Jalan Cifor Rawajaha, Situ Gede, Bogor Barat, Kota Bogor, Jawa Barat, 16115, Indonesia
[e]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia.
[f]Forschungszentrum Jülich GmbH, Agrosphere Institute (IBG-3), 52425 Jülich
[g]Center for International Forestry Research, Jalan CIFOR, Situ Gede, Sindang Barang, Bogor 16115, Indonesia
[h] Ruhr-University Bochum, Institute of Geography, Soil Science/Soil Ecology, Universitätsstrasse 150, 44801 Bochum, Germany
[i]Natural Resources Institute Finland (Luke), FI-31600 Jokioinen, Finland
[j]College of Science, Technology and Engineering and Centre for Tropical Environmental and Sustainability Science, James Cook University, PO Box 6811, Cairns, Queensland.
[k]Institute for Applied Ecology, University of Canberra, ACT 2617, Australia.
[l]Programa Nacional de Conservación de Bosques, Ministerio del Ambiente (National Forest Conservation Program, Ministry of the Environment), Lima, Perú.
[m]Grupo de Estudios Ambientales, IMASL, Universidad Nacional de San Luis y CONICET, Ejercito de los Andes 950, San Luis 5700 Argentina.
[n]Environmental Studies Program. Department of Geography. Institute of Ecology & Evolution. University of Oregon, Eugene, OR, USA 97403.

[o]Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia.
[p] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)/ Facultad de Ciencias Agrarias, Universidad Nacional del Litora. Kreder 2805, Esperanza Santa Fe.
[q]IFEVA, Facultad de Agronomía, Universidad de Buenos Aires, CONICET. Laboratorio de Análisis Regional y Teledetección (LART-FAUBA). San Martin 4453, Buenos Aires, Argentina and Facultad de Agronomia, Universidad de la Republica, Garzon 780, Montevideo, Uruguay.
[r]Escuela de Agroforestería, Universidad Científica del Sur. Lima 33, Perú.
[s]NSW Department of Primary Industries, Armidale, NSW, Australia.
[t]School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia.

* Correspondence: Sandra Duarte, +511 986966059, Universidad Nacional de la Patagonia Austral (UNPA) (9400) Rio Gallegos, Santa Cruz, Argentina. s.duarteguardia@gmail.com.

Better estimates of soil carbon from geographical data: a revised global approach

Abstract

Soils hold the largest pool of organic C on Earth, yet, we are unable to include the soil organic carbon (SOC) reservoir into climate change mitigation strategies, because our database for ecosystems where human impacts are minimal is still fragmentary. The aim of this study is to provide a tool for generating a global baseline of SOC stocks. We use partial least square (PLS) regression on measured SOC and freely available geographic datasets that describe climate, topography, productivity and soil C across biomes. The accuracy of the model was determined by the root mean square deviation (RMSD) of predicted SOC against 100 independent measurements. The best predictors were primary productivity, climate, topography, biome classification and soil type. The largest C stocks were found in boreal forests and tundra, averaging 254 ± 14.3 and 310 ± 15.3 t ha^{-1} in the top 1m of soil, respectively. Deserts had the lowest observed C stocks (53.2 ± 6.3 t ha^{-1}). Solar radiation, potential evapotranspiration, and mean annual temperature were negatively correlated with SOC stocks, whereas soil water content was positively correlated with SOC stocks. RMSD (0.68) represented approximately 14% of observed soil C stock variation with overestimation for extremely low C stocks and underestimation for extremely high C stocks. Using PLS regression can provide baseline predictions of SOC stocks, which may serve for estimating the maximum SOC that may be sequestered across biomes.

Keywords: soil organic carbon; geographic information systems; climate; global; pristine ecosystems; baseline

Introduction

Scientists and policy makers recognise that conservation of ecosystems – where historical anthropogenic land use change is minimal – hereafter “pristine ecosystems” - offers an effective means for preserving terrestrial carbon stocks and abating greenhouse gas emissions (Ladd and Peri 2013; Stockmann et al. 2013; Scharlemann et al. 2014). As a result, initiatives to reduce greenhouse gas emissions from deforestation and enhance ecosystem carbon stocks such as REDD+ (reducing emissions from deforestation and forest degradation in developing countries) have gained increasing traction, providing financial incentives for the protection of existing terrestrial carbon stocks in conservation and production forests (UN-REDD 2016). However, such initiatives focus primarily on aboveground carbon stocks (Scharlemann et al. 2014), for which precise geospatial data are being produced (Asner et al. 2010; Ryan et al. 2012; Ladd and Peri 2013). Much larger amounts of carbon, however, are stored in soil.

It is estimated that the amount of C stored in soils (1500- 2400 Pg) is three times larger than that found aboveground biomass (450 – 600 Pg) (Houghton 2014, Scharlemann et al. 2014). Nevertheless, soil organic carbon (SOC) stocks are notoriously difficult to predict, especially in pristine ecosystems, which are often difficult to access. Carbon storage depends on the existing interactions between all processes in the soil-vegetation-atmosphere system (Silva 2017). Therefore, soil organic C stocks vary depending on i) carbon inputs resulting from ecosystem factors such as climate, soil, and vegetation and ii) mechanisms controlling C mean residence times (Jobbagy and Jackson 2000; Lützow et al. 2006; Laganiere et al. 2010; Don et al. 2011; Kögel-Knaber and Amelung 2014). In pristine ecosystems mean residence times of C stored in plant biomass depends on the vegetation type (Mahli and Grace 2000; Galbraith et al. 2013) while in soils mean residence times are C-pool specific and may span decades or even centuries (Kuzyakov 2006; Schmidt et al. 2011). Therefore, it is not surprising that SOC has become a focus for climate change mitigation policy. For example, the recent COP21 initiative aimed at increasing SOC stocks to reduce atmospheric CO_2 concentration by 4 parts per mill per year (URL: http://4p1000.org, Minasny et al. 2017). On the other hand, other strategies to mitigate climate change such as REDD+ still lack a consideration of soils. A major challenge for preserving or achieving SOC enhancement goals is the paucity of reliable data on current carbon stocks of pristine ecosystems. Therefore, this study was designed to improve our understanding and ability to predict SOC stocks in pristine ecosystems across biomes.

Quantifying SOC stocks of natural systems requires several challenges to be overcome. Accurately quantifying SOC stocks at landscape level is difficult because landscapes are heterogeneous, dynamic, and multiple factors interact to impact both formation and degradation of SOC (Bui et al. 2009; Don et al. 2011; Powers et al. 2011; Ladd et al. 2013; Manning et al. 2015). In addition, a precise estimation of SOC stocks is challenging, particularly in pristine ecosystems that are remote and difficult to access (Scharlemann et al. 2014). However, these ecosystems are expected to store large SOC pools at a long-term steady state equilibrium maximum (Sanderman et al. 2017). This provides target baseline values for the amount of SOC that could be sequestered in soils. Already Minasny et al. (2017) point out, as also discussed by many other authors (e.g., Post and Kwon 2000; Lal 2004b; Lam et al. 2013; Sanderman et al. 2010), that it is mostly managed agricultural lands that provide potential for sequestering SOC, yet, certainly not more than the equilibrium value than exists under climax vegetation in pristine environments (Lal 2004b). Only when knowing this maximum amount of C stored in soils, policy makers may have a clue to value or quantity of soil C that any attempt for mitigating climate change through better soil management could achieve.

To address the difficulties of accurately quantifying SOC stocks, global maps of SOC have been developed. For example, the maps developed by the European Union, and the International Soil Reference and Information Centre (ISRIC, SoilGrids products) (Hiederer and Köchy 2011; Hengl et al. 2014; Hengl et al. 2017). The European Union´s global SOC map is largely based on the FAO system of soil taxonomy (Hiederer and Köchy 2011). This taxonomy is limited as it accounts for soil forming processes, but it does not integrate climatic controls on SOC stabilization in a quantitative manner. In contrast, the SoilGrids products (at 1 km and 250m resolution) are based on automated mapping and machine learning techniques and consider the CLORPT soil forming factors (i.e., Climate, Organisms, Relief, Parent material, and Time; Jenny, 1941). This approach has the drawback that it calculates SOC from soil properties, such as bulk density, soil organic matter, and particle distribution, which has much higher uncertainty than SOC calculated directly from empirical data (Hengl et al. 2014; Hengl et al. 2017). Yet another drawback is that these maps were developed to predict current SOC stocks without considered degree of human impact, leaving the question of predictions for pristine ecosystems, which could be used as baseline C reference, aside.

