The permafrost carbon inventory on the Tibetan Plateau: a new evaluation using deep sediment cores

Running title: Carbon stock across alpine permafrost

Jinzhi Ding[1,2], Fei Li[1,2], Guibiao Yang[1,2], Leiyi Chen[1], Beibei Zhang[1,3], Li Liu[1,2], Kai Fang[1,2], Shuqi Qin[1,2], Yongliang Chen[1], Yunfeng Peng[1], Chengjun Ji[4], Honglin He[5], Pete Smith[6], and Yuanhe Yang[1*]

[1]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Inner Mongolia University of Technology, Inner Mongolia 010051, China

[4]Department of Ecology, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing 100871, China

[5]Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[6]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen, AB24 3UU, UK

[*]**Corresponding author:** Dr. Yuanhe Yang, tel.: + 86 10-6283 6638, fax: + 86 10-6283 6632, E-mail: yhyang@ibcas.ac.cn

Key words: alpine permafrost, carbon-climate feedback, deep sediment, soil organic carbon stock, support vector machine

Article type: Primary Research Article

Revised manuscript submitted to *Global Change Biology*

17th-February-2016

Manuscript information: 39 pages, two tables, six figures, five supplementary tables, six supplementary figures, and three appendixes.

Abstract The permafrost organic carbon (OC) stock is of global significance because of its large pool size, and the potential positive feedback to climate warming. However, due to the lack of systematic field observations and appropriate upscaling methodologies, substantial uncertainties exist in the permafrost OC budget, which limits our understanding of the fate of frozen carbon in a warming world. In particular, the lack of comprehensive estimates of OC stocks across alpine permafrost means that current knowledge on this issue remains incomplete. Here we evaluated the pool size and spatial variations of permafrost OC stock to 3 metres depth on the Tibetan Plateau by combining systematic measurements from a substantial number of pedons (*i.e.*, 342 three-metre-deep cores and 177 50-cm-deep pits) with a machine learning technique (*i.e.*, support vector machine, SVM). We also quantified uncertainties in permafrost carbon budget by conducting a Monte Carlo simulation. Our results revealed that the combination of systematic measurements with the SVM model allowed spatially explicit estimates to be made. The OC density (OC amount per unit area, OCD) exhibited a decreasing trend from the southeastern to the northwestern plateau, with the exception that OCD in swamp meadows was substantially higher than that in surrounding regions. Our results also demonstrated that Tibetan permafrost stored a large amount of OC in the top 3 metres, with the median OC pool size being 15.31 Pg C (interquartile range: 13.03-17.77 Pg C). 44% of OC occurred in deep layers (*i.e.*, 100-300 cm), close to the proportion observed across the northern circumpolar permafrost region. The large carbon pool size, together with significant permafrost thawing suggests a risk of carbon emissions and positive climate feedback across the Tibetan alpine permafrost region.

Introduction

The permafrost region is widely distributed across high-latitude and high-altitude regions, covering 24% of the total land area of the Northern Hemisphere (Zhang *et al.*, 1999). Due to accumulation of remnants of plants and animals over thousands of years, these frozen soils have sequestered substantial quantities of organic matter (Hugelius *et al.*, 2014). It has been reported that permafrost OC stock accounts for more than half of the global soil OC stock (Jobbagy & Jackson, 2000; Carvalhais *et al.*, 2014), equivalent to the sum of OC stored in vegetation and the atmosphere (Hugelius *et al.*, 2014). Because of cryogenic (freeze-thaw) mixing and repeated sediment deposition, more than half of the organic carbon (OC) is buried in deep permafrost sedimentary deposits (Schirrmeister *et al.*, 2002; Grosse *et al.*, 2011; Hugelius *et al.*, 2014). The tremendous amount of OC, especially those stored at depth (> 1 m) in permafrost is of global significance because of its potential positive feedback to climate change with warming and associated permafrost thawing (Koven *et al.*, 2015a). During the last few decades, permafrost regions have experienced significant climate warming (Schuur *et al.*, 2015), with widespread occurrence of permafrost thawing such as active layer deepening and thermal erosion (Grosse *et al.*, 2011). These thawing processes could result in a high emission risk of frozen carbon stored at greater depths (Strauss *et al.*, 2013). The conversion of a small fraction of this frozen carbon stock into greenhouse gases and their release into the atmosphere could trigger significant positive feedback to climate warming (Whiteman *et al.*, 2013; Koven *et al.*, 2015b). Therefore, reliable evaluation of the permafrost OC stock will help to quantify the risk and to better predict the direction and strength of the carbon-climate feedback (Schuur *et al.*, 2015).

79 The permafrost carbon stock has received increasing attention among the global
80 change research community, with three representative studies available at regional and
81 circumpolar scales. Regionally, using actual measurements derived from 117 locations
82 together with previously published data from 22 sites, Ping *et al.* (2008) presented the
83 one-metre-deep OC stock estimation of 98.2 Pg C across the North American Arctic
84 region. Over a broader geographical scale, Tarnocai *et al.* (2009) used the Northern
85 Circumpolar Soil Carbon Database (NCSCD), to estimate for the first time, OC stock
86 to a depth of 3 metres across the northern circumpolar permafrost region to be 1024
87 Pg C. Recently, Hugelius *et al*. (2014) updated the estimate to 1035 Pg C by adding
88 more three-metre-deep pedons and also a detailed uncertainty analysis. These budget
89 studies have clearly depicted the basic characteristics of permafrost carbon stock;
90 however, large uncertainties remain because of the following three constraints. First,
91 despite decades of effort, inadequate and uneven distributions of pedon observations
92 are the most important limiting factor, especially for the deep soil or deposits. For
93 instance, the widely-used NCSCD database contains only 46 three-metre-deep pedons
94 (Tarnocai *et al.*, 2009). The recently updated version of this database (NCSCDv2)
95 includes 524 and 356 pedons for the 200 and 300 cm depths, respectively (Fig. 1a),
96 which could be expected to reduce uncertainties in permafrost OC budget (Hugelius *et*
97 *al.*, 2014). However, given the vast area and diverse landscapes across permafrost
98 regions, as well as the clustered patterns of the sampling sites (Mishra *et al.*, 2013),
99 the scarcity of pedon observations, especially at three-meter depth, remains the largest
100 source of uncertainty in the permafrost OC budget (Mishra *et al.*, 2013; Hugelius *et*
101 *al.*, 2014).
102
103 Second, the lack of effective upscaling approaches introduces substantial uncertainties

in both magnitude and spatial patterns of the estimated OC stock. Most previous budget studies were performed by grouping and averaging point observations of OC density (OCD) into thematic classes according to landscape units (Ping *et al.*, 2008) or soil mapping units (Tarnocai *et al.*, 2009; Hugelius *et al.*, 2014), and then multiplying by the areal extent of the thematic classes. Nonetheless, large spatial heterogeneity, along with scarce pedon observations can lead to significant errors with this kind of simple averaging (Mishra *et al.*, 2013). With the rapid development of computers and related technology, machine learning techniques such as artificial neural networks (Li *et al.*, 2013) and support vector machines (SVM) (Ueyama *et al.*, 2013; Were *et al.*, 2015), have become effective tools to predict spatial patterns of soil physical and chemical properties. Compared with traditional statistical models, machine learning techniques are advantageous because they do not depend on the assumption that data should be drawn from a given probability distribution (Drake *et al.*, 2006). Machine learning techniques are able to overcome non-linearity and over-fitting problems that usually occur in multivariate regression models (Were *et al.*, 2015). Despite these recognized strengths, applications of these techniques in regional carbon budgets are seldom reported (Tor-Gunnar & Leigh, 2013; Were *et al.*, 2015).

Third, most previous budget studies were confined to high-latitude regions, with significant knowledge gaps for high-altitude regions. Although some studies have been conducted (*e.g.*, Mu *et al.*, 2015), a comprehensive evaluation of the pool size and spatial patterns of permafrost OC stock across alpine regions remains elusive. In particular, the permafrost OC stock across alpine regions has not been combined with the Northern Circumpolar Soil Carbon Database (NCSCD) to evaluate the Northern Hemisphere permafrost carbon stock (Fig. 1a). Given the extensive distribution of

permafrost in alpine regions (Zhang *et al.*, 1999), exclusion of this component from northern hemisphere permafrost carbon syntheses renders our understanding of global permafrost carbon stock incomplete.

The Tibetan Plateau is the largest high-altitude permafrost region, accounting for approximately three quarters of the total area of alpine permafrost in the Northern Hemisphere (Wang & French, 1995). Similar to other permafrost regions around the world, significant climate warming (Wang *et al.*, 2008) and consequent permafrost thawing (Li *et al.*, 2012; Wu *et al.*, 2015) have occurred on the plateau during recent decades. These characteristics, together with high sensitivity of permafrost OC to climate warming (Strauss *et al.*, 2013; Schaedel *et al.*, 2014), make the plateau an ideal platform for permafrost carbon studies. To date, several studies have evaluated OC stocks at various depths across the plateau. For instance, by combining extensive field observations with a high-resolution satellite dataset, Yang *et al.* (2008) evaluated OC stock in the top one metre across alpine grasslands on the plateau, amounting to 7.36 Pg C. That study provided the scientific basis for understanding the pool size of the permafrost carbon stock in alpine regions, and has been widely cited by the global change research community. However, this estimation is confined to the top soil layer (0-100 cm), leaving deep OC in frozen sediment unquantified. In fact, OC stored in deep layers (> 100 cm) could account for nearly half of the total OC stock in the top three metres (Hugelius *et al.*, 2014). To address this issue, a recent study by Mu *et al*. (2015) integrated 11 new deep sediment cores with previously published measurements, and estimated the 25-metre-deep OC stock to be 160 Pg C on the plateau, with a three-metre-deep stock of 33 Pg C. Nevertheless, the deep OC stock estimated by Mu *et al*. (2015) still suffers from considerable uncertainties, since deep

pedon observations are absent for most parts of the plateau (Fig. 1b). Moreover, the deep OC stock estimated by Mu *et al*. (2015) lacks a spatially explicit pattern and a detailed uncertainty analysis. Therefore, a large-scale systematic field investigation on permafrost carbon stock, including those within deep layers, is necessary to gain a spatially explicit estimation of OC stock on the plateau.

To evaluate the size and spatial patterns of three-metre-deep permafrost OC stocks on the plateau, we conducted two-year field sampling campaigns and obtained samples from 342 three-metre-deep sediment cores and 177 50-cm-deep soil pits across 173 sampling sites. By combining a high-resolution satellite dataset with interpolated meteorological and edaphic datasets, we then extrapolated site-level measurements of OCD to the regional scale using a support vector machine (SVM) model. We further evaluated uncertainties in regional carbon budget by conducting Monte-Carlo simulations. Specifically, our objectives were to: (1) develop a suite of methodologies to quantify the three-metre-deep OC stock with a detailed uncertainty analysis, and (2) quantify the pool size and spatial distribution of permafrost carbon stock on the Tibetan Plateau.

