Supplementary material

1 Mass conservative aggregation

In order to unify the spatial resolution and geographic coordinate system of dataset from different sources, we need to make sure that the total amount of stock for soil, vegetation, etc. doesn't change during aggregation and transformation, i.e. the variable need to be mass conservative. However, 'state' variable such as temperature, vegetation types do not need to fulfill the mass conservative requirement, nor they should. In our study, we developed a mass conservative method to maintain mass for carbon stocks. We first multiply the variable that need to be aggregated (X_{fine}) by corresponding land area (A_{fine}) at grid cell level represented by equation (1), then aggregate the product (XA_{fine}) by summing the values in N×N grids cell depending on the target resolution (equation (2)). The land area is also sum to the target resolution (equation (3)). Finally, the area-weighted variable is derived by dividing aggregated product (XA_{coarse}) by corresponding land area (A_{coarse}) as illustrated by equation (4).

$$XA_{fine} = X_{fine} \times A_{fine} \quad (1)$$

$$XA_{coarse} = SUM(XA_{fine}) \quad (2)$$

$$A_{coarse} = SUM(A_{fine}) \quad (3)$$

$$XA_{coarse} = \frac{XA_{coarse}}{A_{coarse}} \quad (4)$$

We applied the method to all datasets that requires aggregation including soil, vegetation and GPP that were used in the study.

2 Bulk density correction

The bulk density (BD) in SoilGrids and LandGIS are too high due to two reasons. First, the measurements of BD are less and missing in many horizons (Hengl et al., 2017). And the measurements of BD in permafrost region, especially in Canada forest soil and Russian, are problematic (personal communication with Tomislav Hengl). In this study, we applied a pedotransfer function from Köchy et al. (2015) to make correction based on organic carbon concentration (we only applied the function to the grid cells where carbon > 8%):

$$BD = (1.38 - 0.31 \times \log\left(\frac{OC}{10}\right)) \times 1000$$

3 Model selection for extrapolation of soil

In this section, we introduce the framework that we used to select the models for extrapolating soil from 0 – 2m to full soil depth.

3.1 Different characteristics of permafrost and non-permafrost soil

The amount and vertical distribution of soil organic carbon are largely influenced by vegetation which fixes atmospheric CO2 and transport carbon into the land ecosystem. However, the SOC stock have a much more complicated relationship with productivity of plants than a simple linearly one (Jackson et al., 2017). The higher biomass, which implies more carbon sequestration by aboveground biomass, however, does not necessarily lead to increases in SOC storage. Although the processes of soil formation, accumulation, and stabilization have been intensively studied and debated, the mechanisms that determine the soil carbon stock, especially in deeper soil, are still unclear. Instead of using process modelling approach, we chose statistical approach to extrapolate each soil profiles in the gridded dataset from 2m to full depth. The reason of performing soil carbon stock extrapolation is that we have little knowledge on how much the carbon stored in the soil that is deeper than 2m, although deeper soil is a crucial component in the climate-carbon cycle feedback. The other reason is the different dataset report SOC stock at different depths. The advantage of using statistical method is that we do not need to know the mechanisms that control the soil processes. Instead, we select simple empirical mathematical models that can represent and predict the in-situ soil profiles.

We used 425 permafrost peatland profiles from ISCN soil database and 1000 profiles from WOSIS soil database to study the characteristics of vertical distribution of SOC. Figure S1 shows the accumulated SOC stock profiles with depths in permafrost and non-permafrost region. The vertical distribution of carbon with depth in permafrost soil has a distinguished feature that the SOC has a high linear relationship with depth. This fact implies the soil carbon keeps increasing even after 3 meters in permafrost soil (Figure S1b). However, we have no idea to what depth can soil carbon keep increasing and the total amount of the storage in permafrost peatland due to the limited observational depth of SOC. In contrast, soil profiles in non-permafrost region stop increasing mostly before 2 meters. The results demonstrate the necessity of extrapolating soil to full depth, especially for permafrost soil.

2.2 Selection of models

We included 12 models (Table S1) for predicting SOC stock to full soil depth. Figure S2 shows an example result in which the data points that is shallower than 1m were used to fit all the models and predict the point that is deeper than 2m for a typical soil profile. Due to the different mathematical characteristics of the models, the prediction has quite a spread. Relatively ‘conservative’ models including model ensemble BHIJKL tend to underestimate the carbon stock while the more ‘aggressive’ ones ACDEFG tend to overestimate the stock.

