Reply to comments by Petr Capek (RC1)

General comments

Zhao and co-authors present very interesting world-wide statistical analysis of soil physical and chemical characteristics variability in respect to climatic drivers. They create "biome" plots and maps of soil properties. The selected statistical approach is very innovative, which makes the manuscript sound and worth publishing. I am very skeptical about the direct causality. This issue is, however, relatively well covered in the discussion section of the manuscript. Nevertheless, there is one issue that is not covered in the discussion section at all and that is very important according to my opinion. Authors completely ignore soil orders and associated soil horizons. They report data for top 30 cm of soil. These 30cm can contain either single horizon or several very different horizons with very different physical and chemical properties. My major concern surrounds the results presentation. Authors need to provide more information about the various statistical analyses they used to make Figures 4 – 6 and they also need to clarify various threshold they defined. The manuscript often contains either very vague or very strong statements unsupported by the results (see specific comments). For this type of presented results I think it is especially important to present uncertainty in quantitative terms. Any potential user of the extrapolated maps information/database should be aware of the limitations.

Response: Thanks for your helpful comments. First, we fully agree that ignoring the role of soil orders and horizons might result in uncertainties. In different countries/regions, soil classification was originally based on several different soil classification systems, such as the Unified Soil Classification System, FAO system, USDA Soil Taxonomy, Russia Soil Classification system, Australian Soil Classification system, and Chinese Soil Classification System. These soil classification systems are based on different standards (Carter and Bentley, 2016) and it is difficult to harmonize them (Batjes et al., 2007) and thus to quantify the role of soil orders. Additional, it is the same case for data of soil horizons and we are not able to consider the role of soil horizons. In our database, soil depth was well documented, while some literature data (15% profiles) did not report horizon information. We thus estimated the soil properties by a fixed depth of 30 cm. Nevertheless, the depth of 0-30 cm has been frequently used in the mapping and modelling of surface soil properties at regional and global scales (e.g., Batjes, 1997; Yang et al., 2010; Saiz et al., 2012; Wieder et al., 2013; Shangguan et al., 2014). In the revised manuscript we have added more descriptions and discussed the uncertainties in both Methods and Discussions sessions. Thanks for your understanding!

Second, more details on the statistical analysis and the climate thresholds have been included for Figures 4-6 in the Methods session as follows: "To explore the roles of MAT and MAP as well as their interactions, we averaged soil property values for each MAT×MAP combination by a division of 1°C×100mm and explored quantitative linkages between soil properties and climate variables (MAT/MAP) for different climate types (humid vs. arid; warm vs. cold). Specifically, we used a MAP threshold of 500 mm to differentiate relatively humid versus arid climates (Holdridge, 1967), while a MAT threshold of 10 °C was used to separate relatively warm and cold climates (Trewartha and Horn, 1980). To explore the role of MAT in regulating the critical MAP for a shift from alkaline to acidic soil, we further plotted the critical levels of MAP (100 mm division) at soil pH=7.0 with MAT."

Additionally, we have carefully revised our manuscript according to your suggestions. Please see more details in our reply to your specific comments.

Specific comments:

Page 1, Lines 16 – 17: What is the "critical MAP for the transition from alkaline to acidic soil"? This is some model parameter?

Response: This sentence has been modified as follows: "Soil pH decreased with higher MAP and lower MAT, and the 'critical MAP', which means the corresponding MAP at a soil pH of =7.0 (a shift from alkaline to acidic soil), decreased with lower MAT."

Page 1, Line 18: two dots

Response: Corrected.

Page 1, Lines 18 – 19: I do not understand the meaning of the last sentence of the abstract. Can authors clarify its meaning?

Response: Thanks for your suggestion. Our soil-climate-biome diagram implies strong effects of climate and biota on soil properties. Here we mean that soil properties may shift under global climate warming and land cover change. We have revised this sentence accordingly.

Page 1, Lines 21 – 23: This is very vague statement that deserves more clarity.

Response: Thanks. We have revised the sentence as "As a critical component of the Earth system, soils influence many ecological processes that provide fundamental ecosystem services (Amundson et al., 2015; Milne et al., 2015; Adhikari and Hartemink, 2016)". The next sentences explain this statement by using specific examples: "Soil physical properties, such as bulk density and soil texture, are important for water retention and the preservation of carbon (C) and nutrients (Hassink, 1997; Sposito et al., 1999; Castellano and Kaye, 2009; Stockmann et al., 2013; Jilling et al.,2018), whereas soil chemical properties, such as soil acidity (pH), organic C, and nutrient contents, are essential regulators of nutrient availability and plant growth, further affecting C and nutrient cycling as well as vegetation-climate feedbacks (Davidson and Janssens, 2006; Chapin et al., 2009; Milne et al., 2015) ".

Page 1, Lines 25 – 26: Again, the statement is very vague. I would suggest more specific statement.

Response: Thanks. We have revised the sentence as "soil chemical properties, such as soil acidity (pH), organic C, and nutrient contents, are essential regulators of nutrient availability and plant growth, further affecting C and nutrient cycling as well as vegetation-climate feedbacks (Davidson and Janssens, 2006; Chapin et al., 2009; Milne et al., 2015)."

Page 1, Lines 29 – 30: What doesn't mean "soil stewardship for societal well-being"

Response: We meant soil stewardship to secure soil function and thus sustainable ecosystem services for the human society. We have revised this sentence accordingly.

Page 2, Lines 25 – 26: Please check the superscripts.

Response: Thank you. Typo corrected.

Page 2, Line 30: Third and very important soil-forming factor is the bedrock. I think authors should mention it right away in the introduction, not only in discussion section.
Response: Thanks for your suggestion. We have now mentioned the role of bedrock/parental material in the introduction section as follows: "Although parent material (e.g., bedrock) also plays an important role in affecting soil properties, it affects soil formation at a relatively long time scale (Chesworth, 1973), particularly in the subsoil (Gentsch et al., 2018). In addition, our soil-climate-biome diagram thus focuses on soil properties in the surface layer, given that surface soils are dynamic in time and likely interacting more instantly with climate and vegetation than deeper soils (Weil et al., 2016).".

Page 3, Line 20: Please specify the type of pedologic data that GSD contains. Why these data wasn't used in analysis?
Response: Thanks for the suggestion. Our data base includes pedologic information on soil orders and soil horizons of sampled soil profiles. However, soil pedologic information were originally reported based on several different soil classification systems, such as the Unified Soil Classification System, FAO system, USDA Soil Taxonomy, Russia Soil Classification system, Australian Soil Classification system, and Chinese Soil Classification System. These soil classification systems define soil orders by different standards (Carter and Bentley, 2016) and it is difficult to harmonize them (Batjes et al., 2007) and thus quantify the role of soil orders. In our database, soil depth was well documented, while some literature data (15% profile) did not report horizon information. Therefore, we were not able to consider the role of soil horizons and simply estimated the soil properties by a fixed depth of 30 cm. In fact, the depth of 0-30 cm has been frequently used in the mapping and modelling of surface soil properties (e.g., Batjes, 1997; Yang et al., 2010; Saiz et al., 2012; Wieder et al., 2013; Shangguan et al., 2014). By considering essential climate (mean annual temperature, mean annual precipitation, seasonality of air temperature, seasonality of precipitation), vegetation (mean annual normalized difference vegetation index, and land use type) and topography factors (elevation, slope) that are key to soil formation (Jenny, 1941), our regional Random Forest analysis may partially constrain the uncertainties due to ignoring soil orders and associated soil horizons. In the revised manuscript we have discussed the uncertainties. Thanks for your understanding!

Page 4, Line 4: Authors excluded 10% of all observations (those above and below 95% and 5% quantile respectively). Is there a specific reason for that? In the case of this dataset, it is very difficult to identify outliers. First 30 cm of soil can include one or several different soil horizons with very different physical and chemical properties. Values identified as outliers might be very likely correct and reflect the difference between different soil horizons of different soil orders. Quick look on Fig. 4a suggest to me that all peatlands were very likely removed from the dataset. For that reason I would strongly suggest to keep all data in the dataset.
Response: Thanks for your suggestion. We are sorry that there is a misunderstanding for the proportion of data excluded as outliers. In fact, we only excluded 5% of the data as outliers. We understand that it is difficulty to identify outliers, but outliers were usually excluded based

on a certain criteria to reduce noises and improve the accuracy of prediction (Pleijsier, 1989; Batjes et al., 2007; Jiménez-Muñoz et al., 2015). For instance, two detection criteria are frequently used for outlier identification: 1) samples deviating more than three standard deviations from the mean (±3σ criterion, Jiménez-Muñoz et al., 2015); 2) samples falling outside the range above a certain upper quartile and/or below a certain lower quartile. The latter is useful when the dataset is not normally distributed. In our analysis, we divided the global land into 11 regions to overcome spatial biases of the database and samples above 97.5% and below 2.5% quantile were excluded in each region to obtain a robust variogram. We have mapped the excluded sites and found that these sites are distributed relatively random in space (Fig. R1). This implies that excluding the outliers doesn't bias the datasets. In the revised manuscript, we have included maps for these excluded sites of observation in the supplements.

As you have suggested, we have spent two weeks to conduct a reanalysis by using all data samples and compared the results with those excluded the outliers. We found similar patterns and means of global surface soil properties when using all data samples (Fig. R2; Table R2), but the cross-validated R^2 was obviously decreased for SOCD and STND, especially in US and Russia (Table R1). This implies that excluding the outliers can improve the accuracy of prediction. We thus present the results based on the analysis excluding outliers in the main text and also include the results by using all data samples in the supplement. If the reviewer prefers to present the results based on all data samples in the revised main text, we will do so accordingly.

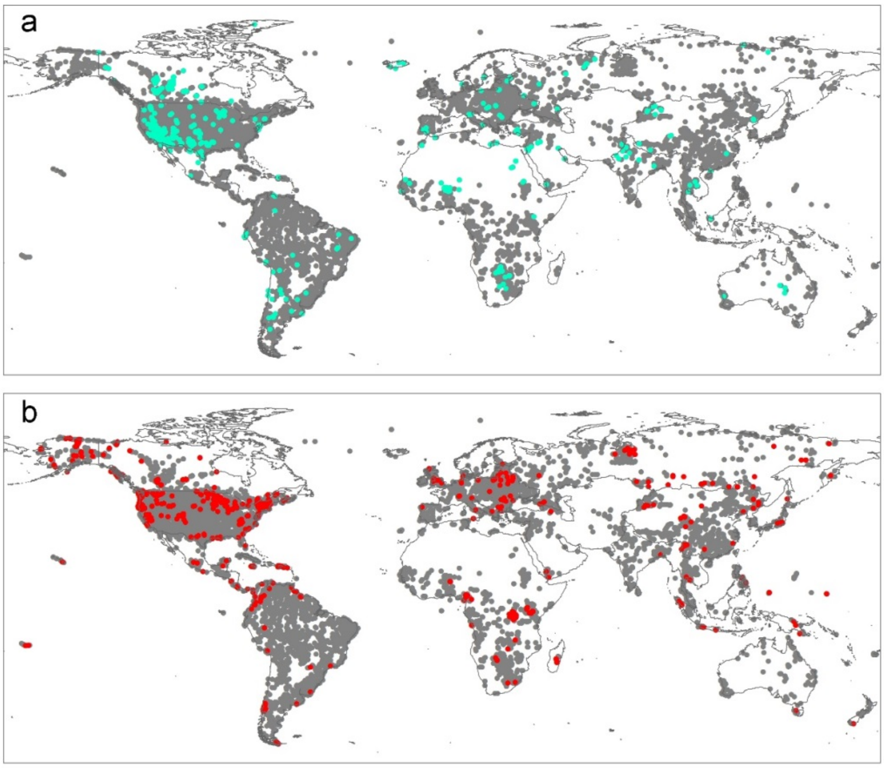

Fig R1. Distribution of global soil profiles. Outliers are plotted as a) green (<2.5%), and b) red (>97.5%) circles (taking soil organic carbon data as an example).

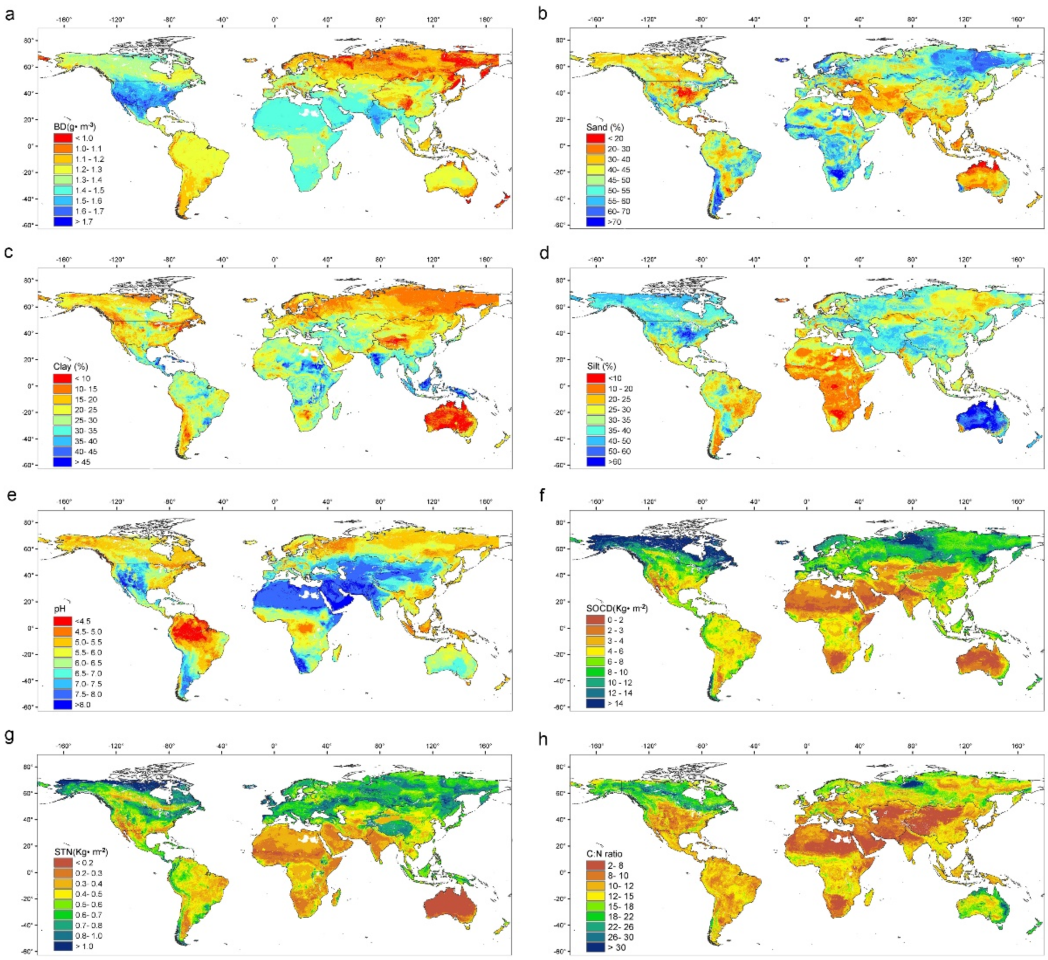

Figure R2. Map of worldwide soil properties in the upper 30-cm soil layer based on analysis using whole datasets. a, BD (bulk density, g cm^{-3}); b, Sand fraction (%); c, Silt fraction (%); d, Clay fraction (%); e, pH; f, SOCD (soil organic carbon density, kg m^{-2}); g, STND (soil total nitrogen density, kg m^{-2}); and h, C:N ratio.

Table R1 Coefficient of determination (R^2) of the Random Forest models excluded the outliers and using all data samples.

Region	95%						ALL DATA					
	Bulk density	Content of sand	Content of clay	pH	SOCD	STND	Bulk density	Content of sand	Content of clay	pH	SOCD	STND
Tropical Asia	0.56	0.27	0.28	0.67	0.37	0.33	0.56	0.33	0.31	0.71	0.42	0.32
Mexico	0.55	0.39	0.41	0.63	0.55	0.49	0.52	0.41	0.45	0.67	0.52	0.48
Africa	0.50	0.44	0.41	0.64	0.512	0.50	0.44	0.49	0.43	0.64	0.58	0.43
Continental US	0.27	0.50	0.44	0.62	0.50	0.46	0.28	0.54	0.46	0.66	0.39	0.43
Canada & Alaska	0.33	0.43	0.49	0.56	0.46	0.40	0.37	0.48	0.53	0.59	0.41	0.39
Russia	0.28	0.21	0.39	0.48	***0.29***	***0.10***	0.24	0.25	0.44	0.47	***0.16***	***0.06***
South America	0.23	0.24	0.16	0.57	0.32	0.24/	0.28	0.22	0.19	0.6	0.3	0.17
Europe	0.29	0.20	0.36	0.49	***0.32***	***0.20***	0.23	0.25	0.37	0.53	***0.15***	***0.08***
East Asia	0.51	0.18	0.39	0.51	0.47	0.28	0.54	0.25	0.33	0.54	0.4	0.18
Australia	0.46	0.65	0.47	0.28	0.47	0.31	0.47	0.67	0.5	0.31	0.51	0.32
West Asia	0.55	0.31	0.41	0.62	0.58	0.36	0.49	0.3	0.37	0.66	0.51	0.41

Notes: Values are the averaged R^2 and RMSE from test dataset of 10-fold cross-validation.

Table R2. Mean values of surface (0-30 cm) soil properties by biome of the world based on analysis using whole datasets.

Biome	Area (10^6 ha)	Bulk density ($g \cdot cm^{-3}$)	Sand (%)	Silt (%)	Clay (%)	pH	SOCD ($kg \cdot m^{-2}$)	STND ($kg \cdot m^{-2}$)	C:N ratio	SOC stock (Pg)	STN stock (Pg)
TroF	1877	1.27±0.10	42.26±9.39	27.39±7.69	30.36±6.11	5.29±0.73	5.71±1.66	0.48±0.13	11.83±1.73	107±0.56	9±0.05
TemF	992	1.28±0.21	45.47±9.27	35.08±7.15	19.45±5.62	5.82±0.81	8.84±3.48	0.64±0.19	14.09±4.33	88±0.47	6±0.04
BF	1435	1.16±0.13	49.82±9.03	33.07±6.15	17.11±4.57	5.39±0.32	11.09±3.60	0.67±0.16	16.90±5.21	159±1.94	10±0.21
TSG	1915	1.34±0.10	48.07±13.17	25.80±16.65	26.12±9.02	6.26±0.79	3.83±1.66	0.32±0.13	12.33±4.30	73±0.62	6±0.04
TGS	1148	1.30±0.16	45.48±10.87	34.51±10.22	20.01±6.72	7.01±0.59	5.33±2.47	0.55±0.20	10.14±3.43	61±0.41	6±0.05
Deserts	2674	1.40±0.12	43.89±10.49	33.22±12.87	22.89±7.27	7.50±0.63	2.80±1.11	0.33±0.13	9.00±3.14	75±1.19	9±0.10
Tundra	644	1.16±0.17	47.76±9.00	36.86±7.43	15.38±3.18	5.40±0.32	13.68±4.55	0.82±0.21	16.95±4.85	88±2.07	5±0.10
Croplands	1984	1.34±0.15	40.68±11.31	33.27±10.10	26.05±7.03	6.41±0.74	6.57±2.83	0.58±0.23	11.38±2.42	130±0.50	12±0.09
PW	159	1.23±0.12	40.61±9.32	34.25±8.16	25.15±6.20	5.85±0.53	9.81±4.16	0.73±0.24	13.58±4.37	16±0.16	1±0.04
Total	12829	1.29±0.17	45.29±10.89	32.24±10.99	22.47±7.96	6.20±1.02	6.95±4.42	0.53±0.23	12.67±4.77	797±4.10	64±0.41

Notes: We include croplands and permanent wetlands in this table, although they are not single biomes. Abbriations: TroF, Tropical forests; TemF, Temperate forests; BF, Boreal forests; TSG, Tropical savannahs and grasslands; TGS, Temperate grasslands and shrublands; PW, Permanent wetlands. Spatial variability of soil properties within each biome was estimated as standard deviations. Uncertainties of total SOC and STN stocks were estimated as standard deviations based on10-fold cross-validation.

