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

JINZHI DING1,2, FEI LI1,2, GUIBIAO YANG1,2, LEIYI CHEN1, BEIBEI ZHANG1,3, LI LIU1,2, KAI FANG1,2, SHUQI QIN1,2, YONGLIANG CHEN1, YUNFENG PENG1, CHENGJUN JI4, HONGLIN HE5, PETE SMITH6 and YUANHE YANG1

1State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China, 2University of Chinese Academy of Sciences, Beijing 100049, China, 3Inner Mongolia University of Technology, Inner Mongolia 010051, China, 4Department of Ecology, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing 100871, China, 5Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China, 6Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK

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 m 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 Monte Carlo simulations. 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 south-eastern to the north-western plateau, with the exception that OCD in the swamp meadow 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 m, 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.

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

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 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).

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

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 machine (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 nonlinearity and overfitting 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 the 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 1 m 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 3 m (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-m-deep OC stock to be 160 Pg C on the plateau, with a 3-m-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 3-m-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 SVM model. We further evaluated uncertainties in regional carbon budget by conducting Monte Carlo simulations. Specifically, our objectives were to (i) develop a suite of methodologies to quantify the 3-m-deep OC stock with a detailed uncertainty analysis and (ii) estimate 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 \( \sim 1.35 \times 10^6 \) 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 m along the Qinghai–Tibetan Highway (Pang et al., 2009). Similar to other permafrost regions around the world, Tibetan permafrost has experienced significant thawing, such as increased active layer thickness (Wu & Zhang, 2010) and thermal erosion (Fig. 1d). The climate is characterized as cold and dry across the main body of the plateau. Mean annual temperature (MAT) ranges between −4.9 and 6.1 °C, and mean annual precipitation (MAP) ranges from 84.3 to 593.9 mm, about 90% of which falls within the growing season from May to September. The permafrost regions are mainly covered by three grassland types, including the alpine steppe, alpine meadow and swamp meadow. The dominant species are *Stipa purpurea* and *Carex moorcroftii* in the alpine steppe, *Kobresia pygmaea* and *K. humilis* in the alpine meadow and *K. tibetica* in the swamp meadow, respectively (Yang et al., 2015).

**Site-level measurements**

During the summers (July and August) of 2013 and 2014, we collected 519 pedons from 173 sampling sites throughout the geographical extent of alpine grasslands on the Tibetan Plateau (Fig. 1b). The sampling sites covered broad climatic gradients and major grassland types across the study area. Of all sampling sites, 91 sites were from the alpine steppe, 75 sites were from the alpine meadow, and 7 sites were from the swamp meadow. To better characterize the spatial heterogeneity of OCD in swamp the meadow, we supplemented unpublished data from extra 36 profiles at 12 sampling sites on the plateau (three replications within each site, collected by Prof. Jingyun Fang’s group). At each site, we set up five 1 m × 1 m quadrats located at each corner and the centre of a 10 m × 10 m plot. For each quadrat, aboveground biomass was clipped at the ground level and pooled. Pedon samples were collected within three quadrats along a 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 north-western plateau. Bulk density samples were obtained for each pit using a standard container with a fixed volume size of 100 cm³, 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 cm³ 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, Malvern, Worcestershire, UK) after removal of organic matter and 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, Giesbeek, 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:

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

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 density 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 (RMSEs). 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). In this study, 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 compositon (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 nonlinear 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 3.2.1, R Development Core Team, 2015), and the source code was provided in Appendix S1.

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 ≥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 was 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.

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

where RMSE represents the root-mean-square error, \(n\) is sample size, and \(y_i\) 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, 2008). 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 \([0.02 + 0.02 \times \text{value}]\), 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 \(0.02 + 0.02 \times \text{value}\) 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 interquartile (difference between the 75th and 25th 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 interquartiles by median values for each pixel (Fig. S4). It should be noted that the uncertainty of the OC stock in 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). Interquartiles were also calculated to assess the relative uncertainties of the OCD in the swamp meadow.

