Science of the Total Environment 669 (2019) 844–855

Contents lists available at ScienceDirect

Science of the Total Environment

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

Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms

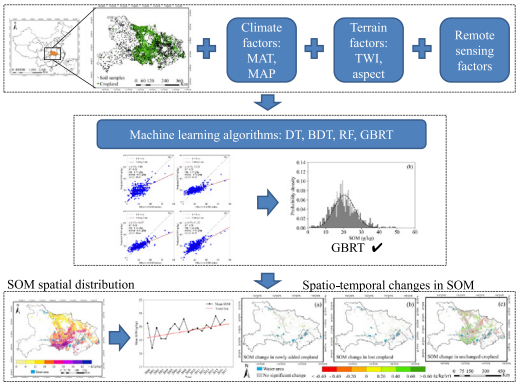
Di Chen [a,b,c,1], Naijie Chang [a,c,1], Jingfeng Xiao [c,*], Qingbo Zhou [a,b], Wenbin Wu [a,b,*]

[a] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[b] Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture, Beijing 100081, China
[c] Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, 03824, USA

HIGHLIGHTS

- GBRT was a better algorithm for spatially predicting and mapping SOM content in Hubei, China than DT, BDT, and RF.
- Remote sensing reflectance and vegetation indices were proved to be key factors for predicting SOM content.
- The SOM content in the topsoil in 2017 varied from 0.89 to 58.86 g/kg, with a mean value of 20.52 g/kg.
- The mean cropland SOM content of Hubei exhibited a slight increasing trend from 2000 to 2017.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 31 December 2018
Received in revised form 2 March 2019
Accepted 10 March 2019
Available online 11 March 2019

Editor: Jose Julio Ortega-Calvo

Keywords:
Digital soil mapping
Multi-year
Soil organic carbon
MODIS
Machine learning algorithms
Cropland

ABSTRACT

As an important indicator of soil quality, soil organic matter (SOM) significantly contributes to land productivity and ecosystem health. Accurately mapping SOM at regional scales is of critical importance for sustainable agriculture and soil utilization management and remains a grand challenge. Many studies used soil sampling data and machine learning algorithms to predict SOM at regional scales for a given year, while few studies mapped SOM for multiple years and examined its temporal dynamics. We compared the performance of four machine learning algorithms: decision tree (DT), bagging decision tree (BDT), random forest (RF), and gradient boosting regression trees (GBRT) in mapping SOM in Hubei province, China over the 18-year period from 2000 to 2017. Our results showed that RF and DT had the highest coefficient of determination (R^2) (0.61) and the lowest potential bias (9.48 g/kg), respectively, while GBRT had the lowest mean error (ME) (1.26 g/kg), root mean squared error (RMSE) (5.41 g/kg) and Lin's concordance correlation coefficient (LCCC) (0.72). The SOM map based on GBRT better captured the distribution of the soil sample data than that based on RF. The trained GBRT model and the spatially explicitly data on explanatory variables (e.g., climate, terrain, remote sensing) were used to predict SOM for each 500 m × 500 m grid cell in Hubei for the period from 2000 to 2017. Our results showed that the SOM content of cropland was relatively high in the southeast and relatively low in the north. The SOM content in the topsoil varied from 0.89 to 58.86 g/kg and was averaged at 20.52 g/kg. The mean cropland SOM content of the province exhibited an increasing trend from 2000 to 2017 with an increase of 0. 26 g/kg and a growth

* Corresponding authors.
E-mail addresses: j.xiao@unh.edu (J. Xiao), wuwenbin@caas.cn (W. Wu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scitotenv.2019.03.151