Page 4, Line 18: Please check the superscript.
Response: Typo corrected. Thanks.

Page 4, Lines 24 - 25: Can authors clarify the uncertainty estimation of C to N ratio? The uncertainty of the ratio composed of two variables, each with its uncertainty, should be calculated differently than the uncertainty of a single variable.
Response: Uncertainties of SOCD and STND were both assessed based on the bootstrap method. Specifically, a robust estimate was derived by averaging the 10-fold cross-validation samples, and the uncertainty of the estimates was calculated as the standard deviation of the 10-fold cross-validation. We only reported the uncertainties of SOCD and STND in Figure S3 and we believe they could jointly indicate the uncertainty of soil C:N ratio, which was calculated based on predicted SOCD and STND. We have clarified this information in the revised manuscript.

Page 4, 2.4. Statistical analysis: Based on presented results, this section requires more detailed information. Most importantly, authors should clarify the statistics reported in figures 4 – 6. See also the specific comments bellow.
Response: Thanks for your suggestion. Regarding figures 4-6, we first averaged soil property values using a MAT division of 1°C and a MAP division of 100mm. To explore the roles of MAT and MAP as well as their interactions, we then plotted soil properties with climate variables (MAT/MAP) by roughly distinguishing climate types (humid vs arid; warm vs cold). Specifically, a MAP at 500 mm is used to indicate a threshold for arid climates, while a MAT at 10 °C is used to separate relatively warmer and colder climates (see more information in our reply to the general comments above). We have added this information in the revised manuscript.

Page 5, Line 15: Can authors also report here calculated total amounts of SOC and STN (Tab. 1)? It would be very interesting to compare this estimate with previous estimates in respect to chosen statistical approach in the discussion section. The estimates reported in Tab. 1 doesn't seem to me very different from previous estimates. Does it mean that the approach selected by authors is not so different in terms of the outputs?
Response: Thanks for your suggestion. We have reported total storage of SOC and STN in the revised text. We also compare these values with previous estimates in the discussion section. Generally, our results of global SOC and STN storage in surface soils are similar to previous estimates (797 vs 716 Pg, Scharlemann et al., 2014; 63 vs 63-67 Pg, Batjes, 1996). The regional RF analysis has several advantages over traditional approach, such as the ability to model non-linear relationships, handle both categorical and continuous predictors, and resist overfitting and noise features (Breiman, 2001). By using climate, vegetation, topography and land use variables as predictors, our region-specific RF approach likely produces more robust global maps of soil properties in a finer spatial resolution. Moreover, uncertainties of the prediction have also been estimated.

Page 5, Line 25: "soillayer"
Response: Typo corrected.

Page 5, Line 26: The “saturation curve” is mentioned here for the first time. Why saturation curve, what does it mean and how it was calculated/estimated? All that should be thoroughly explained in the statistical analysis section. Also explain the term “saturation threshold”.

Response: Precipitation favors net primary productivity (Del Grosso et al. 2008) and thus the C inputs into the soil. Moreover, precipitation intensifies weathering of the parent material and increases soil acidification, thus increasing formation of SOC-stabilizing minerals (Chaplot et al., 2010; Doetterl et al., 2015) and reducing decomposition of soil organic matter (Meier and Leuschner, 2010). Accordingly, our results indicated that SOCD increased with MAP, while it didn’t exceed a certain threshold because of a constraint of C inputs (Del Grosso et al. 2008). Thus, we used a saturation curve to indicate the relationship between surface SOCD and MAP (cold climate: SOCD = 0.0737×MAP/(1+0.0049×MAP); warm climate: SOCD = 0.0144×MAP/(1+0.0016×MAP)). Based on these curves, the saturation threshold for cold climate and warm climate were 14.5 kg C m^{-2} and 8.0 kg C m^{-2} (Fig R3), respectively. We have included more details in the revised manuscript.

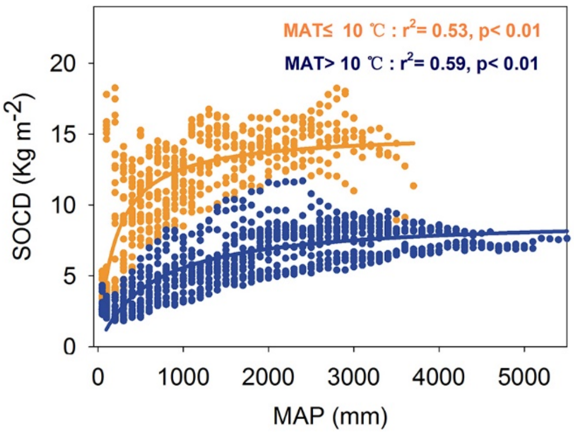

Fig R3. Changes in upper 30-cm soil organic carbon density (SOCD) with mean annual precipitation (MAP). We used MAT of 10 °C as a threshold of transition from cool to warm climate. Each dot shows the average value within each 1 °C MAT and 100 mm MAP.

Page 5, Line 29: Authors should also explain the difference between “cold” and “warm” climates. How was the temperature threshold defined? How the arbitrary selected threshold affects the results?

Response: Many thanks for your comments. We have now included more details on the definition and uncertainties of the MAT thresholds. In the revised manuscript, we used a MAT at 10 °C to separate relatively warmer and colder climates. The threshold MAT was based on a universal thermal scale that used mean monthly temperature (approximately MAT) 10 °C to differentiate “cool” and colder climate from “mild” and warmer climate (Trewartha and Horn, 1980). We have tested the robustness of the trends by using different values of MAT (e.g., 8, 10 and 12°C) and found similar trends (Fig. R4; taking SOC, STN and C:N ratio as an example). We have revised the manuscript accordingly.

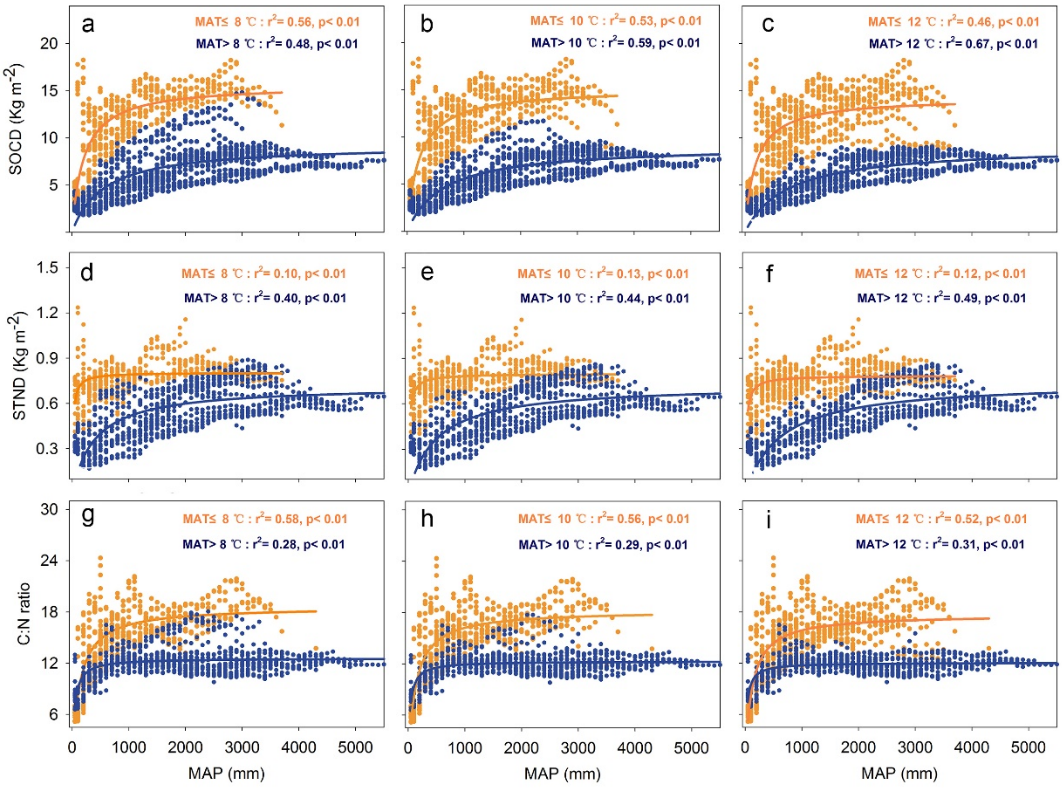

Figure R4. Trends in SOCD (a, b, c), STND (d, e, f) and C:N (g, h, i) ratio with MAP with different segmentations of MAT(8, 10, and 12 °C).

Page 6, Line 11 - 12: Again, this is very strong statement. Given the issues surrounding the true causality in the found relationships discussed in this section I would suggest to make a less strong statement.

Response: We have revised the sentence as "The soil-climate-biome diagram demonstrates the quantitative linkages between surface soil physical properties and climate variables at the global scale".

Page 6, Line 30 - 31: The transition definition and calculation is not explained in statistical analysis section nor reported in results section. It is impossible to review these results without detailed explanation.

Response: Thanks for your suggestion. Our results show that soil pH decreases with increasing MAP. The 'critical MAP' here means the corresponding MAT at soil pH=7.0, which indicates a shift from alkaline to acidic soil. We then plotted the critical levels of MAP (100 mm division) at soil pH=7.0 with MAT. We have included these details in the revised statistical analysis section.

Page 7, Line 10: Please check the subscript.

Response: Typo corrected.

Page 7, Line 12 - 20: C to N ratio does not represent very good proxy for substrate quality so the discussion is very speculative at this point.

Response: Thanks for your comments. We fully agree that C:N ratio in soil organic matter is not a good proxy of substrate quality. We have thus removed this speculative statement in the revised manuscript.

Page 7, Line 21 - 24: This is very speculative. Isn't the correlation simple given by the fact that plant derived organic material always contain C and N no matter what the limitation is?
Response: Thanks for your comments. We revised the sentence as "Previous meta-analyses indicated that C:N ratio in the soil was well-constrained at the global scale (Cleveland and Liptzin, 2007). Accordingly, our results indicated a strong correlation between STN density and SOC density (Fig. 8) and demonstrated a similar pattern of STN density as SOC density across biome and climate regimes (Figs. 3 & 6)" (Page 9, line 21-23).

.

Page 8, 4.4 Uncertainties in mapping soil properties at the global scale: Can authors quantify the uncertainty? According to Tab. S4, statistical model explains sometimes less than 20% of variability. In addition, 10% of all data were removed. I believe that the uncertainty is very important to state unambiguously so any potential user of the database and maps knows the limitations.
Response: Thanks for your suggestions. Modelling uncertainty was calculated as the standard deviation (SD) of the 10-fold cross-validation. Figure R5 shows that uncertainties were relatively higher in regions with less recorded soil profiles, such as high-latitude Russia and Canada. Moreover, soil property data below 2.5% quantile and above 97.5% quantile were excluded as outliers and we have mended our description (see our reply above). Although the values of R^2 is acceptable in most cases, we have also mentioned that our approach sometimes explains less than 20% of the variability due to several potential reasons, including limited sample size for some regions, data quality of original data, and limited independent variables used for modelling analysis. This uncertainty has been discussed in detail in the revised manuscript.

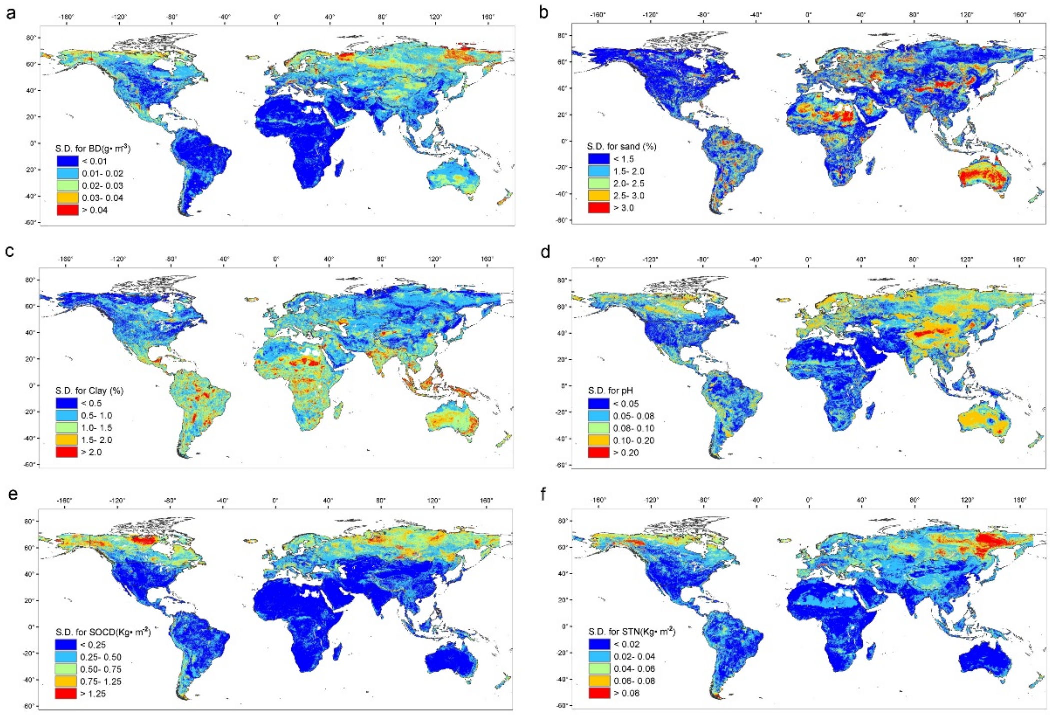

Figure R5. Spatial pattern of uncertainties (standard deviation, SD, n=10) of soil properties in the upper 30-cm layer estimated by 10-fold cross-validation. a: Bulk density ($g \cdot cm^{-3}$); b: Sand (%); c: Clay (%); d: pH e: SOCD ($kg \cdot m^{-2}$); f: STND ($kg \cdot m^{-2}$).

Page 8, Line 28: "Ourregion-specific"

Response: Thanks. Typo corrected.

Page 8 and 9, Conclusions: Again, very strong statements unsupported by the analysis.

Response: We have avoided strong statements and revised the concluding paragraph as "By compiling a comprehensive global soil database, we mapped eight surface soil properties based on machine learning algorithms and assessed the quantitative linkages between soil properties, climate, and biota at the global scale. Our region-specific random forest model generated high-resolution (1km) predictions of surface soil properties, which can be potentially used as inputs for regional and global biogeochemical models. Our results also produced a global soil-climate-biome diagram, which indicates the quantitative linkages between soil, climate, and biomes. Given that significant changes in major soil properties may occur in view of global environmental change (Trumbore and Czimczik, 2008; Chapin et al., 2009; Todd-Brown et al., 2013; Luo et al., 2016, 2017), more efforts should be made in future to understand the dynamics of the global soil-climate-biome diagram."

Tables:

I would suggest to show only R2 in table S4. It would improve the table alignment and reading.

Response: Thanks for your suggestions. We would like to keep both because they are essential indicators of model prediction accuracy.

Figures:

Fig. 4: I have concerns regarding MAP and MAT thresholds. Authors should clarify their definition and use. Specifically, arctic have typically low MAP but because of low MAT, their soils are very often water saturated.

Response: We have now included more details on the definition and uncertainties of these thresholds. In the revised manuscript, we have now used a MAP at 500 mm to roughly indicate a threshold for humid and arid climates, and a MAT at 10 °C to separate relatively warmer and colder climates (see our reply above). The MAP threshold was based on the diagram of Holdridge life zone which used <500 mm to indicate arid climates (Holdridge, 1967). The threshold MAT was based on a universal thermal scale that used mean monthly temperature (approximately MAT) 10 °C to differentiate "cool" and colder from "mild" and warmer climates (Trewartha and Horn, 1980). These thresholds were defined at global scale and we understand that these thresholds are not precise in some cases because of interactions between temperature and precipitation. For instance, soils can be relatively wet in regions with low MAP if low MAT doesn't result in too much evaporation (e.g., arctic ecosystems). Nevertheless, using these thresholds can help us to demonstrate the interactions of MAT and MAP in affecting soil properties (Figs. 4, 5, 6 in the revised manuscript). Thanks for your understanding.

Fig. 5: Please explain the part c of the figure in statistical analysis section.

Response: Thanks for your comments. Our results show that soil pH decreases with increasing MAP. The "critical MAP" here means the corresponding MAT at soil pH=7.0, which indicates a shift from alkaline to acidic soil. To explore the role of MAT in regulating the critical MAP, we then plotted the critical levels of MAP (100 mm division) at soil pH=7.0 with MAT. We have described this in the revised section of statistical analysis.

Fig. 7: Results presented in this figure also require more information. How it was calculated? Was the increase of explained variability by a specific explanatory variable compared to null or some standard model?

Response: Thanks for your comments. We have added the more detailed description in the methods section. Figure 7 indicated the relative importance of variables, denoted by the percent increase in mean-squared error (%IncMSE), which is estimated based on a permuting out-of-bag (OOB) method (Strobl et al., 2009 a &b). For each tree of the random forest, we compared the prediction error on the OOB portion of the data (MSE for regression) with that after permuting each predictor variable. The difference are then averaged over all trees, and normalized by the standard deviation of the differences. The relative percent (mean/SD) increase in MSE as compared to the out-of-bag rate (with all variables intact) was used to indicate the relative importance of each variable (Breiman, 2001). The results indicate that climates (MAT, MAP, TS, PS) and elevation are the most important variables for the prediction of each soil properties.

Fig. 8: I did not find any reference to this figure in the main text.

Response: Thanks a lot. We have referred Figure 8 in the revised manuscript. For instance, "Our results demonstrate that STND as SOCD showed a similar pattern (Figs. 3, S6) and a tight link (Fig. 8) across the global scale".

References

Adhikari, K., and Hartemink, A. E. (2016) Linking soils to ecosystem services—A global review. Geoderma, 262, 101–111.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., and Sparks, D. L. (2015) Soil and human security in the 21st century. Science, 348(6235), 1261071,

Batjes, N.H. (1996) Total carbon and nitrogen in the soils of the world. European Journal of Soil Science, 47,151–163.

Batjes, N.H. (1997) A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. Soil Use and Management, 13(1), 9-16.

Batjes, N.H., Al-Adamat, R., Bhattacharyya, T., et al. (2007) Preparation of consistent soil data sets for modelling purposes: Secondary SOTER data for four case study areas. Agriculture Ecosystems & Environment 122(1): 26-34.

Breiman, Leo. (2001) Random Forests. Machine Learning, 45(1): 5–32.

Carter, M., & Bentley, S. P. (2016) Soil properties and their correlations. John Wiley & Sons.

Castellano, M. J., & Kaye, J. P. (2009) Global within-site variance in soil solution nitrogen and hydraulic conductivity are correlated with clay content. Ecosystems, 12(8), 1343–1351.

Chapin III, S., McFarland, J., McGuire, D. A., Euskirchen, E. S., Ruess, R. W., and Kielland, K. (2009) The changing global carbon cycle: linking plant–soil carbon dynamics to global consequences. J. Ecol., 97(5), 840–850.

Chaplot, V., Bouahom, B., Valentin, C., (2010) Soil organic carbon stocks in Laos: spatial variations and controlling factors. Global Change Biology 16 (4), 1380–1393.

Chesworth, W. (1973) The parent rock effect in the genesis of soil. Geoderma, 10(3), 215-225.

Davidson, E. A., and Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature, 440(7081), 165.

Del Grosso, S., Parton, W., Stohlgren, T., Zheng, D., Bachelet, D., Prince, S., ... & Olson, R. (2008) Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology, 89(8), 2117-2126.

Doetterl, S., et al., (2015) Soil carbon storage controlled by interactions between geochemistry and climate. Nature Geoscience, 8 (10), 780–783.

Gentsch N, Wild B, Mikutta R, et al. (2018) Temperature response of permafrost soil carbon is attenuated by mineral protection. Global Change Biology 24:3401–3415.

Hassink, J. (1997) The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil, 191(1), 77–87.

Holdridge, L.R. (1967) Life zone ecology. Tropical Science Center. San Jose, Costa Rica.