Results

Vertical distributions of OC stock down the 3-m profiles

The OCD decreased as soil depth increased in all of the 3-m 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 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 50 cm 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 were 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 Tibetan permafrost regions

The OCD exhibited large spatial variability across permafrost regions on the Tibetan Plateau, with a decreasing trend from the south-east to the north-west (Fig. 5). Such a pattern was contingent on geographical extents of grassland types (Fig. 1b), that is 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 was 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

---

**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) |

---

**Fig. 5** Estimated OCD for different depths at a resolution of 10 km \(\times\) 10 km across alpine grasslands on the Tibetan Plateau. (a) 0–50 cm, (b) 0–100 cm, (c) 0–200 cm and (d) 0–300 cm. All related data are available in Appendix S2.
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 S3). Using a SVM model, we upscaled 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 nonparametric 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 may be partly derived from the extrapolation of surface to deep OC stock using nonlinear models, since the relationships between surface and deep OCD could be compromised by processes in periglacial environments (e.g. cryoturbation, slope mass movement of material) (Grosse et al., 2011; Harden et al., 2012). Nevertheless, our nonlinear models successfully quantified the relationships of OC stock among various layers for both the alpine steppe and alpine meadow (Fig. 3b–d), possibly due to the unique characteristics of alpine permafrost on the Tibetan Plateau. For instance, the overall arid climate on the plateau tends to suppress periglacial processes (Wang, 1997), which could then result in limited distributions of periglacial landforms (Yang et al., 2010). Consequently, vertical extrapolation may not induce large uncertainties into regional carbon budget. Even so, some uncertainties could still be generated due to the lack of deep soil cores across the western plateau for practical reasons (remoteness, road conditions, etc.). Nevertheless, both the eastern and western Tibetan Plateau share common soil type (cambisols) (Wu et al., 2003; Yang et al., 2015) and comparable permafrost types (Brown et al., 1998). To further test whether regression models derived from the eastern half of the plateau are applicable in the western part of the plateau, we constructed the regression model of OCD between the depths of 0–30 and 0–50 cm of the eastern plateau and validated the model using actual measurements from the western part of the plateau. The validation results showed good performance of the model constructed in the eastern part of the plateau ($r^2 = 0.98$, RMSE = 0.81; Fig S5), demonstrating the reliability of the regression model in both parts of the plateau. Consequently, extrapolation of topsoil OC

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

| Grassland type   | Area ($10^3$ km$^2$) | 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)** |
The lack of ice content may also lead to potential uncertainties in our budget, since ice content is an important parameter for estimating permafrost carbon stock, especially in ice-rich permafrost regions (e.g., the Yedoma region; Strauss et al., 2013; Hugelius et al., 2014). However, ice content might have less impact on our 3-m-deep OC stock estimation, since alpine permafrost on the Tibetan Plateau is generally characterized by the thick active layer (generally more than 2.4 m; Pang et al., 2009) and poor ice as a consequence of the arid climate, high evaporation and glacial history (Wang et al., 2003; Yang et al., 2010). It has been reported that ice segregation, ground heave and subsidence, and related periglacial landforms on the plateau are rare compared with high-latitude permafrost regions (Yang et al., 2010).

To further evaluate the potential effects of ice content on OC estimation across the study area, a correction by subtracting mean gravimetric ice content (mean = 12.19%, based on measured ice content derived from 697 boreholes; Zhao et al., 2010) from the bulk density was conducted in permafrost layers. Our additional analyses indicated deep permafrost carbon stock reduced 19.9%, but the total 3-m-deep OC stock only reduced 2.0% after the correction (Table S3), largely due to the thick active layer (generally more than 2.4 m) on the Tibetan Plateau and small contribution of OCD in permafrost layer to total OCD at depth of the 3 m (Fig. 4b). In addition, given that the arid climate on the main part of the Tibetan Plateau would suppress the development of underground ice (Zhou & Guo, 1982), ice wedges would be only confined in very limited areas if exist. All of these evidences suggest that the effects of segregated ice, wedge ice and other ice types on the 3-m OC stock could be relatively small in our case.

