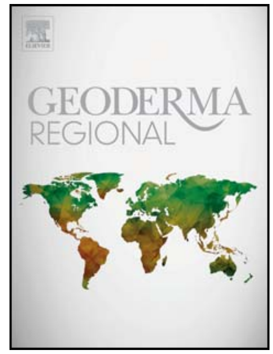
Accepted Manuscript



Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka

R.R. Ratnayake, S.B. Karunaratne, J.S. Lessels, N. Yogenthiran, R.P.S.K. Rajapaksha, N. Gnanavelrajah

PII: S2352-0094(16)30013-X
DOI: doi: 10.1016/j.geodrs.2016.03.002
Reference: GEODRS 82

To appear in:

Received date: 9 November 2015
Revised date: 17 March 2016
Accepted date: 23 March 2016

Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka

R. R. Ratnayake[1*], S. B. Karunaratne[2,5*], J. S. Lessels[3], N. Yogenthiran[4], R. P. S. K. Rajapaksha[1], N.Gnanavelrajah[3]

[1]National Institute of Fundamental Studies, Hantana Road, Kandy, Sri Lanka.

[2]Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Makandura, Gonawila NWP, Sri Lanka

[3]School of Geoscience, University of Aberdeen, UK.

[4]Faculty of Agriculture, University of Jaffna, Jaffna, Sri Lanka.

[5] Hawkesbury Institute for the Environment, Western Sydney University, Penrith NSW, Australia. (*Current address*)

* *Corresponding authors*: senawusl@gmail.com (S. B. Karunaratne); renukar@ifs.ac.lk (R. R. Ratnayake)

Abstract

In this study we explore the environmental covariates that are useful in mapping SOC contents in tropical paddy growing soils in Northern Sri Lanka. We carried out digital mapping of SOC contents across the study area using a spatial soil prediction function with auto correlated errors *via* linear mixed models (LMM). Two separate LMMs were fitted considering two depth intervals, 0 to 0.15 m and 0.15 to 0.30 m respectively. Results revealed that environmental covariates of SOC content for the two depth intervals considered in this study were different despite being situated within a zone typically considered as the top soil (0 to 0.3 m). Landsat 8 bands and its products derived from those bands (namely Landsat 8 band 2, band ratios 4/3 and 4/7) were identified as SOC significant environmental covariates for both depth intervals. In terms of spatially auto correlated residuals, optimum spatial models for both depth intervals were identified as a spherical model. Residuals were found to be spatially auto correlated up to 10000 m and 5000 m for 0 to 0.15 m and 0.15 to 0.30 m depth intervals respectively. Cross validation results suggest that fitted LMMs are adequate to carry out mapping across the study area with little difference between estimated bias (- 0.006 % for 0 to 0.15 m and -0.004 % for 0.15 to 0.30 m depth intervals) and accuracy (0.441 % for 0 to 0.15 m and 0.367 % for 0.15 – 0.30 m depth intervals) for fitted LMMs considering two depth intervals. However, in case of Lin's

concordance correlation value, 0 to 0.15 m depth interval reported higher value (0.81) compared to 0.15 to 0.30 m depth interval (0.60) suggesting upper depth interval model is much superior to the lower one. In conclusion, this study provides; a) firsthand information on current status of SOC contents in northern paddy growing soils in Sri Lanka, b) generated information that are useful to optimize spatial sampling of SOC in future applications and c) provided firsthand information vital to establishment of a national carbon accounting system in future.

Keywords: Soil organic carbon, digital soil mapping, spatial modeling, tropical paddy fields

1. Introduction

SOC is continuously subjected to decomposition and other biologically mediated transformations that generate or consume greenhouse gases (Baldock et al., 2012). In the global context, soils and their management have the potential to either increase or reduce atmospheric concentrations of greenhouse gases and the magnitude of any associated climate change (Baldock et al., 2012). Additionally, SOC plays a vital role in maintaining different biological, chemical and physical properties and processes that exist in soils and are important in sustaining the ecosystem services (Baldock and Broos, 2011). Therefore, a decline in SOC will have negative consequences in relation to soil health and productivity.