Jilling A., Keiluweit M., Contosta A.R., Frey S., Schimel J., Schnecker J., Smith R.G., Tiemann L., Grandy A.S. (2018) Minerals in the rhizosphere: overlooked mediators of soil nitrogen availability to plants and microbes. Biogeochemistry 139, 103–122.

Manzoni, S., Jackson, R. B., Trofymow, J. A., & Porporato, A. (2008) The global

stoichiometry of litter nitrogen mineralization. Science, 321(5889), 684-686.
Meier, I.C., Leuschner, C., (2010) Variation of soil and biomass carbon pools in beech forests across a precipitation gradient. Glob. Chang. Biol. 16 (3), 1035–1045.
Milne, E., Banwart, S. A., Noellemeyer, E., Abson, D. J., Ballabio, C., Bampa, F., and Black, H. (2015) Soil carbon, multiple benefits. Environ. Dev., 13, 33–38.
Pleijsier, 1989. Variability in soil data. In: J. Bouma and A.K. Bregt (Editors), Land Qualities in Space and Time. PUDOC, Wageningen, pp. 89-98.
Saiz, G., Bird, M. I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T. R., ... & Diallo, A. (2012) Variation in soil carbon stocks and their determinants across a precipitation gradient in W est A frica. Global change biology, 18(5), 1670-1683.
Scharlemann, J. P., Tanner, E. V., Hiederer, R., & Kapos, V. (2014) Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management, 5(1), 81-91.
Shangguan, W., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014) A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems, 6(1), 249-263.
Sposito, G., Skipper, N. T., Sutton, R., Park, S. H., Soper, A. K., and Greathouse, J. A. (1999) Surface geochemistry of the clay minerals. P. Natl. Acad. Sci. U.S.A., 96(7), 3358–3364.
Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., and Wheeler, I. (2013) The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agr., Ecosyst. Environ., 164, 80–99.
Strobl, Carolin; James Malley; and Gerhard Tutz. (2009) b. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods. 14(4): 323–348.
Strobl, Carolin; Torsten Hothorn; and Achim Zeileis. (2009) a. Party on! A new, conditional variable importance measure for random forests available in party package. The R Journal. 1(2): 14-17.
Trewartha GT, Horn LH (1980) Introduction to climate, 5th edn. McGraw Hill, New York, NY.
Weil, R. R., and Brady, N. C. (2016) The nature and properties of soils. Pearson Education, USA.
Wieder, W. R., Bonan, G. B., & Allison, S. D. (2013) Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change, 3(10), 909-912.
Yang, Y., Fang, J., Ma, W., Smith, P., Mohammat, A., Wang, S., & Wang, W. E. I. (2010). Soil carbon stock and its changes in northern China's grasslands from 1980s to 2000s. Global Change Biology, 16(11), 3036-3047.

Reply to comments by Anonymous Referee #4 (RC2)

Comment #1: The article by Zhao et al., present an interesting global dataset for some soil parameters, linking these properties with climate and biota. Nevertheless, there are several issues that should be clarified and discussed in much more details. The mentioned databases report row data for soil profiles, while the authors use also some parameters which are derived from these data (e.g. SOC and SON stocks). How these data were derived and harmonized should be better explained, since in the paper they are used to derive the linkages between soil, climate and biota. For the soil profiles in the different databases, were used only the soil layers having all the necessary parameters useful to calculate the stocks of C and N? I am referring in particular to Bulk density and rock fragments content. If not, how the authors were dealing with this fact? They were using pedotransfer functions to derive bulk density? And if rock fragments content was missing? Since these two parameters are affecting very much the stock the authors should make an effort in explaining how the database were harmonized. The discussion is sometimes weak. For instance the authors found a correlation between bulk density, MAT and MAP. Similarly the all variation in relation to MAT and MAP? The discussion on the observed differences between ecosystems is quite poor. Not so many recent references are considered for the discussion. The effect of the vegetation on the selected soil parameters should be better considered and discussed.

Response: Thanks for your helpful comments. We have revised the manuscript according to your suggestions:

First, we have included more details on the method to compile our global soil database (also see our reply to comment #2). Specially, SOC/SON stocks were calculated based on bulk density and concentrations of SOC/SON. We directly calculated the stocks of SOC and SON when all the necessary parameters were available. In the case that bulk density was not measured while SOC content was reported, we made estimates of bulk density based on regional-specific pedotransfer functions (Yang et al. 2007; Abdelbaki, 2018) and further estimated SOC/SON stocks. We first established empirical relationship between bulk density and SOC content in each regions (Table R1) and further estimated bulk density based on measured SOC for the soil profiles with missing data for bulk density. Overall, there were 42% profiles with measured data on bulk density and 58% profiles with estimated data on bulk density. We agree that correction for rock fragment is important to estimate soil C stocks, but it remains a global challenge because existing databases usually contain limited information on gravel fractions than bulk density and SOC concentrations (Jandl et al., 2014). Nevertheless, the inclusion of gravel and roots > 2 mm has been evidenced to exert a relatively low influence on the calculation of SOC stocks in the surface soil layer (0-30 cm), mainly due to the fact that surface soil usually contains a low proportion of gravels (Saiz et al., 2012). Currently, we assumed no rock fragment or rock issue had been handled if it was not reported, but we might use the mean gravel fractions of each vegetation type or soil orders as a potential correction factor. Nevertheless, this approach might also result in new uncertainty if used at the global scale. We may conduct such an analysis to deal with the gravel issue if the reviewer support this idea.

Second, we have improved the discussion section by 1) discussing the potential causes for the correlations between soil physical properties (bulk density and soil texture) and climate (MAT and MAP), 2) discussing the shifts of soil properties across biomes and the interactions between soil and vegetation, and 3) including more recent references. Please find more details in our reply to comments #6, 7, 8, 10 and associated references.

Table R1 Empirical regression models for relationship between bulk density (BD, g cm^{-3}) and soil organic carbon content (SOC, %) for each region.

Region	Model	R^2	Num
Tropical Asia	$BD = 1.336e^{-0.054\ SOC}$	0.26	765
Mexico	$BD = 1.380e^{-0.061\ SOC}$	0.63	1243
Africa	$BD = 1.480e^{-0.073\ SOC}$	0.30	3770
Continental US	$BD = -0.173\ln(SOC) + 1.382$	0.45	1239
Canada	$BD = 1.507e^{-0.027\ SOC}$	0.20	163
Russia	$BD = -0.222\ln(SOC) + 1.287$	0.59	777
South America	$BD = -0.07\ln(SOC) + 1.233$	0.15	2105
Europe	$BD = 1.4661e^{-0.041\ SOC}$	0.60	2391
East Asia	$BD = 1.4719e^{-0.08\ SOC}$	0.35	634
Australia	$BD = 1.3319e^{-0.062\ SOC}$	0.74	167

Comment #2: Specific comment: Page 3 Line 5-10: "Compiled". And what about harmonization of the data?

Response: Thanks for your reminder! We have included more details on the methods of data screening and compiling in the revised manuscript and supplement. Along with ground-truth soil profile data (Table S1), we have also derived general information of soil sampling (site location, sampling time, source of data), pedologic information on soil orders and the horizons of the sampled soil profiles, mean annual temperature (MAT), mean annual precipitation (MAP), seasonality of air temperature (TS, calculated as $100\times SD_{monthly}/Mean_{monthly}$), seasonality of precipitation (PS), mean annual normalized difference vegetation index (NDVI), elevation (global digital elevation map [DEM]), slope, and land use type for each recorded site (Table R1). Specifically for each profile, we recorded data on the number of horizon, top and bottom depth, and values of soil physical properties (sand/silt/clay fraction [%], gravel content [>2mm, %], bulk density [g/cm^3]), and chemical properties (pH, organic carbon content [%]; and total nitrogen content [%]) (Table R2).

Data harmonization was conducted by four steps. First, we screened sampling and measurement approaches of each soil property and excluded data those were not comparable to others in methodology. For instance, geographic coordinate data were included only when WGS84 or a geographic coordinate system that could be converted to WGS84 projection was used; Soil texture data were included only when the internationally accepted particle size class were used (clay < 2 μm < silt < 50 μm < sand < 2000 μm). This allowed us to construct a database of soil properties with comparable methodology. Second, we excluded records with no information on the target soil depth (0-30cm). In case that soil organic matter was measured instead of SOC, we used a Bemmelen index (0.58) to convert soil organic matter into SOC. If data of bulk density were not provided, we estimated them based on regional-specific pedotransfer functions (Schaap and Leij, 1998; Yang et al., 2007; Abdelbaki, 2018) (Table R1). Specifically, we established empirical relationship between bulk density and SOC content to estimate bulk density based on measured SOC for those cases with missing data of bulk density. There were 42% profiles with measured data on bulk density and 58% profiles with estimated data on bulk density. It is true that the correction for rock fragment is important for the estimation of soil C stocks, but it remains a global challenge because existing databases usually contain limited information on gravel fractions (Jandl et al., 2014). Nevertheless, the inclusion of gravel has been evidenced to exert a relatively low impact on the calculation of SOC stocks in the surface soil layer (0-30 cm), mainly due to the fact that surface soil usually contains a low proportion of gravels (Saiz et al., 2012). Therefore, we assumed no rock fragment or the rock issue had been handled once it was not reported. Finally, we extracted data on soil properties of the 0-30cm soil depth and calculated the means of each soil property. SOC (STN) density was calculated based on bulk density and SOC (STN) content. We have also revised the section on data set in the revised manuscript.

Table R2 Information recorded in GSD.

Site Data	Horizon Data
Profile ID[a]	***Profile ID & Horizon Id***[b]
General:	***General:***
Source of data	Horizon number
Description of year	depth, top
Soil classification	depth, bottom
Site location and information:	***Physical attributes:***
Location (description, region/ country)	Sand/Silt/Clay fraction (%)
Latitude & Longitude	Gravel content (>2mm, %)
Climate (MAT & MAP)	Bulk density (g/cm^3)
Elevation/ slope/ aspect	
Parent material	***Chemical attributes:***
Land use	Organic carbon (%)
	Total Nitrogen (%)
	pH-H_2O

a. unique identifier for profile in GSD. b. Unique reference number for horizon within a profile. c. sand, 2.0-0.05mm; silt, 0.05-0.002mm, and clay, <0.002mm.

Comment #3: Page 3 Line 20-30: Since most of the soil profiles were collected a very different range of years, how the climate was related to the properties? What you mean with pedological information? The fact the soil profiles data are presented by horizons?

Response: Thanks for your comments. First, we know that soil profile data were measured across a very different range of years, and we used multiple-year mean values of climate variables in our analysis because soil properties were formed by subjecting to a climate for a long term. As 96% of soil profiles in GSD were sampled during 1950 to 2000, we used multiple-year (1950-2010) averages of climatic variables from WorldClim database. Second, our database includes pedological information on soil orders and soil horizons of sampled soil profiles. We calculated surface soil properties (0-30 cm) based on data for each horizon. We have extended the discussion accordingly in the revised manuscript.

Comment #4: Page 5 line 20-25: What is the meaning of providing a mean global value for SOC and SON?
Response: We realized that this sentence doesn't belong here because this paragraph presents results on spatial patterns of soil properties. In the revised manuscript, we have moved this sentence to the end paragraph of section 3.2, which demonstrated results of the density and stocks of SOC and STN at global scale.

Comment #5: Page 6 line 5: In brackets (MAT < 400 mm) is probably MAP rather than MAT?
Response: Typo corrected.

Comment #6: Page 6 line 20-30: the fact that bulk density is affected by precipitation and temperature should be better discussed. Similarly the increases in clay content in relation to MAT and MAP. How soil erosion affect the clay fraction? Is soil erosion selective for the clay? And Silt and Sand? An effect of the actual land use on bulk density should also be pointed out in the discussion.
Response: Thanks for your comments and suggestions. In the revised manuscript, we have discussed the effects of climate, soil erosion and land use on soil physical properties (e.g., bulk density and soil texture).

First, the increase of bulk density with higher MAT and lower MAP is likely due to an accompanying decrease of SOCD (Ruehlmann and Körschens, 2009), which is jointly regulated by MAT and MAP (Fig. R1; see more discussion on the effect of climate on SOCD in section 4.3; Wiesmeier et al., 2019). Higher MAT and MAP can accelerate the rate of weathering (Jenny, 1941; Lal, 2018), thus resulting in lower sand fraction and higher soil clay fraction.

Second, previous studies indicate that silt is most sensitive to soil erosion, while sand is less mobile due to high weight and clay is protected by soil aggregates (Wischmeier and Mannering, 1969; Torry et al., 1997; Wang et al., 2013).

Third, the effect of land use is important at a local scale. For instance, a change of forest or grassland to croplands can significantly decrease SOCD and thus decrease soil bulk density, while reforestation generally increases SOCD and thus decreases soil bulk density (Don et al., 2011). However, our static mapping of global soil properties is not able to account for the effect of temporal land use change.

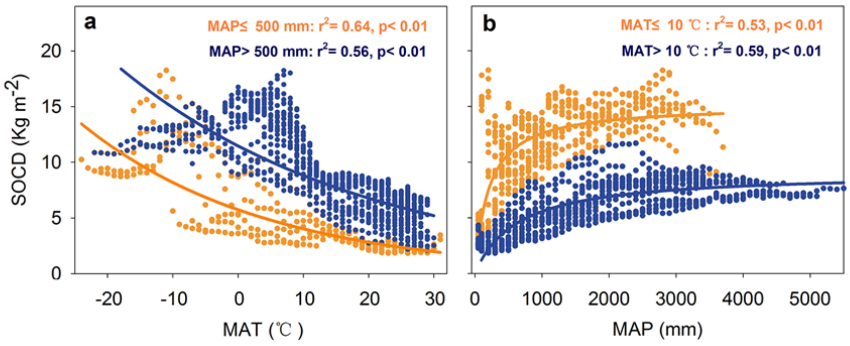

Figure R1. Changes in SOCD with MAT and MAP.

Comment #7: Page 7 line 5-10: The fact that in the tropical area Clay and bulk density decrease with altitude how can be explained? Which is the meaning of this decrease?

Response: Thanks for your suggestion. We have discussed the possible causes for the altitudinal trends of bulk density and clay, which is similar to the trends across latitudes. First, the decrease of clay fraction with higher altitude is likely due to 1) a younger soil age (Waite and Sack, 2011), 2) lower weathering rate under lower temperature (Grieve et al., 1990; Kramer and Chadwick, 2016), and 3) a downslope translocation of surface soil to lower altitude. Second, the decrease of bulk density with altitude is likely due to an increase in SOC retention (Fig. R2f), which mainly results from low rate of decomposition along with lower temperature (Grieve et al., 1990 ; Kramer and Chadwick, 2016).

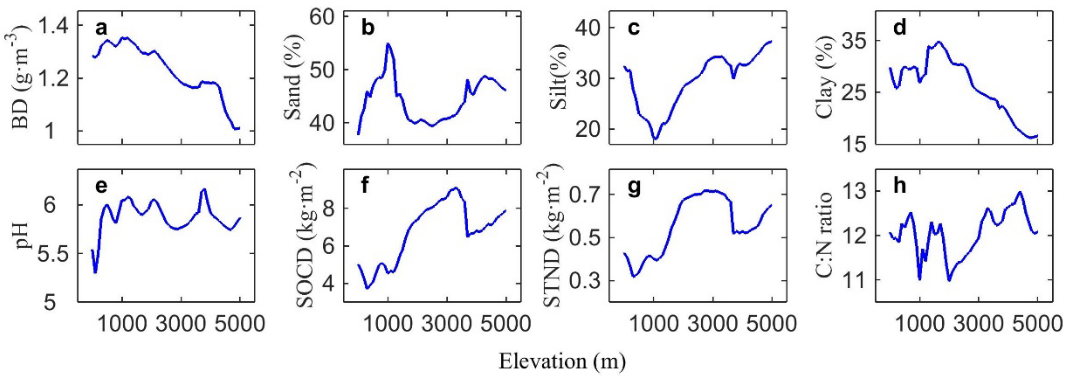

Figure R2. Changes in surface soil properties with elevation in tropical regions. a: Bulk density (g·cm^{-3}); b: Sand (%); c: Silt (%); d: Clay (%); e:Ph; f: SOCD (kg C·m^{-2}); g: STND (kg N·m^{-2}); h: C:N ratio.

Comment #8: Figure 2: SOC density box Looking at the SOC density it appear that there is quite a lot of C in the North Mediterranean area, which is usually quite poor in SOC due to the continuous use of the land for agricultrue since millennia. On the other side also the area covered by tropical primary forests in Africa (e.g. Congo basis) seems to be relatively poor? How they authors can explain these facts?

Response: Thanks for your comments. We have mapped the original records of SOCD on the map in North Mediterranean croplands and found similar results as the mapped values (Fig.

R3). As indicated by a meta-analysis, croplands have significantly lower SOCD as compared with local plantation, forest and grassland (Don et al., 2011). In the North Mediterranean region, an increase in the area of olive plantation and vineyard in last decades might have contributed to the relatively high values of SOCD (Parras-Alcántara et al., 2013). We have separately mapped SOCD for global croplands (Fig. R3) and the values of SOCD in North Mediterranean area were not as high as the impression by Figure 2 in the manuscript. This is likely a visual illusion due to a mix of croplands with natural vegetation.

Due to fast turnover with rapid decomposition of organic matter, SOC content has been evidenced to be relatively poor in tropical forests (e.g., Congo and Amazon tropical forests) (Wang et al., 2018). Accordingly, previous mappings of SOCD have also shown relatively low values in tropical forests (Köchy et al., 2015; Jackson et al., 2017).

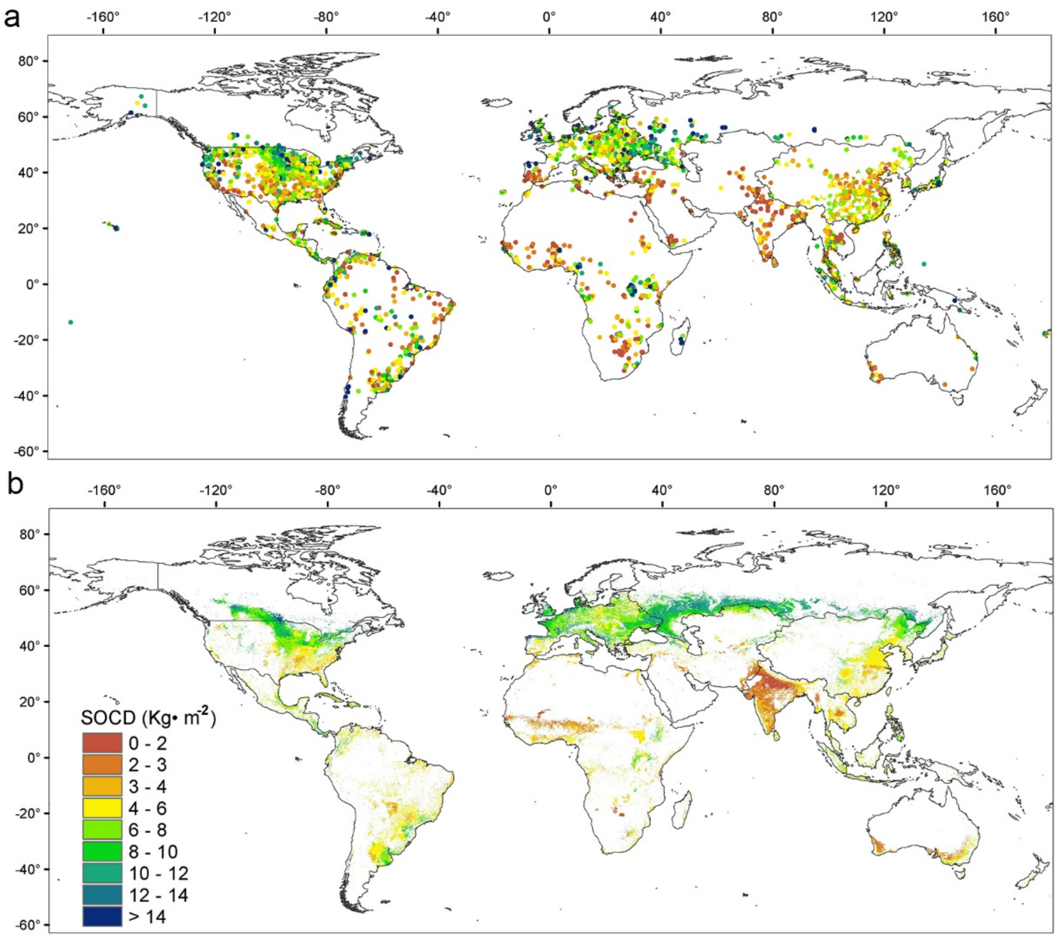

Figure. R3. Site records (a) and spatial variations (b) of SOCD in croplands.