In the sense that we do not know which one or group of models can best predict the accumulated carbon storage, we conducted a selection process (see Methods) by grouping all the models into all possible combination and rank the performance for all the model averaging results as shown in Table 2 All the models were used to fit the WOSIS data which covers most of the biomes and ISCN database which covers only permafrost soil. We conducted three batch of experiments in the same manner but used data points within different depths. The data points lower than 50cm, 100cm and 200cm were used to predict the SOC that is higher than 200cm. Our goal is to find the ensemble of models that has the highest model performance, the best coverage, the minimum error and AIC.

2.3 Extrapolating soil with different method

The main goal of using several models is to search for the best group of models that can best predict the vertical distribution of soil carbon stock and we compared the below approaches for that purpose:

1. The Bayesian Model Averaging (BMA) method is used in this study to find the best model ensemble for the prediction of soil carbon storage to full depth. The MODELAVG Matlab toolbox (Vrugt, 2016) which implemented many different model averaging techniques including BMA method. The advantage of BMA method is that it considers explicitly the uncertainty of prediction of a target variable which can provide a probabilistic distribution of weight for each model instead of only a weighted-average, deterministic prediction. By maximize the likelihood function from the training dataset, the weights $\beta = \{\beta 1, \ldots, \beta k\}$ and standard deviation $\sigma = \{\sigma 1, \ldots, \sigma k\}$ are estimated.
2. Equal weights averaging (EWA) which consider each the participating model have the same weight and the prediction is derived by equal-weighted averaging the model results.

The complete combination among different models are also compared and the best model ensembles are obtained by maximizes MEF, minimizes KL and minimizes AIC. The results show that EWA and BMA methods have similar performances (Table S2). We choose EWA method due to it have a slightly better coverage of observations.
Two model ensembles were selected from the model selection framework that can best represent circumpolar and non-circumpolar region based on observational datasets in the two regions, respectively. The performance of the chosen ensemble is synthesized in Figure S3. It shows the ensemble DIJKL overall can well predict the carbon stock in non-circumpolar region that is deeper than 200cm only using points lower than 50cm. Model efficiency is 0.83 and the residue between observation and prediction is little biasd in the prediction (Figure S3c). The histogram (Figure S3b) shows the prediction has the same distribution as the observations. The ensemble was also used to predict observations within different percentiles and over 70% of observations can be included in the uncertainty ($[-\sigma, +\sigma]$). The results show that the selected models can well represent the vertical distribution of C_{soil} thus we used them to extrapolate the global gridded datasets in order to obtain the total soil carbon storage in the soil. The selected model ensemble ACDEF for the circumpolar soil have lower model efficiency and less well represent the soil in the region (Figure S4). We then applied extrapolation on three

global datasets which are Sanderman, SoilGrids and LandGIS. The averaged results of ensemble DIJKL is used to extrapolate non-circumpolar soil from 2m to full soil depth and ensemble ACDEF to extrapolate circumpolar soil.

References

Jackson, R. B., Lajtha, K., Crow, S. E., Hugelius, G., Kramer, M. G., and 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.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., and Bauer-Marschallinger, B.: SoilGrids250m: Global gridded soil information based on machine learning, PloS one, 12, e0169748, 2017.

Köchy, M., Hiederer, R., and 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, 351-365, 2015.