Due to the importance of SOC, much attention has been given to the digital mapping of SOC and its fractions or conceptual pools in space (*e.g.* Karunaratne et al., 2014a), space and depth (3D) (*e.g.* Poggio and Gimona, 2014) and, space and time *(e.g.* Karunaratne et al., 2014b). Digital soil mapping (DSM) techniques are used to estimate SOC contents/concentration and stock/density (Minasny et al., 2013). Methods of mapping soil properties including SOC using the DSM approach can be broadly categorized under three main approaches *i.e.* geostatistical; deterministic and hybrid geostatistical approaches. Out of these three approaches, the most common approach practiced is the hybrid geostatistical approach. Among the different hybrid geostatistical approaches, such as co-kriging, universal kriging and regression-kriging, latter described by Odeh et al. (1995) is the most commonly applied method. In this study we adopted a similar approach to regression-kriging approach in terms of linear mixed model (LMM) as described by Lark et al. (2006). The application of LMMs for DSM of carbon stocks and the associated measureable fractions were successfully demonstrated by Karunaratne et al. (2014b) and Karunaratne et al. (2014a).

As with many countries in South and South East Asia, rice is the staple food of Sri Lanka. It has been reported that paddy fields have higher SOC storage and sequestration capacity compared to dry land cropping systems (Pan et al., 2004). Rice cropping under waterlogged conditions enhances the soil organic matter (SOM) accumulation (Lal, 2002). The decomposition rates of SOM are considered to be smaller under anaerobic conditions than under aerobic conditions (Sahrawat, 2003), resulting in SOM accumulation and ensuring the carbon remains sequestered in the soils. Wu (2011) reported that SOC accumulation in paddy ecosystems was faster and more pronounced than in other arable ecosystems. Additionally, higher content of silt and clay in paddy soils compared with upland soils contributes to the larger SOC accumulation (Lal, 2002). Therefore, the maintenance of SOC in paddy fields is important not only for improving agricultural productivity but also for the reduction of greenhouse gas emission. However, little information is available on the SOC contents/stocks in tropical and sub-tropical paddy soils. It has been reported that SOC in China's paddy growing top soils contribute approximately 1.3 Pg, which is about 2% of the total SOC storage in the topsoil (Pan et al., 2004). A study conducted in Indo Gangetic Plains of India reported the total organic carbon (TOC) concentration is 6.8 g kg^{-1} in the surface 0 to 0.15 m soil layer (Nayak et al., 2012). Few studies have mapped the SOC contents and stocks in paddy growing soils (Sumfleth and Duttmann, 2008; Minasny et al. 2012). So far, no studies have investigated the spatial distribution of SOC in paddy growing soils at a regional scale within Sri Lanka.

In Sri Lanka, land cultivated under paddy covers approximately 34 % of the total cultivated lands (Central Bank of Sri Lanka, 2014). Rice is cultivated mainly using two cropping systems *i.e.* mono crop and crop rotation using either irrigated or rainfed. Rain is received from two major monsoon namely: South West monsoons (*Yala season*) [May to September] and North West monsoons (*Maha season*) [October to February]. In northern dry zone of Sri Lanka, rice is cultivated mainly in the *Maha* season and in the *Yala* season those lands either kept as fallow lands or crop with other field crops.

Currently the Sri Lankan rice sector is facing numerous issues such as low yield, increasing fertilizer cost, adverse weather conditions and insufficient supply of irrigation (Central Bank of Sri Lanka, 2014). Therefore, firsthand information vital for regional scale land resource planning will be provided through DSM of SOC which includes current status and information on soil health in the paddy growing regions.

In this study we have applied DSM techniques to quantify the SOC content across the paddy growing soils covering the Northern Province of Sri Lanka. We have concentrated on mapping SOC contents due to absence of bulk density data, which prevented the calculation of SOC stocks/densities. Significantly this is the first detailed study on digital mapping of any soil property covering a large area of extent using DSM approach in Sri Lanka. Specific aims of this study were to; (a) develop an appropriate spatial soil prediction function with an auto correlated error (SSPF) model for the identification of environmental drivers and to carry out prediction of SOC contents across the study area; (b) validate the developed SSPF and finally (c) discuss the possible use of derived DSM products in regional and national level planning related to soil and land resources.

2. Methods

2.1. Study area

This study was carried out in paddy growing soils in the Northern Dry Zone of Sri Lanka. Samples were selected from the following districts (administrative units) namely: Jaffna, Kilinochchi, Vavuniya, Mannar and Mullaitivu. Major soil types in the study area that grow rice incudes; Low Humic Gley Soils (USDA Soil taxonomy order: Alfisols; Word reference base Major group: Gleysols), Calcic Red Yellow Latosol (USDA Soil taxonomy order: Oxisols; Word reference based Major group: Ferralsols), Solodized solonetz (USDA Soil taxonomy order: Alfisols; Word reference base Major group: Solonetz) and Alluvials (USDA Soil taxonomy order: Entisols; Word reference base Major group: Fluvisols) (Panabokke, 1996). The average mean temperature of the study area varies in between 25.0 to 27.5 ^{o}C while annual cumulative rainfall is reported in between 1000 to 1500 mm (Survey Department of Sri Lanka, 2007). Rainfall of the study area is seasonal and mainly governed by the North East Monsoon. Terrain of the study is characterized with undulating terrain which is common in the defined dry zone catena of Sri Lanka.

