Agriculture, Ecosystems and Environment 254 (2018) 213–223

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Baseline map of organic carbon stock in farmland topsoil in East China

Xunfei Deng[a,b], Xiaojia Chen[a,b], Wanzhu Ma[a,b], Zhouqiao Ren[a,b], Minghua Zhang[c], Michael L. Grieneisen[c], Wenli Long[a], Zhihua Ni[d], Yu Zhan[e,*], Xiaonan Lv[a,b,**]

[a] Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China
[b] Key Laboratory of Information Traceability for Agricultural Products, Ministry of Agriculture, Hangzhou, Zhejiang 310021, China
[c] Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA
[d] Agricultural Technology Extension Service Center of Zhejiang Province, Hangzhou, Zhejiang 310029, China
[e] Department of Environmental Science and Engineering, Sichuan University, Chengdu, Sichuan 610065, China

ARTICLE INFO

Keywords:
Soil organic carbon
Baseline map
Machine learning
Carbon accounting
Harmonized world soil database
Climate change

ABSTRACT

Soil organic carbon (SOC) is important to soil fertility and the global carbon cycle. Accurate estimates of SOC stock and its dynamics are critical for managing agricultural ecosystems and carbon accounting under climate change, especially for highly cultivated regions. We extensively surveyed the SOC levels in 29,927 sites in Zhejiang province, an intensively cultivated region of East China, from year 2007 to 2008. We then estimated the spatial distribution of topsoil (0–30 cm) organic carbon stock using a random forest (RF) model, which is a powerful machine learning algorithm with superior predictive performance over parametric statistical models. The final RF model contained 23 predictor variables, covering soil properties, vegetation, climate, topography, land cover, farming practices, and locations. The RF model showed high performance in predicting the SOC stock, with a coefficient of determination (R^2) of 0.76 and a root mean square error (RMSE) of 10.63 t C ha^{-1}. This performance was superior to the General Linear Model (GLM) (R^2 = 0.35, RMSE = 19.93 t C ha^{-1}) and the ordinary kriging (OK) method (R^2 = 0.57, RMSE = 14.44 t C ha^{-1}), and was equivalent to Boosted Regressing Trees (BRT) (R^2 = 0.73, RMSE = 11.26 t C ha^{-1}). According to the variable importance evaluation, soil properties were the most important predictor variables, followed by climate and location, with relative importance values of 61%, 17%, and 14%, respectively. The predicted SOC stock ranged from 14.8 to 125.5 t C ha^{-1}, with an average ± standard deviation of 50.1 ± 12.3 t C ha^{-1}. The mean SOC level obtained from this survey was considerably lower than the value of 60.5 t C ha^{-1} reported for the same region in the Harmonized World Soil Database (HWSD), which is the most commonly used soil database worldwide. A large spatial discrepancy of SOC stock was observed between this survey and HWSD in regional and sub-regional levels. This study provided an updated regional baseline map of SOC levels for improving farmland management and refining carbon accounting under climate change.

1. Introduction

There are increasing concerns over intensified carbon dioxide (CO_2) emissions and climate change (Bell and Worrall, 2009). Soil is an important carbon pool in the global carbon cycle since it stores the most terrestrial organic carbon (Lal, 2008). Soil carbon sequestration mitigates the greenhouse gas effect and improves soil fertility (Viscarra Rossel et al., 2014). Recently, increased attention has been paid to the SOC dynamics, particularly with disturbances caused by anthropogenic activities under climate change (Pan et al., 2010; Muñoz-Rojas et al., 2015). Around the world, baseline maps of soil organic carbon (SOC) are extremely important for understanding the soil carbon balances on regional and global scales, and these maps have become indispensable components of Intergovernmental Panel on Climate Change (IPCC) Guidelines for Greenhouse Gas Inventories (IPCC, 2006).