Comment #9: Bulk density box How the authors explain the very high values of BD for the United states? Why they are so high compared to other regions. Apparently in the USA there are not so many differences in BD in relation to the different ecosystems (e.g. Forests vs. grassland vs cropland)

Response: Thanks for the comments. We have summarized the original records of bulk density for each 11 regions (Table R3). The results showed that mean regional bulk density was also relatively high in the continental United States. We have further summarized the original records of bulk density for forests, grassland and cropland in the US. We also found that bulk density of forests, grassland and cropland didn't show much difference (Table R4). Overall, our mapping of bulk density is in consistent with the pattern based on raw data and is similar to previous mapping on global bulk density (Hengl et al., 2014; Shangguan et al., 2014).

Table R3 Mean values of sampled bulk density data for each region in the GSD.

	Bulk density ($g\cdot cm^{-3}$)		
Region	Mean	SD	Number
Tropical Asia	1.33	0.23	860
Mexico	1.22	0.26	316
Africa	1.37	0.16	3740
Continental US	1.57	0.22	9322
Canada	1.25	0.32	790
Russia	1.12	0.28	386
South America	1.21	0.19	1764
Europe	1.27	0.30	1527
East Asia	1.29	0.20	2762
Australia	1.12	0.27	162
West Asia	1.48	0.20	333
Alaska	1.07	0.35	79
Total	1.33	0.23	860

Table R4 Mean values of sampled bulk density data for each biomes in the Continental US.

	Bulk density ($g\cdot cm^{-3}$)		
Continental US	Mean	SD	Number
Forest	1.57	0.28	1588
Shrub	1.64	0.27	1096
Grassland	1.56	0.20	2560
Cropland	1.54	0.15	3084
All*	1.57	0.22	9322

Note: Mean measured bulk density was not shown for savanna, wetlands and sparse vegetation because of limited sample size (<100). However, these biomes were also used to calculated regional mean of all biomes.

Comment #10: Table 1. The BD of cropland appear to be similar to those of savanna and grassland. How it can be explained? Similarly, concerning the SOC stock how it can be explained that cropland have similar values of tropical forests?

Response: Thanks. Table 1 shows global means of soil property across biomes. However, bulk density shows significant spatial variation within savanna and grasslands (Fig. R4a) as well as croplands (Fig. R4b). Generally, bulk density ranged from ~1.0 to ~1.7 $g\cdot cm^{-3}$ in savanna and grasslands (Fig. R4a), and it ranged from ~1.1 to ~1.7 $g\cdot cm^{-3}$ in croplands (Fig. R4b). Considering the large spatial variation in soil properties and limited overlap in spatial distribution, it is difficult to attribute reasons to the difference of global means between croplands and other biomes. This is the same for the comparison of global mean SOCD between croplands and tropical forests (Fig. R5). SOCD in tropical forests generally ranged from 3 to 10 kg·m^{-2}, while it ranged from 2 to 12 kg·m^{-2} in croplands. When comparing values at a same region (e.g., southeast Asia), SOCD is obviously lower in croplands than in tropical forests (compare Fig. 5a and Fig. 5b). This difference has been also evidenced by meta-analysis based on field observations (Don et al., 2011).

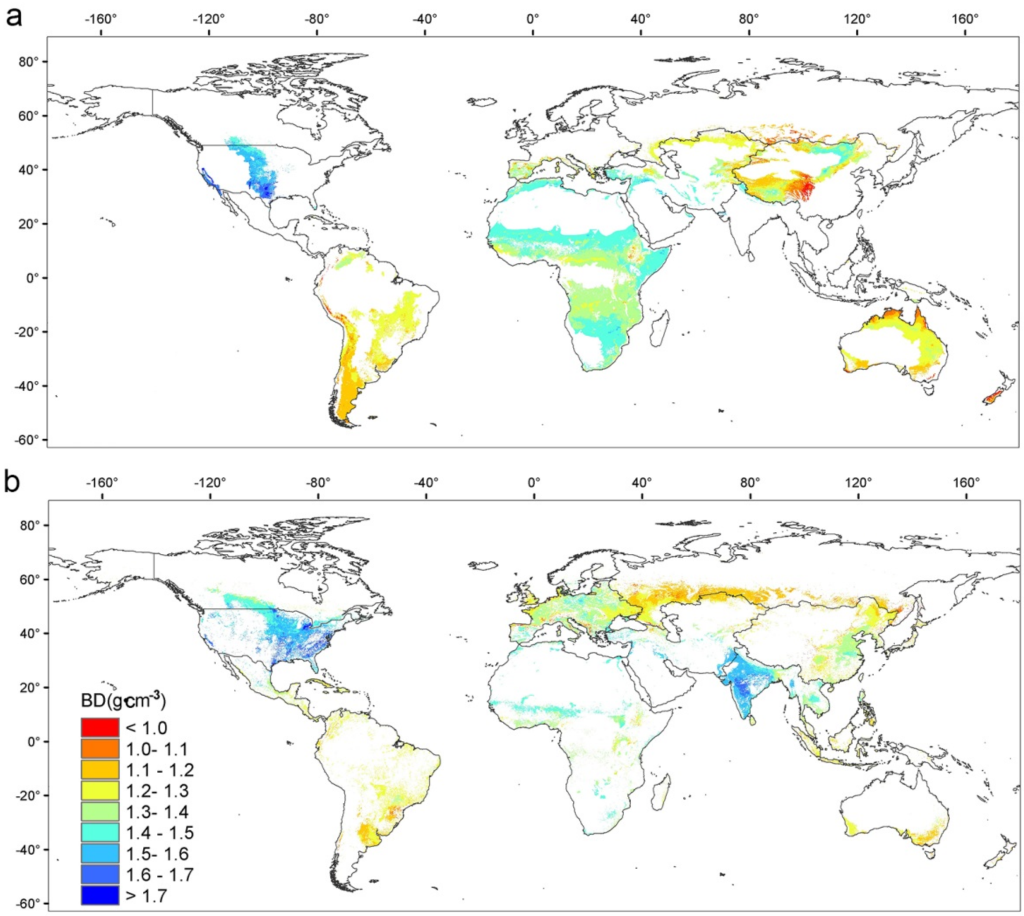

Figure. R4. Spatial variations of bulk density in (a) savanna and grassland, and (b) croplands.

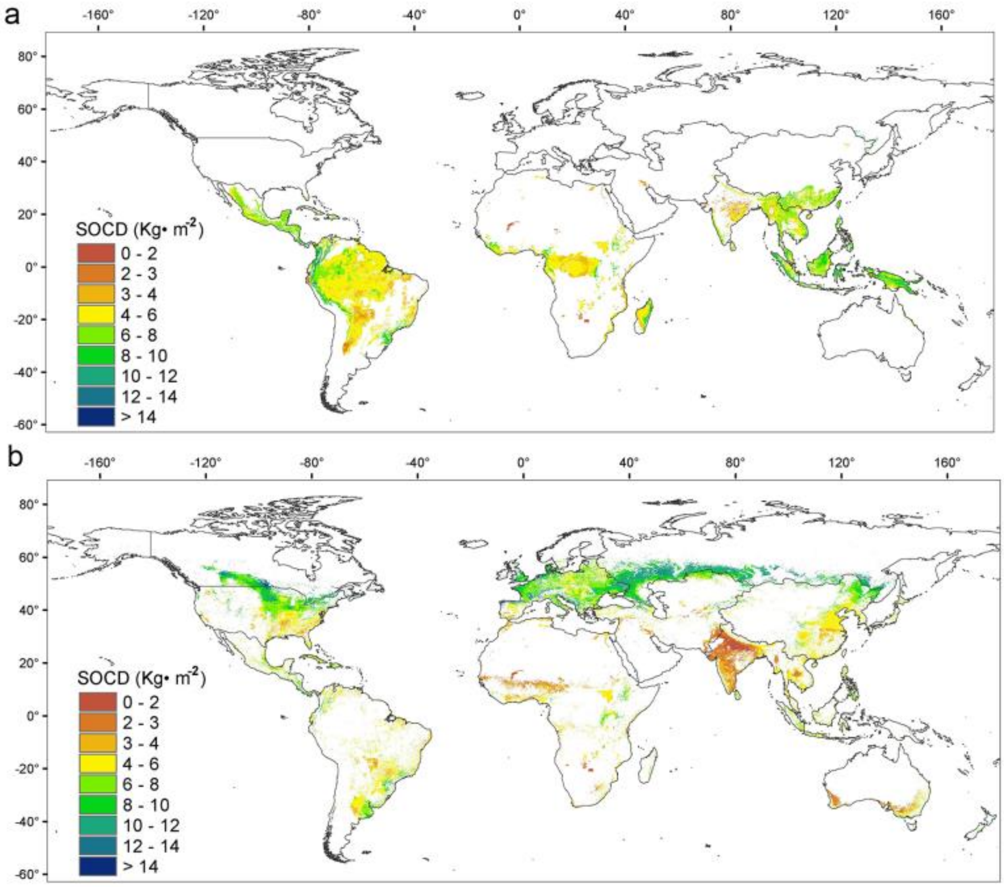

Figure. R5. Spatial variations of SOCD in (a) tropical forests, and (b) croplands.

Reference

Abdelbaki, A. M. (2018) Evaluation of pedotransfer functions for predicting soil bulk density for US soils. Ain Shams Engineering Journal, 9, 1611-1619.

Batjes, N. H. (1997) A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. Soil Use and Management, 13(1), 9-16.

Carter, M., & Bentley, S. P. (2016) Soil properties and their correlations. John Wiley & Sons.

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.

Grieve, I. C., Proctor, J., & Cousins, S. A. (1990) Soil variation with altitude on volcan Barva, Costa Rica. Catena, 17(6), 525-534.

Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, et al. (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS ONE 9(8): e105992. doi:10.1371/journal.pone.0105992.

Jackson, R. B., Lajtha, K., Crow, S. E., Hugelius, G., Kramer, M. G., Piñeiro, G. (2017) The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls. Annual Review of Ecology, Evolution, and Systematics, 48, 419-445.

Jandl, R., Rodeghiero, M., Martinez, C., Cotrufo, M. F., Bampa, F., van Wesemael, B., & Lorenz, K. (2014) Current status, uncertainty and future needs in soil organic carbon monitoring. Science of the Total Environment, 468, 376-383.

Jenny, H. (1941) Factors of soil formation: a system of quantitative pedology. McGraw-Hill Book Company Inc, New York.

Köchy, M., Hiederer, R., & Freibauer, A. (2015) Global distribution of soil organic carbon–Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. Soil, 1(1), 351-365.

Kramer, M. G., & Chadwick, O. A. (2016) Controls on carbon storage and weathering in volcanic soils across a high-elevation climate gradient on Mauna Kea, Hawaii. Ecology, 97(9), 2384-2395.

Lal, R. (2018) Soil and climate. In Soil and Climate (pp. 1-8). CRC Press.

Parras-Alcántara, L., Martín-Carrillo, M., & Lozano-García, B. (2013) Impacts of land use change in soil carbon and nitrogen in a Mediterranean agricultural area (Southern Spain). Solid Earth, 4(1), 167-177.

Pleijsier, J. (1989) Variability in soil data. In: Land Qualities in Space and Time (eds Bouma J., Bregt, A.K.), pp. 89-98. Pudoc, Wageningen.

Ruehlmann, J., & Körschens, M. (2009) Calculating the effect of soil organic matter concentration on soil bulk density. Soil Science Society of America Journal, 73(3), 876-885.

Saiz, G., Bird, M. I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T. R., & Diallo, A. (2012) Variation in soil carbon stocks and their determinants across a precipitation gradient in West Africa. Global Change Biology, 18(5), 1670-1683.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014) A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems, 6(1), 249-263.

Torri, D., Poesen, J., Borselli, L. (1997) Predictability and uncertainty of the soil erodibility factor using a global dataset. Catena, 31(1-2), 1-22.

Waite, M., & Sack, L. (2011) Shifts in bryophyte carbon isotope ratio across an elevation× soil age matrix on Mauna Loa, Hawaii: do bryophytes behave like vascular plants?. Oecologia, 166(1), 11-22.

Wang, B., Zheng, F., Römkens, M. J., Darboux, F. (2013) Soil erodibility for water erosion: A perspective and Chinese experiences. Geomorphology, 187, 1-10.

Wang, J., Sun, J., Xia, J., He, N., Li, M., & Niu, S. (2018) Soil and vegetation carbon turnover times from tropical to boreal forests. Functional ecology, 32(1), 71-82.

Wieder, W. R., Bonan, G. B., & Allison, S. D. (2013) Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change, 3(10), 909-912.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Luetzow, M., Marin-Spiotta, E., ... & Wollschlaeger, U. (2019) Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. Geoderma, 333, 149-162.

Wischmeier, W. H., & Mannering, J. V. (1969) Relation of soil properties to its erodibility 1. Soil Science Society of America Journal, 33(1), 131-137.

Yang Y.H., Mohammat A., Feng J.M., Zhou R., Fang J.Y. (2007) Storage, patterns and environmental controls of soil organic carbon in China. Biogeochemistry, 84, 131–141.

Global soil-climate-biome diagram: linking surface soil properties to climate and biota

Xia Zhao[1], Yuanhe Yang[1], Haihua Shen[1], Xiaoqing Geng[1], Jingyun Fang[1,2]*

[1]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China 100093
[2]College of Urban and Environmental Sciences, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing, China 100871

Correspondence to: Jingyun Fang (jyfang@urban.pku.edu.cn)

Abstract. Surface soils interact strongly with both climate and biota and provide fundamental ecosystem services that maintain food, climate, and human security. However, the quantitative linkages between soil properties, climate, and biota~~ at the global scale~~ remain unclear at the global scale. By compiling a comprehensive global soil database, we mapped eight major soil properties (bulk density; clay, silt, and sand fractions; soil pH; soil organic carbon [SOC] density; soil total nitrogen [STN] density; and soil C:N mass ratios) in the surface ~~(0-30 cm)~~ soil layer (0-30 cm) based on machine learning algorithms, and demonstrated the quantitative linkages between surface soil properties, climate, and biota at the global scale, which we call~~ (i.e.,~~ the global soil-climate-biome diagram~~)~~. On the diagram, bulk density increased significantly with higher mean annual temperature (MAT) and lower mean annual precipitation (MAP); soil clay fraction increased significantly with higher MAT and MAP; Soil pH decreased with higher MAP and lower MAT, and the 'critical MAP', which means the corresponding MAP at a soil pH of =7.0 (a shift from alkaline to acidic soil),~~ for the transition (pH=7) from alkaline to acidic soil~~ decreased with ~~decreasing~~lower MAT; SOC density and STN density both were jointly affected by MAT and MAP, showing an increase at lower MAT and a saturation~~ tendency~~ towards higher MAP. Surface soil physical and chemical properties also showed remarkable variations across biomes. The soil-climate-biome diagram suggests~~ the co-evolution of the~~ ~~a~~ the shifts in soil properties~~,~~ ~~climate, and biota~~ under global ~~warming~~climate and land cover ~~environmental~~ change.

1. Introduction

As a critical component of the Earth system, soils influence many ecological processes ~~which~~that provide fundamental ecosystem services ~~soils influence all aspects of ecosystem processes and provide fundamental ecosystem services that maintain food, climate, and human security~~ (Amundson et al., 2015; Milne et al., 2015; Adhikari and Hartemink, 2016). Soil physical properties, such as bulk density and soil texture, are important for ~~green~~ water retention and the preservation of carbon (C) and nutrients (Hassink, 1997; Sposito et al., 1999; Castellano and Kaye, 2009; Stockmann et al., 2013; Jilling et al.,2018), whereas soil chemical properties, such as soil acidity (pH), organic C, and nutrient contents, are essential regulators of nutrient availability and plant growth, further affecting C and nutrient cycling as well as vegetation-climate feedbacks ~~biogeochemical cycles and climate feedbacks~~ (Davidson and Janssens, 2006; Chapin et al., 2009; Milne et al., 2015). As the most biogeochemically active soil layer, surface soil dominates the soil function and interacts strongly with climate and vegetation (Jenny, 1941; Alexander, 2013; Weil and Brady, 2016). Therefore, assessing the physical and chemical properties in surface soil could provide insights of global soil functions and support soil stewardship to secure sustainable ecosystem services ~~for societal well-being in the context of unprecedented pressure on soils~~ (Batjes, 2009; Sanchez et al., 2009; Koch et al., 2013).

In the context of rapid environmental change, there is an increasing need for high-quality, high-resolution, and timely updated global mapping of soil properties (Grunwald et al., 2011). Based on the global database of soil properties (e.g., the Harmonized World Soil Database [HWSD]), multiple linear regression models have been widely used for soil mapping (Batjes 2009; Hengl et al. 2014). Although recent progress has been made by compiling larger numbers of soil profiles and performing accuracy assessments, the ~~corresponding~~ maps of global soil properties are subject to weak relationships between soil properties and the corresponding ~~explanatory variables~~predictors (Hengl et al., 2014). Moreover, some attempts have been made to predict global soil properties based on Earth system models, but these predictions frequently showed large variation among different models and agreed poorly with observational data (Todd-Brown et al., 2013; Tian et al., 2015). Recently, machine learning algorithms, such as random forest (RF) analyses have been successfully applied to develop spatially explicit estimates of soil organic C (SOC) (Grimm et al., 2008; Wiesmeier et al., 2011; Ding et al., 2016; Hengl et al., 2017). Compared with multiple linear regression models, RF analysis has several advantages, such as the ability to model non-linear relationships, handle both categorical and continuous predictors, and resist overfitting and noise features (Breiman, 2001). ~~It is thus necessary to re-evaluate global patterns of soil properties using machine learning algorithms.~~

The underlying stability of soil systems is controlled by their inherent balance between mass inputs and losses ~~of C and nutrients~~, which strongly feeds back on climate and biota (Amundson et al., 2015; Weil and Brady, 2016~~; Amundson et al., 2015~~). By overlapping the ~~global~~ spatial distribution of climate types, biome types, and soil orders, Rohli et al. (2015) first quantified the percentage of global land surface that is covered by the combinations of climate types, biomes, and soil orders. However, ~~appropriate quantititive~~quantitative linkages of soil properties, climate, and biota ~~remain to be~~have not yet been developed in a common diagram. ~~In the context of~~Encouragingly, significant progress in digital soil mapping techniques and the rapidly growing quantity of recorded soil information (Sanchez et al., 2009; Grunwald et al., 2011; Arrouays et al., 2014; Hengl et al., 2014; Shangguan et al., 2014), provide a~~n~~ great opportunity ~~is present~~ to assess the quantitative linkages between soil properties, climate, and biota at the global scale.