2.2. Soil sampling collation design

A purposive sampling design was adopted in this study considering the limited budget and accessibility. A brief description on the sampling design adopted in this study is described as follows. Twenty one (n = 21) main sampling sites were selected considering the percentages of land area under paddy in each administrative district (Table 1). These 21 main sites were selected depending on accessibility for the paddy sites. In order to determine the local variation of SOC, additional sampling site was selected for each main sampling site. This is similar to approach adopted by Dorji et al. (2014) where paired sampling was carried out for main sampling locations at approximately distance of 1 km which resulted total of 42 sampling sites. Finally, considering those resulted 42 sampling sites, another close range sample was taken at approximate distance of 100 m away similar to approach described by Karunaratne et al. (2014b) except in one location. In total, 83 samples were selected to represent the paddy growing soils of northern Sri Lanka (Table 1). Soil samples were collected at two depth intervals namely 0 to 0.15 m and 0.15 to 0.30 m respectively targeting SOC content of top soils. Sampling was carried out in January 2015. Figure 1 depicts the distribution of sampling locations and paddy fields within the study area.

Tables 1 here

Figure 1 here

2.3. Preparation of soil sample and analysis

All visible organic debris, stones and plant roots were removed and large soil aggregates were crushed prior to sieving of samples. Then samples were sieved using a 2 mm mesh sieve. Prepared samples were analysed for SOC, soil moisture content, soil pH (1:2 soil: water suspension) and conductivity. For the current study we concentrated on the analytical values of SOC contents. For the SOC analysis, samples were air dried and ground to a powder to less than 0.15 mm and determination of TOC was carried out using 'wet' oxidation by acidified dichromate of organic carbon (Baker, 1976). In this study TOC is considered as equal to SOC.

2.4. Preparation of environmental covariates

Table 2 summarises different environmental covariates used in the development of SOC spatial models in this study. These environmental covariates can be grouped into four main categories; topographic, climatic, biological, and spatial which represent some elements of the *SCORPAN* prototype DSM model described by McBratney et al. (2003). In preparation of the environmental covariates all the covariates were re-projected to Sri Lankan national grid (projected coordinate system). Prior to spatial modeling all the environmental covariates were re-sampled to a common grid which is equivalent to 100 m spatial resolution.

The digital elevation model (DEM) for the study area was obtained from the NASA Shuttle Radar Topography Mission. The DEM was processed to ensure that there were no artificial sinks, using SAGA version 2.0.6 software as described by Wang and Liu (2006) (www.saga-gis.org/en/index.html). A topographic wetness index (TWI) was created using the corrected DEM employing TWI module within the SAGA software. The TWI represent the spatial distribution of potential soil moisture. It is a secondary terrain attribute and widely used in DSM.

Climate data including annual cumulative rainfall, annual mean maximum temperature and annual mean minimum temperature data were obtained from the WorldClim climate database (http://www.worldclim.org/). Annual mean temperature was derived using annual mean maximum and annual mean minimum temperature layers.

Three Landsat 8 image tiles available from United States Geological Survey (USGS) covering the entire study area, acquired on 14th June 2014 (Path - 141, Row – 54), 05th June 2014 (Path - 142, Row – 53) and 21st June 2014 (Path - 142, Row – 54) were selected considering minimum impact on cloud cover which is a difficult factor to avoid in tropical environments. Acquired images were converted to top of atmospheric reflectance using the information available in the associated meta data files. Individual Landsat 8 bands namely: band 2 - blue (0.45 – 0.51 µm), band 3 - green (0.53 – 0.59 µm), band 4 – red (0.64 – 0.67 µm), band 5 – near infrared (0.85 – 0.88 µm), band 6 – short wave infrared 1 (1.57 – 1.65 µm), and band 7 - short wave infrared 2 (2.11 – 2.29 µm) were included as environmental covariates. Additionally, Landsat 8 band ratios namely band 4 / band 3, band 4 / band 7, band 6 / band 7 were derived.

Furthermore, normalized vegetation index (NDVI = [band 5 – band 4] / [band 5 + band 4]) was derived and included as an environmental covariate. Landsat image processing was carried out using Erdas Imagine software version 11. Additionally, X and Y coordinates were included as environmental covariates in the current modeling.