Materials and Methods

Study area

This study was conducted on the Tibetan Plateau (Fig. 1b), which is a vast elevated plateau on the Earth, with an average elevation of 4000 m above sea level (Yang *et al.*, 2008). The plateau has the largest extent of permafrost in the low-middle latitudes of the world, with a permafrost area of ~1.35×10^6 km^2, covering 67% of the plateau area (Mu *et al.*, 2015). The average active layer thickness is 2.4 m with a range of 1.3-3.5

179 m along the Qinghai-Tibetan Highway (Pang *et al.*, 2009). Similar to other permafrost
180 regions around the world, Tibetan permafrost has experienced significant thawing,
181 such as increased active layer thickness (Wu & Zhang, 2010) and thermal erosion (Fig.
182 1d). The climate is characterized as cold and dry across the main body of the plateau.
183 Mean annual temperature (MAT) ranges between -4.9 and 6.1 ^{o}C, and mean annual
184 precipitation (MAP) ranges from 84.3 to 593.9 mm, about 90% of which falls within
185 the growing season from May to September. The permafrost regions are mainly
186 covered by three grassland types, including the alpine steppe, alpine meadow and
187 swamp meadow. The dominant species are *Stipa purpurea* and *Carex moorcroftii* in
188 the alpine steppe, *Kobresia pygmaea* and *K. humilis* in the alpine meadow, and *K.*
189 *tibetica* in the swamp meadow, respectively (Yang *et al.*, 2015).
190
191 **Site-level measurements**
192 During the summers (July and August) of 2013 and 2014, we collected 519 pedons
193 from 173 sampling sites throughout the geographical extent of alpine grasslands on
194 the Tibetan Plateau (Fig. 1b). The sampling sites covered broad climatic gradients and
195 major grassland types across the study area. Of all sampling sites, 91 sites were from
196 alpine steppe, 75 sites were from alpine meadow, and 7 sites were from swamp
197 meadow. To better characterize the spatial heterogeneity of OCD in swamp meadow,
198 we supplemented unpublished data from extra 36 profiles at 12 sampling sites on the
199 plateau (three replications within each site, collected by Prof. Jingyun Fang's group).
200 At each site, we set up five 1 m × 1 m quadrats located at each corner and the centre
201 of a 10 m × 10 m plot. For each quadrat, aboveground biomass was clipped at the
202 ground level and pooled. Pedon samples were collected within three quadrats along a
203 diagonal line of the plot. Specifically, of all the 173 sampling sites, 342 boreholes

from 114 sites were drilled at depths of 0-10, 10-20, 20-30, 30-50, 50-70, 70-100, 100-150, 150-200, 200-250 and 250-300 cm (Fig. 1c). Due to the pretty high cost of deep pedon sampling, 177 pits from 59 sites, at depths of 0-10, 10-20, 20-30 and 30-50 cm, were then excavated on the northwestern plateau. Bulk density samples were obtained for each pit using a standard container with a fixed volume size of 100 cm^3, while bulk density samples were not available for those deep cores due to practical constraints. To obtain bulk density for pedon samples derived from 342 boreholes, we additionally sampled 51 natural soil vertical sections at the same depths as the boreholes from 17 sites using a standard container with 100 cm^3 in volume. Bulk density was calculated as the ratio of the oven-dry soil mass to the container volume. We then developed an empirical relationship between measured bulk density and the related OC content derived from the 51 natural soil profiles (Fig. S1) to predict bulk density for deep cores (*e.g.*, Post *et al.*, 1982; Yang *et al.*, 2009).

In the laboratory, all samples were indoor air-dried, sieved (2 mm mesh), and handpicked to remove fine roots for subsequent measurements. Soil texture (*i.e.* clay content, silt content, and sand content) was measured using a particle size analyser (Malvern Masterizer 2000, UK) after removal of organic matter and calcium carbonates. Total carbon (TC) content was measured using an elemental analyser (Vario EL III, Elementar, Germany). Inorganic carbon (IC) content was determined with a carbonate content analyser (Eijkelkamp 08.53, Netherlands). Organic carbon (OC) content was then obtained by subtracting IC from TC. OC density (OCD) for a given depth was calculated using Eq.1:

$$OCD = \sum_{i=1}^{n} T_i \times BD_i \times OC_i \times \frac{(1 - C_i)}{100} \qquad (1)$$

where OCD, T_i, BD_i, OC_i, and C_i are OC density (kg C m^{-2}), soil thickness (cm), bulk density (g cm^{-3}), OC (g kg^{-1}) content, and volume percentage of the fraction > 2 mm at layer i, respectively (Yang *et al*., 2008). Notably, volumetric OC content in kg C m^{-3} was also calculated for easy comparison with other studies (Table S1).

Model inputs

To obtain climate dataset across the study area, we compared four widely-used interpolation methods based on r^2 and root mean square errors (*RMSE*s). Given that the Cokriging interpolation displayed the best performance across the four methods (Fig. S2), we chose this method using altitude as a covariant to retrieve MAT and MAP from 2010 to 2014 for each sampling site, and also their spatial distributions at a resolution of 10 km ×10 km. The climate records from 73 weather stations on the plateau were obtained from the China Meteorological Administration (http://cdc.nmic.cn/home.do). Likewise, the spatially gridded database of soil texture was obtained using Kriging interpolations. The interpolation analyses were performed using the Geostatistical Toolbox of ArcMap 10.0 (Environmental Systems Research Institute, Inc., Redlands, CA, USA).

Enhanced vegetation index (EVI) is a remotely-sensed vegetation index, which is designed to represent vegetation biomass that can be measured from earth-orbiting satellites, aircraft, or with field instruments (Goward *et al.*, 1985; Huete *et al.*, 2002). It has been reported that the growing season's EVI was also closely correlated with OC density in Tibetan alpine grasslands (Yang *et al.*, 2008). The moderate resolution imaging spectroradiometer (MODIS) EVI data were obtained from the United States Geological Survey (USGS) (http://modis.gsfc.nasa.gov/), with a spatial resolution of

250 m × 250 m for every 16-day interval, over the period 2010-2014. We then developed the monthly composites from the original EVI data using the Maximum Value Composition (MVC) method proposed by Holben (1986), and further resampled to 10 km × 10 km resolution. The monthly EVI data were subsequently averaged over the growing season from May to September to generate seasonal values.

Model predictions

Support vector machine (SVM), a machine learning method, uses kernel functions to construct an optimal hyperplane in a high- or infinite-dimensional space (Burges, 1998; Drake *et al.*, 2006). The new optimal hyperplane, where complex non-linear patterns can be simply represented, can be used to separate classes (*i.e.*, classification), or fit data and make predictions (*i.e.*, support vector regression) with minimal empirical risk and complexity of the modelling function (Nello & John, 2000). In this study, combined with the MODIS-EVI dataset, climatic and edaphic properties, support vector regression was conducted to predict permafrost carbon stock on the Tibetan Plateau. The 'e1071' package in the software R was used to perform SVM analysis (R version 3.1.1, R Development Core Team, 2014), and the source code was provided in Appendix 1.

To remove the effects of outliers on the models, we excluded outliers as detected by Q-Q normal test of the residuals and Cook's distance of the SVM model and the subsequent nonlinear model. Two sites were detected and excluded from the final simulation. Given that SVM model underestimated OCD in the azonal swamp meadow, we stratified this grassland type out of the model prediction, averaged pedon

observations of OCD (57 pedons from 19 sites) and multiplied by its area to estimate OC stock for the swamp meadow. For the alpine steppe and alpine meadow, we first simulated spatial distributions of OCD in the top 50 cm using a SVM model, and then extrapolated them to obtain OCD at depths of 0-100, 0-200 and 0-300 cm based on the relationships between OCD at 50-cm-depth and the corresponding values at other depths (Fig. 2).

To construct input variables for the SVM training, a statistical screening was performed to select potential regressors. Specifically, we first quantified relative contributions of the regressors using the Lindeman-Merenda-Gold (LMG) method (Fig. S3). It decomposes r^2 into non-negative contributions that automatically sum to the total r^2, with bootstrap confidence intervals to assess stability of the ranking (Lindeman *et al.*, 1980). We then detected the collinearity among the input variables. Both clay and silt content were excluded from the SVM model due to high variance inflation factors (VIFs$\geqslant$10) (Were *et al.*, 2015). After screening, the final predictors entered the SVM model included EVI, sand content, MAT, and MAP.

During the model training, the grid search method (Kavzoglu & Colkesen, 2009) was carried out to identify the best parameters (cost = 100, gamma = 0.001). A final SVM model was then developed using these best parameters. To test the predictive ability of the model, "leave-one-out" cross-validation was conducted. The high r^2 and small *RMSE* suggest the efficiency and fidelity of the SVM model (Fig. 3a). The SVM model were further used to predict OCD in the top 50 cm based on the spatially gridded input datasets covering the whole region of interest, except for the swamp meadow.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2)$$

where *RMSE* represents the root mean square error, n is sample size, y_i and $\hat{y}_i$ display measured and predicted values, respectively.

To further generate spatial distributions of OCD in deep layers across the alpine steppe and alpine meadow, we established nonlinear models between OCD in the top 50 cm and OCD at 0-100, 0-200 and 0-300 cm depths. Given the close relationships between them (r^2 ranges from 0.68 to 0.94, Fig. 3b-d), these nonlinear models were then used to extrapolate OCD from surface to deep layers.

Uncertainty analysis

Monte Carlo methods are a broad class of computational algorithms which rely on repeated random sampling to obtain numerical results (Rubinstein & Kroese, 2007). Of them, Monte Carlo simulation relies on the process of explicitly representing uncertainties by specifying inputs as probability distributions (Chew & Walczyk, 2012). In this study, Monte Carlo simulation was used to quantify potential errors derived from the following three sources: (i) uncertainties introduced by interpolating site-level meteorological and edaphic measurements to obtain the spatially gridded datasets, (ii) the measured errors of the EVI dataset, and (iii) uncertainties in predicting OCD in the top 50 cm using the SVM model and subsequent extrapolation to deep layers.

We first quantified the uncertainties of the input variables, including MAT, MAP, sand content, and EVI. Specifically, the standard errors of the MAT, MAP and sand content

were generated during the interpolation process. For the MODIS EVI data, measured errors would originate from aerosol optical thickness error, aerosol model error, and reflectance approximation error as well as calibration error (Lin *et al*., 2011). It has been reported that 94% of globally retrieved EVI values fall within the theoretical MODIS one-sigma error bar (± (0.02 + 0.02 × value)), indicating that the error in a given index value is 0.02 plus 2% of the index value (Vermote & Kotchenova, 2008). Similar to other studies (Lin *et al*., 2011; He *et al*., 2014), we thus used 0.02 + 0.02 × value as standard error for EVI data.

We then performed 1000 Monte Carlo simulations of OCD for each pixel. For each simulation, the input variables were randomly generated based on the normal distributions using the standard error of MAT, MAP, sand content, and EVI as the standard deviation (SD), and then subjected to the SVM model (to calculate OCD of the 0-50 cm layer) and the nonlinear models (to extrapolate 0-50 cm layer to deep layers). To account for uncertainties introduced by the nonlinear models, we used values that were randomly generated from a normal distribution of the 95% confidence interval of regression results for each grid pixel. Finally, we calculated the inter-quartile (difference between the 75^{th} and 25^{th} percentiles) of the 1000 iterations of the simulated OCD for each pixel, and the summed quartiles were used to assess the uncertainty of the OC stock over the alpine steppe and alpine meadow. We also obtained the relative uncertainty by dividing the inter-quartiles by median values for each pixel (Fig. S4). It should be noted that, the OC stock uncertainty of the swamp meadow was addressed using a 1000-iteration bootstrap resampling method at various depths (0-50, 0-100, 0-200 and 0-300 cm). Inter-quartiles were also calculated to assess the relative uncertainties of the OCD in the swamp meadow.