In this study, we first compiled a global soil database (GSD, see Materials and Methods) that contains more than 28000 soil profiles for seven soil physical and chemical properties in surface soil layer (0-30 cm), including bulk density (g cm^{-3}~~);~~), sand, silt and clay fractions ~~(%);~~ (%), soil pH~~;~~ , SOC density (kg m^{-2}), and soil total nitrogen (STN) density (kg m^{-2}). ~~Using~~ Using regional RF ~~machine learning~~ algorithms, we then established global soil maps for eight soil properties (the above mentioned seven soil properties plus C:N ratios, being estimated based on SOC density and STN density) at a 1-km resolution and evaluated their corresponding uncertainties. On the basis of Whittaker biome diagram which illustrates ~~T~~the essential role of climate in shaping the spatial pattern of global biomes ~~has been well demonstrated by Whittaker biome diagram~~ (Whittaker, 1962)~~. As climate and vegetation are two key soil-forming factors (Jenny, 1941)~~, we further developed a global soil-climate-biome diagram by ~~plotted~~ plotting each soil property on a climate basis ~~within a modified Whittaker biome diagram~~ as climate and vegetation are two key soil-forming factors (Jenny, 1941)~~(Whittaker, 1962)~~. Although parent material (e.g., bedrock) also plays an important role in affecting soil properties, it affects soil formation at a relatively long time scale (Chesworth, 1973), particularly in the subsoil (Gentsch et al., 2018). In addition, our soil-climate-biome diagram thus focuses on soil properties in the surface layer, ~~However,~~given that surface soils are dynamic in time and likely interacting more instantly with climate and vegetation than deeper soils (Weil et al., 2016). ~~Specifically~~Overall, our objectives were to (i) map the physical and chemical properties of global surface soils, and (ii) determine the linkages between surface soil properties, climate and biota at the global scale.

2. Materials and Methods

2.1 Data set

We compiled ground-truth ~~S~~soil property data ~~were compiled~~ to establish a comprehensive database of worldwide soil profile information (Global soil database, GSD). Our GSD includes existing sources of soil profile data from the International Soil Reference and Information Centre-World Inventory of Soil Emission (ISRIC-WISE) Potential database (version 3.2; Batjes, 2009), soil reference profiles of Canada (Pan et al., 2011), the Land Resources of Russia/International Institute for Applied Systems Analysis (IIASA) (http://nsidc.org/data/ggd601.html), the International Soil Carbon Network (ISCN 2012, http://www.fluxdata.org/nscn/Data/AccessData/SitePages/Carbonto1M.aspx), the Soil Profile Analytical Database of Europe (SPADE), the Northern Circumpolar Soil Carbon Database (NCSCD, Tarnocai et al., 2009), the Second State Soil Survey of China (National Soil Survey Office, 1998), literature-retrieved soil data on the forests of China (Yang et al., 2014), field campaign data on the grasslands of northern China (from our research team; Yang et al., 2008, 2010), and field survey data of Australia (Wynn et al., 2006) (see Table S1 for more detailed information on these data sources). Overall, ~~T~~the GSD ~~compiled ground-truth soil data in the world and~~ includes more than 28000 soil profiles (Fig. 1; Table S1). Although the total sample number and spatial distribution of the profile data are similar to those of the WISE30sec (Batjes, 2016), the GSD includes more specific soil data from China. Nonetheless, both databases include limited profiles for some regions of the world, notably Australia, Sahara and the northern territories of both Canada and Russia (Fig. 1).

The GSD includes ~~observed~~field measured data ~~on~~ of four soil physical properties (bulk density [g cm^{-3}]~~;~~, and sand, silt and clay fractions [%]) ~~and~~, three chemical properties (soil pH,~~;~~ SOC density [kg C m^{-2}]~~;~~, and STN density [kg N m^{-2}]) in the surface soil layer (see Table S2 and Fig. S1 for more details~~ed information~~), and general information on~~f~~ soil sampling (site location, sampling time, and data source ~~of data~~). Data harmonization was conducted by three~~four~~ steps. First, we screened sampling and measurement approaches of each soil property and excluded data those were not comparable to others in methodology. For instance, geographic coordinate data were included only when WGS84 or a geographic coordinate system that could be converted to WGS84 projection was used; Soil texture data were included only when the internationally accepted particle size class were used (clay < 2 μm < silt < 50 μm < sand < 2000 μm). This allowed us to construct a database of soil properties with comparable methodology. Second, we excluded records with no ~~measured data~~information on the target soil depth (0-30cm). In case that soil organic matter was measured instead of SOC, we used a Bemmelen index (0.58) to convert soil organic matter into SOC. If data of bulk density were not ~~measure~~provided, we ~~made~~ estimatesd them based on regional-specific pedotransfer functions (Schaap and Leij, 1998; Yang et al., 2007; Abdelbaki, 2018) (Table S3). Specifically, ~~W~~we ~~first~~ established empirical relationship between bulk density and SOC content ~~and~~ to ~~further~~ estimated bulk density based on measured SOC ~~in~~for those cases with ~~data were~~ missing data of~~or~~ bulk density. There were 42% profiles with measured data on bulk density and 58% profiles with estimated data on bulk density. It is true that the correction for rock fragment is important for the estimation of soil C stocks, but it remains a global challenge because existing databases usually contain limited information on gravel fractions (Jandl et al., 2014). Nevertheless, the inclusion of gravel has been evidenced to exert a relatively low impact on the calculation of SOC stocks in the surface soil layer (0-30 cm), mainly due to the fact that surface soil usually contains a low proportion of gravels (Saiz et al., 2012). Therefore, we assumed no rock fragment or the rock issue had been handled once it was not reported. ~~Third~~Finally, we extracted data on soil properties of the 0-30cm soil depth ~~based on their horizon in a profile~~ and calculated the means ~~for~~of each soil property. SOC (STN) density was calculated based on bulk density and ~~contents of~~ SOC (STN) content. ~~Finally, we divided the global land into 11 regions to lower spatial biases of the database and samples above 97.5% and below 2.5% quantile were excluded in each region to obtain a robust variogram (Pleijsier, 1989; Jiménez-Muñoz et al., 2015). We have mapped the excluded sites and found that these sites are distributed relatively random in space (Fig. S2) .). The remaining data were used for statistical analyses in order to reduce the influence of outliers.~~

The GSD also contains pedologic information on soil orders and the horizons of the sampled soil profiles, mean annual temperature (MAT), mean annual precipitation (MAP), seasonality of air temperature (TS, calculated as $100\times SD_{monthly}/Mean_{monthly}$) (Xu & Hutchinson, 2011), seasonality of precipitation (PS), mean annual normalized difference vegetation index (NDVI), elevation (global digital elevation map [DEM]), slope, and land use type for each recorded site (see Table ~~S3~~ S4 for more details). ~~Data on soil orders were originally reported based on several different soil classification systems with different standards (Carter and Bentley, 2016), so~~ Notably, it was difficult ~~for us~~ to harmonize data of soil orders~~them~~ and further quantify their roles~~, of soil orders~~ because data on soil orders were originally reported based on several different soil classification systems with different standards (Carter and Bentley, 2016). It was the same case for~~ and~~ soil horizon. Additionally, horizon information was not reported in some cases (accounting for 15% profile), while soil depth was well documented in our database. Therefore, we were not able to consider the role of soil horizons and instead we simply estimated the mean soil properties by a fixed depth of 30 cm. Nevertheless, the depth of 0-30 cm has been frequently used in the mapping and modelling of surface soil properties at regional and global scales (e.g., Batjes, 1997; Yang et al., 2010; Saiz et al., 2012; Wieder et al., 2013; Shangguan et al., 2014). As 96% of soil profiles in GSD were sampled during 1950 to 2000, we thus used multiple-year (1950-2010) averages of climatic variables from WorldClim database. For sites with missing reports on climate or topographical data, profile coordinates were used to derive data at each site using a selection of GIS layers, from the WorldClim database for MAT and MAP and GTOP30 DEM-derived surfaces.

2.2 Region-specific random forest model

The random forest (RF) model is a data mining algorithm to make predictions based on an ensemble of randomized classification and regression trees (Breiman, 2001). We mapped soil properties based on a region-specific RF approach that yields spatially explicit estimates of each pixel (see Fig. S~~3~~2 for more details on the workflow of this approach). To overcome spatial biases of the database (for example, heavy sampling in the USA), we divided the global land into 11 regions: Africa, Australia, Canada and Alaska, East Asia, Europe, Mexico, Russia, South America, tropical Asia, the USA, and West Asia (Table S2). In each region, we first constructed a RF model using the regional datasets ~~of GSD~~ and then used the model to estimate the spatial distribution of each soil property at a resolution of 1 km. Predictions were based on eight environmental variables, including MAT, MAP, TS, PS, vegetation cover conditions (NDVI), elevation, slope, and land use type (see Table ~~S3~~ S4 for more details on the data sources of each variable). To obtain a robust variogram, ~~Soil~~ soil property data below 2.5% quantile and above 97.5% quantile were excluded as outliers and were not used for modeling in each region (Pleijsier, 1989; Jiménez-Muñoz et al., 2015). Notably, these excluded samples were distributed relatively randomly in space (Fig. S3). By conducting an analysis using all data samples and comparing the results with those excluding outliers, we found similar spatial patterns and means of global surface soil properties (Fig. 2 & Fig. S4) but lower cross-validated R^2 when including all samples (Table 6 & Table S5). This implies an improvement of prediction by excluding outlier samples.~~In addition, soil C:N ratio for each pixel was calculated based on predicted SOC and STN densities.~~

Because a large number of regression trees are constructed, one major advantage of RF model is that the risk of overfitting can be reduced. Another advantage is that the prediction depends on only three user-defined parameters: the number of trees (ntree), the minimum number of data points at each terminal node (nodesize), and the number of features sampled for splitting at each node (mtry). We used ntree = 1000 (default ntree = 500) in order to achieve more stable results (Grimm et al., 2008). For nodesize and mtry, we used the default set for RF regression. Also called a “black box” approach, one major disadvantage of RF model is that the relationships between the response and predictor variables cannot be interpreted individually for every RF tree. The relative importance of variables, denoted by the percent increase in mean-squared error (%IncMSE), was estimated based on a permuting out-of-bag (OOB) method (Strobl et al., 2009 a & b). For each tree of the random forest, we compared the prediction error on the OOB portion of the data (MSE for regression) with that after permuting each predictor variable. The differences ~~a~~were then averaged over all trees, and normalized by the standard deviation of the differences. The relative percent (mean/SD) increase in MSE as compared to the out-of-bag rate (with all variables intact) was used to indicate the relative importance of each variable (Breiman, 2001). ~~Nevertheless, the importance of each environmental variable was estimated by the mean change in prediction accuracy before and after permuting each variable.~~

2.3 Uncertainty analysis

In each region, we used 10-fold cross-validation to estimate the average mapping accuracy for each target soil property. The modelling accuracy for each bootstrap sample was evaluated by the amount of variation explained by the models (R^2) and by the root mean square error (RMSE) calculated based on the observational and predicted soil property in the independent validation dataset (Table ~~S4~~S5). Model uncertainties were assessed based on the bootstrap method. A robust estimate was derived by averaging the 10-fold cross-validation samples, and the uncertainty of the estimates was calculated as the standard deviation (SD) of the 10-fold cross-validation (Fig. ~~S3~~S4).

2.4 Statistical analysis

Based on the results of the ensemble models, ~~the spatial distribution of~~we mapped each soil property (bulk density, sand, silt, clay, pH, SOC density, and STN density ~~and C:N ratio~~) and their uncertainty ~~were mapped~~ at a resolution of 1 km. Soil C:N ratios were mapped based on values of SOCD and STND, and its~~The~~ uncertainty ~~of C:N ratio can~~ could be jointly indicated by the uncertainties of SOCD and STND. We also plotted each soil property on ~~a climate space illustrated by~~ a modified Whittaker biome diagram. ~~Moreover, we explored quantitative linkages between each soil property and climate parameters.~~ To explore the roles of MAT and MAP as well as their interactions, w~~W~~e averaged soil property values for each MAT×MAP combination by a division of 1°C×100mm and explored quantitative linkages~~plotted~~ between soil properties ~~with~~and climate variables (MAT/MAP) for different~~by roughly distinguishing~~ climate types (humid *vs.* arid; warm *vs.* cold). Specifically, we used a MAP threshold of ~~at~~ 500 mm to differentiate relatively humid versus arid climates~~is used to indicate a threshold for arid climates~~ (Holdridge, 1967), while a MAT threshold of~~at~~ 10 °C ~~is~~was used to separate relatively warm~~er~~ and cold~~er~~ climates (Trewartha and Horn, 1980). To explore the role of MAT in regulating the critical MAP for a shift from alkaline to acidic soil, we further plotted the critical levels of MAP (100 mm division) at soil pH=7.0 with MAT. We ~~further~~ then compared ~~these~~ the soil properties across main biomes, including tropical forest, temperate forest, boreal forest~~s~~, tropical savannahs and grasslands, temperate grasslands and shrublands, tundra, permanent wetlands, deserts and croplands. All statistical analyses were performed using Matlab 2015a (The MathWorks Inc., Natick, MA, USA). Values were~~are~~ presented as mean ± standard deviations, if not specially noted.

3. Results

3.1 Global mapping of soil properties

Our results ~~indicate that model predictions (Fig. S3S4)~~ agreed well with the observed data across most regions (Fig. S5), and ~~that~~ the ensemble models generally explained 30~60% of the variation in soil properties (Table ~~S4~~S5). The eight soil properties showed great spatial heterogeneity across the globe in the upper 30-cm layer (Fig. 2). For instance, bulk density showed low values in the northern latitudes in the Eurasian continent, whereas high values occurred in the USA, North Africa, West Asia, and India (Fig. 2a). The clay fraction exhibited lower values at higher latitudes, whereas higher levels of sand fraction occurred at lower latitudes (Figs. 2b, 2c). The pH value of the surface soil was high (generally > 7.0) in arid regions and it was relatively low (generally < 6.0) in most forested regions (Fig. 2e). The spatial patterns of SOC density and STN density were generally similar,~~;~~ both showing~~ed~~ greater values at higher latitudes in the northern hemisphere and no consistent change with latitude in the southern hemisphere (Figs. 2f, 2g). Specifically, SOCD and STND both showed highest values in the northern high latitudes, while low values occurred in semiarid and desert regions. Soil C:N ratio showed the highest values at high latitudes in northern hemisphere, while~~ it was~~ the lowest values occurred in arid regions in Northern Africa, West Asia and Southern Europe (Fig. 2h). ~~On average, the global means of SOC density and STN density were 6.94 (SD= 4.42) kg C m^{-2} and 0.53 (SD= 0.23) kg N m^{-2} in surface soils, respectively (Table 1). In the surface soil layer, global stocks of SOC and STN were estimated to be summing up to a global total storage of 797 ± 4.1 Pg C (10^{15} g, or billion tons) and 64 ± 0.4 Pg N, respectively (Table 1).~~

3.2 Global soil-climate-biome diagram

By ~~putting~~placing data of surface soil properties on the Whittaker climate-biome diagram (Whittaker, 1962), we then documented the linkages between soil properties and climate across global biomes. We call this as the global soil-climate-biome diagram (Fig. 3). ~~As showed in Fig. 3~~Specifically, bulk density ~~generally~~ decreased with lower MAT and higher MAP (Fig. 3a~~, also see~~, Figs. 4a, 4b); sand fraction was inversely related to MAP and MAT (Fig. 3b~~; also~~, Figs. 4c, 4d), whereas the clay fraction showed an opposite pattern (Fig. 3c, Figs. 4e, 4f); and soil pH increased with higher MAT in arid climate (MAT ≤ ~~400~~ 500 mm) (Figs. 3e, 5a), while it decreased significantly with higher MAP both in cold (MAT ≤ 10°C) and warm (MAT > 10°C) climates (Figs. 3e, 5b). The critical MAP for the transition from alkaline to acidic soil (pH = 7.0) showed a non-linear increase with MAT and reached to a maximum of 400-500 mm when MAT exceeded 10 °C (Fig. 5c).

~~Generally,~~ SOC density in the upper 30-cm soil layer decreased significantly with MAT at both arid (MAT ≤ ~~400~~ 500 mm) and~~or~~ humid climates (MAT > ~~400~~ 500 mm) (Figs. 3f), whereas it increased with MAP in accordance with a saturation curve (cold climate, MAT ≤ 10°C: SOCD = 0.0737×MAP/(1+0.0049×MAP); warm climate, MAT > 10°C: SOCD = 0.0144×MAP/(1+0.0016×MAP)), showing a higher saturation threshold ~~(14.5 kg C m^{-2})~~ in cold climates (14.5 kg C m^{-2})

~~(MAT ≤ 10°C)~~ compared with that in warm climates (8.0 kg C m^{-2}) ~~(MAT > 10°C)~~ (Figs. 6b). Similarly, STN density decreased significantly with MAT (Figs. 6c) and increased with MAP in accordance with a saturation curve (cold climate: STN = 0.0401×MAP/(1+0.0502×MAP); warm climate: STN = 0.0015×MAP/(1+0.0021×MAP)), showing a higher saturation threshold ~~(0.80 vs. 0.65 kg N m^{-2})~~ in cold (0.80 kg N m^{-2}) than warm climates(0.65 kg N m^{-2}) (Figs. 6d). Combining the trends of SOC density and STN density, the C:N ratio of the upper 30-cm layer increased with MAT at a climate of MAT < 0 °C and then decreased (Fig. 6e). In contrast, the C:N ratio increased with MAP in accordance with a saturation curve (cold climate: C:N = 0.1450×MAP/(1+0.0080×MAP); warm climate: C:N = 0.3781×MAP/(1+0.0308×MAP)), showing a higher saturation threshold ~~(18:1 vs. 12:1)~~ in cold climates (18:1) compared with warm climates (12:1) (Fig. 6f).

Soil properties showed ~~different~~ varied values across and within biomes throughout the world (Table 1; Fig. 3). Mean bulk density was lowest in ~~the~~ tundra and boreal forest, ~~but~~ and it was highest in the desert and tropical thorn scrub and woodland (Tables 1). Mean sand fraction was highest in boreal forest, whereas mean clay fraction was highest in the tropical rainforests (Tables 1). Soil pH was generally lower than 5.5 in tropical forest, boreal forest and tundra, but mean pH values could approach and even exceed 7.0 in dry biomes, such as the desert, grassland and savanna (Table 1). Moreover, means of SOC and STN densities both showed high values in boreal forest and tundra, but they were extremely low in the desert and tropical thorn scrub and woodland (Table 1). ~~Overall, mean~~ Mean soil C:N ratio showed the highest values in tundra and boreal forest (> 15:1), while it was lowest in desert, temperate shrubs and grasslands (≤ 10:1) (Table 1; Figs. 3h). On average, the global means of SOC density and STN density were 6.94 (SD= 4.42) kg C m^{-2} and 0.53 (SD= 0.23) kg N m^{-2} in surface soils, summing up to a global total storage of 797 ± 4.1 Pg C (10^{15} g, or billion tons) and 64 ± 0.4 Pg N, respectively (Table 1).

4. Discussion

4.1 ~~Global soil-climate-biome diagram for~~Linkages between climate and surface soil physical properties

The soil-climate-biome diagram demonstrate~~s~~d ~~strong~~ the quantitative linkages between surface soil physical properties and climate variables at the global scale. Compared with variables associated with topography (e.g., elevation and slope), vegetation activity (i.e., NDVI) and land cover (i.e., land use type), climate variables (such as MAT, MAP, TS and PS) were~~are~~ stronger predictors of bulk density and soil texture (Fig. 7a-c). This wa~~i~~s likely due to the essential ~~effect~~ role of temperature and precipitation ~~on~~ in physical, chemical and biological processes during soil formation (Weil and Brady, 2016). ~~For instance,~~Specifically, bulk density showed an increase with higher MAT and lower MAP, ~~being~~ likely due to an accompanying decrease of SOCD (Ruehlmann and Körschens, 2009) which was driven by stronger microbial decomposition under warmer and wetter conditions~~as jointly regulated by MAT and MAP~~ (Fig. 6; see more discussion on the effect of climate on SOCD in section 4.3; Wiesmeier et al., 2019). In addition, ~~H~~higher MAT and MAP can accelerate the rate of weathering (Jenny, 1941; Lal, 2018), thus resulting in lower sand fraction and higher soil clay fraction (Fig. 4). Along with topographical variables, climate may also affect soil physical properties via erosion processes. For example~~instance~~, soil erosion is highly selective to silt, while sand is less mobile due to high weight and clay is protected by soil aggregates (Wischmeier and Mannering, 1969; Torry et al., 1997; Wang et al., 2013).