Table 2 here

2.5. Development of spatial soil prediction function with auto correlated error

The environmental covariates were intersected with the sampling sites and values were extracted. Ranges of values of extracted for the environmental covariates at sampling sites and their original range values are depicted in Table 2. The DSM modeling in this study was performed using LMMs. LMMs belong to hybrid geostatistical modeling class and more details on them and how they differ from popular regression-kriging approach can be found in Lark et al., 2006. The main difference between regression-kriging and LMM is that, in the latter estimation of parameters both fixed effect terms (i.e. regression coefficients) and random effect terms (i.e. variogram parameters for residuals) are done simultaneously rather than in two separate steps. The general form of LMM used for spatial prediction can be written as (after Lark et al., 2006);

$$\mathbf{z} = \mathbf{X\tau} + \mathbf{Zu} + \boldsymbol{\varepsilon} \qquad 1$$

where **z** is a vector of observed responses (*n*) (SOC content values for respective depth interval), **X** is an *n* x *p* design matrix that associates each of the *n* observations with a value of each *p* environmental covariates/fixed effect terms (See Table 2 and selected based on most the parsimonious model as described below), and **τ** is a vector that contains the *p* fixed effect terms coefficients (parameters that describe the relationship between the fixed effect terms, **X** and the response, **z)**. The vector **u** contains *q* random effect terms, realisations of a variable *u*, that are associated with the *n* observations by the *n* x *q* design matrix **Z**. The *u* is spatially dependent random variable and the term **ε** is a vector of independent random errors. These terms are independent of each other and; contain random errors which are spatially correlated such that;

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 \xi \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix}\right) \qquad 2$$

where **G** is the correlation matrix where the correlation depends only on the relative locations of the observations, **I** is the identity matrix, σ^2 is the variance of the independent error and ξ is the ratio of the variance of **u** to σ^2 (Lark et al., 2006). Here $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ refers to the vector of means of $\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\varepsilon} \end{bmatrix}$. Also, $\begin{bmatrix} \sigma^2 \xi \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I} \end{bmatrix}$ refers to the variance-covariance matrix. So the variance-covariance matrix of **u** is $\mathrm{var}(\mathbf{u}) = \sigma^2 \xi \mathbf{G}$, the variance-covariance matrix for $\boldsymbol{\varepsilon}$ is $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}$, *i.e.* independent random errors, and cov(**u**, **ε**) = **0**, means a matrix of all zeros, implying that the spatial effects **u** are independent of the random errors $\boldsymbol{\varepsilon}$.

The most parsimonious spatial LMM model for two depth intervals was selected using backward elimination similar to Karunaratne et al. (2014,ab). Backward elimination stopped when all P-values were equal or less than 0.1.

2.6. Assessment of fitted model quality

Assessment of the model quality was performed using leave-one-outcross-validation (LOOCV) where each observation was removed from the data set and the SOC content at that location was predicted using the remaining observations. Four measures of model quality were calculated:

Mean error (ME)

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) \qquad 3$$

where y_i is the measured SOC content while $\hat{y}_i$ is the predicted SOC content from LOOCV.

Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \qquad 4$$

Mean standardized squared deviation ratio (MSDR)

$$MSDR = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{\sigma_i^{\,2}} \qquad 5$$

where $\sigma_i^{\,2}$ is the prediction variance from empirical best linear unbiased predictor (E-BLUP).

Lin's concordance correlation coefficient (LCC)

$$\rho_c = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2} \qquad 6$$

where; ρ_c is the estimated LCC, μ_x and μ_y are the means for the measured and predicted SOC contents and σ_x^2 and σ_y^2 are the corresponding variances of measured and predicted SOC contents and ρ is the Pearson correlation coefficient between the measured and predicted SOC contents.

The ME gives an estimate of the bias in the predictions while the RMSE indicates the accuracy of the predictions. The LCC indicates how well a plot of the measured SOC contents values *vs.* predicted SOC content values follows a 45 degree line (Lin, 1989). A LCC value close to 1 indicates that the measured *vs.* predicted values closely match each other.

The bootstrapping approach outlined by Lark (2002) was used to estimate the 95% confidence interval (CI) around the median and mean MSDR. A mean value of 1.00 and median of 0.455 indicates that the prediction variance accurately reflects the actual errors and in that case our models for spatial variation of SOC contents is considered as acceptable (Lark, 2002).

Fitting of the LMMs for two depth intervals, cross validation of derived optimum spatial models including bootstrapping were carried out using geoR package (Ribeiro Jr and Diggle, 2001) using a customized routine programmed in R statistical programming language (R Development Core Team, 2014).