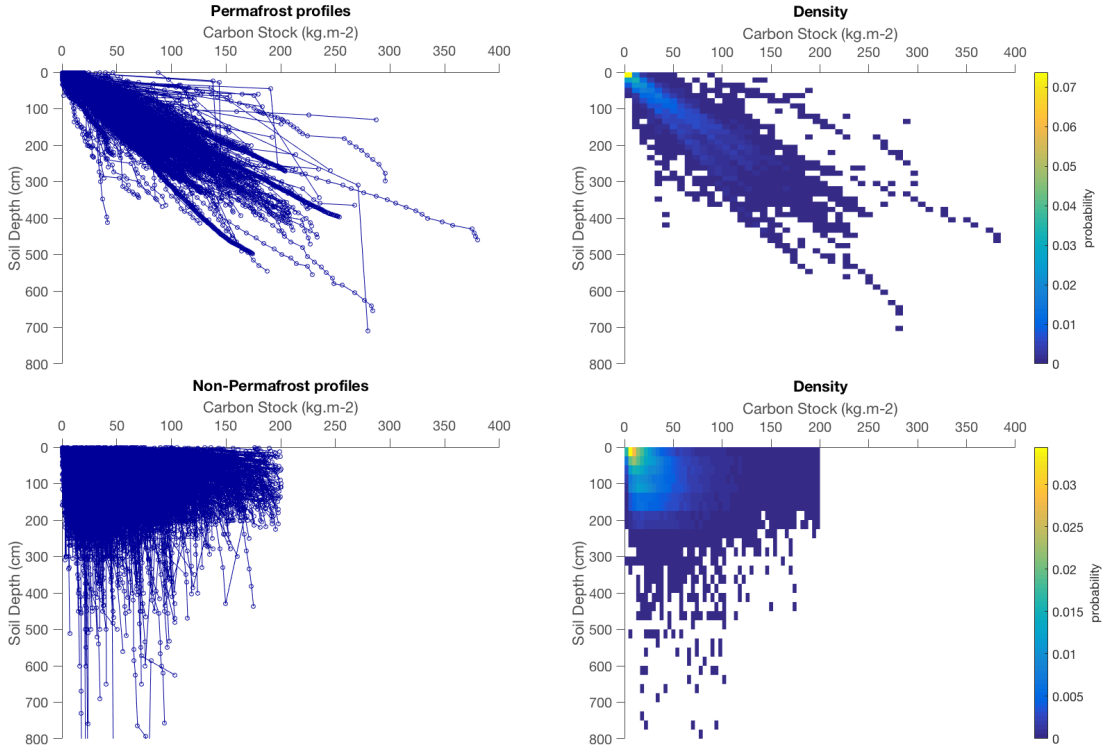
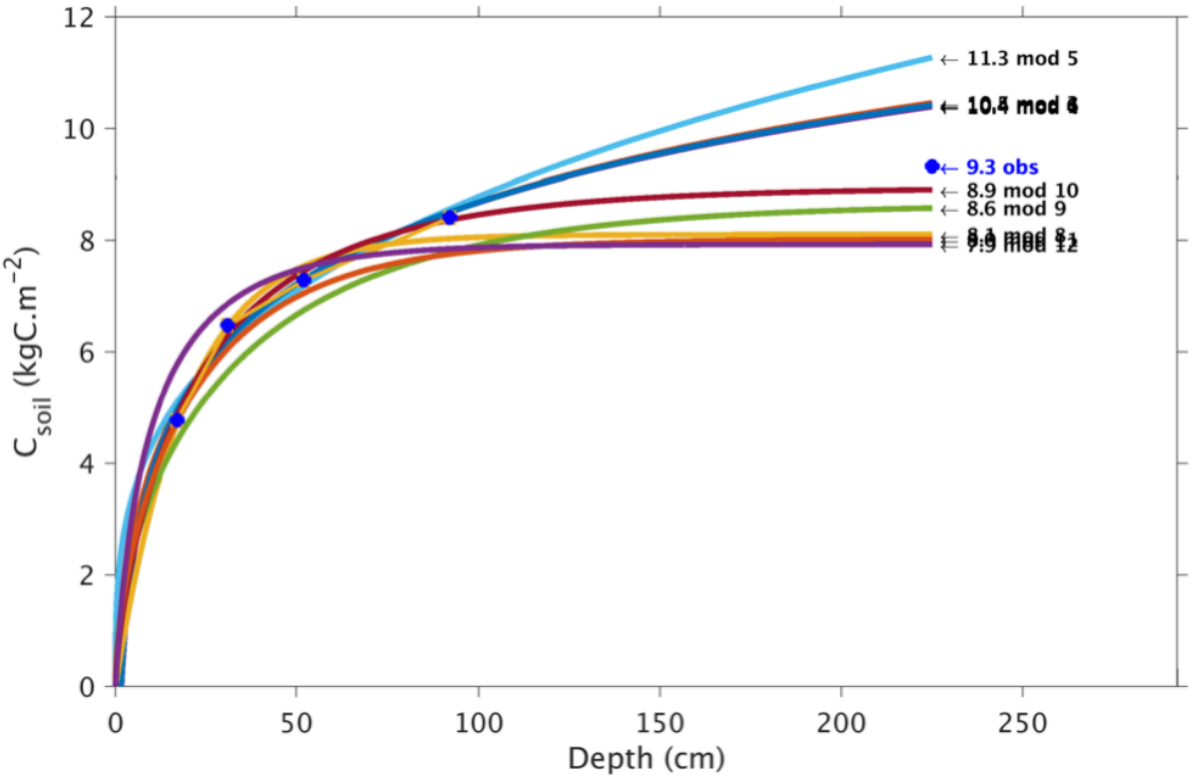


Figure S1: **The vertical distribution of accumulated SOC stock (kg.m-2) with depth (cm).** (a) 425 soil profiles of permafrost peatland region and (c)1000 soil profiles of non-permafrost region. The probability distribution density of SOC for (b) permafrost, (d) non-permafrost. The blue open circle represents observational data points in each profile.