Other factors, such as historical tectonics, glaciations and soil ages, could also affect soil physical properties (Jenny, 1941; Weil and Brady, 2016), but they are often spatially correlated with ~~current~~ climate variables ~~and biome distribution~~, making it difficult to separate their role from the latter. For instance, the effect of glaciations is stronger, the soil age is younger and air temperature is lower towards higher latitudes. Likewise, the role of tectonics in rejuvenating younger soils might also be mixed by corresponding climat~~e~~ic conditions across altitudinal gradients. ~~By exploring variations in soil physical properties with elevation i~~In tropical regions, we found a significant decrease in bulk density and clay fraction with higher elevation (Fig. ~~S4~~S~~5~~6a&d). Thi~~s~~e decrease of bulk density ~~with~~ along the altitude gradient ~~i~~was likely due to an increase in SOC retention (Fig. S~~5~~6f), ~~which mainly~~being resulted~~s~~ from low rates of soil organic matter decomposition along with lower temperature (Grieve et al., 1990; Kramer and Chadwick, 2016). Meanwhile, ~~T~~the decrease of clay fraction with higher altitude ~~i~~was likely due to a younger soil age (Waite and Sack, 2011), lower weathering rate under lower temperature (Grieve et al., 1990; Kramer and Chadwick, 2016), and a downslope translocation of surface soil to lower altitude. Interestingly, ~~T~~these altitudinal gradients were consistent with the results of field studies (Dieleman et al., 2013) and also mirrored a similar trend across latitudes.

4.2 Key role of climate in determining global patterns of surface soil ~~pH~~chemical properties

Our results indicated that MAP was the most important surrogate for soil pH prediction (Fig. 7d). Such a pattern ~~may~~might be due to the increased leaching of exchangeable base cations across large-scale precipitation gradients (Jenny et al., 1941). Interestingly, ~~our~~ further analysis showed that the critical levels of MAP for the transition from alkaline to acidic soil decreased non-linearly with lower MAT owing to changing water balance (Fig. 5). Specifically, the critical MAP ranged from 400-500 mm when the MAT exceeded 10 °C and could decrease to 50-100 mm when MAT was close to 0 °C, highlighting significant interactions between MAP and MAT. Such a pattern was supported by a recent study, which revealed that the transition from alkaline to acidic soil occurred when the MAP began to exceed the mean annual potential evapotranspiration (Slessarev et al., 2016). It should be noted that, other factors besides climate variables, such as acid deposition may also contribute to regional-scale patterns of soil pH, especially in Europe, eastern North America and southern China, where have received high-level acid deposition (Bouwman et al., 2002; Vet et al., 2014).

~~4.3 Climate as drivers of SOC and STN in global surface soils~~

Our analysis also indicated~~s~~ that climate variables (e.g., MAT, MAP) ~~a~~were the strongest predictors of SOC density (Fig. 7e) ~~SOC density with higher MAP and lower MAT (Fig. 6)~~, being in agreement with the ~~pattern~~ findings of previous ~~estimates~~ studies (Post et al., 1982; Gray et al., 2009). Such a pattern reflects the fact that soil C stock depends on the balance between plant inputs (i.e.~~g.~~, litterfall and other plant debris) and microbially mediated metabolic losses of CO_2 to atmosphere (Stockmann et al., 2013), which ~~we~~are strongly controlled by climate (Davidson and Janssens, 2006; Bond-Lamberty and Thomson, 2010). ~~For instance~~In general, precipitation ~~favors~~favours net primary productivity (Del Grosso et al. 2008) and the consequent C inputs into the soil, while it intensifies weathering of the parent material and soil acidification, thus increasing formation of SOC-stabilizing minerals (Chaplot et al., 2010; Doetterl et al., 2015) and reducing decomposition of soil organic matter (Meier and Leuschner, 2010). These processes could then explain the~~Our results thus indicate that~~ increase of SOCD ~~increased~~ with MAP (Fig. 6), while it did n~~'~~ot exceed a certain threshold because of a constraint of C inputs (Del Grosso et al., 2008). Compared with precipitation, ~~T~~temperature largely affects the rate and degree of microbial decomposition of soil organic matter (Wiesmeier et al., 2019). ~~Our results indicate that~~Consequently, SOCD increased with lower MAT (Fig. 6), while it reached saturation due to a threshold of SOM stabilization (Doetterl et al., 2015).

Further analysis also ~~evidences~~revealed an interaction between MAT and MAP in shaping the patterns of SOC density ~~density~~. For instance, SOC density showed a tendency of saturation with higher MAP, while the saturation thresholds were higher under MAT ≤ 10 °C compared with MAT > 10 °C (Fig. 6). Specifically, the saturation threshold for SOC density under MAT ≤ 10 °C (14.5 kg C m^{-2}) were nearly twice of that under MAT > 10 °C (8.0 kg C m^{-2}) (Fig. 6b). These critical levels of SOCD imply a saturation threshold of SOC stocks under certain climate regime (Stewart et al., 20~~1~~07). Soil C saturation has also been evidenced by experimental studies, which indicate that ~~the~~ SOC pool has an upper limit with respect to C input levels because of a threshold of SOM stabilization efficiency (Stewart et al., 2008; Kimetu et al., 2009). These thresholds of soil C saturation can help to estimate soil C sequestration potential and provide important guidelines for regional soil steward and ecosystem management.

Previous meta-analyses indicated that C:N ratio in the soil ~~i~~was well-constrained at the global scale (Cleveland and Liptzin, 2007). ~~Due to a prevailing N limitation in terrestrial ecosystems, the C and N cycles are tightly linked (LeBauer and Treseder, 2008; Gruber and Galloway, 2008; Chapin et al., 2009).~~ Accordingly, our results indicated a strong correlation between STN density ~~as~~ and SOC density (Fig. 8) and demonstrated a similar pattern of STN density as SOC density across ~~space~~biome and climate regimes (Figs. 3 & 6). Based on a synthesis of long-term experimental results, Manzoni et al. (2008) demonstrated that the C:N ratio of the litter decreased throughout decomposition. Because soil organic matter is a result of long-term decomposition, surface soil C:N ratio is thus negatively correlated with decomposition degree while positively correlated with SOC content and turnover time (Carvalhais et al., 2014). ~~Accordingly, o~~Our analysis also indicated that higher soil C:N ratios was associated with higher SOC density ~~was associated with higher soil C:N ratios~~ (Figs. 2, 3).

4.43 Shifts in soil properties across biomes and land use types

Our analysis indicate~~s~~d that soil properties var~~y~~ied significantly ~~global~~ across global biomes (Table 1). For example~~instance~~, SOC density showed high values in boreal forests and tundra due to the slower microbial decomposition compared with biomass inputs (Hobbie et al., 2000; Hashimoto et al., 2015; Bloom et al., 2016), but these values were extremely low in drylands due to low plant cover and productivity (Delgado-Baquerizo et al., 2013). Due to fast turnover with rapid decomposition of organic matter, SOC content is relatively poor in tropical forests (e.g., Congo and Amazon tropical forests in Fig. 2f) (Carvalhais et al., 2014; Wang et al., 2018). Accordingly, previous mappings of SOC~~D~~ density have also shown relatively low values in tropical forests (Köchy et al., 2015; Jackson et al., 2017). In view of a strong~~ly~~ and negative correlation between SOC and bulk density, bulk density showed an opposite shifts across biomes (Table 1). Moreover, we also found an increase in SOC density and a decrease of soil bulk density with elevation (Fig. S~~5~~6a & Fig. S6f), likely due to a shift in climate regime and vegetation type.

The effect of land use is important for the SOC stock at a regional to local scale. ~~For instance, a~~A change of forest or grassland to croplands can significantly decrease SOC density and thus decrease soil bulk density, while reforestation generally increases SOC density and ~~thus~~ decreases soil bulk density (DeGryze et al., 2004; Machmuller et al., 2015). When comparing values at ~~a~~the same region (e.g., Southeast Asia), SOC density is obviously lower in croplands than in forests (~~compare~~ Fig. S~~6~~7a and Fig. S~~6~~7b). This difference has been also evidenced by meta-analysis based on field observations (Don et al., 2011). In the ~~the~~ Mediterranean region, an increase in the area of olive plantation and vineyard in last decades have likely contributed to a consequent increase of SOC density (Parras-Alcántara et al., 2013). Moreover, a~~n~~ re~~n~~cent assessment indicates that ecological restoration projects (e.g., Three-North Shelter Forest Program, Natural Forest Protection Project, Grain for Green Program, Returning GrazingLand to Grassland Project) in China has ~~substantinaly~~substantially increased soil and biomass C storage in the corresponding regions (Lu et al. 2018). However, our static mapping of global soil properties is not able to account for the effect of temporal land use change on SOC density.

4.4 Global carbon and nitrogen stocks in surface soils

Earlier estimates of global SOC and STN stocks were based on either an area-weighted extrapolation or an empirical model of the soil profile data according to climate, vegetation type or soil order (Post et al. 1982, Batjes, 1996; Batjes 2009; Hengl et al. 2014; Scharlemann et al. 2014). In the range of these estimates, our results based on RF modeling indicated that the global stocks of SOC were 788 ± 39.4 Pg in the upper 30-cm soil layer. Reports of global STN stocks are relatively rare compared with those of SOC stocks. Based on information on measured soil profiles, Batjes (1996) estimated global STN stocks to be 63-67 Pg N in the upper 30-cm layers. Similar to the estimates by Batjes (1996), our results indicate that global STN stocks were 63 ± 3.3 Pg in the upper 30-cm soil layer (Table 1). Despite similar estimation of global total SOC and STN stocks, our regional RF analysis has several advantages, such as the ability to model non-linear relationships, handle both categorical and continuous predictors, and resist overfitting and noise features (Breiman, 2001). By using climate, vegetation, topography and land use variables as predictors, our region-specific RF approach likely produces more robust global maps of soil properties at a finer spatial resolution.

4.5 Uncertainties in mapping surface soil properties at the global scale

In this study, we used machine learning algorithms to map global surface soil properties at a 1-km resolution. Although this approach could overcome uncertainties derived from large variations in mapping unit, several limitations still exist~~ed~~ in our analysis. First, the limited sample size in certain area may lead to estimation uncertainties. Particularly, the accuracy of the region-specific RF model partially depends on the number of sampling sites and the evenness of the spatial pattern. The limited number and uneven distribution of the soil profile may thus constrain the accuracy of region-specified RF models, especially in regions such as Russia and South America (~~Fig. S3~~Table S45).

Second, ~~the various approaches used to measure soil properties among regions and nations may also generate uncertainties in global-scale mapping. Specifically,~~ soil properties ~~were~~ have been measured using various approaches and compiled for several decades, ~~and~~ while there are no straightforward solutions ~~exist for~~ to accurately harmoniz~~eing~~ the data at the global level (Maire et al., 2015; Batjes, 2016). ~~Similar to other studies, t~~The errors due to varied sampling and measurement methods ~~across~~ over time may lead to uncertainties in our analysis ~~and also hinder reliable hindcast and forecast estimates at the global scale (Grunwald et al., 2011)~~. Moreover, ~~O~~our data base includes pedologic information on soil orders and soil horizons of sampled soil profiles~~. However,~~ and these data were originally reported based on several different soil classification systems ~~based on~~using different standards (Batjes et al., 2007; Carter and Bentley, 2016). ~~and i~~It is thus a challenge for us ~~difficult~~ to harmonize the data on soil orders and soil horizons and quantify their impacts on surface soil properties. Nevertheless, the depth of 0-30 cm has been frequently used in the mapping and modelling of surface soil properties at regional and global scales (e.g., Batjes, 1997; Yang et al., 2010; Saiz et al., 2012; Wieder et al., 2013; Shangguan et al., 2014)~~(e.g., Batjes, 1997; Wieder et al., 2013)~~. By considering essential climat~~e~~ic variables (~~mean annual temperature~~MAT, ~~mean annual precipitation~~MAP, seasonality of air temperature, seasonality of precipitation), vegetation parameters (mean annual NDVI, and land use type), and topograph~~y~~ic factors (elevation, slope) that are key to soil formation (Jenny, 1941), our Random Forest analysis ~~is~~ may have partially constrained the uncertainties due to ~~ignoring~~ the lack of information on soil orders and associated soil horizons.

Finally, ~~the~~ uncertainties may also arise from the limited independent variables used in this study. Although essential surrogate variables of climate, topography, vegetation activity, and land cover ~~(see method section)~~ were incorporated in our analysis, we still could not account for the role of soil horizons, soil ages and parental material characteristics due to the lack of global-scale dataset. For instance, surface soil (top 30 cm) can contain either a single horizon or several very different horizons with very different physical and chemical properties. ~~Due to an absence of global data for soil horizons, we were not able to consider soil horizons in our global soil mapping. soil~~Soil mineralogy, being a function of parent material, climate and soil age (Jenny, 1941), has been demonstrated to be important in determining the quantity of SOC storage and its turnover time during long-term soil development (Torn et al., 1997). Soil age may also play an important role in forming soil property (Jenny, 1941), but it is hard to evaluate its individual role in regulating spatial patterns of soil properties due to its strong interactions with climate variables. Therefore, future studies should make more efforts to consider these variables when predicting spatial patterns of soil physical and chemical properties at the global scale.

5. Conclusion

By compiling a comprehensive global soil database, we mapped eight surface soil properties based on machine learning algorithms and assessed the quantitative linkages between soil properties, climate, and biota at the global scale. Our region-specific random forest model generated high-resolution (1km) ~~reliable~~ predictions of surface soil properties, which can be potentially used as inputs for~~, and can thus provide useful inputs to~~ regional and global biogeochemical models. Our results also produced a global soil-climate-biome diagram, which indicates the quantitative linkages ~~improved the understanding of the strong correspondence~~ between soil, climate, and biomes~~biota~~. Given that significant changes in major soil properties may ~~have~~ occur~~red and will continue due to~~ in view of global environmental change (Trumbore and Czimczik, 2008; Chapin et al., 2009; Todd-Brown et al., 2013; Luo et al., 2016, 2017), more efforts should be made in future to understand the dynamics~~the co-evolution of soil, climate and biota in view~~ of the global soil-climate-biome diagram.

Acknowledgement

This work is supported by National Key Research and Development Program (2017YFC0503901), the National Natural Science Foundation of China (31321061 and 31330012) and Key Research Program of Frontier Sciences, CAS (QYZDY-SSW-SMC011). All data will be available online as a supplement when this paper is published.

References

Abdelbaki, A.: Evaluation of pedotransfer functions for predicting soil bulk density for US soils. Ain Shams Engineering Journal, 9, 1611–1619, 2018.

Adhikari, K., and Hartemink, A. E.: Linking soils to ecosystem services—A global review. Geoderma, 262, 101–111, 2016.

Alexander, E. B.: Soils in natural landscapes. CRC Press, 2013.

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., and Sparks, D. L.: Soil and human security in the 21st century. Science, 348(6235), 1261071, 2015.

~~Amundson, R.: The carbon budget in soils. Annu. Rev. Earth Pl. Sc., 29(1), 535–562, 2001.~~

Arrouays, D., McKenzie, N., Hempel, J., de Forges, A. R., and McBratney, A. B. (Eds.).: GlobalSoilMap: basis of the global spatial soil information system. CRC Press, 2014.

Batjes, N. H.: Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma, 269, 61–68, 2016.

Batjes, N. H.: Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use and Manage~~-~~ment, 25(2), 124–127, 2009.

Batjes, N.H., Al-Adamat, R., Bhattacharyya, T., et al.: Preparation of consistent soil data sets for modelling purposes: Secondary SOTER data for four case study areas. Agriculture Ecosystems & Environment, 122(1), 26–34, 2007.

Batjes, N.H.: A world dataset of derived soil properties by FAO–UNESCO soil unit for global modelling. Soil Use and Manage~~-~~ment, 13(1), 9–16, 1997.

Batjes, N.H.: Total carbon and nitrogen in the soils of the world. European Journal of Soil Science, 47, 151–163, 1996.

Bloom, A. A., Exbrayat, J. F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. Proceedings of the National Academy of Sciences of the USA, 113(5), 1285–1290, 2016.

Bond-Lamberty, B., and Thomson, A.: Temperature-associated increases in the global soil respiration record. Nature, 464(7288), 579–582, 2010.

Bouwman, A. F., Van Vuuren, D. P., Derwent, R. G., and Posch, M.: A global analysis of acidification and eutrophication of terrestrial ecosystems. Water Air Soil Pollution, 141(1), 349–382, 2002.

Breiman, L.: Random forests. Machine learning, 45(1), 5–32, 2001.

Carter, M., Bentley, S.P.: Soil properties and their correlations. John Wiley & Sons, 2016.

Carvalhais, N., et al.: Global covariation of carbon turnover times with climate in terrestrial ecosystems. Nature, 514, 213–217, 2014.

Castellano, M. J., and Kaye, J. P.: Global within-site variance in soil solution nitrogen and hydraulic conductivity are correlated with clay content. Ecosystems, 12(8), 1343–1351, 2009.

Chapin III, S., McFarland, J., McGuire, D. A., Euskirchen, E. S., Ruess, R. W., and Kielland, K.: The changing global carbon cycle: linking plant–soil carbon dynamics to global consequences. Journal of Ecology, 97(5), 840–850, 2009.

Chaplot, V., Bouahom, B., Valentin, C.: Soil organic carbon stocks in Laos: spatial variations and controlling factors. Global Change Biology. 16 (4), 1380–1393, 2010.

Chesworth, W.: The parent rock effect in the genesis of soil. Geoderma, 10, 215–225, 1973.

Cleveland, C. C., Liptzin, D.: C: N: P stoichiometry in soil: is there a “Redfield ratio” for the microbial biomass? Biogeochemistry, 85(3), 235-252, 2007.

Davidson, E. A., and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. Nature, 440(7081), 165, 2006.

DeGryze, S., Six, J., Paustian, K., Morris, S. J., Paul, E. A., Merckx, R.: Soil organic carbon pool changes following land-use conversions. Global Change Biology, 10(7), 1120–1132, 2004.

Del Grosso, S., Parton, W., Stohlgren, T., Zheng, D., Bachelet, D., Prince, S., et al.: Global potential net primary production predicted from vegetation class, precipitation, and temperature. Ecology, 89(8), 2117–2126, 2008.

Delgado-Baquerizo, M., Maestre, F. T., Gallardo, A., Bowker, M. A., Wallenstein, M. D., Quero, J. L., and García-Palacios, P.: Decoupling of soil nutrient cycles as a function of aridity in global drylands. Nature, 502(7473), 672–676, 2013.

Dieleman, W. I., Venter, M., Ramachandra, A., Krockenberger, A. K., and Bird, M. I.: Soil carbon stocks vary predictably with altitude in tropical forests: implications for soil carbon storage. Geoderma, 204, 59–67, 2013.

Ding, J., Li, F., Yang, G., Chen, L., Zhang, B., Liu, L., and Ji, C.: The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores. Global Change Biology, 22(8), 2688–2701, 2016.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Pinto, M. C., et al.: Soil carbon storage controlled by interactions between geochemistry and climate. Nature Geoscience, 8(10), 780–783, 2015.

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

Gentsch, N., Wild, B., Mikutta, R., Čapek, P., Diáková, K., Schrumpf, M., et al.: Temperature response of permafrost soil carbon is attenuated by mineral protection. Global Change Biology, 24(8), 3401–3415, 2018.

Gray, J. M., Humphreys, G. S., and Deckers, J. A.: Relationships in soil distribution as revealed by a global soil database. Geoderma, 150(3), 309–323, 2009.

Grieve, I. C., Proctor, J., Cousins, S. A.: Soil variation with altitude on volcan Barva, Costa Rica. Catena, 17(6), 525–534, 1990.

Grimm, R., Behrens, T., Märker, M., and Elsenbeer, H.: Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. Geoderma, 146(1), 102–113, 2008.

~~Gruber, N., and Galloway, J. N.: An Earth-system perspective of the global nitrogen cycle. Nature, 451(7176), 293–296, 2008.~~

Grunwald, S., Thompson, J. A., and Boettinger, J. L.: Digital soil mapping and modeling at continental scales: Finding solutions for global issues. Soil science society of America journal, 75(4), 1201–1213, 2011.