To address these important limitations of previous SOC mapping efforts, here we propose a solution or complementary approach that considers both CLORPT soil formation factors and uses empirical SOC data to refine SOC stock predictions at large scales. Specifically, this study predicts global SOC stocks for the top 1 m of the soil profile, by using freely available geographic data sets [climate, topography, primary productivity (indicated by normalized difference vegetation index: NDVI), and soil characteristics] on SOC measurements in pristine ecosystems across biomes. Our objective was to ease the challenge of predicting and assessing SOC baselines; this may help guide land management and restoration efforts, e.g. in pursuit of Land Degradation Neutrality (LDN), where SOC is one of the LDN indicators (Lal 2016; Cowie et al. 2018).

Materials and methods

The SOC stock dataset

A database of SOC stocks in pristine ecosystems was compiled from peer-reviewed publications. Data were obtained from 38 sources (see appendix S1) compiled with electronic search engines (i.e. Google Scholar and ISI web of science) and by reference to relevant meta-analyses (Don et al. 2011; Powers et al. 2011; Deng et al. 2016). In addition a subset of sites from the ISRIC-WISE international soil profile dataset were included in this study (Batjes 1995). To ensure the best coverage of the globe, we also included sites from the National Soil Survey Center (NSSA, www.nrcs.usda.gov), which is operated by the United States Department of

Agriculture (USDA), and comprises a large dataset of soil properties in several sites across the globe. For the sites derived from the NSSA, we calculated the C stocks using the following equation:

$$C = \%C \times \delta a \times h \times Cm$$

Where C is the C stocks in t ha^{-1}, %C is the percentage of C in soil, δa is the bulk density (g cm^{-3}), h is the depth of measurement (in cm) and Cm is the correction for coarse fraction measured for each soil pit (Soil Survey Staff 2011).

For this compilation we only included data when SOC stocks were measured to 1 m depth. And, as the aim of the study was to generate a baseline for C stocks, we only included sites that could represent pristine ecosystems, i.e., where no information indicated that they had been used for arable cropping in the past or which showed unusual climax vegetation. Hence, we equate these sites as pristine, i.e., with minimal anthropogenic impacts. Therefore, all observed sites are considered to approximate SOC stock at equilibrium. Assessing the effects of land use on SOC stocks was not the objective of this study.

The final dataset included 1346 observations of SOC stocks in pristine ecosystems, plus an independent dataset of 100 sites for validation (Figure 1). The 1346 sites represent adequately every biome (Figure 2) and are distributed across all 5 ice-shield free continents (Figure 1). Moreover, the 100 sites used for model accuracy assessment are also distributed across the 5 continents and 11 of the 12 biome classes used for the analysis. The only biome type that was not represented in the validation dataset was tropical conifer forests, however, this is the biome with the lowest coverage on Earth (0.3%, Figure 2).

Data sources used to compile the GIS-derived independent variables

The selection of freely available, GIS-derived, independent variables was guided by the state factor model (Dokuchaev 1879; Jenny 1941), where soils are described as a function of Climate, Organisms, Relief, Parent material and Time (CLORPT, see introduction and discussion).

Estimates of climate parameters and solar radiation for each site (i.e. 1346 observed sites for modelling and 100 sites for accuracy calculation) were derived from the WorldClim data set (Hijmans et al. 2005; Fick and Hijman 2017). See Table 1 for a description of the WorldClim parameters and columns H to Z in appendix S1 for parameter estimates. WorldClim contains geographic surfaces for 19 different climatic parameters that describe rainfall, temperature and variation in those parameters at a resolution of 30 arc seconds (approximately 1 km^2 at the equator). The second version of WorldClim (Fick and Hijman 2017) contains monthly values of solar radiation (in kJ m^{-2} day^{-1}) at a spatial resolution of approximately 1 km^2 at the equator. We used ArcGIS 10.1 to obtain estimates of the mean yearly values (column AA, appendix S1).

Global potential evapotranspiration and aridity index (Trabucco and Zomer 2009) were downloaded from the CGIAR-CSI GeoPortal (http://www.csi.cgiar.org) at a spatial resolution of 30 arc seconds (~1 km^2) (columns AH and AI in appendix S1).

The period for this dataset corresponds to 1950-2000. Soil water balance (SWB) variables (Columns AE to AG in appendix S1) for the 1346 sites were also downloaded from the CGIAR-CSI GeoPortal (http://www.csi.cgiar.org). These variables also have a spatial resolution of approximately 1 km^2 (30 arc seconds) (Trabucco and Zomer 2010).

Biome classification was obtained either from the source publications (column AL in appendix S1) or from the Global terrestrial ecoregions map, based on the World Wildlife Fund biome classification (https://www.worldwildlife.org). Biome classifications have proven useful as a variable for modelling global ecological processes (Prentice et al. 1992).

Primary production at each site was approximated using normalized difference vegetation index (NDVI). We used the global map of the European Space Agency (ESA), (http://maps.elie.ucl.ac.be, see their Land Surface Seasonality products). This map provides NDVI values for the period 1999-2012 at 1 km^2 spatial resolution divided in two measurements: Aggregated Mean (AggMean: greenness dynamic over a period of 7 days) and standard deviation (Std of the mean NDVI, representing the variability of NDVI over a 7 day period) (columns AB and AC in appendix S1).

Elevation data (column AD, appendix S1) were downloaded from the SRTM30 project at approx. 1 km^2 resolution (https://dds.cr.usgs.gov/srtm/). They comprise a combination of data from the Shuttle Radar Topography Mission and the U.S. Geological Survey's GTOPO30 data sets.

Information on the geology was obtained from the world geologic maps available from the USGS database (http://energy.usgs.gov) to determine geologic units and age (columns AJ, AK and AS in appendix S1). The geochemical, mineralogical, and physical properties of rocks (column AR in appendix S1) for the 1346 sites were derived from the Global Lithological Map (Hartmann and Moosdorf 2012).

Soil taxonomic class of the soil at each site (Column AM, appendix S1) was obtained directly from the source publication whenever this information was provided. When this information was not reported we extracted soil types from the SoilGrids soil taxonomy map at 250m of spatial resolution, based on the World Reference Base (WRB) (Hengl, et al. 2017). Global Landform Classification data (Columns AN to AQ, appendix S1) were downloaded from the European Commission (http://eusoils.jrc.ec.europa.eu). These maps provide information on relief classes, steepness, soil texture and local convexity following Meybeck et al. (2001) and Iwahashi and Pike (2007) at a spatial resolution of 1 km^2.

GIS processing

The extraction and calculation of the required GIS-derived independent variables was done using ArcGIS 10.1 (Redlands, California). All processed geographical datasets where uploaded to ArcGIS. Using the extraction tool of ArcGIS, we extracted the corresponding pixel value for each observed site, or nearest neighbor pixel if the corresponding pixel of target site was not indexed, and added them to the raw dataset (Appendix S1). For the NDVI values, we obtained estimates of the mean values (of the 7 day-period) for the AggMean and Std of the NDVI per pixel, developing new raster images. Then, we extracted the value per site. Annual average values of solar radiation, were calculated based on the monthly average for each pixel, and then the corresponding value was extracted. Finally, we calculated the total area of each biome.

Analyses

The dependent variable, SOC stock to 1 m depth, was natural log transformed prior to analysis to reduce the influence of outliers (Quinn and Keough 2002). The resulting dataset was analysed using Partial Least Squares (PLS) Regression (XlStat, AddinSoft, Paris) because this type of analysis allows the inclusion of both quantitative and categorical independent variables (see appendix S2). This analysis transforms the original variables to new ones (or components), which are linear combinations from the originals. To evaluate the strength of correlation between our observed SOC stocks and the model predictions, we performed a linear regression (XlStat, AddinSoft, Paris) on the naturally log transformed values of the observed and the predicted SOC stock values (appendix S2).

To determine de relation between the most representative variables and the naturally log transformed values of the observed SOC stocks, we performed linear regressions and ANOVA (XlStat, AddinSoft, Paris).

To estimate the accuracy of the model, we calculated the predicted values (using the model equation) for the dataset of 100 sites, then, used root mean square deviation ($RMSD = [\sum(\text{observed-predicted})^2/n]^{1/2}$) (Figure 1, appendix S1).

Results

The values of SOC stocks (to 1 m depth) in published literature ranged from 3 to 3069 t ha^{-1} and varied considerably across biomes: boreal forests (253.6 ± 14.3 all values in t ha^{-1} ± standard error), deserts (53.2 ± 6.3), Mediterranean forests (141.5 ± 20.4), montane grass and shrubs (99.1 ± 6.5), temperate broadleaf and mixed forests (151.6 ± 15.4), temperate conifer forests (220.9 ± 23.7), temperate grasslands (104 ± 4.6), tropical and subtropical coniferous forests (93.9 ± 20.6), tropical and subtropical dry forests (130.7 ± 40.5), tropical and subtropical grasslands (100.9 ± 10), tropical and subtropical moist forests (143.1 ± 9), tundra (310.2 ± 15.3) (see also appendix S1 for further detail).