SOC stocks and their spatial distributions have been reported on various spatial scales in the past. At national or continental levels, the SOC stocks were estimated in Europe (Lugato et al., 2014), Africa (Hengl et al., 2015), North America (Mishra et al., 2010), and China (Ni, 2001). Most of these SOC baseline maps were derived from limited soil survey data collected in previous decades, resulting in coarse spatial resolutions and a tendency to be very out of date. Globally, the Harmonized World Soil Database (HWSD) with the world soil map (FAO-90) has aimed to meet the requirements for soil carbon

* Corresponding author.
** Corresponding author at: Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China.
E-mail addresses: yzhan0@gmail.com (Y. Zhan), luxn@mail.zaas.ac.cn (X. Lv).

https://doi.org/10.1016/j.agee.2017.11.022
Received 29 June 2017; Received in revised form 22 November 2017; Accepted 23 November 2017
0167-8809/

measurements under the Kyoto Protocol and the FAO/IIASA global agro-ecological assessment study. It is the most commonly used SOC baseline in the world, and has a spatial resolution of approximately 1 km (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). HWSD has become the main data source for estimating large-scale SOC storage and is currently widely used by climate change models (Jobbagy and Jackson, 2000; Batjes, 2009; Perlman et al., 2014). These large-scale SOC baselines provide valuable bases for carbon accounting as well as modeling interactions between the global carbon cycle and climate change (Liu et al., 2006; Mishra et al., 2010; Romero et al., 2012).

While SOC levels generally remain steady for decades, dramatic land use changes and intensive agricultural activities lead to considerable changes of local SOC. This results in high discrepancies between the current situation and the SOC levels recorded in HWSD, as the data in that source were derived from the soil surveys conducted decades ago (FAO/IIASA/ISRIC/ISSCAS/JRC, 2012). SOC contents are continuously influenced by environmental conditions and land use changes (Fantappiè et al., 2010; Martin et al., 2010), especially for intensively cultivated farmlands (Six et al., 2000; Halvorson et al., 2002; Yan et al., 2012). SOC turnover might also be induced by shifts in soil management practices (West and Post, 2002; Hengl et al., 2015). For instance, increases of regional SOC stock were found in China and the USA due to enhanced agricultural production or land use changes (Liao et al., 2009; Pan et al., 2010; Xiong et al., 2014). Furthermore, the quality of the SOC baseline map is highly affected by sampling densities. Data interpolation or extrapolation from a limited number of soil samples on the basis of soil map units often overlook the effects of climate, land cover, and soil management practices (Smith et al., 2012). In addition, the uncertainties associated with the SOC baselines tend to be propagated in downstream analyses, such as carbon accounting for the global carbon cycle or modeling greenhouse gas emissions (e.g., Batjes, 2009; Perlman et al., 2014; Ma et al., 2016).

Many studies have derived accurate spatial SOC stock on regional or national scales using various approaches (Mishra et al., 2010; Elbasiouny et al., 2014). Calculating SOC averages by soil types and spatial interpolation from site-specific observations have been commonly used (Batjes, 2009; Chaplot et al., 2010). In addition, regression models, such as general linear model (GLM) with a set of predictor variables (e.g., land uses, soil types, and management practices), were developed to capture the SOC spatial variation (Mishra et al., 2010; Meersmans et al., 2012). Moreover, various machine leaning algorithms (e.g., random forest and boosted regression trees) have been used to develop SOC baselines at fine spatial resolutions based on existing soil databases and auxiliary data, such as satellite retrievals and climatic conditions (Viscarra Rossel et al., 2014; Were et al., 2015; Hengl et al., 2017; Schillaci et al., 2017). These regional estimates of SOC stock tend to be more accurate than the estimates solely derived from the site observations, as the auxiliary information carried in the predictor variables improves the interpolation accuracies. The estimates at higher resolutions (e.g., ~250 m or ~90 m) have greatly facilitated sophisticated regional simulations (Lugato et al., 2014; Hengl et al., 2015). However, the large-scale assessment of baseline SOC stock at fine spatial resolutions are typically unavailable due to the lack of extensive soil survey data and the sophisticated statistical modeling that is required (Viscarra Rossel and Chen, 2011).