2.7. Predictions: mapping across the study area

The optimal models identified in the model fitting process were used to predict SOC contents across the study area. Estimated fixed effect terms and variance parameters from the geoR package were extracted and included as suitable for model fitting process in gstat package. Finally, prediction was carried out using E-BLUP using the gstat package in R statistical programming language (Pebesma, 2004).

3. Results

3.1. Summary statistics for SOC contents

Summary statistics revealed that mean SOC content for the 0 to 0.15 m reported as 1.78 % (±0.78) while 1.03 % (±0.47) for the 0.15 to 0.30 m depth interval. SOC in the upper depth interval also had a higher standard deviation (Table 3). Reported skewness values for both depth intervals are less than one (1) which is proxy for normal distribution of the measured data (Webster and Oliver, 2007). Many SOC studies reported to have a skewed distributions and most instances negative skewed (Karunaratne et al., 2014b; Minasny et al., 2013).

Summary of SOC values with respect to major soil types are depicted in Table 4. Highest mean SOC for two depth intervals was recorded for Low Humic Gley Soils which are situated in low slopes and valley bottoms in the characteristics undulating terrain in the dry zone of Sri Lanka. These soils further characterize with poorly drained soils. Lowest mean SOC content for two depth intervals are reported for Alluvials soils.

Table 3 here

Table 4 here

3.2. Estimated fixed and random effect terms

Estimated fixed effect terms of the LMMs for the two depth intervals are depicted in Table 5. Based on the results; main environmental covariates that were significant and explains the variability of SOC across the landscape for the 0 to 0.15 m depth interval SOC were elevation, Landsat 8 bands 2,4,7, X coordinate, Landsat 8 band ratios4/3 and 4/7. For 0.15 to 0.30 m depth interval main environmental covariates that were significant and explains the SOC variability across the landscape were Landsat 8 bands 2,5,6, Y coordinate, Landsat 8 band ratios

4/3 and 4/7 and NDVI. Landsat band 2 and band ratios; 4/3 and 4/7 were identified as common environmental covariates that were related to of both depth intervals.

Estimated spatially correlated random effect terms of the LMM residuals of the two depth intervals are depicted in the Table 6. Results revealed that for 0 to 0.15 m depth interval residuals are spatially auto correlated to up to 10000 m while only 5000 m for the depth interval 0.15 to 0.30 m depth interval. The Nugget to Sill ratio was 41.21 % and 61.00 % for 0 to 0.15 m and 0.15 to 0.30 m depth intervals respectively.

Table 5 here

Table 6 here

3.3. Model validation

Model validation results based on LOOCV are depicted in Table 7. Both the ME and RMSE values for two depth intervals were close to zero (0) (Table 7). The reported CCC value for upper depth interval SOC content was higher than the lower depth interval (Table 7). The graphical summary of the LOOCV results for each depth interval is depicted in the Figure 2.

Table 7 here

Figure 2 here

The results of the bootstrap analysis are summarized in Table 8. Results revealed the estimated median MSDR value for both depth intervals were quite close to 0.455. However, the mean MSDR value of 0.98 and 0.99 reported for depth interval 0 to 0.15 and 0.15 to 0.30 m respectively. Furthermore, median MSDR values for the lower 5 % and upper 95 % confidence interval for both depth interval reported similar values (Table 8).

Table 8 here

3.4. Mapped distribution of SOC across the study region

Figure 3 shows the spatial distribution of SOC content (Figure 3a,b) and the associated uncertainty in terms of prediction variance (Figure 3c,d). It is evident that the northern part of the study area reported lower SOC content for both depth intervals (Figure 3a,b). Additionally, variance maps provide estimates of the uncertainty of the created maps. It is therefore, essential to report this prediction uncertainty since these maps are later used for policy decision making in country's agricultural context. Predictions were carried out across a land extent of 1169.95 km^2 covering all the paddy growing regions identified in the study area based on 1:50,000 map sheets produced by Survey Department of Sri Lanka.

Figure 3 here

4. Discussion

4.1. Comparison of SOC content values with paddy growing soils vs. dryland cropping systems in the study area

Any vegetation will possess the ability to sequester atmospheric carbon and store it in their above ground biomass. This biomass will eventually fall onto the soil and be sequestered into soil carbon primarily through the continuous process of microorganisms activities. Rates of decomposition of SOM in low land conditions are considered to be slower under continuous water logging resulting a relatively greater accumulation of SOC in soil (Kölbl et al., 2014). A study conducted by Ratnayake et al., (2014) in the Jaffna district of Sri Lanka which is situated within the current study area focused on upland land uses with annual crops (such as Chili and Onion crop rotation) and perennial crops (such as banana) showed that the average SOC contents are considerably lower (<0.4%) in comparison to SOC contents reported in our current study (Table 3). Additionally, these low levels of SOC content are explained by the amount of crop residues returned to the soil, excessive tillage and imbalance in fertilizer use (Ratnayake et al., 2014). Therefore, compared with the upland cropping systems, rice fields situated in the low land landscape accumulated more SOC which exhibits greater potential to store SOC in terms of carbon sequestration.