The model derived from the PLS regression proved effective in predicting measured SOC (Figure 3, y = -0.09+1.01x, R^2 = 0.49, appendix S2). The predictive model included all the independent variables (qualitative and quantitative) (see appendix S2, cell B7099). From the 100 independent sites used to evaluate the accuracy of the model, we found that the RMSD (0.68) represented approximately 14% of the average observed C stock values. However, it overestimated values for extremely low C stocks (e.g. deserts) and underestimated values for extremely high C stocks (e.g. tundra or boreal forests) (Figure 3).

The PLS regression analysis indicated that the 10 most important variables in the model were solar radiation, the standard deviation of normalized difference vegetation index (i.e. NDVI-std), potential evapotranspiration, soil water balance variables (i.e. soil water content and aridity stress on vegetation), biome classification, landform, soil type and temperature-related climate variables (i.e.

mean diurnal range and mean annual temperature) (Figure 4); see appendix S2 for statistical details of the PLS regression analysis and model. The further evaluation of the effect of solar radiation, potential evapotranspiration, soil water content, biome classification and mean annual temperature on soil C stocks is shown in Figure 5. Linear regressions for the quantitative variables and the ANOVA for biome classification gave the following results.

Higher values of solar radiation (y = 6.4-0.0001x, R^2=0.27, Figure 5a), potential evapotranspiration (y = 5.33-0.001x, R^2=0.10, Figure 5c) and mean annual temperature (y = 4.86-0.02x, R^2=0.09, Figure 5e) were related to low C stocks. However, high soil water content (y = 3.95+0.01x, R^2=0.12, Figure 5b) was related to high C stocks.

Further, the highest stored C stocks in the top 1m were found in boreal forests and tundra, with no detectable statistical difference between these biomes. Temperate conifer forests and mediterranean forests were statistically similar to both tundra and boreal forests, but also similar to temperate broadleaf and mixed forests. All tropical and subtropical forests (i.e. moist, dry and coniferous) had no detectable statistical difference. Tropical and subtropical grasslands were similar to both temperate grasslands and montane grass and shrub, although the latter were statistically different from each other. In contrast, deserts had the lowest observed C stocks and were statistically different from all other biomes (Figure 5d).

Discussion

Our results show that baseline SOC stocks can be reasonably well predicted from freely available geographic and climatic datasets. Using a combination of environmental variables. Across 100 different pristine ecosystem locations we found that 49% of average variance in SOC stocks across biomes was predicted. The most powerful predictive variables in our model were: solar radiation, variability in plant productivity as approximated by NDVI-std, temperature-related variables, soil water balance, landform, soil type and biome classification. Sanderman et al. (2017), who developed two predictive models, one for historical and one for current soil C stocks (i.e. prior to and after land use changes), also found that the most important variables in these models (historical and current C stocks) were related to climate and topographic attributes, such as temperature, elevation, landform and precipitation.

Within and across biomes, variation in aboveground biomass is known to be positively correlated with variations in temperature and precipitation (Sala et al. 1988; Chapin et al. 2002). We also found that these variables affected SOC stocks, though not necessarily in the same simple direction (Figure 5e). Specifically, our PLS analysis indicated that SOC stock was negatively correlated with solar radiation, potential evapotranspiration and temperature, but was positively correlated with soil water content and primary productivity. This is explained by the fact that beside primary productivity (see above), soil temperature and moisture also control the rate of litter decomposition and soil C accumulation (Stockmann et al. 2013). High temperatures increase microbial activity and decomposition rates, releasing C from soil to the atmosphere (Chapin et al. 2002; Gilmanov et al. 2007) at rates that depend on interactions between microbial and plant communities (Winsome et al. 2017). On the other hand, low temperatures decrease both primary productivity and decomposition, explaining why cold environments preserve ancient C-rich soil horizons under conditions that currently limit plant growth (Adams et al. 2011). We also found that elevated soil water content correlated with elevated C stocks. This relationship could again be explained by the increase in primary productivity due to higher soil moisture and higher precipitation values, lower soil temperatures at elevated water supply, and enhanced SOC storage where water stagnates, e.g., in depressions (e.g., Batjes 1996; Chapin et al. 2002; Doetterl et al. 2013; Adhikari et al. 2014). In any case, we were able to describe significant predictable relationships that show how SOC stocks might change with climate.

The key driver behind soil temperature and moisture is solar radiation. It generally promotes plant productivity (Chapin et al. 2002; Gilmanov et al. 2007; Ladd, et al. 2009; Ladd, et al. 2014) and therefore soil C inputs, in ecosystems that are not water-limited. The negative effect of solar radiation on SOC levels could, in part, be related to the role of solar radiation in the decomposition of surface litter (Austin and Vivanco 2006; Borchard et al. 2014) or to litter quality, which may change with primary production (Bouwman 1990). Beside temperature and geography, litter quality (e.g. nutrients, C:N ratio or lignin:N ratio) is a key regulator of initial surface litter decomposition rates (Zhang et al. 2008) and thus likely also for formation of stable organic matter (Neff et al. 2002).

Variability in primary productivity (i.e NDVI-std) was an important predictive variable but this may be due to vegetation characteristics influenced by topography instead of primary productivity itself (Mulder et al. 2011). For example, in Argentinean

Nothofagus forests, evergreen species (i.e. less variability in primary productivity) are more common in lowlands, near lakes and locations with alluvial soils, while upland forests are generally deciduous (i.e. more variability in primary productivity; Peri et al. 2012). Similarly, Brazilian Cerrado evergreen forests are typically found on alluvial and riparian sites whereas drought deciduous species occur in dryer upland Cerrado sites (Silva et al. 2008). Because SOC tends to accumulate in depressions (Roman-Sanchez et al. 2018, Figure 1S Appendix) and because plant productivity is more stable in these same locations, this may mean that continuous growth and less variability in primary productivity, leads to a positive correlation between NDVI-std and SOC. More research is required to elucidate causal relationships and the mechanisms that determine them across the globe. However, despite the questions related to mechanisms the data do clearly show that it is possible to predict SOC storage from freely available climatic and GIS data, thus offering scientist and policy makers a tool for valuing the success of conservation measures or for estimating the potential of a given degraded biome to sequester C; i.e. up until the baseline in pristine ecosystems is reached.

Implications for large-scale C mapping

We have shown that using now readily available geographic datasets can lead to a good ability to model, predict and assess SOC stocks that could in turn provide better targeting of habitat protection to help maximise both carbon storage and conservation of nature (Sheil et al. 2016). Our analysis provides a clear demonstration of a strategy that could help improve the current generation of global SOC maps in pristine ecosystems and also shows that independent variables that can be derived simply, such as simple ecosystem classifications, can add significant power to empirical models that aim to predict soil C stocks.

As the quality and availability of geospatial and satellite data that relate to the mechanistic factors that drive soil formation (Climate, Organisms, Relief, Parent material and Time; i.e. CLORPT, see first paragraph of methods) (Dokuchaev 1879; Jenny 1941) advance, more accurate predictions should become possible (Ladd et al. 2014). Additionally, advances in GIS capability, data handling and statistical methods, mean that we are now in a position to develop a quantitative version of the CLORPT model that allows accurate prediction of SOC stocks in pristine ecosystems on a large scale. With better resolution of NDVI, soil and topographic data within a given landscape, producing highly resolved SOC maps at catchment scale is then feasible based on advanced statistical or machine learning tools like random forest regression (e.g., Grimm et al. 2008; Wiesmeier et al. 2014; Hounkpatin et al. 2018). Here, we found that within the biomes mean standard deviation ranged from 36 t C ha^{-1} in tropical and subtropical coniferous forests to 277 t C ha^{-1} in tundra biome, thus tropical and subtropical coniferous forest were the least represented biome both in the dataset (N=3) and of biomes in general, representing only 0.3% of the total ice free-terrestrial surface. While tundra, was the best represented biome (N=325) and also the one with the largest area (~40%) of the total ice free terrestrial surface (Figure 2). This could indicate that larger biome coverage means more variability in C stocks.

Yet, large difference between C stocks across extreme biomes, for example, the ones existing between tundra or boreal forests and deserts, includes the risks of an overestimation and underestimation of extreme values (very low and high SOC stocks) in our model, thus calling for even more sophisticated non-linear learning algorithms for potential improvements to achieve more exact SOC estimations for extreme environments.