Bootstrapping method has often been used to evaluate the uncertainties in regional carbon budget (e.g., Strauss et al., 2013; Hugelius et al., 2014). To compare with uncertainty analysis approach used in this study, we also evaluated the median OC stock using bootstrapping methods. Our results revealed that the newly estimated OC stock was significantly inflated compared with that derived from SVM methods (Table S4). Such a difference is induced by these two different algorithms. Given the inherent heterogeneity of the statistical population of interest, the estimated stock size and its uncertainty using bootstrapping methods are highly dependent on sample size (Fig. S6). In contrast, the SVM combined with Monte Carlo simulations is able to incorporate independent input variables (i.e., MAT, MAP, sand content and EVI) over the entire target area, leading to reduced uncertainty in regions with less sampling sites.
The OC stock in the top 1 m across Tibetan alpine grasslands was estimated to be 8.51 Pg C, larger than the previous estimate of 7.36 Pg C (Yang et al., 2008). Such a difference could result from the following two considerations. First, different definitions of the swamp meadow are probably the major reason for this difference. Specifically, this study evaluated OC stock in the swamp meadow independently, while Yang et al. (2008) treated it as a part of the alpine meadow. The swamp meadow belongs to azonal vegetation which is mainly located in low-lying and poorly drained regions (Chinese Academy of Sciences, 2001). Despite occupying a small proportional area, soils in the swamp meadow are characterized by the highest OC density among the three major grassland types on the plateau (Fig. 6a, b), since anaerobic conditions slow down microbial decomposition rates. Consequently, unless treated separately from the alpine meadow, OC stock in the swamp meadow would be underestimated, which could be responsible for the observed difference in regional OC budget between Yang et al. (2008) and this study. Besides, the upscaling approach used in this study (i.e. SVM) would be expected to provide better prediction than the linear model used in Yang et al. (2008) (Table S2).

The total OC stock to the depth of 300 cm across alpine grasslands on the Tibetan Plateau was estimated to be 15.31 Pg C, less than half of an earlier estimate (33.0 Pg C) by Mu et al. (2015). Such a difference could be due to several improvements that we have made in this study, including broad geographical coverage of deep soil cores, and a more reliable upscaling methodology. Specifically, the larger number and more extensive distribution of deep sediment cores used in this study (Fig. 1b) could be more representative and thus reduce the uncertainties induced by large spatial heterogeneity. In contrast, the deep OC stock estimated by Mu et al. (2015) was mainly derived from the extrapolation of actual measurements from only 11 deep sediment cores in the north-eastern Tibetan Plateau, and 7 of the 11 cores belong to the alpine meadow. On the parameter level, their average OC content is about 10-fold greater than ours at depths of 100–300 cm (Table S5). This difference in OC content overwhelms the larger bulk density at depths of 100–200 and 200–300 cm across our measurements (Table S5). The confined distribution of the 11 deep cores is thus likely to be the major reason for such a significant difference. In addition, different algorithms could also introduce differences between the two studies. The previous estimate by Mu et al. (2015) used a simple averaging approach to obtain OCD for each vegetation type, while this study adopted a machine learning approach to upscale site-level measurements to the regional scale. To further test whether different algorithms would lead to significant difference in OC budget, we re-evaluated OC stock in the top 3 m 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 across Tibetan alpine permafrost at both the 0–1 and 0–3 m depths are considerably lower than that across the northern circumpolar permafrost region, with 7.44 vs. 26.52 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 be 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 °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 °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 good drainage and aeration (Gao et al., 1985). Further, it has been reported that the active layer in high-latitude regions (mean = 0.71 m reported by the Circumpolar Active Layer Monitoring programme; Brown et al., 2000) is much thinner than that on the Tibetan Plateau (mean = 2.41 m reported by Pang et al., 2009). Consequently, lower temperature, poor drainage and anaerobic conditions, long-lasting cryoturbation and repeated sediment deposition could inhibit microbial activities and lead to more OC accu-
mulation across the northern circumpolar permafrost region (Ping et al., 2008; Natali et al., 2015).