115 Figure S2: **An example of soil profile vs models.** Overlay the observational points and model results.

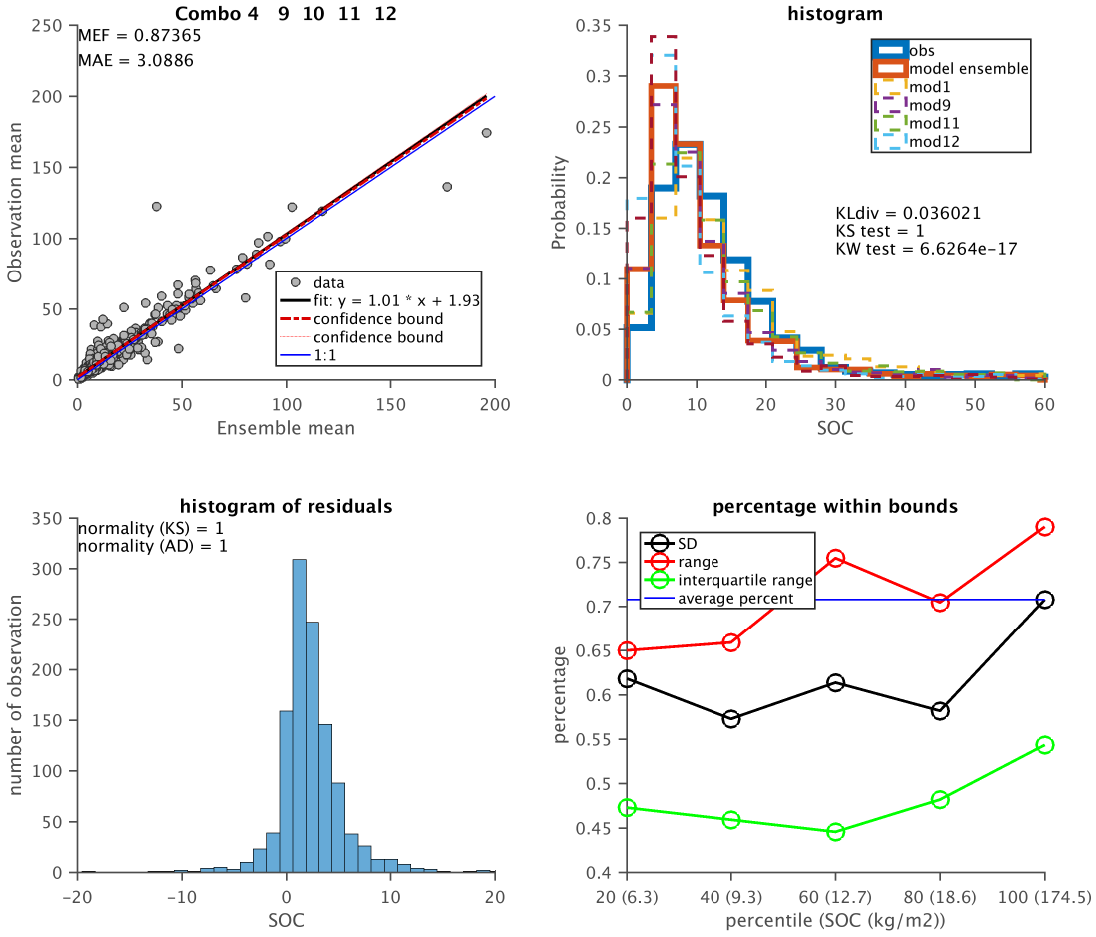


Figure S3: Performance of the averaged results of model D, I, J, K and L in predicting soil carbon storage from 50cm to 200cm using WOSIS data. (a) Ensemble mean vs. observation, 1:1 line in blue. (b) The histogram of observation, model ensemble and each model. It shows the Kullback-Leibler distance from model ensemble mean to observation, the two-sample Kolmogorov-Smirnov test (1 represent the model ensemble mean and the observation come from the same distribution, 0 otherwise), the p-value of Kruskal-Wallis test (significant if $p<0.05$). (c) residue between model ensemble mean and observation. KS represents the one-sample Kolmogorov-Smirnov test (1 represent the model ensemble mean and the observation come from the same distribution, 0 otherwise). AD represents Anderson-Darling test (1 represent the model ensemble mean and the observation come from the same distribution, 0 otherwise). (d) The coverage of observation data points within [-σ, +σ], [min, max], [25%, 75%] and average.