4.2 Sampling, sampling density and its representativeness of environmental covariates

Soil sampling surveys are costly and time consuming. Therefore, it is important to allocate soil samples in an optimal manner especially if it is intended to collect data for mapping purpose. Due to the associated cost of sampling and laboratory analysis it is common for DSM studies to have low sampling densities especially in developing counties where soil information is rarely collected across the landscape. In this study sampling density was 14.1 km^2 for one sample.

The sampling scheme adopted in this current study was designed to capture the inherent variation of SOC, specifically the short range variation based on the work carried out by Dorji et al. (2014) and Karunaratne et al. (2014a,b), whilst considering the limited budget and accessibility of the sampling sites. In this study the main sampling locations were selected without considering the inherent variation of the environmental covariates. Therefore, the adoptation of sampling methods such as conditional latin hypercube sampling (cLHS) will enable the allocation of a limited number of sampling sites optimally considering distribution of respective environmental covariates (Minasny and McBratney, 2006). The cLHS algorithm select samples that form a Latin hypercube in the feature space created using environmental covariates (Minasny and McBratney, 2006). This enables covariates at sampling locations to have a similar distribution to the original environmental covariates used in the spatial modeling process. Comparison of summary statistics values of the environmental covariates and those extracted environmental covariate values at sampling locations were closely match with each other in case of their ranges (see Table 2). However, application of algorithms such as cLHS in low sampling density DSM mapping will be useful in future applications. Previous research has successfully applied cLHS for DSM of SOC (Dorji et al., 2014; Kidd et al., 2015). In the work presented by Dorji et al. (2014), a similar sampling method as applied in this current study was adopted, where, the main sampling sites were selected based on cLHS and additional samples within a close range to the main sample site (approximately 1 – 2 km) were also chosen. Additionally, recent modifications performed on cLHS sampling algorithm by Clifford et al. (2014) and Roudier et al. (2012) made possible to incorporate cost constrains and other factors in designing spatially optimum sampling schemes which might be useful in DSM studies in less data rich counties like Sri Lanka.

4.3. Spatial modelling of soil organic carbon and model accuracy

It is evident that environmental covariates that explains SOC contents across the landscape for the two depth interval are different (Table 5). Many SOC studies investigating multiple depth intervals based on *GlobalSoilMap* specifications (Arrouays et al.,, 2014) have reported that environmental covariates that explains the SOC are different depending on the depth interval considered from soil surface (Adhikari et al., 2014; Dorji et al., 2014). In the current study, environmental covariates that explains the SOC contents across the landscape were represented under three main categories; as relief (elevation and SAGA - WI), organism (Landsat 8 bands, ratios and NDVI) and spatial (X and Y coordinates). Both depth intervals reported that either of the spatial coordinates were included as significant fixed effect terms in the most parsimonious LMM (Table 5). Similar to previous DSM research work, spatial coordinates were included in the DSM studies to represent the environmental covariates which are not included in the current modelling approach and were used as a proxy for environmental gradients (Bishop et al., 2015). In this instance the X and Y coordinates can present the environmental variation that is potentially caused by geology and soil types not included in the current modelling.

Landsat 8 bands 2, 4, 7 and 2, 5, 6 were identified as environmental covariates that drivers of SOC for 0 to 0.15 m and 0.15 to 0.30 m depth interval respectively. Landsat bands; 2, 4 (namely blue and red) are within the visible spectrum while 5, 6 and 7 (namely near infrared, short wave infrared 1 and short wave infrared 2) are within infrared and short wave near infrared spectrum regions of the electromagnetic spectrum. These Landsat reflectance data are good indicators of vegetation rather than soil properties (Bishop et al., 2015). Nevertheless, some of these Landsat bands namely red, near infrared and short wave infrared represents biological components and its state (e.g. actively growing vegetation or fallow) of the SCORPAN model outlined by McBratney et al. (2003). In this study we used single images. However, recent work by (Zhao et al., 2015) has highlighted the importance of considering time series images as environmental covariates for predicting SOM in low relief agricultural lands. In tropical countries sourcing cloud free or less optical remote sensing products such as Landsat satellite images from its archive is challenging. Zhao et al. (2015) states that in these environments selecting appropriate archived images based on (a) time of the year and (b) location. Using these data to derive for summary statistics (e.g. mean, median and standard deviation) are important when considered them as environmental covariates. This approach can be further extended by performing quantile analysis (e.g. 0.05, 0.50 and 0.95) on time series satellite data on either selected bands or derived vegetation indices and create innovative environmental covariates for DSM that will enable to capture more variation as opposed to using a single image as applied in the current study.