Implications for conservation and management strategies

The development of a global C stock baseline can facilitate efforts to include SOC in schemes such as REDD+ that aim to conserve ecosystems and their carbon stocks and for achievement of Land Degradation Neutrality (Cowie et al. 2018). Including SOC in such schemes with the aim of increasing SOC stocks, could favour better land management practices. These practices include conservation, biological farming and the use of compost and eventually biochar (Adams et al. 2011; Ladd et al. 2018). Although these practices can increase SOC stocks, there are several limitations in adopting them. These limitations include negative impact on food security (from lower yields of organic systems and from the conversion of agricultural lands to woodlands), limited biomass resources for compost and biochar, limited economic resources of land, among other limiting factors (Adams et al. 2011; Poulton et al. 2017). Additionally, soil management strategies should also focus on the restoration of degraded ecosystems and desertified landscapes, restoration and improved productivity of grasslands, and reforestation (Lal 2004a; Adams et al. 2011; Lal 2013), which in some cases have been shown to increase SOC beyond levels found in the most productive terrestrial ecosystems (Silva et al. 2013; Silva et al. 2015). Also, when developing management or conservation strategies changes in species composition across biomes should be taken into account. This is because such variations cause shifting ecosystem C fixation and water balance, having important consequences for SOC stocks and groundwater recharge (Maxwell et al. 2018).

In our study, we do not include the effects of anthropogenic land use change (LUC) on SOC stocks, since we focus on pristine ecosystems. However, LUC have been linked to important C stocks variations, especially in transitioning from pristine ecosystems to cropping (Smith et al. 2008; Berhongaray et al. 2013; Rabbi et al. 2015; Sanderman et al. 2017). Recent work has shown that combining LUC and freely available GIS data may also predict changes in SOC stocks in croplands (e.g., Sandermann et al. 2017; Hounkpatin et al. 2018), thus providing an option to rate former SOC losses against baseline SOC levels as assessed here – which is basically the inverse of estimating the maximum amount of SOC that might potentially be re-sequestered. Yet, all these changes depend on time, and while changes in SOC losses may be rapid, occuring within a few years to decades (Dalal and Mayer 1986; Solomon et al. 2007), it may take decades until SOC levels restore, frequently not to the level of the pristine environment (Robles and Burke 1998; Post and Kwon, 2000; Lal 2004b; Preger et al. 2010). For future Earth system modelling it will likely be needed to augment the statistical modelling approach used here with process-based models to gain mechanistic insight and facilitate prediction and analysis of possible future scenarios.

Conclusion

We have shown that SOC stocks of pristine ecosystems can be accurately predicted using available data. Further, 49% of average variance for pristine ecosystems across biomes was predicted. Our model provides a theoretical baseline for accurately identifying sites with a high potential for C storage through restoration and rehabilitation of degraded land. Additionally, it shows where the highest SOC stocks are in pristine ecosystems, and where efforts can be best made to conserve soil C. And perhaps most importantly, this information can be used in conjunction with estimates of current SOC stocks (obtained from other sources, e.g. current SOC stocks from Sanderman et al. 2017) to identify sites with the greatest gap between current and potential SOC stocks. This could help guide the selection of sites for restoration, to achieve maximum negative emissions, through carbon sequestration in soil.

Supplementary material

Data to this article can be found online at: http://www.unpa.edu.ar/cecyt/1876/grupo/forestal_silvopastoril/actividades

Acknowledgements

We thank INTA Argentina for supporting our work in Patagonia. Nils Borchard was placed as an integrated expert at the Centre for International Migration and Development (CIM). CIM is a joint venture of the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH and the International Placement Services (ZAV) of the German Federal Employment Agency (BA).

Additional information

Conflict of Interest: The authors declare that they have no conflict of interest.

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

References

Adams M, Crawford J, Field D, Henakaarchchi N, Jenkins M, McBratney A, de Courcelles VdR, Singh K, Stockmann U, Wheeler I (2011) Managing the soil-plant system to mitigate atmospheric CO_2. Discussion paper for the Soil Carbon Sequestration Summit, 31 January – 2 February 2011. The United States Studies Centre at the University of Sydney.

Adhikari K, Hartemink AE, Minasny B, Kheir RB, Greve MB, Greve MH (2014) Digital mapping of soil organic carbon contents and stocks in Denmark. PLoS One 9 (8): e105519.

Asner GP, Powell GVN, Mascaro J, Knapp DE, Clark JK, Jacobson J, Kennedy-Bowdoin T, Balaji A, Paez-Acosta G, Victoria E, Secada L, Valqui M, Hughes RF (2010) High-resolution forest carbon stocks and emissions in the Amazon. Proceedings of the National Academy of Sciences 107: 16738-16742.

Austin AT, Vivanco L (2006) Plant litter decomposition in a semi-arid ecosystem controlled by photodegradation. Nature 442: 555-558.

Batjes NH (1995) A homogenized soil data file for global environmental research: a subset of FAO, ISRIC and NRCS profiles (Version 1.0). Working Paper and preprint 95/10b, International Soil Reference and Information Centre, Wageningen. In.

Batjes NH (1996) Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47: 151–163.

Berhongaray G, Alvarez R, De Paepe J, Caride C, Cantet R (2013) Land use effects on soil carbon in the Argentine Pampas. Geoderma 192: 97-110.

Borchard N, Siemens J, Ladd B, Möller A, Amelung W (2014) Application of biochar to sandy and silty soil failed to increase maize yield under common agricultural practice. Soil and Tillage Research 144: 184-194.

Bouwman AF (1990) Soils and the Greenhouse Effect. John Wiley & Sons Inc. 596 pp.

Bui E, Henderson B, Viergever K (2009) Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. Global Biogeochemical Cycles 23(4).

Chapin FS, Matson PA, Mooney HA (2002) Principles of Terrestrial Ecosystem Ecology, Principles of Terrestrial Ecosystem Ecology. Springer. doi: 10.1007/978-1-4419-9504-9.

Cowie AL, Orr BJ, Sanchez VMC, Chasek P, Crossman ND, Erlewein A, Louwagie G, Maron M, Metternicht GI, Minelli S, Tengberg AE (2018) Land in balance: The scientific conceptual framework for Land Degradation Neutrality. Environmental Science and Policy *79*: 25-35.
Dalal RC, Mayer RJ (1986) Long-term trends in fertility of soils under continuous cultivation and cereal cropping in Southern Queensland. III. Distribution and kinetics of soil organic carbon in particle size fractions. Australian Journal Soil Research 24: 293-300.

Deng L, Zhu G, Tang Z, Shangguan Z (2016) Global patterns of the effects of land-use changes on soil carbon stocks. Global Ecology and Conservation 5: 127–138.

Doetterl S, Stevens A, van Oost K, Quine TA, van Wesemael B (2013) Spatially-explicit regional-scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. Geoderma 204: 31–42.

Dokuchaev V (1879) Mapping the Russian Soils. Imperial Univ. of St. Petersburg, St. Petersburg, Russia.

Don A, Schumacher J, Freibauer A (2011) Impact of tropical land-use change on soil organic carbon stocks – a meta-analysis. Global Change Biology 17(4): 1658-1670.

Fick SE, Hijmans RJ (2017) Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37: 4302-4315.

Galbraith D, Malhi Y, Affum-Baffoe K, Castanho ADA, Doughty CE, Fisher RA, Lewis SL, Peh KSH, Phillips OL, Quesada CA, Sonké B, Lloyd J (2013). Residence times of woody biomass in tropical forests. Plant Ecology and Diversity 6(1): 139–157.

Gilmanov TG, Soussana JF, Aires L, Allard V, Ammann C, Balzarolo M, Barcza Z, Bernhofer C, Campbell CL, Cernusca A, Cescatti A, Clifton-Brown J, Dirks BOM, Dore S, Eugster W, Fuhrer J, Gimeno C, Gruenwald T, Haszpra L, Hensen A, Ibrom A, Jacobs AFG, Jones MB, Lanigan G, Laurila T, Lohila A, Manca G, Marcolla B, Nagy Z, Pilegaard K, Pinter K, Pio C, Raschi A, Rogiers N, Sanz MJ, Stefani P, Sutton M, Tuba Z, Valentini R, Williams ML, Wohlfahrt G (2007) Partitioning European grassland net ecosystem CO_2 exchange into gross primary productivity and ecosystem respiration using light response function analysis. Agriculture Ecosystems and Environment 121: 93-120.

Grimm R, Behrens T, Marker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island - digital soil mapping using Random Forests analysis. Geoderma 146: 102–113.

Hartmann J, Moosdorf N (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. Geochem. Geophys. Geosyst. 13: Q12004. doi:10.1029/2012GC004370.

Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG, Gonzalez MR (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS ONE 9(8): e105992. doi:10.1371/journal.pone.0105992

Hengl T, de Jesus JM, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12(2): e0169748. https://doi.org/10.1371/journal.pone.0169748.

Hiederer R, Köchy M (2011) Global soil organic carbon estimates and the Harmonized World Soil Database. EUR 25225 EN. Publications Office of the European Union. 79 pp.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25(15): 1965-1978.

Houghton RA (2014). The contemporary carbon cycle. In HDHK Turekian (Ed.), *Treatise on Geochemistry (Second Edition)* (pp. 399–435). Oxford: Elsevier. https://doi.org/http://dx.doi.org/10.1016/B978-0-08-095975-7.00810-X.