Results

Vertical distributions of OC stock down the 3-metre profiles

The OCD decreased as soil depth increased in all of the 3-metre profiles, with the maximum occurring in the top 50 cm (Fig. 4a). The vertical distributions of OCD differed among three major grassland types on the plateau. Specifically, the OCD of each soil layer in the swamp meadow was much larger than those in the other two grassland types ($P < 0.05$). However, no significant differences were observed between in the alpine steppe and alpine meadow, except for soil layers of the 0-50 and 50-100 cm.

Similar to vertical patterns of OCD, the highest proportion of OCD occurred in the upper 50cm layer, with significant vegetation-specific differences (Fig. 4b). For the alpine steppe, alpine meadow and swamp meadow, 41%, 50%, and 36% of the total OCD in the top 300 cm was contained in the uppermost 50 cm, respectively. The smallest proportion of OCD in top 50 cm layer was found in the swamp meadow, despite the fact that more OC was stored in the top 50 cm layer (Fig. 4a). This suggests that the swamp meadow had a higher proportional distribution of OCD in deep layers than the other two grassland types. The proportion of OCD below 100 cm depth in the alpine steppe, alpine meadow and swamp meadow was 45%, 34%, and 48%, respectively.

Spatial variations of OC stock across permafrost regions

The OCD exhibited large spatial variability across permafrost regions on the Tibetan Plateau, with a decreasing trend from the southeast to the northwest (Fig. 5). Such a

pattern was contingent on geographic extents of grassland types (Fig. 1b), *i.e.*, relatively larger OCD tended to be found in the alpine meadow while smaller OCD occurred in the alpine steppe. Notably, OCD in the swamp meadow at different depths was significantly higher than that in surrounding regions.

The average OCD across Tibetan alpine grasslands was estimated at 5.45, 7.44, 10.68, and 13.39 kg C m^{-2}, in the depths of 0-50, 0-100, 0-200 and 0-300 cm, respectively (Table 1). Accordingly, total OC stock was equal to 6.23, 8.51, 12.22 and 15.31 Pg C in 0-50, 0-100, 0-200 and 0-300 cm depths, respectively (Table 2). Mean OCD significantly differed among various grassland types ($P < 0.05$), with the order of alpine steppe < alpine meadow < swamp meadow. The OC stock at 0-300 cm depth in the alpine steppe, alpine meadow and swamp meadow were estimated to be 5.48, 6.53, 3.31 Pg C, of which 47%, 38% and 53% occurred in deep soil layers, respectively (Table 2). Compared with measurements across the northern circumpolar permafrost regions, both the alpine steppe and alpine meadow had much lower OCD, but similar vertical proportional distributions along the profile (Fig. 6). Notably, soils in the swamp meadow had comparable OCD values and similar proportional distributions along the profile with those across the northern circumpolar permafrost regions.

Discussion

Combining inventory data with machine learning technique is an effective methodology for spatially explicit carbon estimation

This study offered a large-scale comprehensive investigation on deep OC stock across alpine permafrost on the Tibetan Plateau, and generated 519 pedon observations in total, including 342 three-metre-deep sediment cores (Appendix 3). Using a SVM

model, we up-scaled these site-level observations to regional-scale OCD in the top 50 cm. Our results demonstrated that in combination with other satellite-derived, climatic and edaphic variables, the SVM displayed good performance to predict OCD for the 0-50 cm depth, with high r^2 and low *RMSE*, according to the results of the "leave-one-out" cross-validation (Fig. 3a). The "leave-one-out" cross-validation results also demonstrated that SVM showed better performance in predicting OCD than Kriging interpolation and satellite-derived approach that were frequently used in regional carbon budgets (higher r^2 and smaller *RMSE*, Table S2). This can be ascribed to the several advantages of the SVM approach. First, compared with Kriging interpolation, SVM incorporates high-resolution satellite datasets and other spatially gridded climatic and edaphic variables into the model, and thus overcomes biases due to large spatial heterogeneity and uneven distribution of the sampling sites. Second, compared with satellite-derived approach, SVM can deal with complex multivariate models, including both linear and nonlinear relationships (Nello & John, 2000; Drake *et al.*, 2006), which is a common situation when exploring associations of OCD with biotic, climatic and edaphic properties (Yang *et al.*, 2008). In addition, as a machine learning technique, SVM is not constrained by any statistical premise such as normality and independence, and also overcomes the limitations of parametric and non-parametric statistical methods, such as spatial autocorrelation (Nello & John, 2000; Drake *et al.*, 2006; Were *et al.*, 2015). Hence, combining systematic measurements with SVM is an ideal approach for spatially explicit carbon estimation across permafrost regions.

Although SVM and subsequent extrapolation models provided good performance in regional carbon budget, some uncertainties still existed (Fig. S4). The uncertainties

425 may be partly derived from the extrapolation of surface to deep OC stock using
426 nonlinear models, since the relationships between surface and deep OCD could be
427 compromised by processes in periglacial environments (*e.g.*, cryoturbation, slope
428 mass movement of material) (Grosse *et al.*, 2011; Harden *et al.*, 2012). Nevertheless,
429 our nonlinear models successfully quantified the relationships of OC stock among
430 various layers for both the alpine steppe and alpine meadow (Fig. 3b-d), possibly due
431 to the unique characteristics of alpine permafrost on the Tibetan Plateau. For instance,
432 the overall arid climate on the plateau tends to suppress periglacial processes (Wang,
433 1997), which could then result in limited distributions of periglacial landforms (Yang
434 *et al.*, 2010). Consequently, vertical extrapolation may not induce large uncertainties
435 into regional carbon budget. Even so, some uncertainties could still be generated due
436 to the lack of deep soil cores across the western plateau for practical reasons
437 (remoteness, road conditions, etc.). Nevertheless, both the eastern and western Tibetan
438 Plateau share common soil type (cambisols) (Wu *et al.*, 2003; Yang *et al.*, 2015), and
439 comparable permafrost types (Brown *et al.*, 1998). To further test whether regression
440 models derived from the eastern half of the plateau is applicable in the western part of
441 the plateau, we constructed the regression model of OCD between the depths of 0-30
442 cm and 0-50 cm of the eastern plateau, and validated the model using actual
443 measurements from the western part of the plateau. The validation results showed
444 good performance of the model constructed in the eastern part of the plateau (r^2 =
445 0.98, *RMSE* = 0.81; Fig S5), demonstrating the reliability of the regression model in
446 both parts of the plateau. Consequently, extrapolation of topsoil OC stock to deep
447 layers is deemed both statistically and physically reliable for the alpine steppe and
448 alpine meadow.
449

450 The lack of ice content may also lead to potential uncertainties in our budget, since ice
451 content is an important parameter for estimating permafrost carbon stock, especially
452 in ice-rich permafrost regions (*e.g.*, the Yedoma region) (Strauss *et al.*, 2013; Hugelius
453 *et al.*, 2014). However, ice content might have less impact on our 3-m-deep OC stock
454 estimation, since alpine permafrost on the Tibetan Plateau is generally characterized
455 by the thick active layer (generally more than 2.4 m; Pang *et al.*, 2009) and poor ice as
456 a consequence of the arid climate, high evaporation, and glacial history (Wang *et al.*,
457 2003; Yang *et al.*, 2010). It has been reported that ice segregation, ground heave and
458 subsidence, and related periglacial landforms on the plateau are rare compared with
459 high-latitude permafrost regions (Yang *et al.*, 2010). To further evaluate the potential
460 effects of ice content on OC estimation across the study area, a correction by
461 subtracting mean gravimetric ice content (mean = 12.19%, based on measured ice
462 content derived from 697 boreholes; Zhao *et al.*, 2010) from the bulk density was
463 conducted in permafrost layers. Our additional analyses indicated, deep permafrost
464 carbon stock reduced 19.9%, but the total 3-m-deep OC stock only reduced 2.0% after
465 the correction (Table S3), largely due to the thick active layer (generally more than 2.4
466 m) on the Tibetan Plateau and small contribution of OCD in permafrost layer to total
467 OCD at depth of the three metres (Fig. 4b). In addition, given that the arid climate on
468 the main part of the Tibetan Plateau would suppress the development of underground
469 ice (Zhou & Guo, 1982), ice wedges would be only confined in very limited areas if
470 exist. All of these evidences suggest that the effects of segregated ice, wedge ice, and
471 other ice types on the three-metre OC stock could be relatively small in our case.
472
473 Bootstrapping method has often been used to evaluate the uncertainties in regional
474 carbon budget (*e.g.*, Strauss *et al.*, 2013; Hugelius *et al.*, 2014). To compare with

475 uncertainties analysis approach used in this study, we also evaluated the median OC
476 stock using bootstrapping methods. Our results revealed that the newly estimated OC
477 stock was significantly inflated compared with that derived from SVM methods
478 (Table S4). Such a difference is induced by these two different algorithms. Given the
479 inherent heterogeneity of the statistical population of interest, the estimated stock size
480 and its uncertainty using bootstrapping methods are highly dependent on sample size
481 (Fig. S6). In contrast, the SVM combined with Monte Carlo simulations is able to
482 incorporate independent input variables (*i.e.*, MAT, MAP, sand content and EVI) over
483 the entire target area, leading to reduced uncertainty in regions with less sampling
484 sites.
485
486 **Non-negligible carbon stock across alpine permafrost**
487 The OC stock in the top 1 metre across Tibetan alpine grasslands was estimated to be
488 8.51 Pg C, larger than the previous estimate of 7.36 Pg C (Yang *et al*., 2008). Such a
489 difference could result from the following two considerations. First, different
490 definitions of the swamp meadow are probably the major reason for this difference.
491 Specifically, this study evaluated OC stock in the swamp meadow independently,
492 while Yang *et al*. (2008) treated it as a part of the alpine meadow. The swamp meadow
493 belongs to azonal vegetation which is mainly located in low-lying and poorly drained
494 regions (Chinese Academy of Sciences, 2001). Despite occupying a small
495 proportional area, soils in the swamp meadow are characterized by the highest OC
496 density among the three major grassland types on the plateau (Fig. 6a, b), since
497 anaerobic conditions slow down microbial decomposition rates. Consequently, unless
498 treated separately from alpine meadow, OC stock in the swamp meadow would be
499 underestimated, which could be responsible for the observed difference in regional

500 OC budget between Yang *et al.* (2008) and this study. Besides, the upscaling approach
501 used in this study (*i.e.*, SVM) would be expected to provide better prediction than the
502 linear model used in Yang *et al.* (2008) (Table S2).
503
504 The total OC stock to the depth of 300 cm across alpine grasslands on the Tibetan
505 Plateau was estimated to be 15.31 Pg C, less than half of an earlier estimate (33.0 Pg
506 C) by Mu *et al*. (2015). Such a difference could be due to several improvements that
507 we have made in this study, including broad geographic coverage of deep soil cores,
508 and a more reliable upscaling methodology. Specifically, the larger number and more
509 extensive distribution of deep sediment cores used in this study (Fig. 1b) could be
510 more representative, and thus reduce the uncertainties induced by large spatial
511 heterogeneity. In contrast, the deep OC stock estimated by Mu *et al.* (2015) was
512 mainly derived from the extrapolation of actual measurements from only 11 deep
513 sediment cores in the north-eastern Tibetan Plateau, and 7 out of the 11 cores belong
514 to the alpine meadow. On the parameter level, their average OC content is about 10
515 fold greater than ours at depths of 100-300 cm (Table S5). This difference in OC
516 content overwhelms the larger bulk density at depths of 100-200 cm and 200-300 cm
517 across our measurements (Table S5). The confined distribution of the 11 deep cores is
518 thus likely to be the major reason for such a significant difference. In addition,
519 different algorithms could also introduce differences between the two studies. The
520 previous estimate by Mu *et al.* (2015) used a simple averaging approach to obtain
521 OCD for each vegetation type, while this study adopted a machine learning approach
522 to upscale site-level measurements to the regional scale. To further test whether
523 different algorithms would lead to significant difference in OC budget, we
524 re-evaluated OC stock in the top three metres by simple averaging of OCD from

various sampling sites within three major grassland types and multiplying by their corresponding area. Our analyses revealed that OC stock based on simple averaging approach was substantially larger than that obtained from the machine learning approach (20.17 *vs.* 15.31 Pg C), demonstrating that the difference introduced by different algorithms is non-negligible.