Despite the lower OCD on the Tibetan Plateau than that in the northern circumpolar permafrost region, this region is of critical importance for the regional carbon cycle, given the high proportion of OC stored in deep soil layers (Fig. 6c) and significant warming-induced permafrost changes (Wu & Zhang, 2008). Deep OC stock (~1 m) accounts for 44% of the total 3-m OC stock on the Tibetan Plateau, close to that reported across the northern circumpolar permafrost region (54%, Hugelius et al., 2014; Fig. 6c). The large proportion of deep OC stock highlights that this fraction should not be ignored in regional carbon budgets. Without considering the deep OC component, our knowledge of the size of permafrost OC stock, and its feedback to climate warming, may be incomplete. Evidence from field observations (Schuur et al., 2009; Nowinski et al., 2010; Strauss et al., 2015), laboratory incubation (Schaedel et al., 2014) and regional modelling (Koven et al., 2013, 2015b; Schneider Von Deimling et al., 2015) has consistently demonstrated that deep OC stock is susceptible to climate warming. Given the high proportion and vulnerability of OC in deep soil and sediment layers, deep OC dynamics could potentially influence 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 m 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), Chinese Academy of Sciences-Peking University Pioneer Collaboration Team and the Thousand Young Talents Program.

References

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

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

Burgess 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.

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

Chinese Academy of Sciences (2001) Vegetation Atlas of China. Science Press, Beijing.

Drake JM, Randin C, Guisan A (2006) Modelling ecological niches with support vector machines. Journal of Applied Ecology, 43, 424–432.

Epstein HE, Raynolds MK, Walker DA, Bhatt US, Tucker CJ, Puezo JE (2012) Dynamics of aboveground phytomass of the circumpolar Arctic tundra during the past three decades. Environmental Research Letters, 7, 015506. doi:10.1088/1748-9326/7/1/015506.

Fan Y, Van den Doel H (2008) A global monthly land surface air temperature analysis for 1948-present. Journal of Geophysical Research: Atmospheres, 113, D01103. doi:10.1029/2007JD008470.

Gao Y, Chen H, Wu Z, Sun H, Li M (1985) Soils of Xizang (Tibet). Science Press, Beijing.

Goward S, Tucker C, Dye D (1985) North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. Vegetatio, 64, 3–14.

Grosse G, Harden J, Turetsky M et al. (2011) Vulnerability of high-latitude soil organic carbon in North America to disturbance. Journal of Geophysical Research: Biogeosciences, 116, G00K06. doi:10.1029/2010JG001507.

Harden JW, Koven CD, Ping C-L et al. (2012) Field information links permafrost carbon to physical vulnerabilities of thawing. Geophysical Research Letters, 39, L15704. doi:10.1029/2012GL051998.

He H, Liu M, Xiao X et al. (2014) Large-scale estimation and uncertainty analysis of gross primary production in Tibetan alpine grasslands. Journal of Geophysical Research: Biogeosciences, 119, 2013JG002449. doi:10.1002/2013JG002449.
Holben BN (1986) Characteristics of maximum-value composite images from tempo-

Hugelius G, Bockheim JG, Cassill P et al. (2013) A new data set for estimating organic carbon storage to 3 m depth in soils of the northern circumpolar permafrost region. Earth System Science Data, 5, 395–402.

Johannesson EC, Jackson RB (2000) The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications, 10, 423–436.

Kazviri T, Clakesen I (2009) A kernel functions analysis for support vector machin-

Koven CD, Riley WJ, Stern A (2013) Analysis of permafrost thermal dynamics and response to climate change in the CMIP5 earth system models. Journal of Climate, 26, 1877–1900.