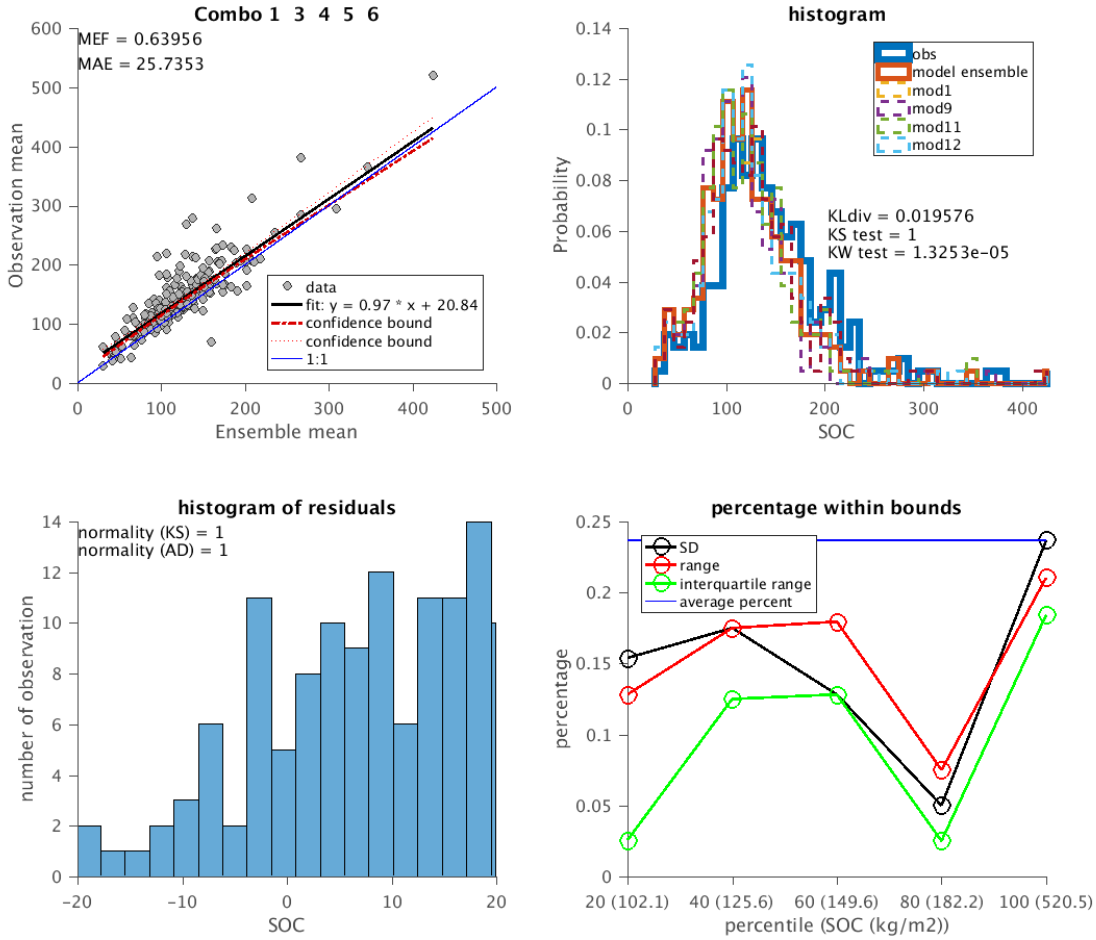


Figure S4: The same as Figure 3 except for using ISCN data and model ensemble of A, C, D, E and F to predict soil carbon storage from 200cm to deep soil.

Table S1: Empirical functions candidates for extrapolation of soil carbon

	Equation

A	$a \cdot D^b + C$
B	$a \cdot e^{b \cdot D} + c \cdot e^{d \cdot D}$
C	$a \cdot log\ (b \cdot D + 1)$
D	$a \cdot \log\ (b \cdot D + c)$
E	$K \cdot log_{10}(D) + I$
F	$(10^I \cdot D^{K+1})/(\mathrm{K} + 1) + \mathrm{c}$
G	$a + b \cdot D$
H	$b \cdot (1 - \beta^D)$
I	$b \cdot (1 - \beta^D)^a$
J	$a \cdot (1 - e^{-(D/b)^c})$
K	$a \cdot (1 - e^{-b \cdot D})^c$
L	$a \cdot (1 - \frac{\log(1 - (1 - b) \cdot e^{-c \cdot D})}{\log\ (b)})$

145

150 **Table S2: Performance of different methods.**

	Circumpolar		Non-circumpolar	
	EWA	**BMA**	**EWA**	**BMA**
RMSE	36.575	37.686	5.482	5.292
AIC	1516.134	1528.526	3977.226	3895.513
KL	0.020	0.020	0.039	0.036
MEF	0.640	0.617	0.862	0.872
Coverage (%)	14.5	13.5	61.9	50.3