Chemosphere 127 (2015) 49–57

Contents lists available at ScienceDirect

Chemosphere

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

Estimating spatial distribution of soil organic carbon for the Midwestern United States using historical database

Sandeep Kumar *

Department of Plant Science, South Dakota State University, Brookings, SD 57007, USA

HIGHLIGHTS

- Estimation of SOC at regional scale is important to address climate change issues.
- The GWRK approach was used in this study for SOC estimations.
- GWRK provided lower estimation errors compared to GWR for estimating SOC.
- GWRK performs better at regional scale as compared to the global regression.
- Local models provide better explanation of spatial distribution of SOC.

ARTICLE INFO

Article history:
Received 3 July 2014
Received in revised form 26 November 2014
Accepted 12 December 2014
Available online 2 February 2015

Handling Editor: X. Cao

Keywords:
Soil organic carbon
Geographically weighted regression
Geographically weighted regression kriging
Spatial distribution
SOC mapping

ABSTRACT

Soil organic carbon (SOC) is the most important parameter influencing soil health, global climate change, crop productivity, and various ecosystem services. Therefore, estimating SOC at larger scales is important. The present study was conducted to estimate the SOC pool at regional scale using the historical database gathered by the National Soil Survey Staff. Specific objectives of the study were to upscale the SOC density (kg C m^{-2}) and total SOC pool (Pg C) across the Midwestern United States using the geographically weighted regression kriging (GWRK), and compare the results with those obtained from the geographically weighted regression (GWR) using the data for 3485 georeferenced profiles. Results from this study support the conclusion that the GWRK produced satisfactory predictions with lower root mean square error (5.60 kg m^{-2}), mean estimation error (0.01 kg m^{-2}) and mean absolute estimation error (4.30 kg m^{-2}), and higher R^2 (0.58) and goodness-of-prediction statistic ($G = 0.59$) values. The superiority of this approach is evident through a substantial increase in R^2 (0.45) compared to that for the global regression ($R^2 = 0.28$). Croplands of the region store 16.8 Pg SOC followed by shrubs (5.85 Pg) and forests (4.45 Pg). Total SOC pool for the Midwestern region ranges from 31.5 to 31.6 Pg. This study illustrates that the GWRK approach explicitly addresses the spatial dependency and spatial non-stationarity issues for interpolating SOC density across the regional scale.

1. Introduction

Being the largest pool in terrestrial ecosystem (Post et al., 1990; Lal, 2002), soil organic carbon (SOC) has gained attention because of its relevance to soil health and the global climate. Apparent changes in global climate and ever increasing atmospheric concentrations of greenhouse gases (GHGs) have increased the major concerns among the researchers for slowing down the rate of increase of these gaseous emissions. Emission of these GHGs can partly be mitigated by sequestering C in the soils (Batjes, 1998). To quantify this C, especially organic C, denoted as SOC density (kg C m^{-2}) at regional or national scales, and to understand how does this density manifestate spatially are the important researchable issues (Vasques et al., 2010). Nevertheless, there are still a lot of uncertainties in the estimations of SOC density because of high spatial variability (Palmer et al., 2002). Since, Midwestern United States is the largest and most intensive crop-producing region of the United States (Kolpin et al., 1999). This region comprises of 21% of the Nation's land, and accounts for 20–30% of the nation's total SOC pool (Guo et al., 2006). The SOC density of this region, therefore, has created concern about its possible effect on the regional climate change. Thus, a better approach is needed to estimate the SOC pool for the regional scale with reduced uncertainties, and to understand how carbon density varies spatially.

* Tel.: +1 605 688 4306; fax: +1 605 688 4452.
E-mail address: Sandeep.Kumar@sdstate.edu

http://dx.doi.org/10.1016/j.chemosphere.2014.12.027
0045-6535/

Traditionally, measurement techniques of SOC content rely on methods involving direct *in situ* sampling, and subsequent laboratory analysis. Such methods involve experiments on field plots, employing discrete point measurements of SOC or other soil properties (Merrill, 1998). These properties are then calculated for a larger area based on the data available from point measurements, providing inaccurate and incomplete information. In order to quantify space–time variability of SOC, a large number of measurements are required at regional scale (Chang and Islam, 2000). Assessment of SOC variability is critical to site-specific management, soil survey and natural resource inventory. Further, spatial distribution of SOC is important to facilitate the regional planning (Piccini et al., 2014).

Geostatistics analysis that integrates auxiliary data with SOC density can improve the ability to resolve finer differences of variability across the space. Nevertheless, there are large uncertainties among different approaches regarding the estimated magnitude of SOC density. For example, approaches such as multiple linear regression (MLR), and regression kriging (RK) (Odeh et al., 1995; Lopez-Granados et al., 2005) can be used for a larger regional scale, reducing time and costs (Tajgardan et al., 2010) and have better predictions as compared to earlier traditional approaches such as ordinary kriging (OK) where estimations are not based on predictors. Major limitation, however, of these (MLR and RK) approaches is that the relationship between target and covariates is assumed to be stationary across the space, and hence provide a misleading information.