The average OCD in Tibetan alpine permafrost at both the 0-1 m and 0-3 m depths are considerably lower than that across the northern circumpolar permafrost region, with 7.44 *vs.* 26.52 kg C m^{-2} and 13.39 *vs.* 58.15 kg C m^{-2}, respectively (Hugelius *et al.*, 2014). These differences may be associated with different carbon inputs from aboveground biomass and different carbon outputs from microbial decomposition between the two permafrost regions. It has been reported that aboveground biomass in circumpolar arctic tundra (315.75 g m^{-2}; Epstein *et al.*, 2012) is significantly larger than that in Tibetan alpine grasslands (68.80 g m^{-2}; Yang *et al.*, 2009), which is likely to the major reason for such a significant difference of OCD. Moreover, different microbial decomposition due to different temperature, soil drainage and levels of oxygenation, as well as active layer thickness is supposed to be another important reason responsible for the difference of OCD between the two permafrost regions. It has been reported that circumpolar regions (-6.94 ^{o}C, data obtained from GHCN_CAMS Gridded 2 m Temperature generated by Fan & van den Dool, 2008) have lower MAT compared to the Tibetan Plateau (1.77 ^{o}C, data obtained from the spatial interpolation using meteorological measurements across the plateau). It has also been documented that wetlands, peatlands and lakes are widely distributed in high-latitude regions where poor drainage and anaerobic conditions frequently persist (Walker *et al.*, 2005), while soils across most of the Tibetan Plateau have reasonably

550 good drainage and aeration (Gao *et al.*, 1985). Further, it has been reported that the
551 active layer in high-latitude regions (mean = 0.71 m reported by the Circumpolar
552 Active Layer monitoring program; Brown *et al.*, 2000) is much thinner than that on
553 the Tibetan Plateau (mean = 2.41 m reported by Pang *et al.*, 2009). Consequently,
554 lower temperature, poor drainage and anaerobic conditions, long lasting cryoturbation
555 and repeated sediment deposition could inhibit microbial activities and lead to more
556 OC accumulation across the northern circumpolar permafrost region (Ping *et al.*, 2008;
557 Natali *et al.*, 2015).
558
559 Despite the lower OCD on the Tibetan Plateau than that in the northern circumpolar
560 permafrost region, this region is of critical importance for the regional carbon cycle,
561 given the high proportion of OC stored in deep soil layers (Fig. 6c) and significant
562 warming-induced permafrost changes (Wu & Zhang, 2008). Deep OC stock (>1 m)
563 accounts for 44% of the total three-metre OC stock on the Tibetan Plateau, close to
564 that reported across the northern circumpolar permafrost region (54%, Hugelius *et al.*,
565 2014; Fig. 6c). The large proportion of deep OC stock highlights that this fraction
566 should not be ignored in regional carbon budgets. Without considering the deep OC
567 component, our knowledge of the size of permafrost OC stock, and its feedback to
568 climate warming, may be incomplete. Evidence from field observations (Schuur *et al.*,
569 2009; Nowinski *et al.*, 2010; Strauss *et al.*, 2015), laboratory incubation (Schaedel *et*
570 *al.*, 2014), and regional modelling (Koven *et al.*, 2013; Koven *et al.*, 2015b;
571 Schneider von Deimling *et al.*, 2015) has consistently demonstrated that deep OC
572 stock is susceptible to climate warming. Given the high proportion and vulnerability
573 of OC in deep soil and sediment layers, deep OC dynamics could potentially influence
574 ongoing climate change (Koven *et al.*, 2015b; Schuur *et al.*, 2015). The conversion of

reactivated frozen carbon into carbon dioxide (or methane) and release to the atmosphere could trigger positive carbon-climate feedback.

Implications for understanding permafrost carbon-climate feedback in Tibetan alpine regions

Our findings have three important implications for understanding permafrost carbon-climate feedback in Tibetan alpine regions. First, this study generates a new valuable database, which contributes 342 three-metre-deep sediment cores across alpine permafrost on the Tibetan Plateau. The combination of our database with NCSCD could be expected to provide a more comprehensive assessment of permafrost carbon stock at the global scale. This database can also be used for benchmarking and parameterization of Earth system models to produce credible projections on the fate of permafrost carbon under warming environment (Walker *et al.*, 2014; Schuur *et al.*, 2015). Second, this study develops a comprehensive methodology which could be used in future for spatially explicit estimation and quantitative uncertainty analysis. Our results demonstrate that the combination of systematic measurements derived from deep sediment cores with machine learning techniques such as a SVM model, is an effective method to evaluate permafrost carbon stock, and could be applied to other regions around the world. Third, this study demonstrates that Tibetan permafrost stores a large amount of carbon, which is usually omitted from global permafrost carbon budgets. If 10% of this carbon pool in the top 3 metres is decomposed by soil microbes, 1.53 (1.30-1.78) Pg C would be released into the atmosphere. This process may counteract the enhanced vegetation carbon sink and consequently trigger positive feedback to climate warming. To gain further insights on this feedback loop, future studies should focus on deep carbon

dynamics across alpine permafrost on the Tibetan Plateau by conducting joint studies including manipulative experiments, laboratory incubation, and model predictions.

Acknowledgements

We are grateful for Dr. Jens Strauss and the other two anonymous reviewers for their insightful comments on an earlier version of this MS, and appreciate members of the IBCAS Sampling Campaign Teams for their assistance in field investigation. This work was supported by the National Basic Research Program of China on Global Change (2014CB954001 and 2015CB954201), National Natural Science Foundation of China (31322011 and 41371213), and the Thousand Young Talents Program.

References

Brown J, Hinkel KM, Nelson FE (2000) The circumpolar active layer monitoring (CALM) program: Research designs and initial results. *Polar Geography,* **24**, 166-258.

Brown J, Jr., Ferrians OJ, Heginbottom JA, Melnikov ES (1998) Circum-arctic map of permafrost and ground ice conditions. Boulder, CO: National Snow and Ice Data Center. Digital media.

Burges CC (1998) A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery,* **2**, 121-167.

Carvalhais N, Forkel M, Khomik M *et al.* (2014) Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature,* **514**, 213-217.

Chew G, Walczyk T (2012) A Monte Carlo approach for estimating measurement uncertainty using standard spreadsheet software. *Analytical and Bioanalytical Chemistry,* **402**, 2463-2469.

625 Chinese Academy of Sciences (2001) *Vegetation Atlas of China,* Beijing, Science
626 Press.
627 Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support
628 vector machines. *Journal of Applied Ecology,* **43**, 424-432.
629 Epstein HE, Raynolds MK, Walker DA, Bhatt US, Tucker CJ, Pinzon JE (2012)
630 Dynamics of aboveground phytomass of the circumpolar Arctic tundra during
631 the past three decades. *Environmental Research Letters,* **7**.
632 doi:10.1088/1748-9326/7/1/015506.
633 Fan Y, Van Den Dool H (2008) A global monthly land surface air temperature analysis
634 for 1948-present. *Journal of Geophysical Research*: *Atmospheres,* **113**,
635 D01103, doi:10.1029/2007JD008470.
636 Gao Y, Chen H, Wu Z, Sun H, Li M (1985) *Soils of Xizang (Tibet)*, Beijing, Science
637 Press.
638 Goward S, Tucker C, Dye D (1985) North American vegetation patterns observed
639 with the NOAA-7 advanced very high resolution radiometer. *Vegetatio*, **64**,
640 3-14.
641 Grosse G, Harden J, Turetsky M *et al.* (2011) Vulnerability of high-latitude soil
642 organic carbon in North America to disturbance. *Journal of Geophysical*
643 *Research: Biogeosciences*, **116**, G00K06. doi: 10.1029/2010JG001507.
644 Harden JW, Koven CD, Ping C-L *et al.* (2012) Field information links permafrost
645 carbon to physical vulnerabilities of thawing. *Geophysical Research Letters*,
646 **39**, L15704. doi: 10.1029/2012GL051958.
647 He H, Liu M, Xiao X *et al.* (2014) Large-scale estimation and uncertainty analysis of
648 gross primary production in Tibetan alpine grasslands. *Journal of Geophysical*
649 *Research: Biogeosciences*, **119**, 2013JG002449. doi: 10.1002/2013JG002449.

650 Holben BN (1986) Characteristics of maximum-value composite images from
651 temporal AVHRR data. *International Journal of Remote Sensing,* **7**,
652 1417-1434.
653 Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of
654 the radiometric and biophysical performance of the MODIS vegetation indices.
655 *Remote Sensing of Environment,* **83**, 195-213.
656 Hugelius G, Bockheim JG, Camill P *et al.* (2013) A new data set for estimating
657 organic carbon storage to 3 m depth in soils of the northern circumpolar
658 permafrost region. *Earth System Science Data,* **5**, 393-402.
659 Hugelius G, Strauss J, Zubrzycki S *et al.* (2014) Estimated stocks of circumpolar
660 permafrost carbon with quantified uncertainty ranges and identified data gaps.
661 *Biogeosciences,* **11**, 6573-6593.
662 Jobbagy EG, Jackson RB (2000) The vertical distribution of soil organic carbon and
663 its relation to climate and vegetation. *Ecological Applications,* **10**, 423-436.
664 Kavzoglu T, Colkesen I (2009) A kernel functions analysis for support vector
665 machines for land cover classification. *International Journal of Applied Earth*
666 *Observation and Geoinformation,* **11**, 352-359.
667 Koven CD, Riley WJ, Stern A (2013) Analysis of Permafrost Thermal Dynamics and
668 Response to Climate Change in the CMIP5 Earth System Models. *Journal of*
669 *Climate,* **26**, 1877-1900.
670 Koven CD, Schuur EaG, Schädel C *et al.* (2015a) A simplified, data-constrained
671 approach to estimate the permafrost carbon–climate feedback. *Philosophical*
672 *Transactions of the Royal Society of London A: Mathematical, Physical and*
673 *Engineering Sciences,* **373**.
674 Koven CD, Lawrence DM, Riley WJ (2015b) Permafrost carbon-climate feedback is