NDVI was identified as a significant environmental covariate for SOC content for 0.15 to 0.30 m depth interval. Landsat bands and NDVI are often used as a proxy to represent spatial variation of vegetation (Malone et al., 2009). In our study the target land use was paddy-growing areas, but there can be differences in vegetation within those considered areas. For instances, paddy fields in Northern Sri Lanka, is manly cultivated as seasonal crops in the *Maha* season based on North East monsoon and as a rain fed crop. In the dry season *i.e*. *Yala* season field crops are cultivated. Therefore, most of the paddy growing soils in Northern Sri Lanka cannot be considered as a paddy mono-culture, but are considered as paddy - field crop rotation farming system. Additionally, due to 30 years of civil war in Northern Sri Lanka, paddy field boundaries extracted from the 1:50,000 maps sheets produced by Survey Department of Sri Lanka, might have not been cultivated for a long time period and are potentially subjected to conditions of long fallow.

For both depth intervals Landsat 8 band ratios 4/3 and 4/7 were included as environmental covariates for the SOC in Northern paddy fields in Sri Lanka. Landsat band ratios have been reported as soil enhancement ratios and used to represent the carbonate radicals, ferrous iron, and hydroxyl radicals in soil (Saunders and Boettinger, 2006).

Cambardella et al. (1994) categorized spatial dependence based variogram parameters namely on the nugget/sill ratio as; high (0-25 %), moderate (25-75 %) and low (>75 %). The results of this study indicated that the spatial dependence of spatially auto-correlated residuals can be considered as moderate spatial dependence for the both considered depth intervals (See Table 6). In this case, variogram parameters were estimated for the residuals. It is well documents that SOC inherent large variation which can be correlated spatially from ~ 100 m to 3000 m (Hengl et al., 2004; Malone et al., 2009; McBratney and Pringle 1999).

Minasny et al. (2013) reported that many SOC mapping studies do not carry out any form of validation. The results of both LOOCV (Table 7) and bootstrap analysis (Table 8) revealed that the fitted spatial models adequately and accurately represent the SOC distribution across the landscapes. The results of the model validation are acceptable and provide a baseline for SOC modelling in the study region. Nevertheless, fully independent validation based on designed-based sampling schemes as outlined by Brus et al. (2011) will provide unbiased estimates about the prediction quality in future studies as shown in DSM application of clay mapping by Bishop et al (2015).

Interestingly the reported LCC value for the lower depth interval was lower than the upper depth interval (Table 7). This may be due to the fact that environmental covariates considered (Table 2) represent the variation of top depth interval rather than bottom depth interval more adequately. Additionally, the confidence interval for the LCC of the lower depth interval is much wider compared with the upper depth interval (Table 7). Bishop et al. (2015) reported that the environmental covariates such as gamma radiometric data are capable to penetrate to much lower depth intervals compared to optical remote sensing products such as Landsat images similar to the images used in the current study. The inclusion of gamma radiometric data may have improved the accuracy of the predictions of the lower depth interval. However, in developing counties like Sri Lanka, rich datasets like gamma radiometric are rarely available. Recently, Gray et al. (2015) outlined methods to derive cost effective environmental covariates such as using terrain attributes (*e.g.* topo – slope index) and the use of regenerated coarse scale geology maps (considering the silica contents). They demonstrated application of these newly derived environmental covariates in DSM of SOC and showed improvement of prediction capabilities which might be vital to explore in data limited DSM studies similar to current study.

Lark (2002) reported that it is better to calculate median instead of mean to overcome the effect of outliers in calculation of MSDR. In the current study, both the mean and median of the MSDR were calculated. Lark (2002) further reported that a mean value close to 1.00 and median value close to 0.455 reflects the prediction variance is an accurate representation of the actual errors. The estimated mean and median MSDR values in this study were close to the desired values stated by Lark (2002) (Table 8).

4.5. Uses of derived SOC digital soil maps

According to reviewed literature this is the first detailed study on DSM undertaken in Sri Lanka. Therefore, this study provides firsthand information for soil scientists and policy makers who are seeking fine resolution digital soil data. The outputs derived in this study can be coupled with different agricultural planning activities from regional and to national scales. At a policy level this information is the representation of the current soil health condition or status for paddy growing soils in the Northern Sri Lanka.