East China is one of the most intensively cultivated regions in the world. The most recent national soil survey was conducted in early 1980s, and while it is the data source for HWSD it is woefully inadequate to represent the current status of SOC in this region (Pan et al., 2010). It is therefore important to update the SOC baseline map. This study mapped the spatial distribution of SOC stock (0–30 cm) in the farmlands of Zhejiang province, East China. Random forest is an ensemble of classification or regression trees which has shown higher predictive performance than other statistical models in various environmental and ecological research fields (Cutler et al., 2007; Reid et al., 2015). Recently, random forest has been employed to estimate various soil properties in digital soil mapping (DSM) (Grimm et al., 2008; Were et al., 2015; Hengl et al., 2017; Zhang et al., 2017). In our region, a random forest model was developed with a comprehensive set of predictor variables, including vegetation, topography, climate, land use, locations, farming practices, and soil properties. The regional SOC stock distributions were estimated at a fine spatial resolution (~250 m), and estimation uncertainties were also assessed. On the basis of the estimated SOC distributions, we then evaluated the differences in the SOC stock distributions between this study and the HWSD. The high-resolution SOC baselines mapped in this study are critical for improving regional farmland management by promoting good practices associated with elevated SOC stock, and refining carbon accounting under climate change with the updated SOC stock data.

2. Materials and methods

2.1. Study area

Zhejiang province (27°02′N-31°12′N, 118°01′E-122°58′E; ~102,646 km^2) is one of the most highly developed regions in East China. The primary land covers are forest and farmland. The farmlands are mainly located in the Hangjiahu plain, central basin, and east coast region, where the elevations are generally lower than 100 m (Fig. S1). Paddy rice, vegetables, dry-land crops (e.g., broad bean, maize, and oil rape) and their rotations are the dominant cropping patterns in these agricultural areas. The main soil types are Fluvisols, Anthrosols, Acrisols, Alisols, and Regosols (ZJSSO, 1994). This region is located in the subtropical climate zone, with an annual mean temperature of 15–18 °C and an annual mean precipitation of 980–2000 mm. The precipitation and temperature exhibit increasing gradients from north to south, with local variations occurring based on topography (CMDC, 2015).

2.2. Field survey and laboratory analysis

A total of 29,927 topsoil (0–20 cm) samples were collected from May 2007 to April 2008 by employing the standard soil sampling protocol (Fig. 1) (Wang and Zhou, 2011). Note that the 20cm-depth SOC measurements were used to estimate the 30 cm-depth SOC by using the method employed in a previous study (Yu et al., 2012) in order to compare our data with data in the HWSD. Please refer to S1 in the Supplementary material for the details. The digital maps of soil types, cropping patterns, and land use were spatially overlaid, and their intersected polygons were identified as the spatial units for designing the field survey. This spatial-unit map was used to determine the locations of sampling sites in order to best capture the spatial variations in SOC. In the field survey, one soil sample was collected per 0.25–0.4 km^2. In farmlands with frequent crop rotations, resulting in intense SOC turnover, the sampling density was increased to approximately one soil sample per 0.15–0.25 km^2. Each sample consisted of 6–8 soil sub-samples, which were collected by following an "S" shape at intervals of a few meters. All of the sub-samples were collected at a depth of 20 cm and had a similar weight of approximately 0.5 kg. The methods of coning and quartering were employed to adequately mix these sub-samples in order to obtain one representative soil sample per site with the weight of 1 kg. Note that soil sampling was conducted at least a week after the previous rotation crop was harvested in order to exclude the effects of fertilization.

During the field survey, the previous rotation crop, topography of the farmland, and the longitude/latitude of each representative soil sample were also recorded. The soil texture of the sampling site was obtained from the field estimation. A commonly used method for estimating soil bulk density was conducted with five parallel measurements at each sampling site. To collect a fixed volume (100 cm^3) of topsoil, a metal ring with an intact core was pressed into the soil, followed by careful excavation of the soil around the ring without disturbing or