Hashimoto, S., Carvalhais, N., Ito, A., Migliavacca, M., Nishina, K., and Reichstein, M.: Global spatiotemporal distribution of soil respiration modeled using a global database. Biogeosciences, 12, 4121–4132, 2015.

Hassink, J.: The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant Soil, 191(1), 77–87, 1997.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., and Guevara, M. A.: SoilGrids250m: Global gridded soil information based on machine learning. PloS One, 12(2), e0169748, 2017.

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., et al.: SoilGrids1km—global soil information based on automated mapping. PLoS One, 9(8), e105992, 2014.

Hobbie, S. E., Schimel, J. P., Trumbore, S. E., and Randerson, J. R.: Controls over carbon storage and turnover in high-latitude soils. Global Change Biology, 6(S1), 196–210, 2000.

Holdridge, L.R.: Life zone ecology. Tropical Science Center. San Jose, Costa Rica, 1967.

Jackson, R. B., Lajtha, K., Crow, S. E., Hugelius, G., Kramer, M. G., Piñeiro, G.: The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls. Annual Review of Ecology, Evolution, and Systematics, 48, 419–445, 2017.

Jandl, R., Rodeghiero, M., Martinez, C., Cotrufo, M. F., Bampa, F., van Wesemael, B., and Lorenz, K.: Current status, uncertainty and future needs in soil organic carbon monitoring. Science of the Total Environment, 468, 376–383, 2014.

Jenny, H.: Factors of soil formation: a system of quantitative pedology. McGraw-Hill Book Company Inc, New York, 1941.

Jilling A., Keiluweit M., Contosta A.R., Frey S., Schimel J., Schnecker J., Smith R.G., Tiemann L., Grandy A.S.: Minerals in the rhizosphere: overlooked mediators of soil nitrogen availability to plants and microbes. Biogeochemistry 139, 103–122, 2018

Kimetu, J. M., Lehmann, J., Kinyangi, J. M., Cheng, C. H., Thies, J., Mugendi, D. N., and Pell, A.: Soil organic C stabilization and thresholds in C saturation. Soil Biology and Biochemistry, 41(10), 2100–2104, 2009.

Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., and Angers, D.: Soil security: solving the global soil crisis. Global Policy, 4(4), 434–441, 2013.

Köchy, M., Hiederer, R., Freibauer, A.: Global distribution of soil organic carbon–Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. Soil, 1(1), 351–365, 2015.

~~Kopittke, P. M., Dalal, R. C., Finn, D., and Menzies, N. W.: Global changes in soil stocks of carbon, nitrogen, phosphorus, and sulphur as influenced by long-term agricultural production. Global Change Biol., 23(6), 2509–2519, 2017.~~

Kramer, M. G., Chadwick, O. A.: Controls on carbon storage and weathering in volcanic soils across a high-elevation climate gradient on Mauna Kea, Hawaii. Ecology, 97(9), 2384–2395, 2016.

Lal, R. 2018. Soil and climate. In Soil and Climate (pp. 1—8). CRC Press.

~~LeBauer, D. S., and Treseder, K. K.: Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. Ecology, 89(2), 371–379, 2008.~~

Lu, F., Hu, H., Sun, W., Zhu, J., Liu, G., Zhou, W., et al. Effects of national ecological restoration projects on carbon sequestration in China from 2001 to 2010. Proceedings of the National Academy of Sciences of the USA, 115(16), 4039–4044, 2018.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., et al.: Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles, 30(1), 40–56, 2016.

Luo, Z., Wang, E., and Sun, O.J.: Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. Ecol. Appl., 27(3), 1001–1009, 2017.

Machmuller, M. B., Kramer, M. G., Cyle, T. K., Hill, N., Hancock, D., Thompson, A.: Emerging land use practices rapidly increase soil organic matter. Nature Communications, 6, 6995, 2015.

Maire, V., Wright, I. J., Prentice, I. C., Batjes, N. H., Bhaskar, R., Bodegom, P. M., and Reich, P. B.: Global effects of soil and climate on leaf photosynthetic traits and rates. Global Ecology Biogeography, 24(6), 706–717, 2015.

Manzoni, S., Jackson, R.B., Trofymow, J.A., Porporato, A.: The global stoichiometry of litter nitrogen mineralization. Science, 321(5889), 684–686, 2008.

Meier, I.C., Leuschner, C.: Variation of soil and biomass carbon pools in beech forests across a precipitation gradient. Global Change Biology 16 (3), 1035–1045, 2010.

Milne, E., Banwart, S. A., Noellemeyer, E., Abson, D. J., Ballabio, C., Bampa, F., and Black, H.: Soil carbon, multiple benefits. Environment, Development and Sustainability, 13, 33–38, 2015.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A. and Ciais, P.: A large and persistent carbon sink in the world's forests. Science, 333(6045), 988–993, 2011.

Parras-Alcántara, L., Martín-Carrillo, M., Lozano-García, B.: Impacts of land use change in soil carbon and nitrogen in a Mediterranean agricultural area (Southern Spain). Solid Earth, 4(1), 167–177, 2013.

Pleijsier, J.: Variability in soil data. In: Land Qualities in Space and Time (eds Bouma J., Bregt, A.K.), pp. 89-98. Pudoc, Wageningen, 1989.

Post, W. M., Emanuel, W. R., Zinke, P. J. and Stangenberger, A. G.: Soil carbon pools and world life zones. Nature, 298(5870), 156–159, 1982.

~~Post, W. M., Pastor, J., Zinke, P. J., and Stangenberger, A.G.: Global patterns of soil nitrogen storage. Nature, 317(6038), 613–616, 1985.~~

Rohli, R. V., Joyner, T. A., Reynolds, S. J., and Ballinger, T. J.: Overlap of global Köppen–Geiger climates, biomes, and soil orders. Physical Geography, 36(2), 158–175, 2015.

Ruehlmann, J., & Körschens, M.: Calculating the effect of soil organic matter concentration on soil bulk density. Soil Science Society of America Journal, 73(3), 876–885, 2009.

Saiz, G., Bird, M. I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T. R., and Diallo, A.: Variation in soil carbon stocks and their determinants across a precipitation gradient in West Africa. Global Change Biology, 18(5), 1670-1683, 2012.

Saiz, G., Bird, M.I., Domingues, T., Schrodt, F., Schwarz, M., Feldpausch, T. R., ... & Diallo, A. (2012). Variation in soil carbon stocks and their determinants across a precipitation gradient in W est A frica. Global Change Biology, 18(5), 1670-1683.

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J. and Minasny, B.: Digital soil map of the world. Science, 325(5941), 680–681, 2009.

Schaap, M. G., Leij, F. J.: Database-related accuracy and uncertainty of pedotransfer functions. Soil Science~~.~~, 163(10), 765-779, 1998.

Scharlemann, J. P. W., Tanner, E. V. J., Hiederer, R., and Kapos V.: Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management, 5, 81–91, 2014.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems, 6(1), 249–263, 2014.

Slessarev, E. W., Lin, Y., Bingham, N. L., Johnson, J. E., Dai, Y., Schimel, J. P., and Chadwick, O. A.: Water balance creates a threshold in soil pH at the global scale. Nature, 540(7634), 567–569, 2016.

Sposito, G., Skipper, N. T., Sutton, R., Park, S. H., Soper, A. K., and Greathouse, J. A.: Surface geochemistry of the clay minerals. Proceedings of the National Academy of Sciences of the USA, 96(7), 3358–3364, 1999.

Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F., and Six, J.: Soil carbon saturation: concept, evidence and evaluation. Biogeochemistry, 86(1), 19–31, 2007.

Stewart, C. E., Paustian, K., Conant, R. T., Plante, A. F., Six, J.: Soil carbon saturation: evaluation and corroboration by long-term incubations. Soil Biology Biochemistry, 40(7), 1741–1750, 2008.

Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., and Wheeler, I. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. Agriculture, Ecosystems & Environment, 164, 80–99, 2013.

Strobl, C., Hothorn, T., Zeileis, A.: Party on! A new, conditional variable importance measure for random forests available in party package. The R Journal, 1(2): 14–17, 2009a.

Strobl, C.: Malley J., Tutz, G.: An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods, 14(4): 323–348, 2009b..

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region. Global Biogeochemical Cycles, 23, GB2023, 2009.

Tian, H., Lu, C., Yang, J., Banger, K., Huntzinger, D. N., Schwalm, C. R., et al.: Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. Global Biogeochemical Cycles, 29(6), 775–792, 2015.

Todd-Brown, K. E., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A., and Allison, S. D.: Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences, 10(3), 1717-1736, 2013.

Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M., and Hendricks, D. M.: Mineral control of soil organic carbon storage and turnover. Nature, 389(6647), 170–173, 1997.

Torri, D., Poesen, J., Borselli, L. (1997). Predictability and uncertainty of the soil erodibility factor using a global dataset. Catena, 31(1—2), 1—22.

Trewartha, G.T., Horn, L.H.: Introduction to climate, 5th edn. McGraw Hill, New York, NY, 1980.

Trumbore, S. E., and Czimczik, C. I.: An uncertain future for soil carbon. Science, 321, 1455–1456, 2008.

Vet, R., Artz, R. S., Carou, S., Shaw, M., Ro, C. U., Aas, W., and Hou, A.: A global assessment of precipitation chemistry and deposition of sulfur, nitrogen, sea salt, base cations, organic acids, acidity and pH, and phosphorus. Atmospheric Environment, 93, 3–100, 2014.

Waite, M., Sack, L.: Shifts in bryophyte carbon isotope ratio across an elevation× soil age matrix on Mauna Loa, Hawaii: do bryophytes behave like vascular plants? Oecologia, 166(1), 11–22, 2011.

Wang, B., Zheng, F., Römkens, M. J., Darboux, F.: Soil erodibility for water erosion: A perspective and Chinese experiences. Geomorphology, 187, 1–10, 2013.

Wang, J., Sun, J., Xia, J., He, N., Li, M., & Niu, S.: Soil and vegetation carbon turnover times from tropical to boreal forests. Functional Ecology, 32(1), 71–82, 2018.

~~Wei, S., Dai, Y., Duan, Q., Liu, B., & Yuan, H. (2014). A global soil data set for earth system modeling. Journal of Advances in Modeling Earth Systems, 6(1), 249-263.~~

Weil, R. R., and Brady, N. C.: The nature and properties of soils. Pearson Education, USA, 2016.

Whittaker, R. H.: Classification of natural communities. Botanical Review, 28(1), 1–239, 1962.

Wieder, W.R., Bonan, G.B., Allison, S.D.: Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change, 3(10), 909–912, 2013.

Wiesmeier, M., Barthold, F., Blank, B., and Kögel-Knabner, I.: Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil, 340(1-2), 7–24, 2011.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Luetzow, M., Marin-Spiotta, E.,et al.: Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. Geoderma, 333, 149–162, 2019.

Wischmeier, W.H., Mannering, J.V.:. Relation of soil properties to its erodibility. Soil Science Society of America Journal, 33(1), 131–137, 1969.

Wynn, J. G., Bird, M. I., Vellen, L., Grand-Clement, E., Carter, J., and Berry, S. L.: Continental-scale measurement of the soil organic carbon pool with climatic, edaphic, and biotic controls. Global Biogeochemical Cycles, 20, GB1007, doi:101029/2005GB002576, 2006.

Xu, T., and Hutchinson, M.: ANUCLIM version 6.1 user guide. The Australian National University, Fenner School of Environment and Society, Canberra, Austrialia, 2011.

Yang Y., Mohammat A., Feng J., Zhou R., Fang J.: Storage, patterns and environmental controls of soil organic carbon in China. Biogeochemistry, 84, 131–141, 2007.

Yang, Y., Fang, J., Ma, W., Smith, P., Mohammat, A., Wang, S., & Wang, W. E. I. (2010). Soil carbon stock and its changes in northern China's grasslands from 1980s to 2000s. Global Change Biology, 16(11), 3036-3047.

Yang, Y., Fang, J., Ma, W., Smith, P., Mohammat, A., Wang, S., and Wang, W. E. I.: Soil carbon stock and its changes in northern China's grasslands from 1980s to 2000s. Global Change Biology, 16(11), 3036–3047, 2010.

Yang, Y., Fang, J., Tang, Y., Ji, C., Zheng, C., He, J., and Zhu, B.: Storage, patterns and controls of soil organic carbon in the Tibetan grasslands. Global Change Biology, 14(7), 1592–1599, 2008.

Yang, Y., Li, P., Ding, J., Zhao, X., Ma, W., Ji, C., and Fang, J.: Increased topsoil carbon stock across China's forests. Global Change Biology, 20(8), 2687–2696, 2014.

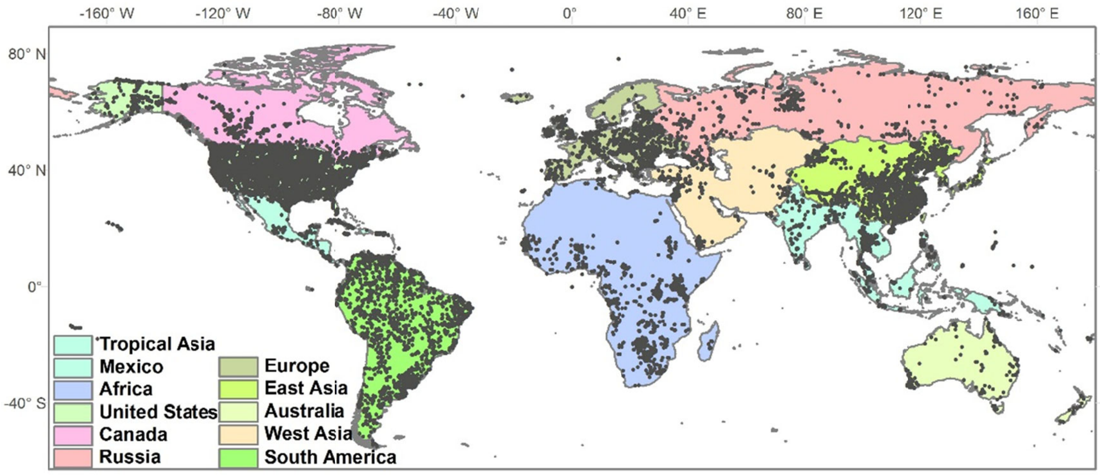

Figure 1. Global distribution of 28222 soil profiles included in the global soil database (GSD).

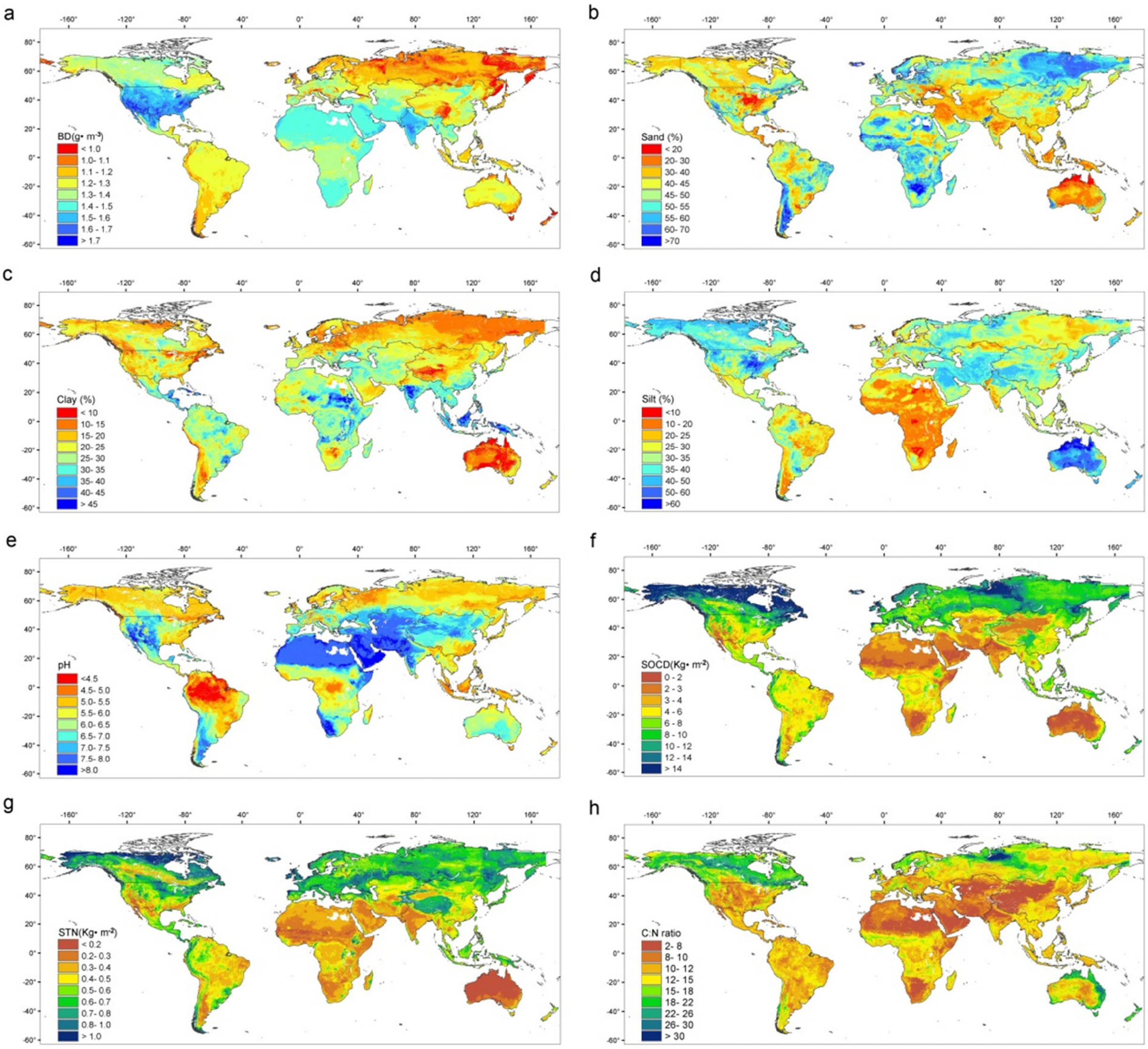

Figure 2. Map<u>s</u> of ~~worldwide~~ <u>surface (0-30cm)</u> soil properties ~~in the upper 30 cm soil layer~~. a, BD (bulk density, g m^{-3}); b, Sand fraction (%); c, Silt fraction (%); d, Clay fraction (%); e, pH; f, SOCD (soil organic carbon density, kg m^{-2}); g, STND (soil total nitrogen density, kg m^{-2}); and h, C:N ratio.