675 sensitive to deep soil carbon decomposability but not deep soil nitrogen
676 dynamics. *Proceedings of the National Academy of Sciences of the United*
677 *States of America*, **112**, 3752-3757.
678 Li Q-Q, Yue T-X, Wang C-Q *et al.* (2013) Spatially distributed modeling of soil
679 organic matter across China: An application of artificial neural network
680 approach. *Catena*, **104**, 210-218.
681 Li R, Zhao L, Ding Y *et al.* (2012) Temporal and spatial variations of the active layer
682 along the Qinghai-Tibet Highway in a permafrost region. *Chinese Science*
683 *Bulletin*, **57**, 4609-4616.
684 Lin JC, Pejam MR, Chan E, Wofsy SC, Gottlieb EW, Margolis HA, Mccaughey JH
685 (2011) Attributing uncertainties in simulated biospheric carbon fluxes to
686 different error sources. *Global Biogeochemical Cycles*, **25**, GB2018. doi:
687 10.1029/2010GB003884.
688 Lindeman RH, Merenda PF, Gold RZ (1980) *Introduction to Bivariate and*
689 *Multivariate*, Scott Foresman & Co.
690 Mishra U, Jastrow JD, Matamala R *et al.* (2013) Empirical estimates to reduce
691 modeling uncertainties of soil organic carbon in permafrost regions: a review
692 of recent progress and remaining challenges. *Environmental Research Letters*,
693 **8**, 035020.
694 Mu C, Zhang T, Wu Q *et al.* (2015) Editorial: Organic carbon pools in permafrost
695 regions on the Qinghai-Xizang (Tibetan) Plateau. *Cryosphere*, **9**, 479-486.
696 Natali SM, Schuur EaG, Mauritz M *et al.* (2015) Permafrost thaw and soil moisture
697 driving CO_2 and CH_4 release from upland tundra. *Journal of Geophysical*
698 *Research: Biogeosciences*, **120**, 2014JG002872. doi: 10.1002/2014JG002872.
699 Nello C, John S-T (2000) *An Introduction to Support Vector Machines and Other*

700 *Kernel-based Learning Methods*, Cambridge University Press.
701 Nowinski NS, Taneva L, Trumbore SE, Welker JM (2010) Decomposition of old
702 organic matter as a result of deeper active layers in a snow depth manipulation
703 experiment. *Oecologia*, **163**, 785-792.
704 Pang Q, Cheng G, Li S, Zhang W (2009) Active layer thickness calculation over the
705 Qinghai–Tibet Plateau. *Cold Regions Science and Technology*, **57**, 23-28.
706 Ping C-L, Michaelson GJ, Jorgenson MT, Kimble JM, Epstein H, Romanovsky VE,
707 Walker DA (2008) High stocks of soil organic carbon in the North American
708 Arctic region. *Nature Geoscience*, **1**, 615-619.
709 Rubinstein RY, Kroese DP (2007) *Simulation and the Monte Carlo Method, 2nd*
710 *Edition*.
711 Schaedel C, Schuur EaG, Bracho R *et al.* (2014) Circumpolar assessment of
712 permafrost C quality and its vulnerability over time using long-term
713 incubation data. *Global Change Biology*, **20**, 641-652.
714 Schirrmeister L, Siegert C, Kuznetsova T *et al.* (2002) Paleoenvironmental and
715 paleoclimatic records from permafrost deposits in the Arctic region of
716 Northern Siberia. *Quaternary International*, **89**, 97-118.
717 Schneider Von Deimling T, Grosse G, Strauss J *et al.* (2015) Observation-based
718 modelling of permafrost carbon fluxes with accounting for deep carbon
719 deposits and thermokarst activity. *Biogeosciences*, **12**, 3469-3488.
720 Schuur EaG, Mcguire AD, Schadel C *et al.* (2015) Climate change and the permafrost
721 carbon feedback. *Nature*, **520**, 171-179.
722 Schuur EaG, Vogel JG, Crummer KG, Lee H, Sickman JO, Osterkamp TE (2009) The
723 effect of permafrost thaw on old carbon release and net carbon exchange from
724 tundra. *Nature*, **459**, 556-559.

725 Strauss J, Schirrmeister L, Grosse G, Wetterich S, Ulrich M, Herzschuh U, Hubberten
726 H-W (2013) The deep permafrost carbon pool of the Yedoma region in Siberia
727 and Alaska. *Geophysical Research Letters,* **40**, 2013GL058088. doi:
728 10.1002/2013GL058088.
729 Strauss J, Schirrmeister L, Mangelsdorf K, Eichhorn L, Wetterich S, Herzschuh U
730 (2015) Organic-matter quality of deep permafrost carbon - a study from Arctic
731 Siberia. *Biogeosciences,* **12**, 2227-2245.
732 Tarnocai C, Canadell JG, Schuur EaG, Kuhry P, Mazhitova G, Zimov S (2009) Soil
733 organic carbon pools in the northern circumpolar permafrost region. *Global*
734 *Biogeochemical Cycles,* **23**. doi: 10.1029/2008gb003327.
735 Tor-Gunnar V, Leigh AW (2013) Mapping of soil organic carbon stocks for spatially
736 explicit assessments of climate change mitigation potential. *Environmental*
737 *Research Letters,* **8**, 015011.
738 Ueyama M, Ichii K, Iwata H *et al.* (2013) Upscaling terrestrial carbon dioxide fluxes
739 in Alaska with satellite remote sensing and support vector regression. *Journal*
740 *of Geophysical Research-Biogeosciences,* **118**, 1266-1281. doi:
741 10.1002/jgrg.20095.
742 Vermote EF, Kotchenova S (2008) Atmospheric correction for the monitoring of land
743 surfaces. *Journal of Geophysical Research: Atmospheres,* **113**, D23S90. doi:
744 10.1029/2007JD009662.
745 Walker AP, Hanson PJ, De Kauwe MG *et al.* (2014) Comprehensive ecosystem
746 model-data synthesis using multiple data sets at two temperate forest free-air
747 CO_2 enrichment experiments: Model performance at ambient CO_2
748 concentration. *Journal of Geophysical Research: Biogeosciences,* **119**,
749 2013JG002553. doi: 10.1002/2013JG002553.

750 Walker DA, Raynolds MK, Fred JaD *et al.* (2005) The Circumpolar Arctic Vegetation
751 Map. *Journal of Vegetation Science,* **16**, 267-282.
752 Wang B, Bao Q, Hoskins B, Wu G, Liu Y (2008) Tibetan Plateau warming and
753 precipitation changes in East Asia. *Geophysical Research Letters*, **35**, L14702.
754 doi: 10.1029/2008GL034330.
755 Wang B, French HM (1995) Permafrost on the Tibet Plateau, China. *Quaternary*
756 *Science Reviews,* **14**, 255-274.
757 Wang S (1997) On the classification of permafrost on Qinghai-Xizang Plateau. *Arid*
758 *Land Geography,* **20**.
759 Wang S, Tian H, Liu J, Pan S (2003) Pattern and change of soil organic carbon storage
760 in China: 1960s-1980s. *Tellus B,* **55**, 416-427.
761 Were K, Bui DT, Dick B, Singh BR (2015) A comparative assessment of support
762 vector regression, artificial neural networks, and random forests for predicting
763 and mapping soil organic carbon stocks across an Afromontane landscape.
764 *Ecological Indicators*, **52**, 394-403.
765 Whiteman G, Hope C, Wadhams P (2013) Vast costs of Arctic change. *Nature*, **499**,
766 401-403.
767 Wu H, Guo Z, Peng C (2003) Distribution and storage of soil organic carbon in China.
768 *Global Biogeochemical Cycles*, **17**, 1048. doi: 10.1029/2001GB001844.
769 Wu Q, Hou Y, Yun H, Liu Y (2015) Changes in active-layer thickness and
770 near-surface permafrost between 2002 and 2012 in alpine ecosystems,
771 Qinghai-Xizang (Tibet) Plateau, China. *Global and Planetary Change*, **124**,
772 149-155.
773 Wu Q, Zhang T (2008) Recent permafrost warming on the Qinghai-Tibetan Plateau.
774 *Journal of Geophysical Research: Atmospheres*, **113**, D13108. doi:

775 10.1029/2007JD009539.
776 Wu Q, Zhang T (2010) Changes in active layer thickness over the Qinghai-Tibetan
777 Plateau from 1995 to 2007. *Journal of Geophysical Research: Atmospheres*,
778 **115**, D09107. doi: 10.1029/2009JD012974.
779 Yang M, Nelson FE, Shiklomanov NI, Guo D, Wan G (2010) Permafrost degradation
780 and its environmental effects on the Tibetan Plateau: A review of recent
781 research. *Earth-Science Reviews*, **103**, 31-44.
782 Yang Y, Fang J, Tang Y, Ji C, Zheng C, He J, Zhu B (2008) Storage, patterns and
783 controls of soil organic carbon in the Tibetan grasslands. *Global Change*
784 *Biology*, **14**, 1592-1599.
785 Yang YH, Fang JY, Pan YD, Ji CJ (2009) Aboveground biomass in Tibetan grasslands.
786 *Journal of Arid Environments*, **73**, 91-95.
787 Yang Y, Ji C, Chen L, Ding J, Cheng X, Robinson D (2015) Edaphic rather than
788 climatic controls over ^{13}C enrichment between soil and vegetation in alpine
789 grasslands on the Tibetan Plateau. *Functional Ecology*, **29**, 839-848.
790 Zhang T, Barry RG, Knowles K, Heginbottom JA, Brown J (1999) Statistics and
791 characteristics of permafrost and ground-ice distribution in the Northern
792 Hemisphere. *Polar Geography*, **23**, 132-154.
793 Zhao L, Ding Y, Liu G, Wang S, Jin H (2010) Estimates of reserves of ground ice in
794 permafrost regions on the Tibetan Plateau. *Journal of Glaciology and*
795 *Geocryology*, **32**, 1-9.
796 Zhou Y, Guo D (1982) Principal characteristics of permafrost in China. *Journal of*
797 *Glaciology and Cryopedology*, **4,** 1-19.
798

Supporting information captions

Table S1 Summary of estimated median volumetric organic carbon (OC) content (with interquartile range) in alpine grasslands on the Tibetan Plateau.

Table S2 Comparison of cross validation results among various upscaling methods to predict organic carbon density (OCD) at the 0-50 cm depth.

Table S3 Summary of estimated median organic carbon (OC) stock (with interquartile range) in alpine grasslands on the Tibetan Plateau after correcting ice content using field measurements from Zhao *et al*. (2010).

Table S4 Summary of estimated median organic carbon (OC) stock (with interquartile range) using 1000-iteration bootstrap resampling in alpine grasslands on the Tibetan Plateau.

Table S5 Comparison of the organic carbon (OC) stock estimated by Mu *et al*. (2015) and this study.

Fig. S1 Relationship between bulk density and organic carbon (OC) content across 51 natural profiles.

Fig. S2 Comparison of different interpolation methods based on cross validation results. *RMSE* represents root mean square error. (a, e) show results from inverse distance weighting (IDW) interpolation; (b, f) represent results from thin plate spline function interpolation; (c, g) illustrate results from Kriging

interpolation; and (d, h) display results from Cokriging interpolation with altitude as a covariate.

Fig. S3 Relative importance of predictor variables, reflected by percentage of r^2, for organic carbon density (OCD) in the top 50 cm. Relative importance metrics were normalized to sum 100% using Lindeman-Merenda-Gold (LMG) method (Lindeman *et al.*, 1980). EVI, Sand, Silt, Clay, MAP and MAT represents, enhanced vegetation index, sand content, silt content, clay content, mean annual precipitation and mean annual temperature, respectively.

Fig. S4 Relative uncertainties of permafrost carbon budget at different depths across the study area, calculated as the ratio between the interquartile range (difference between the 75^{th} and 25^{th} percentiles) of the 1000 estimates and the median.

Fig. S5 Validation of the regression model of OCD between the depths of 0-30 cm and 0-50 cm ($\text{Log OCD}_{0\text{-}50\ cm} = 0.9806\ \text{Log OCD}_{0\text{-}30\ cm} + 0.2911$) constructed in the eastern Tibetan Plateau, using actual measurements in the western part of the plateau. Note that the OCD was log-transformed to meet the normal distribution premise before model construction.