Hounkpatin OKL, Op de Hipt F, Bossa AY, Welp G, Amelung W (2018) Soil organic carbon stocks and their determining factors in the Dano catchment (Soutwest Burkina Faso). Catena 166: 298-309.

Iwahashi J, Pike RJ (2007) Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86(3): 409-440.

Jenny H (1941) Factors of soil formation: a system of quantitative pedology. Dover books on Earth sciences. Dover Publications.

Jobbagy E, Jackson R (2000) The vertical distribution of soil organic carbon and its relation to climate vegetation. Ecological Applications 10(2): 423-436.

Kögel-Knaber I, Amelung W (2014) Dynamics, chemistry, and preservation of organic matter in soils. Reference Module in Earth Systems and Environmental Sciences, from Treatise on Geochemistry. Geochemistry 12: 157-215.

Kuzyakov Y (2006) Sources of CO_2 efflux from soil and review of partitioning methods. Soil Biology and Biochemistry 38(3): 425–448.

Ladd B, Bonser SP, Peri PL, Larsen JR, Laffan SW, Pepper DA, Cendón DI (2009) Towards a physical description of habitat: quantifying environmental adversity (abiotic stress) in temperate forest and woodland ecosystems. Journal of Ecology 97: 964-971.

Ladd B, Laffan SW, Amelung W, Peri PL, Silva LCR, Gervassi P, Bonser SP, Navall M, Sheil D (2013) Estimates of soil carbon concentration in tropical and temperate forest and woodland from available GIS data on three continents. Global Ecology and Biogeography 22(4): 461-469.

Ladd B, Peri PL (2013) REDD+ en Latinoamérica: el caso de Perú. Bosque 34(2): 125-128.

Ladd B, Peri PL, Pepper DA, Silva LCR, Sheil D, Bonser SP, Laffan SW, Amelung W, Ekblad A, Eliasson P, Bahamonde H, Duarte-Guardia S, Bird M (2014) Carbon isotopic signatures of soil organic matter correlate with leaf area index across woody biomes. Journal of Ecology 102(6): 1606-1611.

Ladd B, Dumler S, Loret de Mola E, Anaya de la Rosa R, Borchard N (2018) Incremento de rentabilidad en producción del maíz en perú: N-fertilizantes y biochar. The Biologist (Lima) 15: 23-35.

Laganiere J, Angers DA, Pare D (2010) Carbon accumulation in agricultural soils after afforestation: a meta-analysis. Global Change Biology 16: 439-453.

Lal R (2004a) Soil Carbon sequestration to mitigate climate change. Geoderma 123: 1-22.

Lal R (2004b) Soil carbon sequestration impacts on global climate change and food security. Science 304: 1623-1627.

Lal R (2013) Soil carbon management and climate change. Carbon Management 4(4): 439-462

Lal R (2016) Beyond COP21: Potential and challenges of the "4 per Thousand" initiative. Journal of Soil and Water Conservation 71: 20A-25A

Lam SK, Chen D, Mosier AR, Roush R (2013) The potential for carbon sequestration in Australian agricultural soils is technically and economically limited. Scientific Reports 3: 2179.

Lützow Mv, Kögel-Knabner I, Ekschmitt K, Matzner E, Guggenberger G, Marschner B, Flessa H (2006) Stabilization of organic matter in temperate soils: Mechanisms and their relevance under different soil conditions – A review. European Journal of Soil Science 57: 426–445.

Malhi Y, Grace J (2000) Tropical forests and atmospheric carbon dioxide. Tree 15: 332-337.

Manning P, Vries FT, Tallowin JR, Smith R, Mortimer SR, Pilgrim ES, Harrison KA, Wright DG, Quirk H, Benson J (2015) Simple measures of climate, soil properties and plant traits predict national-scale grassland soil carbon stocks. Journal of Applied Ecology 52(5): 1188-1196.

Maxwell TM, Silva LCR, Horwath WR (2018) Integrating effects of species composition ans soil properties to predict shifts in montane forests carbon-water relations. PNAS 1718864115.

Meybeck M, Green P, Vörösmarty C (2001) A new typology for mountains and other relief classes: an application to global continental water resources and population distribution. Mountain Research and Development 21(1): 34-45.

Minasny B, Malone BP, McBratney AB, Angers DA, Arrouays D, Chambers A, Chaplot V, Chen Z, Cheng K, Das BS, Field D, Gimona A, Hedly CB, Hong SY, Mandal B, Marchant BP, Martin M, McConkey BG, Mulder VL, O'Rourke S, Richer-de-Forges AC, Odeh I, Padarian J, Paustian K, Pan G, Poggio L, Savin I, Stolbovoy V, van Wesemael B, Winowiecki L (2017) Soil carbon 4 per mille. Geoderma 292: 59-86.

Mulder VL, de Bruin S, Schaepman ME, Mayr TR (2011) The use of remote sensing in soil terrain mapping - A review. Geoderma 162: 1-19.

Neff JC, Townsend AR, Gleixner G, Lehman SJ, Turnbull J, Bowman WD (2002) Variable effects of nitrogen additions on the stability and turnover of soil carbon. Nature 419: 915-917.

Peri PL, Ladd B, Pepper DA, Bonser SP, Laffan SW, Amelung W (2012) Carbon (δ13C) and nitrogen (δ15N) stable isotope composition in plant and soil in Southern Patagonia's native forests. Global Change Biology 18: 311-321.

Post WM, Kwon KC (2000) Soil Carbon Sequestration and Land-Use Change: Processes and Potential. Global Change Biology 6: 317-328.

Poulton P, Johnston J, Macdonald A, White R, Powlson D (2017) Major limitations to achieving "4 per 1000" increase in soil organic carbon stock in temperate regions: Evidence from long-term experiments at Rothamsted Research, United Kigdom. Global Change Biology 2018;00: 1-22.

Powers JS, Corre MD, Twine TE, Veldkamp E (2011) Geographic bias of field observations of soil carbon stocks with tropical land-use changes precludes spatial extrapolation. Proceedings of the National Academy of Sciences 108(15): 6318-6322.

Preger AC, Kösters R, Du Preez CC, Brodowski S, Amelung W (2010) Carbon sequestration in secondary pasture soils: a chronosequence study in the South African Highveld. Eur. J. Soil Sci 61: 551-562.

Prentice IC, Cramer W, Harrison SP, Leemans R, Monserud RA, Solomon AM (1992) Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography 19: 117-134.

Quinn GP, Keough MJ (2002) Experimental design and data analysis for Biologists. Cambridge, UK: Cambridge University Press.

Rabbi SMF, Tighe M, Delgado-Baquerizo M, Cowie A, Robertson F, Dalal R, Page K, Crawford D, Wilson BR, Schwenke G, McLeod M, Badgery W, Dang YP, Bell M, O'Leary G, Liu DL, Baldock J (2015) Climate and soil properties limit the positive effects of land use reversion on carbon storage in Eastern Australia. Scientific Reports 5: 17866.

Robles MD, Burke IC (1998) Soil organic matter recovery on conservation reserve program fields in southeastern Wyoming. Soil Science Society of America Journal 62: 725-730.

Roman-Sanchez A, Vanwalleghm T, Peña A, Giráldez JV (2018) Controls on soil Carbon storage from topography andvegetation in a rocky, semi arid landscapes. Geoderma. 311: 159-166.

Ryan CM, Hill T, Woollen E, Ghee C, Mitchard E, Cassells G, Grace J, Woodhouse IH, Williams M (2012) Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. Global Change Biology 18(1): 243-257.

Sala OE, Parton WJ, Joyce LA, Lauenroth W (1988) Primary production of the central grassland region of the United States. Ecology 69: 40-45.

Sanderman J, Farquharson R, Baldock J (2010) Soil Carbon Sequestration Potential: A review for Australian agriculture. A report prepared for Department of Climate Change and Energy Efficiency. CSIRO: EP10121.

Sanderman J, Hengl T, Fiske GJ (2017) Soil Carbon debt of 12,000 years of human land use. PNAS 114: 9575-9580.

Scharlemann JP, Tanner EV, Hiederer R, Kapos V (2014) Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management 5(1): 81-91.

Schmidt MWI, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA, Kleber M, Kögel-Knaber I, Lehmann J, Manning DAC, Nannipieri P, Rasse DP, Weiner S, Trumbore SE (2011) Persistence of soil organic matter as an ecosystem property. Nature 478: 49–56.

Sheil D, Ladd B, Silva LCR, Laffan SW, Van Heist M (2016) How are soil carbon and tropical biodiversity related? Environmental Conservation 43: 231-241.

Silva LCR, Sternberg LSL, Haridasan M, Hoffmann WA, Miralles-Wilhelm F, Franco AC (2008) Expansion of gallery forests into central Brazilian savannas. Global Change Biology 14: 2108-2118.