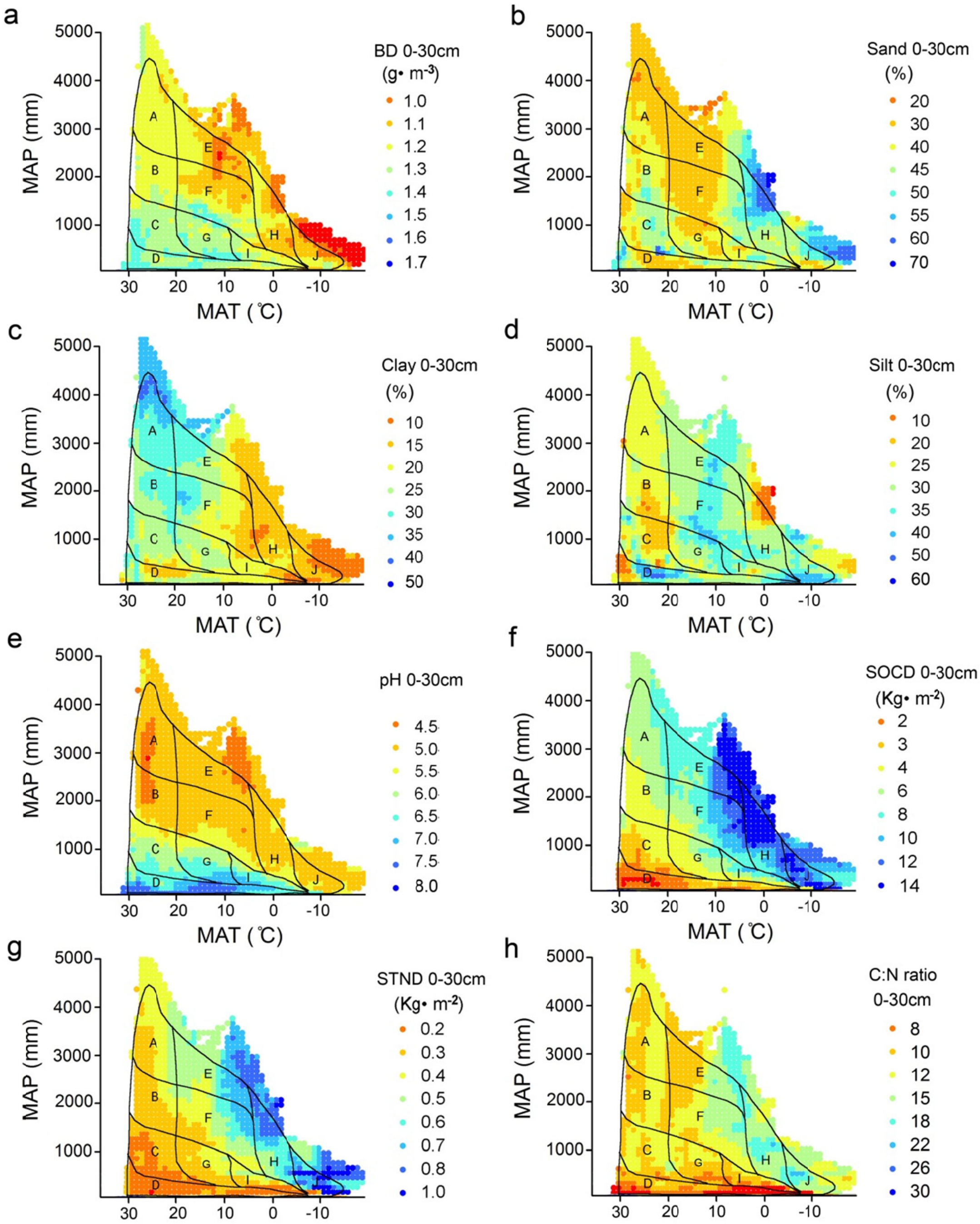

Figure 3. Changes in surface (0-30m) soil properties ~~of the upper 30-cm layer~~ on the Whittaker biome diagram. Each square shows the average C density within each 1 °C of MAT and 100 mm of MAP. Each biome type in the modified Whittaker biome diagram is indicated by a capital letter. A, Tropical rainforest; B, Tropical seasonal forest; C, Tropical thorn scrub and woodland; D, Desert; E, Temperate rainforest; F, Temperate forest; G, Savanna; H, Boreal forest; I, Grassland; and J, Tundra.

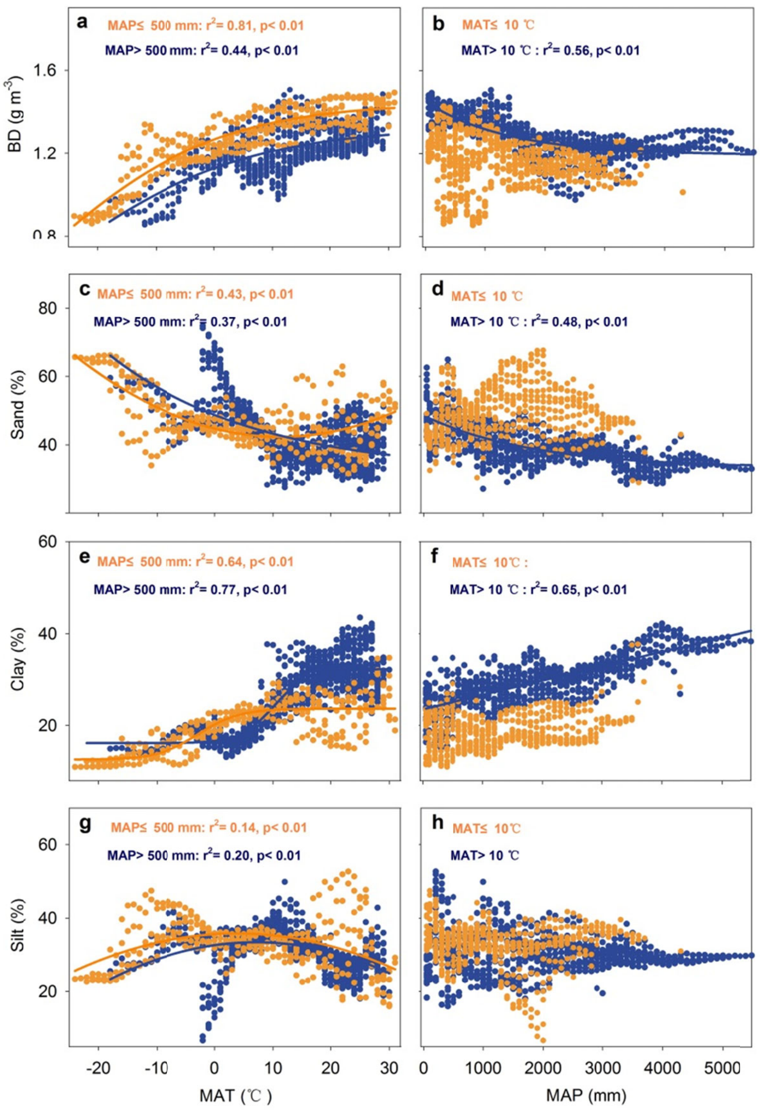

Figure 4. Changes in surface(0-30 cm)~~upper 30 cm~~ soil bulk density (BD, g·m^{-3}), sand fraction (%), clay fraction (%) and silt fraction (%) with mean annual precipitation (MAP) and mean annual temperature (MAT). We used ~~400~~ 500 mm of MAP as a threshold of transition from arid to humid climate, and 10 °C of MAT as a threshold of transition from cool to warm climate.

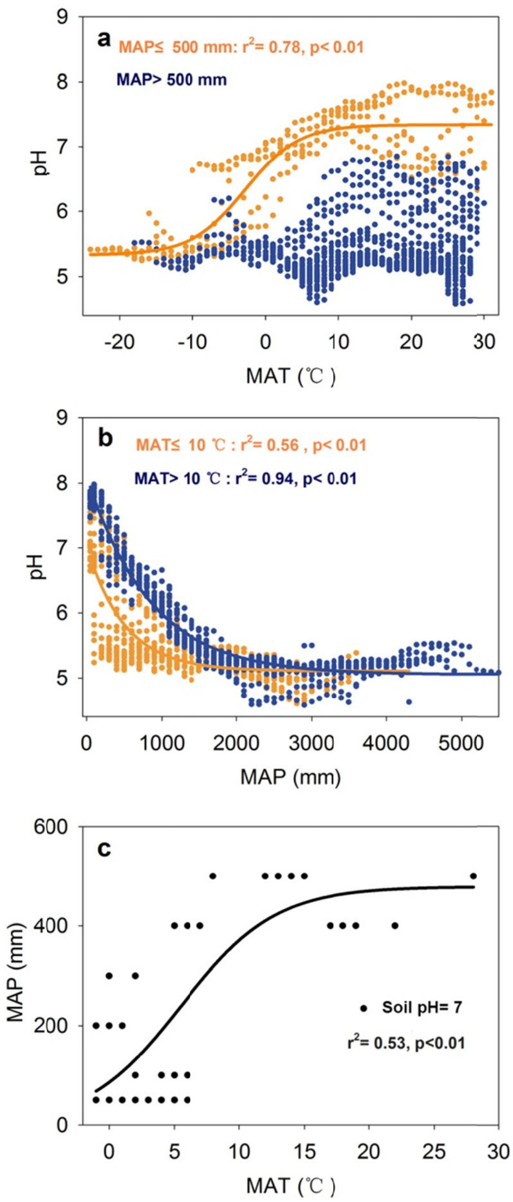

Figure 5. Changes in surface(0-30 cm) soil pH with climate. a, mean annual temperature (MAT); b, mean annual precipitation (MAP); and c, changes in critical ~~levels~~ of MAP at soil pH=7.0 with MAT. The 'critical MAP' here means the corresponding MAP at a soil pH of=7.0, which indicates a shift from alkaline to acidic soil. We used MAP of ~~400~~ 500 mm as a threshold of transition from arid to humid climate, and MAT of 10 °C as a threshold of transition from cool to warm climate.

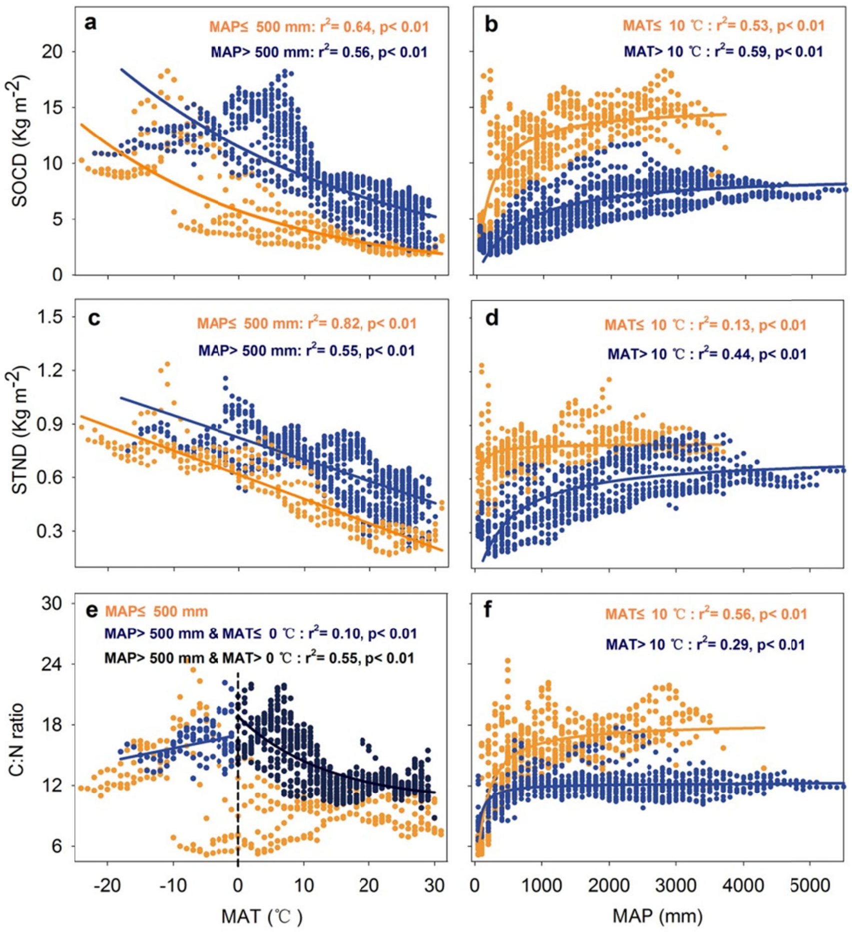

Figure 6. Changes in surface(0-30 cm) ~~upper 30 cm~~ soil organic carbon density (SOCD), soil total nitrogen density (STND) and C:N ratios with mean annual precipitation (MAP) and mean annual temperature (MAT). We used MAP of ~~400~~ 500 mm as a threshold of transition from arid to humid climate, and MAT of 10 °C as a threshold of transition from cool to warm climate. (a) and (b), SOCD; (c) and (d), STND; and (e) and (f), C:N ratios. Each dot shows the average value within each 1 °C MAT and 100 mm MAP.

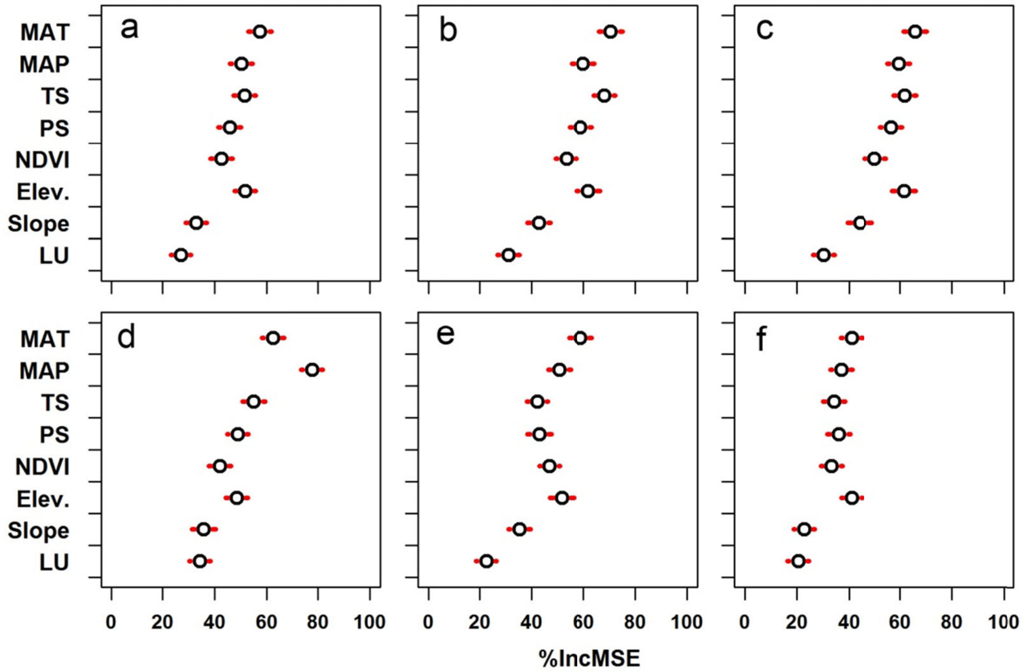

Figure 7. Importance of variables, denoted by the percent increase in mean-squared error (%IncMSE), for each soil property estimation ~~RF model constructed from the training dataset in the top 30-cm layer~~. a, BD (bulk density, g m^{-3}); b, Sand fraction (%); c, Clay fraction (%); d, pH; e, SOCD (soil organic carbon density, kg m^{-2}); f, STND (soil total nitrogen density, kg m^{-2}). MAT, MAP, TS, PS, Elev. and LU indicate mean annual temperature, mean annual precipitation, annual temperature seasonality, annual precipitation seasonality, elevation and land use type, respectively.

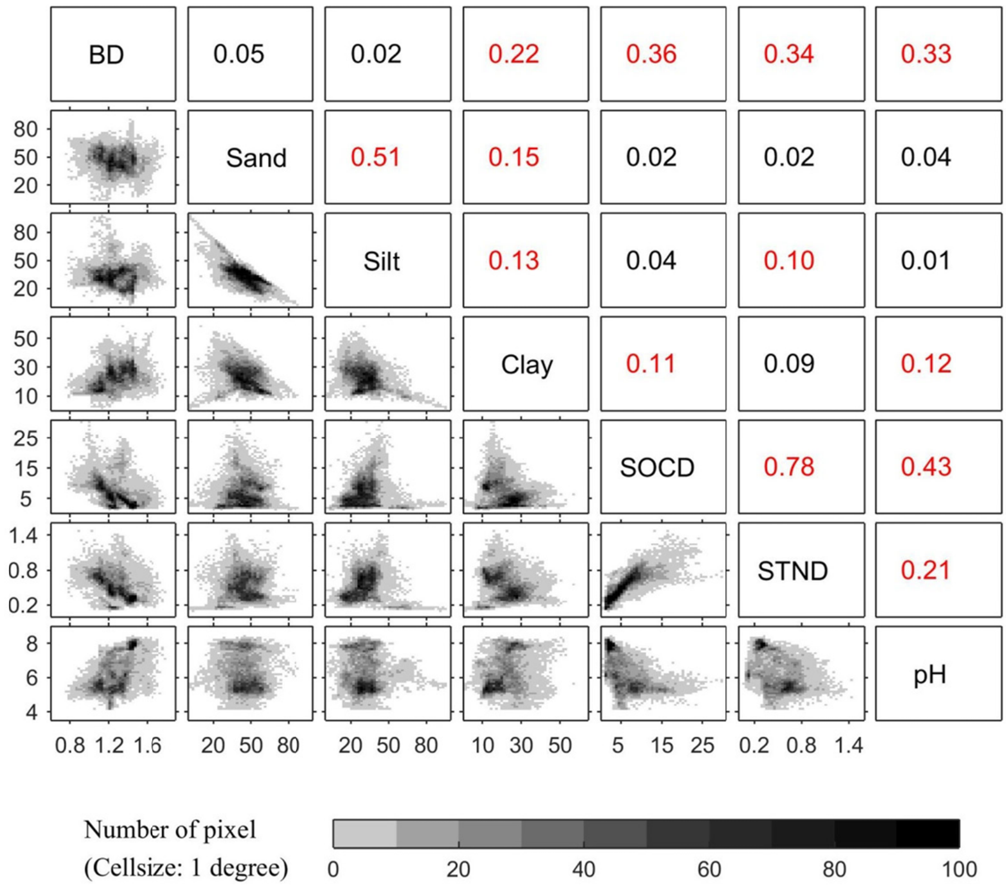

Figure 8. Correlation between surface (0-30cm) soil properties. R^2 between each two soil properties is shown in the upper plots with red color indicating $R^2 > 0.1$. BD, SOCD, and STND indicate bulk density, soil organic carbon density, and soil total nitrogen density, respectively.

Table 1. Mean values of surface (0-30 cm) soil properties ~~by~~ across terrestrial biomes of the world.

Biome	Area	Bulk density	Sand	Silt	Clay	pH	SOCD	STND	C:N ratio	SOC stock	STN stock
	(10^6 ha)	($g \cdot m^{-3}$)	(%)	(%)	(%)		($kg \cdot m^{-2}$)	($kg \cdot m^{-2}$)		(Pg)	(Pg)
TroF	1877	1.27±0.10	42.42±9.37	27.05±7.25	30.53±5.97	5.30±0.74	5.67±1.62	0.48±0.13	11.79±1.53	107±0.56	9±0.05
TemF	992	1.28±0.21	45.43±8.94	34.97±6.89	19.59±5.65	5.80±0.80	8.82±3.46	0.64±0.18	13.99±4.07	87±0.47	6±0.04
BF	1435	1.16±0.12	50.01±8.77	32.77±6.03	17.22±4.52	5.36±0.31	11.11±3.56	0.66±0.15	17.07±4.98	159±1.94	10±0.21
TSG	1915	1.34±0.10	47.92±13.06	25.64±16.68	26.44±9.05	6.25±0.80	3.82±1.65	0.32±0.13	12.28±4.05	73±0.62	6±0.04
TGS	1148	1.30±0.16	45.14±11.13	34.71±10.68	20.15±6.73	6.97±0.58	5.36±2.45	0.55±0.20	10.19±3.27	62±0.41	6±0.05
Deserts	2674	1.40±0.12	43.28±10.66	33.67±12.89	23.05±6.80	7.45±0.63	2.78±1.09	0.34±0.12	8.64±2.47	74±1.19	9±0.10
Tundra	644	1.16±0.16	47.57±8.92	37.02±7.35	15.41±3.12	5.44±0.34	13.78±4.51	0.81±0.19	17.18±5.06	89±2.07	5±0.10
Croplands	1984	1.34±0.15	40.70±11.08	33.13±9.87	26.16±7.04	6.40±0.74	6.54±2.80	0.58±0.22	11.36±2.19	130±0.50	11±0.09
PW	159	1.23±0.12	41.24±8.74	33.45±7.59	25.31±6.16	5.83±0.53	9.77±4.16	0.72±0.23	13.62±4.11	16±0.16	1±0.04
Total	12829	1.29±0.16	45.20±10.84	32.17±10.97	22.63±7.89	6.18±1.01	6.94±4.42	0.53±0.23	12.63±4.68	797±4.10	64±0.41

Notes: We include croplands and permanent wetlands in this table, although they are not single biomes. ~~Abbriations~~Abbreviations: TroF, Tropical forests; TemF, Temperate forests; BF, Boreal forests; TSG, Tropical savannahs and grasslands; TGS, Temperate grasslands and shrublands; PW, Permanent wetlands. Spatial variability of soil properties within each biome was estimated as standard deviations. Uncertainties of total SOC and STN stocks were estimated as standard deviations based on10-fold cross-validation.