Fig. S6 Effect of sample size on the confidence interval (CI) of 3-m-deep organic carbon (OC) stock using bootstrap resampling based on pedon observations from 114 sites.

Appendix 1 Source code for the SVM and uncertainty analyses in R software.

Appendix 2 Spatial organic carbon density (OCD) at different depths at a resolution of 10 km × 10 km in alpine grasslands on the Tibetan Plateau.

Appendix 3 Pedon observations from 173 sites in alpine grasslands on the Tibetan Plateau.

856 **Table 1** Summary of estimated median OC density (with interquartile range) in alpine grasslands on the Tibetan Plateau.

Grassland type	OC density (kg C m^{-2})			
	0-50 cm	0-100 cm	0-200 cm	0-300 cm
Alpine steppe (AS)	3.19 (2.58-3.95)	4.49 (3.66-5.51)	6.55 (5.42-7.89)	8.56 (7.19-10.12)
Alpine meadow (AM)	6.67 (5.71-7.76)	8.99 (7.75-10.40)	11.95 (10.43-13.69)	14.41 (12.70-16.33)
Swamp meadow (SM)	23.05 (21.47-24.73)	30.79 (28.41-33.19)	51.34 (43.53-60.12)	65.00 (52.60-76.36)
Total	5.45 (4.66-6.38)	7.44 (6.38-8.68)	10.68 (9.10-12.51)	13.39 (11.39-15.53)

857

858 **Table 2** Summary of estimated median OC stock (with interquartile range) in alpine grasslands on the Tibetan Plateau.

Grassland type	Area (10^3 km^2)	OC stock (Pg C)			
		0-50 cm	0-100 cm	0-200 cm	0-300 cm
Alpine steppe (AS)	640.0	2.04 (1.65-2.53)	2.88 (2.34-3.53)	4.19 (3.47-5.05)	5.48 (4.60-6.48)
Alpine meadow (AM)	453.1	3.02 (2.59-3.52)	4.07 (3.51-4.71)	5.41 (4.73-6.20)	6.53 (5.75-7.40)
Swamp meadow (SM)	50.9	1.17 (1.09-1.26)	1.57 (1.45-1.69)	2.61 (2.22-3.06)	3.31 (2.68-3.89)
Total	1144.0	6.23 (5.33-7.30)	8.51 (7.30-9.93)	12.22 (10.41-14.31)	15.31 (13.03-17.77)

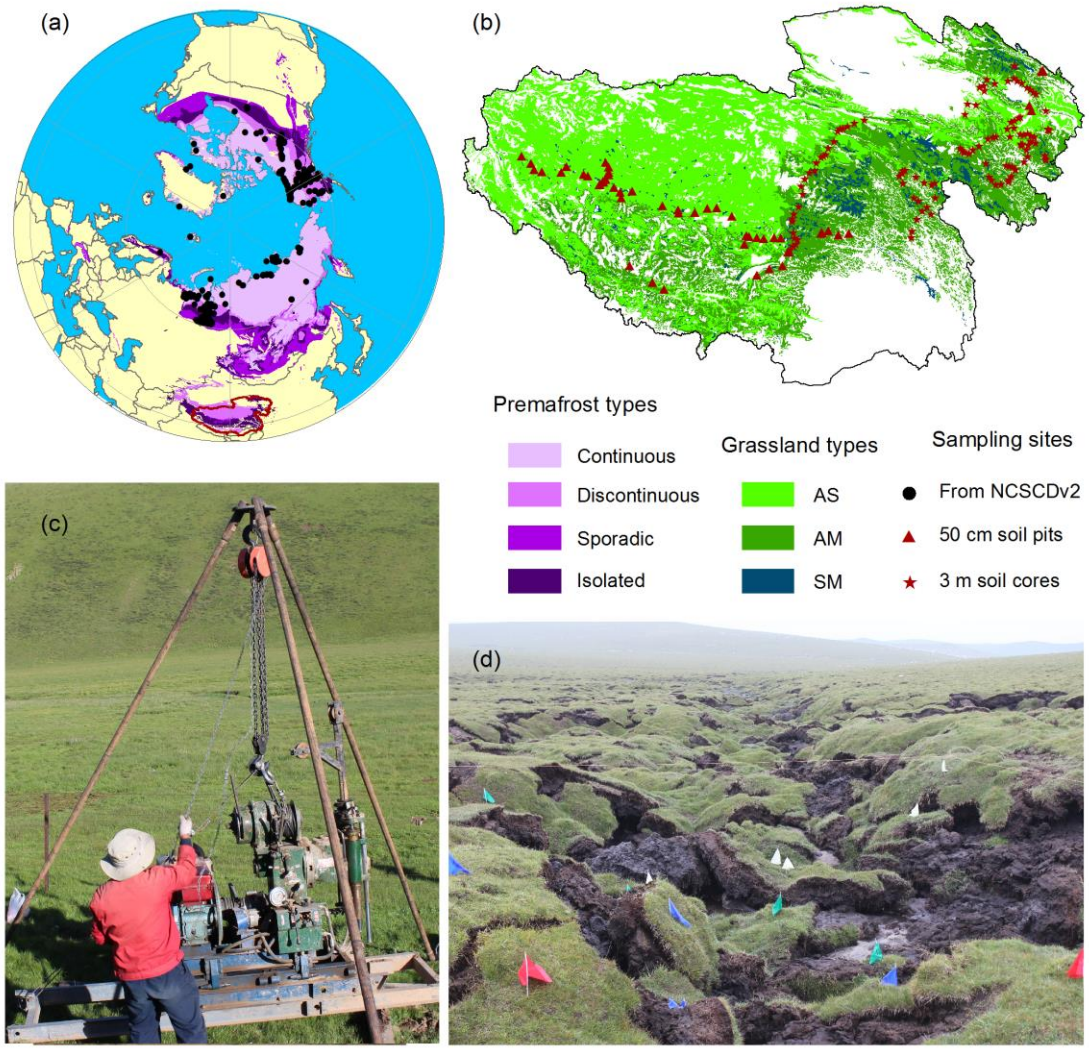
859

860 **Figure captions**

861 **Fig. 1** Existing pedon observations in the Northern Hemisphere permafrost region
862 ranging from 100 to 300 cm depth (a), sampling sites and vegetation map
863 across alpine grasslands on the Tibetan Plateau (b), the machine used to
864 sample deep sediment cores (c), and ground collapse and erosion features in
865 the swamp meadow on the Tibetan Plateau (d). The permafrost map is
866 obtained from the National Snow & Ice Data Center (Brown *et al.*, 1998),
867 point pedon data are derived from the Bolin Centre Database (Hugelius *et al.*,
868 2013), and the vegetation map is generated from China's Vegetation Atlas with
869 a scale of 1: 1 000 000 (Chinese Academy of Sciences, 2001). AS, AM and
870 SM represent alpine steppe, alpine meadow and swamp meadow, respectively.
871
872 **Fig. 2** Methodology used for quantifying permafrost OC stock and associated
873 uncertainty analysis. The dotted boxes in black, blue, and orange indicate
874 site-level observations of the model input variables, the SVM model
875 parameterization process, and the module iterated 1000 times, respectively.
876 The nonlinear models between site-level OCD among various layers were
877 used to extrapolate regional OCD in the top 50 cm to deeper layers OCD.
878
879 **Fig. 3** "Leave-one-out" cross-validation for the SVM model used for predictions of
880 OCD at the 50 cm depth (a), scatter plots and extrapolation functions of OCD
881 in the top 50 cm and OCD at depths of 0-100 cm (b), 0-200 cm (c), and 0-300
882 cm (d). Note that the OCD (b-d) was log-transformed to meet the normal
883 distribution premise of the regression models.
884

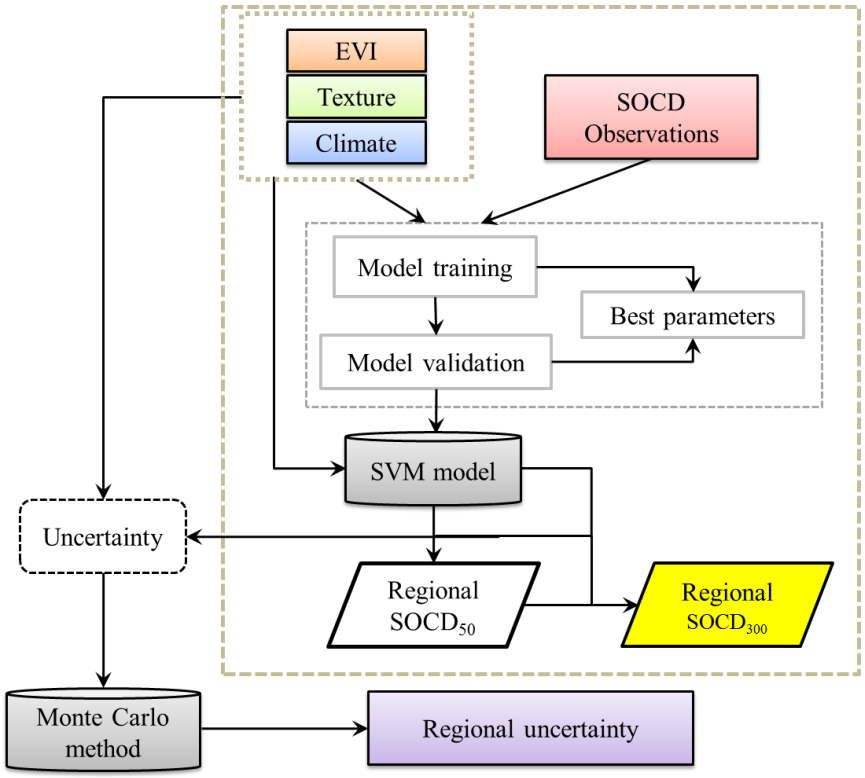
885 **Fig. 4** Vertical distributions of OCD (a) and relative proportions (b) at 50-cm intervals
886 (mean + SD) in alpine grasslands on the Tibetan Plateau. The relative
887 proportion is represented by the proportional contribution of each layer to total
888 OCD at depth of the three metres. AS, AM and SM represent alpine steppe,
889 alpine meadow and swamp meadow, respectively. Mean values with different
890 letters (*e.g.*, a, b, c) indicate significant differences among grassland types at
891 each depth interval. Mean values with the same letters (*e.g.*, a, ab) indicate no
892 significant differences among these grassland types (Kruskal-Wallis test, $P <$
893 0.05).
894
895 **Fig. 5** Estimated OCD for different depths at a resolution of 10 km × 10 km across
896 alpine grasslands on the Tibetan Plateau. (a) 0-50 cm, (b) 0-100 cm, (c) 0-200
897 cm, and (d) 0-300 cm. All related data are available in Appendix 2.
898
899 **Fig. 6** Comparisons of OCD in the 0-100 cm (a), 0-300 cm depths (b), and vertical
900 proportional distributions of OC stock (c) among different grassland types on
901 the Tibetan Plateau, and the counterparts in the northern circumpolar
902 permafrost region (CP). AS, AM and SM represent alpine steppe, alpine
903 meadow and swamp meadow, respectively. The whiskers illustrate the data
904 range, and the box ends indicate the 25^{th} and the 75^{th} quartile (interquartile
905 range). The horizontal lines inside each box show the median, and the notches
906 represent the 95% confidence intervals. The related OC data across the
907 northern circumpolar permafrost region are derived from Hugelius *et al*.
908 (2014).