Silva LCR, Corrêa RS, Doane TA, Pereira EI, Horwath WR (2013) Unprecedented carbon accumulation in mined soils: the synergistic effect of resource input and plant species invasion. Ecological Applications 23: 1345-1356.

Silva LCR, Doane TA, Corrêa RS, Valverde V, Pereira EI, Horwath WR (2015) Iron-mediated stabilization of soil carbon amplifies the benefits of ecological restoration in degraded lands. Ecological Applications 5: 1226-1234.

Silva LCR (2017) Carbon sequestration beyond tree longevity. Science 355: 1141.

Smith P, Martino D, Cai Z, Gwary D, Janzen H, Kumar P, McCarl B, Ogle S, O'Mara F, Rice C (2008) Greenhouse gas mitigation in agriculture. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363(1492): 789-813.

Soil Survey Staff (2011) Soil Survey Laboratory Information Manual. Soil Survey Investigations Report No. 45, Version 2.0. R. Burt (ed.). U.S. Department of Agriculture, Natural Resources Conservation Service.

Solomon D, Lehmann J, Kinyangi J, Amelung W, Lobe I, Ngoze S, Riha S, Pell A, Verchot LL, Mbugua D, Skjemstad J, Schäfer T (2007) Long-term impacts of anthropogenic perturbations on dynamics and speciation of organic carbon in tropical forest and subtropical grassland ecosystems. Global Change Biol. 13: 511-530.

Stockmann U, Adams MA, Crawford JW, Field DJ, Henakaarchchi N, Jenkins M, Minasny B, McBratney AB, de Courcelles VdR, Singh K (2013) The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems and Environment 164: 80-99.

Trabucco A, Zomer R (2009) Global aridity index (global-aridity) and global potential evapo-transpiration (global-PET) geospatial database. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.csi.cgiar.org/(2009). Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. In. CGIAR Consortium for Spatial Information.

Trabucco A, Zomer R (2010) Global soil water balance geospatial database. CGIAR Consortium for Spatial Information. Published online, available from the CGIAR-CSI GeoPortal at: http://www.cgiar-csi.org.

UN-REDD (2016) Technical Resource Series 3: Towards a common understanding of REDD+ under the UNFCCC. UN-REDD Programme Secretariatt.

Wiesmeier M, Barthold F, Spörlein P, Geuß U, Hangen E, Reischl A, Schilling B, Angst G, Lützow Mv, Kögel-Knabner I (2014) Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). Geoderma Reg. 1: 67–78.

Winsome T, Silva LCR, Scow CR, Doane TA, Power RF, Howarth WR (2017) Plant-microbe interactions regulate carbon and nitrogen accumulation in forest soils. Forest Ecology and Management 384: 415-423.

Zhang D, Hui D, Luo Y, Zhou G (2008) Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. J. Plant Ecol. 1: 85–93.

Table legends

Table 1 A description of the WorldClim parameters (i.e. climatic variables) extracted with ArcGIS 10.1 (Redlands, California) for each observed site used for the development of the model (further detail available at http://www.worldclim.com, and in Hijmans *et al.* 2005)

Table 1

Parameter	Description
BIO1	Annual mean temperature
BIO2	Mean diurnal range [mean of monthly (max. temp.–min. temp.)]
BIO3	Isothermality (BIO2/BIO7) (x 100)
BIO4	Temperature seasonality (standard deviation x 100)
BIO5	Max. temperature of warmest month
BIO6	Min. temperature of coldest month
BIO7	Temperature annual range (BIO5–BIO6)
BIO8	Mean temperature of wettest quarter
BIO9	Mean temperature of driest quarter
BIO10	Mean temperature of warmest quarter
BIO11	Mean temperature of coldest quarter
BIO12	Annual precipitation
BIO13	Precipitation of wettest month
BIO14	Precipitation of driest month
BIO15	Precipitation seasonality (coefficient of variation)
BIO16	Precipitation of wettest quarter
BIO17	Precipitation of driest quarter
BIO18	Precipitation of warmest quarter
BIO19	Precipitation of coldest quarter

Figure legends

Fig. 1 The global distribution of the 1346 soil pits that were geo-referenced and used to measure SOC stock (t ha^{-1}) to 1-meter depth and the 100 sites used for validation of the model obtained with partial least square (PLS) regression analysis. Raw data and source publications from which the data were obtained are provided in appendix S1

Fig. 2 Number of observations of C stocks according to each biome classification (following the WWF terrestrial biome classification) and the area (millions of km^2)

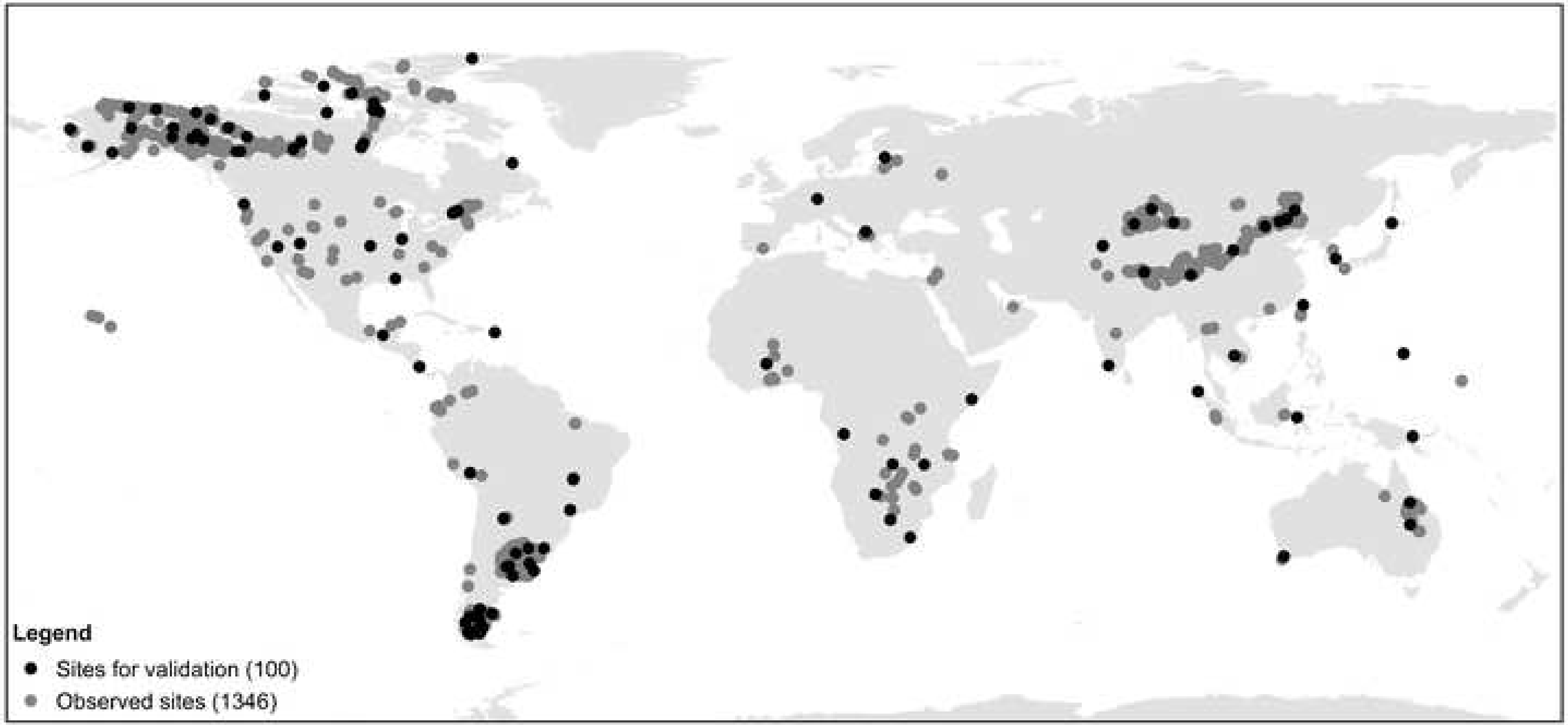
that each biome represent in the globe. BF= boreal forests, D= deserts, MF= Mediterranean forests, MGS= montane grass and shrub, TBF= temperate broadleaf and mixed forest, TmG= temperate grasslands, TrCF= tropical and subtropical coniferous forests, TDF= tropical and subtropical dry forests, TrG= tropical grasslands, TMF= tropical and subtropical moist forest, T= tundra

Fig. 3 The correlation between predictions of the natural log of SOC stocks derived from the PLS regression model constructed with independent variables derived from freely available geographic datasets and the same measured (natural log transformed) values of SOC stock obtained from the peer reviewed literature (y = -0.09+1.01x, R^2 = 0.49). See appendix S2 for further statistical detail of the partial least square (PLS) regression model. RMSD: root mean square deviation

