Geoderma 318 (2018) 148–159

Contents lists available at ScienceDirect

Geoderma

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

Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks

Luigi Lombardo[a,b,*], Sergio Saia[c], Calogero Schillaci[d], P. Martin Mai[b], Raphaël Huser[a]

[a] Computer, Electrical and Mathematical Sciences & Engineering Division, KAUST, Thuwal, Saudi Arabia
[b] Physical Sciences and Engineering Division, KAUST, Thuwal, Saudi Arabia
[c] Council for Agricultural Research and Economics (CREA), Research Centre for Cereal and Industrial Crops (CREA-CI), Foggia, Italy
[d] Department of Agricultural and Environmental Science, University of Milan, Italy

ARTICLE INFO

Keywords:
Quantile Regression
R coding
Topsoil organic carbon
Digital soil mapping
Mediterranean agro-ecosystem

ABSTRACT

Soil organic carbon (SOC) estimation is crucial to manage natural and anthropic ecosystems. Many modeling procedures have been tested in the literature, however, most of them do not provide information on predictors' behavior at specific sub-domains of the SOC stock. Here, we implement Quantile Regression (QR) to spatially predict the SOC stock and gain insight on the role of predictors (topographic and remotely sensed) at varying SOC stock (*0–30cm depth*) in the agricultural areas of an extremely variable semi-arid region (Sicily, Italy, around 25,000km^2). QR produces robust performances (*maximum quantile loss = 0.49*) and allows to recognize dominant effects among the predictors at varying quantiles. In particular, clay mostly contributes to maintain SOC stock at lower quantiles whereas rainfall and temperature influences are constantly positive and negative, respectively. This information, currently lacking, confirms that QR can discern predictor influences on SOC stock at specific SOC sub-domains. The QR map generated at the median shows a Mean Absolute Error of *17 t SOC* ha^{-1} with respect to the data collected at sampling locations. Such MAE is lower than those of the Joint Research Centre at Global (*18 t SOC* ha^{-1}) and at European (*24 t SOC* ha^{-1}) scales and of the International Soil Reference and Information Centre (*23 t SOC* ha^{-1}) while higher than the MAE reached in Schillaci et al. (2017b) (Geoderma, 2017, issue 286, page 35–45) using the same dataset (*15 t SOC* ha^{-1}). The results suggest the use of QR as a comprehensive method to map SOC stock using legacy data in agro-ecosystems and to investigate SOC and inherited uncertainty with respect to specific subdomains. The R code scripted in this study for QR is included.

1. Introduction

Soil Organic Carbon (SOC) plays a key role in various agricultural and ecological processes related to soil fertility, carbon cycle and soil-atmosphere interactions including CO_2 sequestration. Thus, its knowledge has a crucial importance both at global and local scales, especially when aiming at managing natural, anthropic areas and agricultural lands. In this context, the scientific community has spent considerable efforts in mapping SOC, its spatiotemporal variation, and in confirming its primary role in shaping ecosystems functioning (Ajami et al., 2016; Grinand et al., 2017; Ratnayake et al., 2014; Schillaci et al., 2017a).

Spatio-temporal studies can be found in various geographic contexts from Africa (Akpa et al., 2016), Asia (Chen et al., 2016), Australia (Henderson et al., 2005), Europe (Yigini and Panagos, 2016), North-America (West and Wali, 2002), central America (Ross et al., 2013) to South-America (Araujo et al., 2016). The variability of the local landscape, availability of funding, mean gross income of the population in the area, and temporal commitment affect the number of samples, their spatial density and distribution. As a result, experiments have been conducted on almost regular and dense grids, mostly focusing on small areas (Lacoste et al., 2014; Taghizadeh-Mehrjardi et al., 2016) and others using sampling strategies that significantly vary across space (Mondal et al., 2016). The latter studies mainly correspond to regional or larger scales (Reijneveld et al., 2009; Sreenivas et al., 2016), with only few cases having an optimal sample density maintained at a national level (Mulder et al., 2016). The characteristics of the environment under study can require the use of different predictors capable of explaining the variability of soil traits, topography and standing biocoenosis, especially (cropped or natural) phytocoenosis, the latter being efficiently explained by remotely sensed (RS) properties (Morellos et al., 2016; Peng et al., 2015).

Modeling procedures for SOC primarily aim at constructing present, past or predictive maps, and at studying the role of each predictor over the target variable. Regarding the latter, the estimation of predictor

* Corresponding author.
E-mail address: luigi.lombardo83@gmail.com (L. Lombardo).

https://doi.org/10.1016/j.geoderma.2017.12.011
Received 1 August 2017; Received in revised form 5 December 2017; Accepted 9 December 2017
0016-7061/ © 2017 Elsevier B.V. All rights reserved.