909 **Fig. 1**


(a)
(b)
Premafrost types
Continuous
Discontinuous
Sporadic
Isolated
Grassland types
AS
AM
SM
Sampling sites
From NCSCDv2
50 cm soil pits
3 m soil cores
(c)
(d)

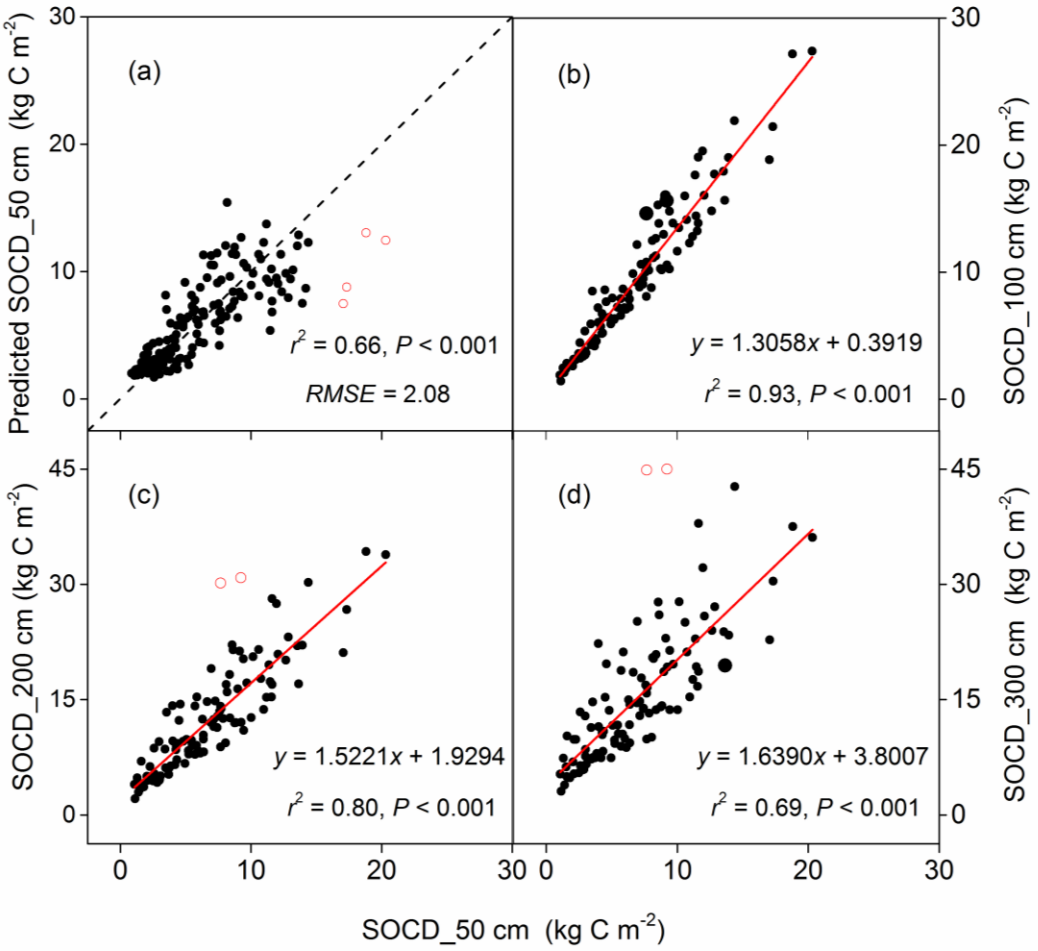
910

911 **Fig. 2**


EVI
Texture
Climate
SOCD Observations
Model training
Best parameters
Model validation
SVM model
Uncertainty
Regional SOCD50
Regional SOCD300
Monte Carlo method
Regional uncertainty

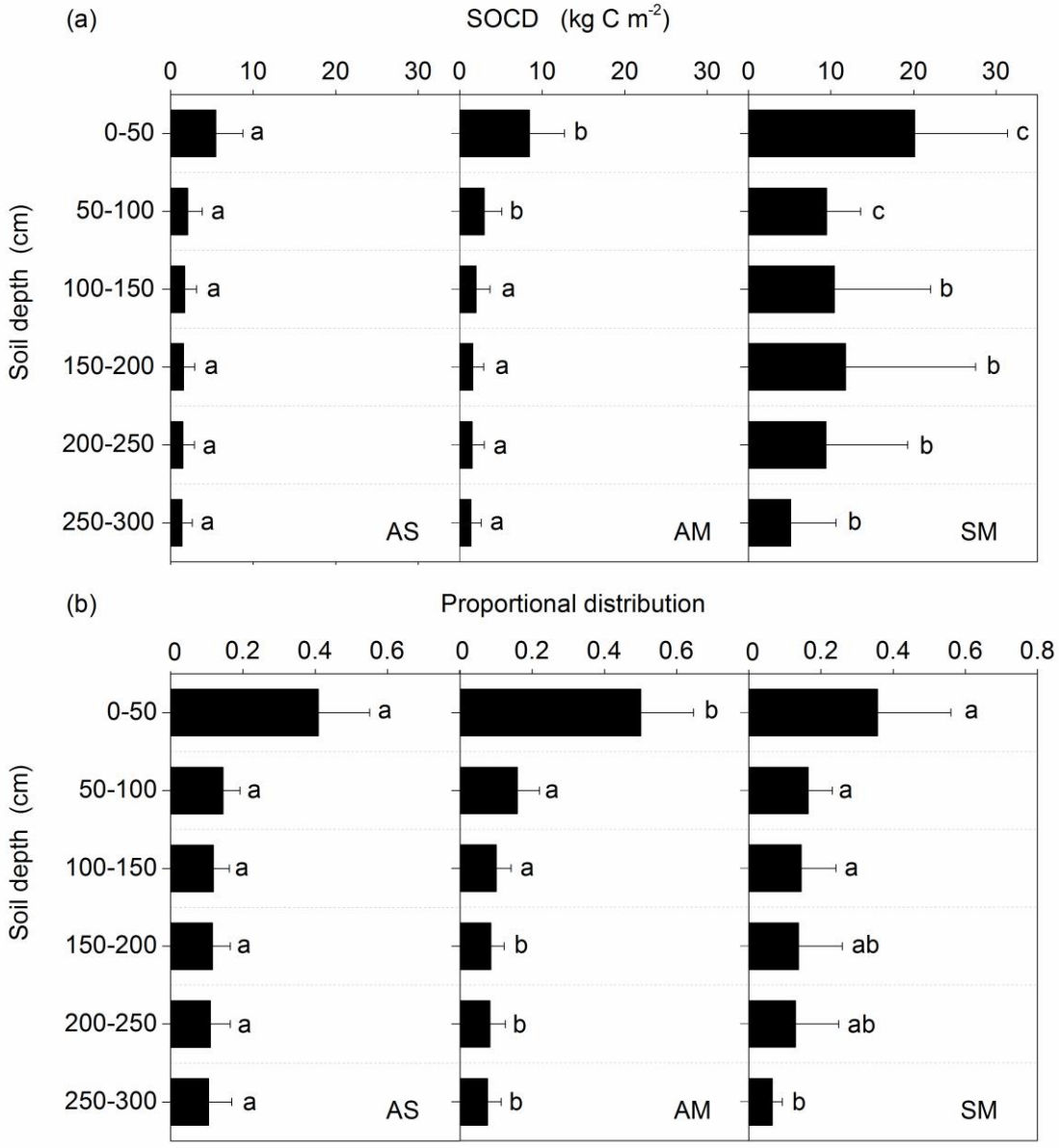
912

913 **Fig. 3**


(a)
(b)
(c)
(d)
Predicted SOCD_50 cm (kg C m-2)
SOCD_100 cm (kg C m-2)
SOCD_200 cm (kg C m-2)
SOCD_300 cm (kg C m-2)
SOCD_50 cm (kg C m-2)
r2 = 0.66, P < 0.001
RMSE = 2.08
y = 1.3058x + 0.3919
r2 = 0.93, P < 0.001
y = 1.5221x + 1.9294
r2 = 0.80, P < 0.001
y = 1.6390x + 3.8007
r2 = 0.69, P < 0.001

914

915 **Fig. 4**


(a)
SOCD (kg C m-2)
0
10
20
30
Soil depth (cm)
0-50
50-100
100-150
150-200
200-250
250-300
a
b
c
AS
AM
SM
(b)
Proportional distribution
0
0.2
0.4
0.6
0.8
ab

916

Fig. 5

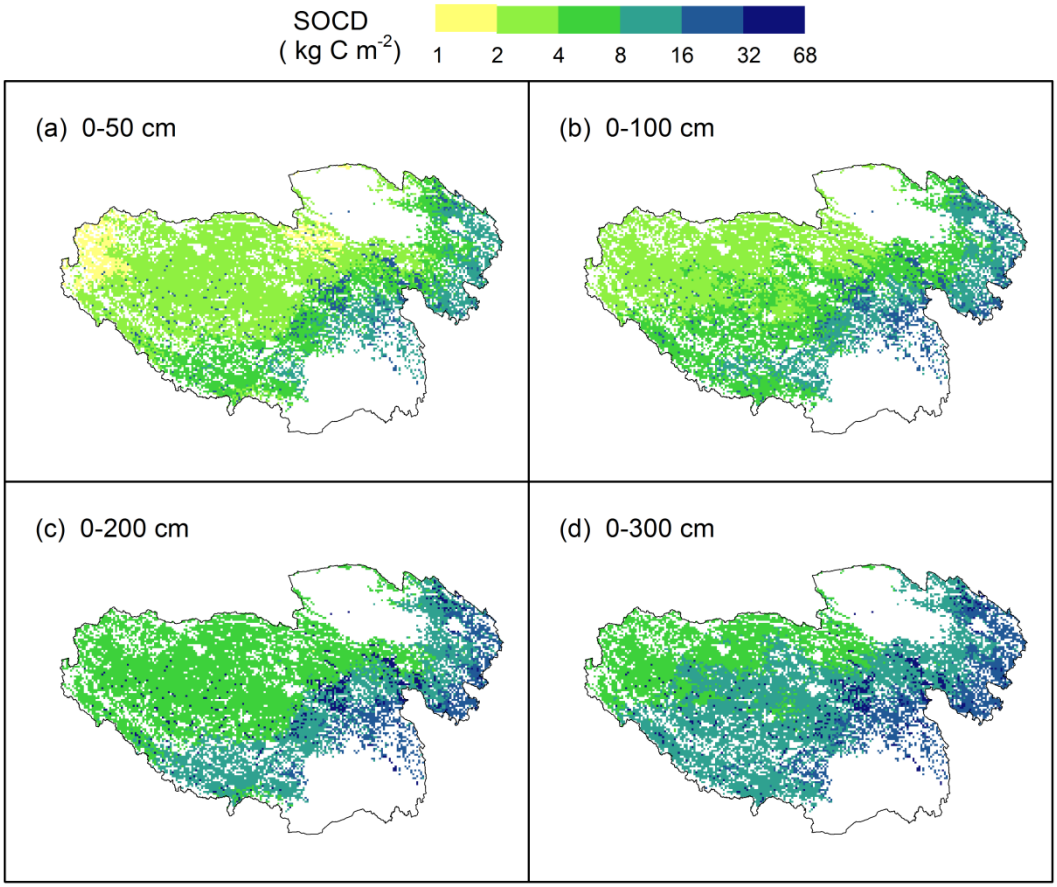
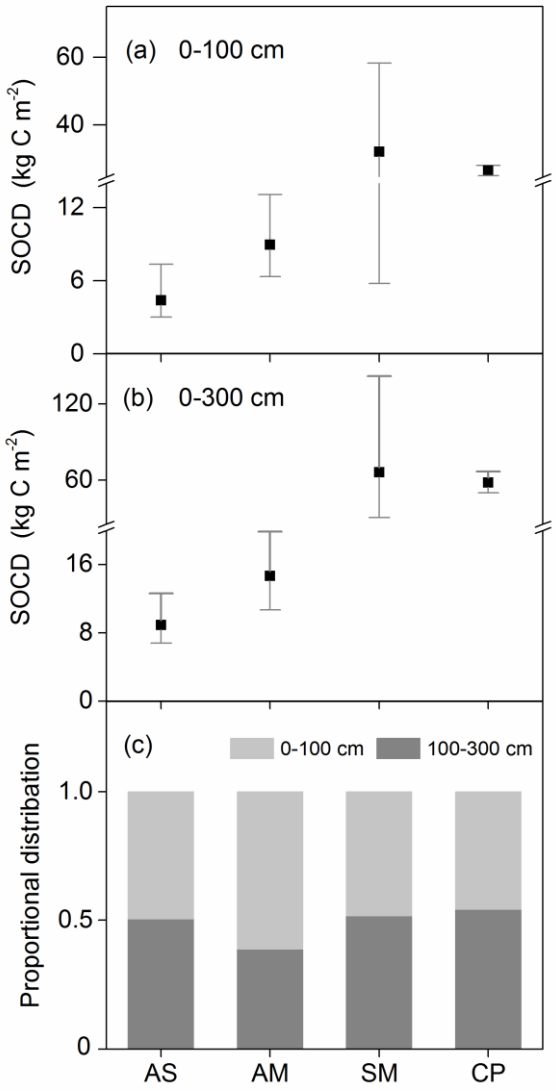