Koven CD, EtAl Schuur, Schädel C et al. (2015a) A simplified, data-constrained approach to estimate the permafrost carbon-climate feedback. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 373, 20140243. doi: 10.1098/rsta.2014.0243.

Koven CD, Lawrence DM, Riley WJ (2015b) Permafrost carbon-climate feedback is sensitive to deep soil carbon decomposability but not deep soil nitrogen dynamics. Proceedings of the National Academy of Sciences of the United States of America, 112, 3752–3757.

Li R, Zhao L, Ding Y et al. (2012) Temporal and spatial variations of the active layer along the Qinghai-Tibet Highway in a permafrost region. Chinese Science Bulletin, 57, 4609–4616.

Li Q, Yue T-X, Wang C-Q et al. (2013) Spatially distributed modeling of soil organic matter across China: an application of artificial neural network approach. Catena, 104, 210–218.

Lin JC, Pejam MR, Chan E, Wofsy SC, Gottlieb EW, Margolis HA, McCaughey JH, Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83, 195–213.

Nowinski NS, Taneva L, Trumbore SE, Welker JM (2010) Decomposition of old organic matter as a result of deeper active layers in a snow depth manipulation experiment. Geophysical Research Letters, 37, 4616.

Qinghai-Tibet Plateau.

Wang B, French HM (1995) Permafrost on the Tibetan Plateau. China. Quaternary Science Reviews, 14, 255–274.

Wang B, French HM (1995) Permafrost on the Tibetan Plateau. China. Quaternary Science Reviews, 14, 255–274.

Wang S (1997) On the classification of permafrost on Qinghai-Xizang Plateau. Arid Land Geography, 20, 56–61.

Wang J, Chen X, Zhao H (2010) The vertical transport of old carbon in permafrost regions and its environmental effects on the Tibetan Plateau: a review of recent research. Journal of Geophysical Research-Biogeosciences, 115, D09107. doi: 10.1029/2009JD012974.

Yang M, Nelson FE, Shiklomanov NI, Guo D, Wan G (2010) Permafrost degradation and its environmental effects on the Tibetan Plateau. A review of recent research. Earth-Science Reviews, 105, 31–44.

Yang Y, Ji C, Chen L, Ding J, Cheng X, Robinson D (2015) Edaphic rather than cli-
matic controls over 13C enrichment between soil and vegetation in alpine grasslands on the Tibetan Plateau. Functional Ecology, 29, 839–848.

Zhao L, Ding Y, Liu G, Wang S, Jin H (2010) Estimates of reserves of ground ice in permafrost regions on the Tibetan Plateau. Journal of Glaciology and Geocryology, 32, 1–9.

Zhou Y, Guo D (1982) Principal characteristics of permafrost in China. Journal of Glaciology and Cryogeo.

© 2016 John Wiley & Sons Ltd, Global Change Biology, 22, 2688–2701.
Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1** Summary of estimated median volumetric organic carbon (OC) density (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.

**Fig. S3** Relative importance of predictor variables, reflected by percentage of $r^2$, for organic carbon density (OCD) in the top 50 cm.

**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 75th and 25th percentiles) of the 1000 estimates and the median.

**Fig. S5** Validation of the regression model of organic carbon density (OCD) between the depths of 0–30 cm and 0–50 cm ($\log_{10}\text{OCD}_{0-50\text{ cm}} = 0.9806 \log_{10}\text{OCD}_{0-30\text{ cm}} + 0.2911$) constructed in the eastern Tibetan Plateau, using actual measurements in the western part of the plateau.

**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 S1** Source code for the support vector machine (SVM) and uncertainty analyses in R software.

**Appendix S2** Spatial organic carbon density (OCD) at different depths at a resolution of 10 km × 10 km in alpine grasslands on the Tibetan Plateau.

**Appendix S3** Pedon observations from 173 sites in alpine grasslands on the Tibetan Plateau.