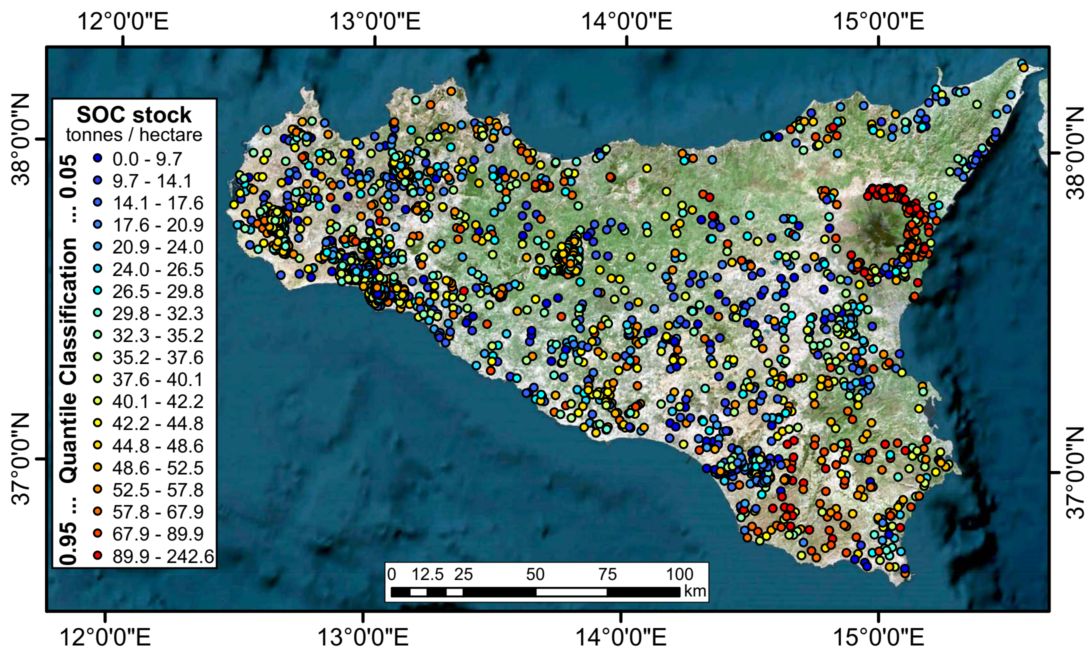

Fig. 1. SOC stock dataset and geographic contextualization.

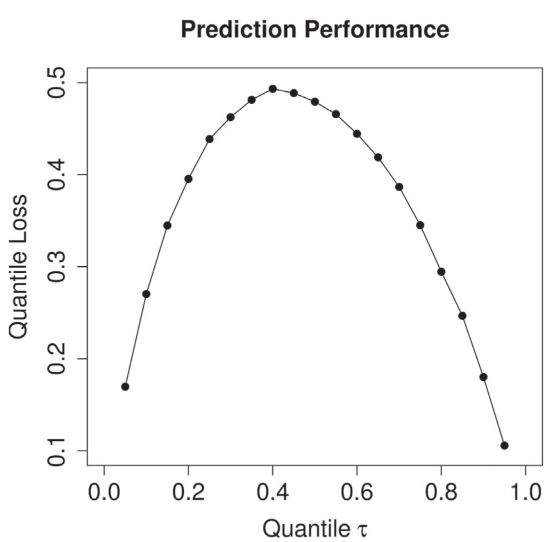

Fig. 2. Leave-one-out performance evaluation using the quantile loss.

contributions on a target variable such as SOC, is of particular interest to efficiently obtain agro-environmental and social benefits (e.g. Viscarra Rossel and Bouma, 2016).

Statistical applications provide quantitative ways to deal with such research questions. The current literature encompasses algorithms that can be clustered into interpolative and predictive. Pure interpolators are broadly used when the density of the samples is sufficient to regularly describe the variation of SOC across a given area. Examples with excellent prediction performances are reported in Hoffmann et al. (2014), Piccini et al. (2014). The weakness of these approaches becomes evident when using datasets with non-regular distribution in space (Dai et al., 2014; Miller et al., 2016). Conversely, regression-based predictive models hardly suffer from the spatial sampling scheme as they do not rely on the distribution across the geographic space in order to derive functional relations between SOC and independent variables (Hobley et al., 2016).

Among these, linear regression models are a well-established tool for estimating how, on average, certain environmental properties affect SOC and SOC stock (Rodríguez-Lado and Martínez-Cortizas, 2015). However, they are bounded by definition to model the conditional mean, thus being unable to explore the effects of the same properties at different C contents or stock of the soil, especially near the tails of the distribution.

In the present work, Quantile Regression (hereafter QR, Koenker (2005)) is used to model SOC stock from a non-homogenously sampled topsoil SOC dataset using soil texture, land use, topographic and remotely sensed covariates. In particular, QR is able to model the relationship between a set of covariates and specific percentiles of SOC stock. In classical regression approaches, the regression coefficients (also often called beta coefficients) represent the mean increase in the response variable produced by one unit increase in the associated covariates. Conversely, the beta coefficients obtained from QR represent the change in a specific quantile of the response variable produced by a one unit increase in the associated covariates. In this way, QR allows one to study how certain covariates affect median (quantile $\tau = 0.5$) or extremely low (e.g., $\tau = 0.05$) or high (e.g., $\tau = 0.95$) SOC stock values. Therefore, it gives a more comprehensive description of the effect of predictors on the whole SOC stock probability distribution (i.e., not just the mean) and may be used to analyze differential SOC stock responses to environmental factors.

Furthermore, when used for mapping purposes, QR allows for soil mapping at given quantiles, providing analogous estimates to more common approaches by using the median instead of the mean. A flexible non-parametric extension of linear QR, known as Quantile Regression Forest (QRF), was used in Malone et al. (2017) to derive conditional quantiles used to perform an efficient spatial downscaling.

In the present experiment we use a nested strategy to model SOC stock in Sicilian agricultural areas with QR: we initially aim at testing the QR overall performances when modeling the SOC stock by segmenting its distribution into 19 quantiles ($\tau = 0.05$ to $\tau = 0.95$). Subsequently, we examine the coefficients of each predictor for each of the quantiles. Ultimately, we compare the median prediction with available SOC benchmarks for the same study area to test the efficiency of QR for soil mapping purposes. The dataset used in this contribution is the same used in Schillaci et al. (2017b) where a Stochastic Gradient Treeboost is adopted.

Download English Version:

https://daneshyari.com/en/article/8894193

Download Persian Version:

https://daneshyari.com/article/8894193

Daneshyari.com