Fig. 6


(a) 0-100 cm
SOCD (kg C m-2)
60
40
12
6
0
(b) 0-300 cm
SOCD (kg C m-2)
120
60
16
8
0
(c)
0-100 cm
100-300 cm
Proportional distribation
1.0
0.5
0
AS
AM
SM
CP

Supporting Information

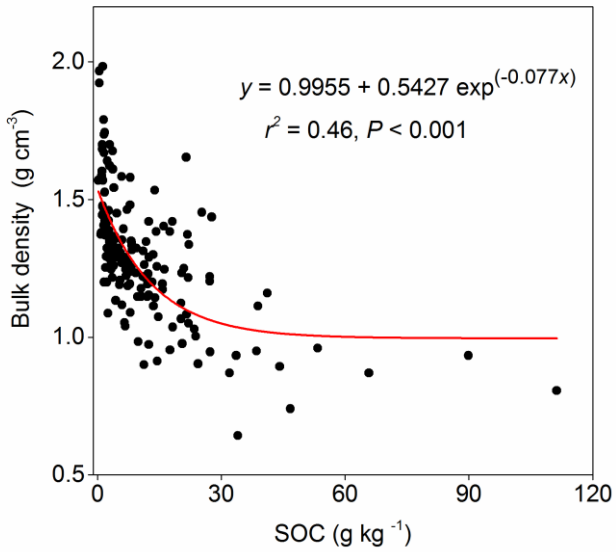
Fig. S1 Relationship between bulk sensitivity and SOC content from 51 natural profiles, which was used to estimate missing bulk density of 300-cm-depth soil cores.

Fig. S2 Relative importance of predictor variables, reflected by percentage of r^2, for SOCD in the top 50 cm. Relative importance metrics were normalized to sum 100% using Lindeman-Merenda-Gold (LMG) method (Lindeman *et al.*, 1980).

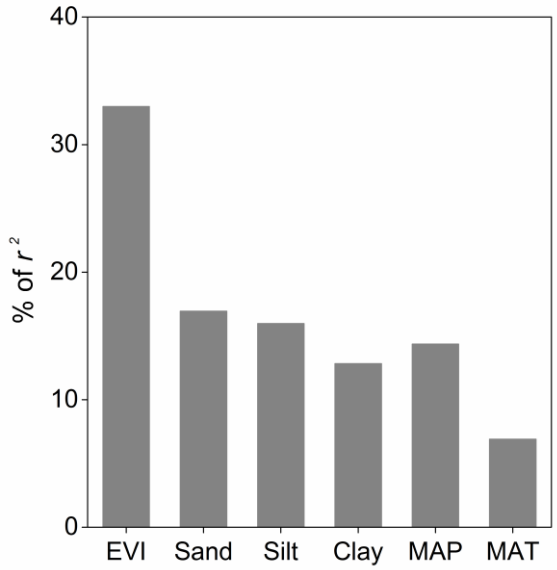
Fig. S3 Relative uncertainties of permafrost C budget at different soil depths across the study area, calculated as the ratio between the interquartile range (difference between the 75^{th} and 25^{th} percentiles) of the 1000 estimates and the mean.

Fig. S4 Spatial uncertainties of the input variables for support vector machine (SVM) modelling, including measurement error (ME) of EVI (a), as calculated as 0.02 + 0.02 × value (He *et al.*, 2014), and the standard errors (SE) generated by Kriging analysis for sand content (b), MAT (c) and MAP (d).

Fig. S1

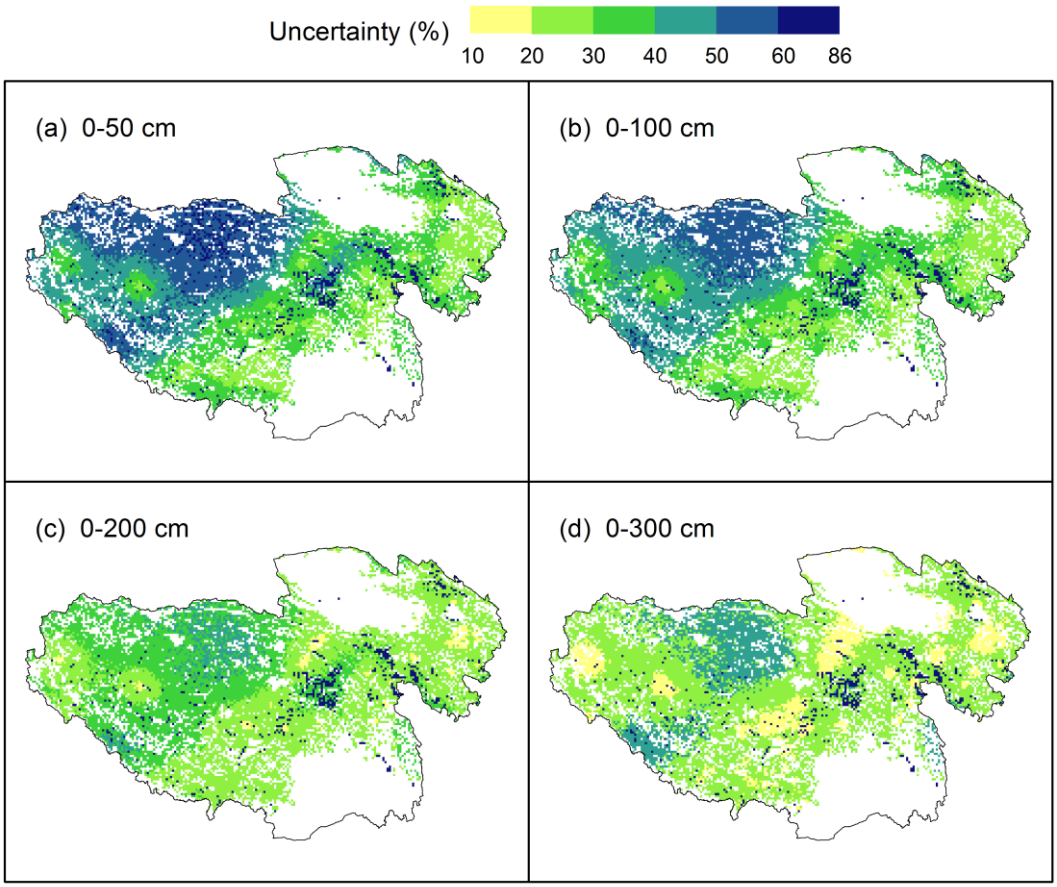

y = 0.9955 + 0.5427 exp(-0.077x)
r2 = 0.46, P < 0.001
Bulk density (g cm-3)
SOC (g kg -1)
2.0
1.5
1.0
0.5
0
30
60
90
120

944 **Fig.S2**



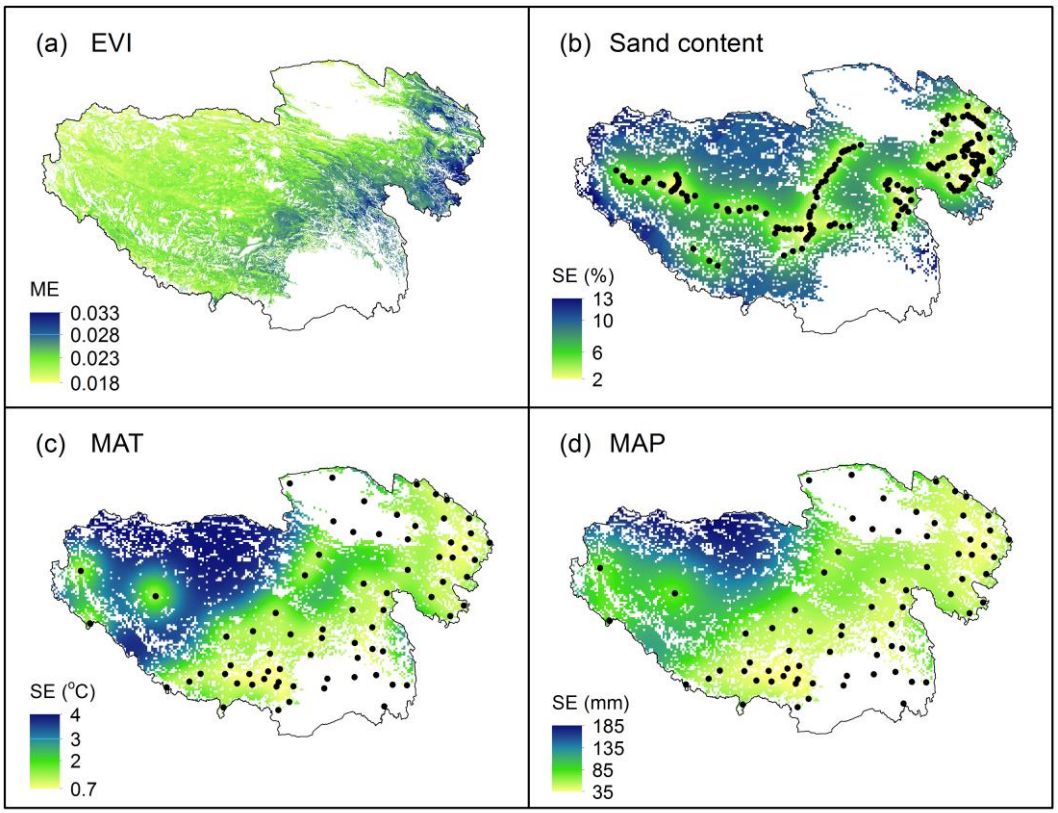
945

946 **Fig. S3**



947

948 **Fig. S4**



949
950