Geographically weighted regression (Fotheringham et al., 2002), a local spatial statistical is the only approach specifically designed for exploring spatial nonstationarity, defined as when the nature and significance of relationships between variables differs from location to location (Fotheringham et al., 2002). Compared to global regression, GWR can be expected to yield smaller residuals, and outputs of GWR can be used to visualize spatial variations in regression diagnostics and model parameters within a study region (Gilbert and Chakraborty, 2011). Geographical relationships in GWR vary across the space (Fotheringham et al., 1996). Geographically weighted regression kriging (GWRK) is an extension of GWR which combined the local regressions (GWR) with kriging of the regression residuals (Harris et al., 2010) for improving the estimations of SOC density at state scale (Kumar and Lal, 2011; Kumar et al., 2012). Use of statistical and geostatistical techniques for mapping SOC density using a series of satellite and other easily accessible data have been conducted globally. The GWRK approach has been successfully attempted at the state scale in our previous study (Kumar and Lal, 2011; Kumar et al., 2012), no attempt, however, has been made using the GWRK approach at regional scale in the U.S. Therefore, this study was conducted to evaluate the statistical correlations between SOC density and land use, bedrock geology, and environmental data from 3485 georeferenced locations extracted from the National Soil Survey Center (NSSC) database for estimating the SOC density at regional (consists 12 states of the USA) scale.

Specific objectives of the present study were to (i) estimate the spatial distribution of SOC density for 1-m depth across the Midwestern region of U.S.A., and (ii) estimate the total SOC pool for 1-m depth stored in different land uses of the study region based on a recent, the GWRK approach.

2. Materials and methods

2.1. Data source and study area

The proposed study was conducted for the Midwestern region of U.S.A., comprises of 12 states including North Dakota, South Dakota, Wisconsin, Minnesota, Iowa, Nebraska, Indiana, Ohio, Illinois, Kansas, Missouri, and Michigan (Fig. 1). Mean annual precipitation (MAP) of the region ranges from 846 to 1098 mm, and mean annual air temperature (MAAT) is 10.1 °C. Geographical area of this region is 1.98×10^6 km^2 (20% of the total nation area) with elevation ranges between 97 and 2023 m above mean sea level. Major soil orders present in the region include Alfisols, Mollisols, Histosols, Entisols, and Spodosols.

A total of 3485 georeferenced data profiles were extracted from the National Soil Survey Lab, Lincoln, NE (NSSC, 2013), of which 2788 (80%) were used for calibration and 697 (20%) for validation. Soil parameters that were extracted from the NSSC database include: genetic features of horizons, horizon depth (cm), SOC concentrations (%), and sand, silt and clay contents (%). Dataset was loaded in Microsoft Access 2007 database, and a query was made to match the different profiles with the soil attributes. Soil profiles having zero value of SOC content were deleted from the dataset and were not used in the interpolation process. Fig. 1 shows the study site and profile locations for calibration and validation sets.

2.2. Explanatory variables, their sources and selection

Predictors used in the present study include: digital elevation model (DEM), slope in degrees (slope), elevation, mean annual air temperature (MAAT), mean annual precipitation (MAP), land use (LULC), bedrock geology, and normalized difference vegetation index (NDVI) (Fig. 2). The LULC was reclassified into 6 classes that include: developed (6.7%), barren (0.20%), wetland (7.6%), shrubs (19.6%), cropland (48.7%), and forests (17.2%). A total of about 100 different types of bedrocks were present in the study region. The DEM and LULC maps were extracted from the U.S. Geologic Survey database and climatic data [long term (1970–2000) MAAT, and MAP] were extracted from the database of Spatial Climatic Analysis Service of the Oregon State University (Daly et al., 2001). A 30-m (1″) resolution for all the variables was used for the study site. The DEM, which represents a raster model of the elevation values (Balkovič et al., 2007), was used for calculating the slope using *3D Analyst tools* of ArcGIS 9.3. The NDVI (Rouse et al., 1973) data were extracted from the Global Land Cover Facility database, and the data derived from Moderate-Resolution Imaging Spectroradiometer (MODIS) Bands 1 (red) and 2 (near infrared). These NDVI values are produced every 16 d (Carroll et al., 2004). The NDVI values ranges from −1.0 to +1.0, values closer to zero suggest low vegetation and values closer to 1.0 suggest higher vegetation. The NDVI is widely used vegetation index due to its wide-spread familiarity, simplicity, and ease of application (Gu et al., 2009), and is described as:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \tag{1}$$

where NIR is the near-infrared band and red is the red band. The NDVI has been available since 1981 on a routine basis at a global scale with coarse resolution. Many studies have used these archived data to identify trends in vegetation phenology and productivity during the last two decades.

For estimating SOC density, a total of 15 predictors were used out of 163 after using stepwise regression analysis with SAS 9.2 software. The SOC density for individual soil profile was calculated by summing up the soil C (kg m^{-2}) in each soil horizon from the surface to 1-m depth as follows (Eq. (1)):

$$\mathrm{SOC}_{\mathrm{density}} = \sum_{i=1}^{n} (\mathrm{SOC}_i \times \rho_b \times z_i) \tag{2}$$

where $\mathrm{SOC}_{\mathrm{density}}$ is the SOC density (kg m^{-2}) up to 1.0-m depth, i is the soil horizon $(1,2,3,\ldots,n)$, SOC_i is the SOC concentration