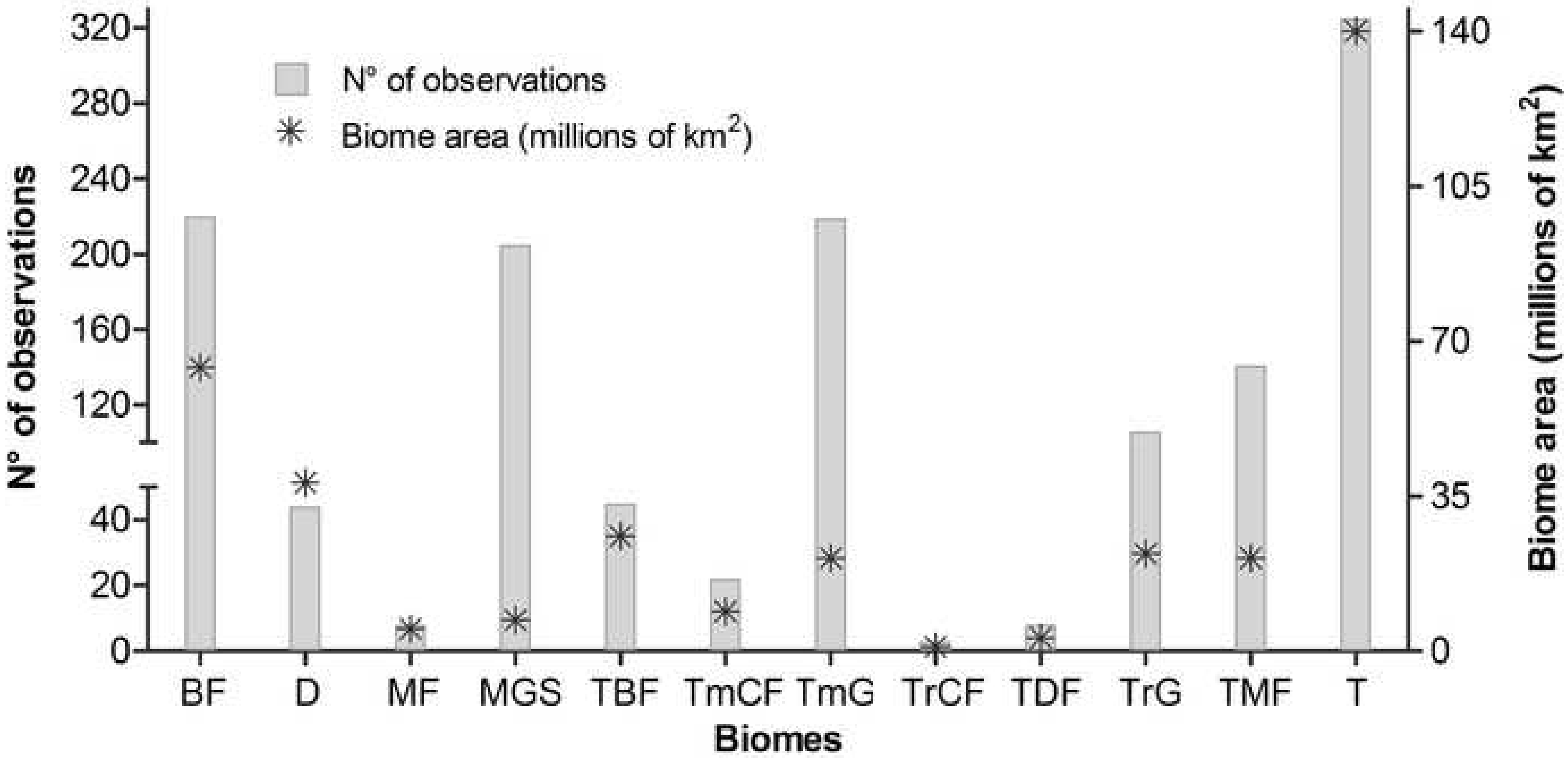
Fig. 4 The top ten variables for SOC stock prediction, based on node frequencies from the PLS regression analysis (see appendix s2 for details). Solar radiation (kJ m^{-2}) for the period 1970-2000, NDVI = Normalized Difference Vegetation Index, in this case the mean value of the standard deviation of the recorded values between 1999 and 2012, Landform = relief class (e.g. lowlands, plains, high altitude mountains, etc.), soil water content in and aridity stress on vegetation for the period 1950-2000. PET is Potential Evapotranspiration (1950-2000), soil type according to WRB classification and biome classification (e.g. tundra, boreal forest, temperate grasslands, etc.)

Fig. 5 Relation between observed C stocks and the most representative variables in the projection. Obtained from the PLS regression model for the 1346 sites with measured C stocks (t ha^{-1}). a) Solar Radiation (kJ m^{-2}) vs. log transformed C stocks, b) Soil water content (%) vs. log transformed C stocks, c) Potential evapotranspiration vs. log transformed C stocks, d) ANOVA for C stocks according to biome classification (Fisher LSD test, α=0.99), different letters mean statistical differences. BF= boreal forests, D= deserts, MF= Mediterranean forests, MGS= montane grass and shrub, TBF= temperate broadleaf and mixed forest, TmG= temperate grasslands, TrCF= tropical and subtropical coniferous forests, TDF= tropical and subtropical dry forests, TrG= tropical grasslands, TMF= tropical and subtropical moist forest, T= tundra. e) Annual mean temperature (°C) vs. log transformed C stocks

line figure

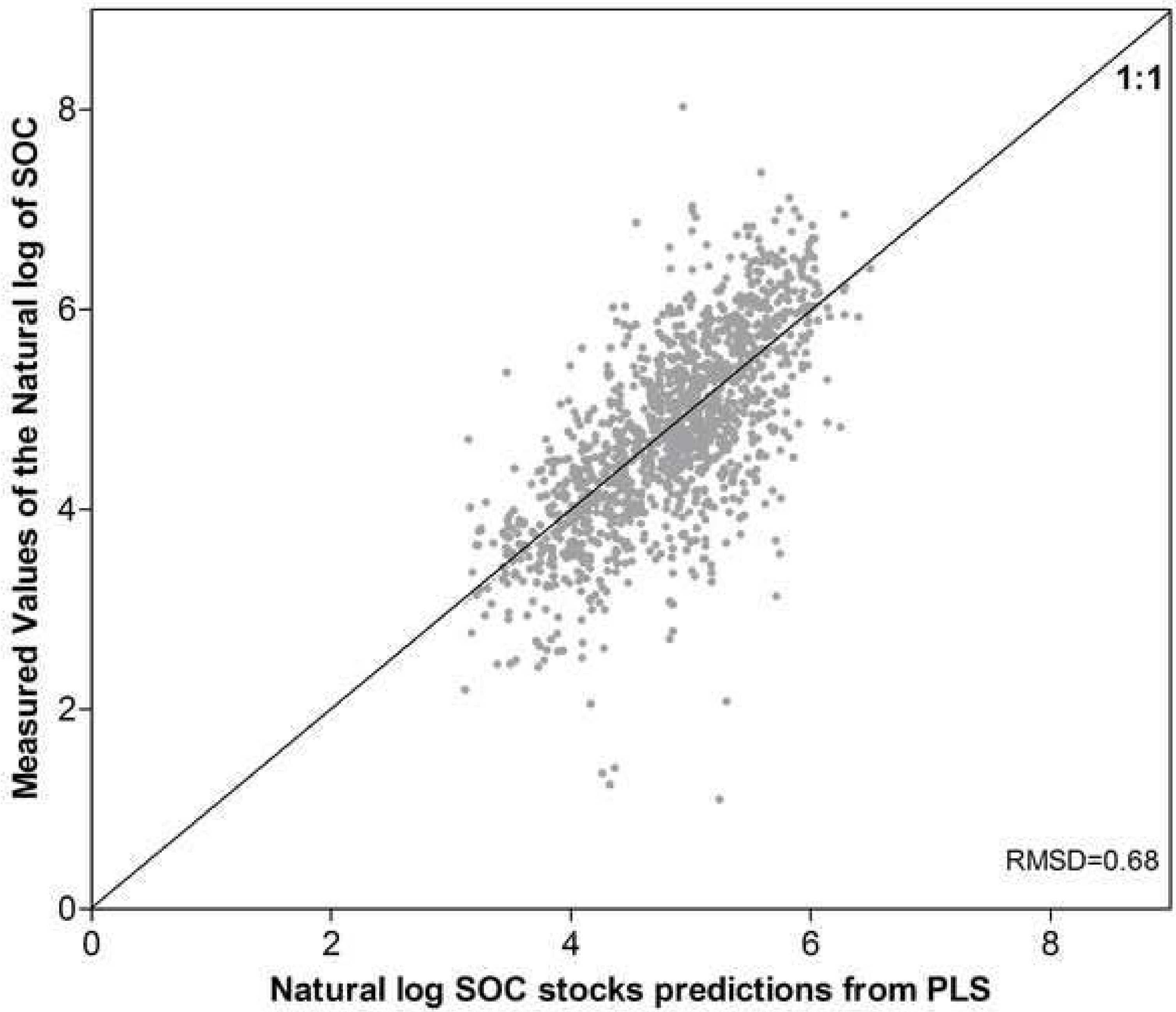


line figure



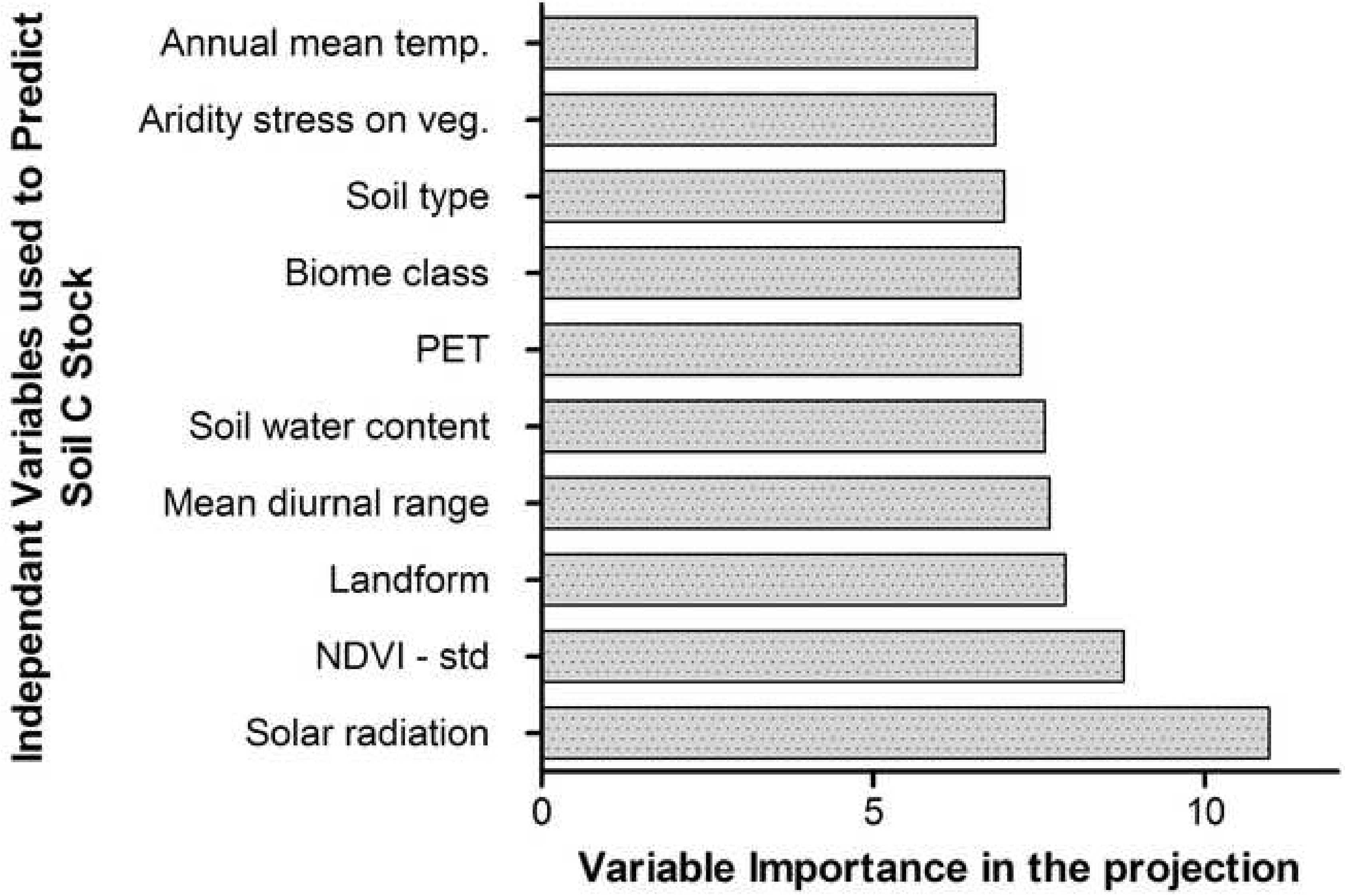
N° of observations
Biome area (millions of km²)
320
280
240
200
160
120
40
20
0
140
105
70
35
0
BF
D
MF
MGS
TBF
TmCF
TmG
TrCF
TDF
TrG
TMF
T
Biomes
Biome area (millions of km²)

line figure



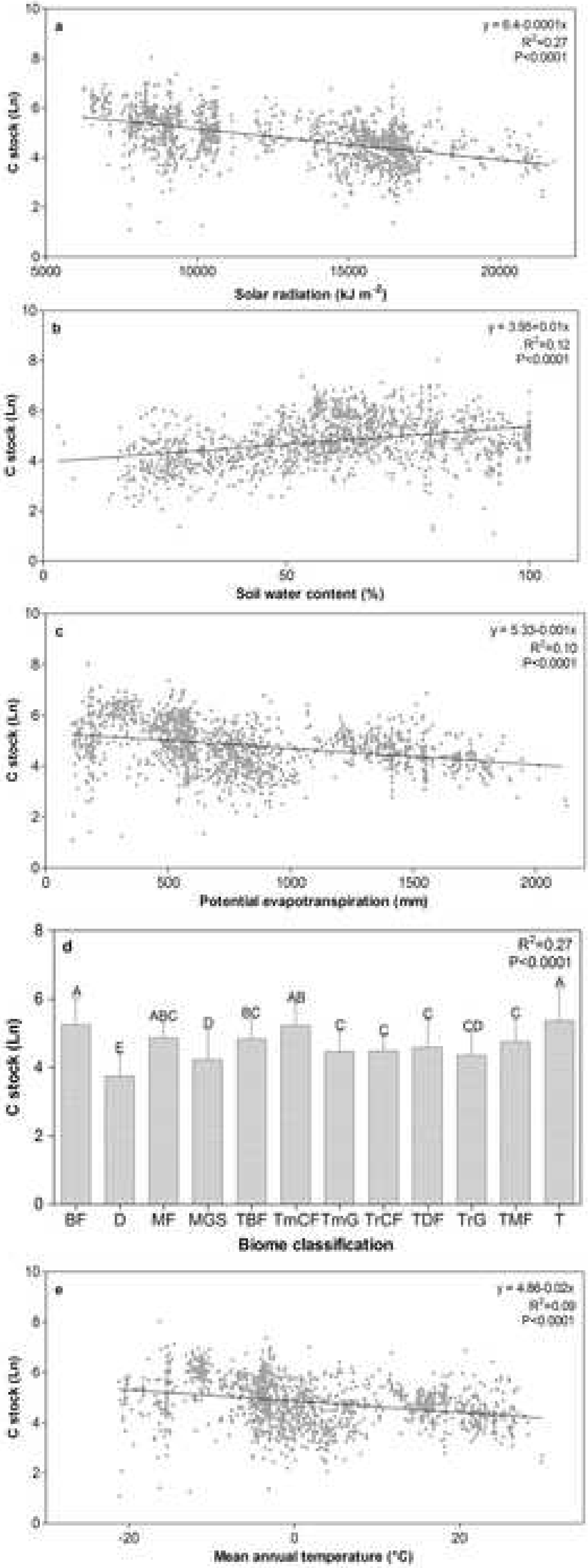
Measured Values of the Natural log of SOC
Natural log SOC stocks predictions from PLS
1:1
RMSD=0.68
0
2
4
6
8

line figure



Independant Variables used to Predict
Soil C Stock
Annual mean temp.
Aridity stress on veg.
Soil type
Biome class
PET
Soil water content
Mean diurnal range
Landform
NDVI - std
Solar radiation
0
5
10
Variable Importance in the projection

line figure



a
y = 6.4-0.0001x
R²=0.27
P<0.0001
C stock (Ln)
Solar radiation (kJ m⁻²)
b
y = 3.96+0.01x
R²=0.12
P<0.0001
C stock (Ln)
Soil water content (%)
c
y = 5.33-0.001x
R²=0.10
P<0.0001
C stock (Ln)
Potential evapotranspiration (mm)
d
R²=0.27
P<0.0001
A
E
ABC
D
BC
AB
C
C
C
CD
C
A
C stock (Ln)
BF
D
MF
MGS
TBF
TmCF
TmG
TrCF
TDF
TrG
TMF
T
Biome classification
e
y = 4.86-0.02x
R²=0.09
P<0.0001
C stock (Ln)
Mean annual temperature (°C)