One of the major cost components of paddy production in Sri Lanka is the application of in-organic fertilizers. Therefore, in regional scale planning of agricultural production, information generated from this study can be used to direct paddy growers to incorporate organic matter to their lands especially in regions where there are low levels of SOC contents are reported.

Baldock et al. (2012) reported the possible use of the derived SOC content maps to be coupled with the design of spatially optimum sampling strategies. Their study reported that DSM products could be clustered using statistical algorithms such as K-means clustering. This approach would enable the identification of homogenous areas based on SOC content map across the landscapes as specific regions or strata. Additionally, these regions or strata could be used to carry out future sampling of SOC.

Currently there is no mechanism to report national carbon accounting in Sri Lanka. Therefore, this information will lay a foundation for such a system. For example, derived DSM can be coupled with process models (*e.g.* RothC carbon model) to predict the space - time variation of SOC which will contribute towards below ground carbon accounting. In Australia, the national carbon accounting system utilises the RothC carbon model to simulate the changes of SOC and assess the below ground carbon (Richards, 2001). Recently, Karunaratne et al. (2015) proposed the “Space-time observation system for SOC” to account for changes in SOC. This novel approach incorporates DSM, digital climatic layers and satellite derived products such as NPP (as a proxy for changes in carbon inputs to ecosystem). Therefore, derived digital SOC content maps can be incorporated to simulate changes of SOC through space and time. The development of new approaches which utilise satellite derived data are important as soil sampling across large regions scale is expensive and time consuming.

5. Conclusions

This study predicted the spatial distribution of SOC contents at two soil depth intervals (*i.e.* 0 to 0.15 and 0.15 – 0.30 m) across all the paddy growing soils in Northern Sri Lanka. Hybrid geostatistical modelling *via* LMMs were used to identify the environmental covariates that were significant and contribute towards spatial variation of SOC across the landscapes. The validation of the results showed that the fitted spatial models were acceptable. It was revealed that, significant environmental covariates of SOC were different in the two considered depth intervals.

This study provides: a) firsthand information on current status of SOC in northern paddy growing soils in Sri Lanka; b) useful information to optimize sampling of SOC in future applications; and c) provide information vital to establishment of national carbon accounting system in future.

Acknowledgement

Authors are grateful to Charitha Adikari Arachchi (Western Sydney University, Australia) for her support in data modelling. Thanks are also due to Ms. Kumuduni Karunaratne and Mr. A. Pathirana for assistance in sampling and chemical analysis.

References

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O. A., Sanchez, P. A., Thompson, J.A., Zhang, G., 2014. Chapter Three-*GlobalSoilMap*: Toward a Fine-Resolution Global Grid of Soil Properties. Advances in agronomy, 125, 93-134.

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark, PloS one, 9(8), p.e105519.

Baker, K., 1976. The determination of organic carbon in soil using a probe-colorimeter. Laboratory practice 25, 82-83.

Baldock, J., Wheeler, I., McKenzie, N., McBrateny, A., 2012. Soils and climate change: potential impacts on carbon stocks and greenhouse gas emissions, and future research for Australian agriculture. Crop and Pasture Science 63(3), 269-283.

Baldock, J.A., Broos, K., 2011. Soil organic matter. In: p. Huang (Ed.), Handbook of soil science. CRC Press/Taylor Francis Group, LLC: Boca Raton, FL, USA, pp. 11.11–11.52.

Bishop, T., Horta, A., Karunaratne, S., 2015. Validation of digital soil maps at different spatial supports. Geoderma 241, 238-249.

Brus, D.J., Kempen, B. and Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science, 62(3), 394-407.

Cambardella, C., Moorman, T., Parkin, T., Karlen, D., Novak, J., Turco, R., Konopka, A., 1994. Field-scale variability of soil properties in central Iowa soils. Soil science society of America journal 58(5), 1501-1511.

Central Bank of Sri Lanka, 2014. Annual Report. Central Bank of Sri Lanka, Colombo.

Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. Computers & Geosciences, 67, 62-68.

De Costa, W., Suranga, H., 2012. Estimation of carbon stocks in the forest plantations of Sri Lanka. Journal of National Science Foundation of Sri Lanka 40, 9-41.

Dorji, T., Odeh, I.O., Field, D.J., Baillie, I.C., 2014. Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. Forest Ecology and Management 318, 91-102.

Gray, J.M., Bishop, T.F. and Yang, X., 2015. Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia. Soil Research, *53*(1), 24-42.

Hengl, T., Heuvelink, G.B., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma,120(1), 75-93.

Karunaratne, S.B., Bishop, T.F.A., Baldock, J.A., Odeh, I.O.A., 2014a. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219–220, 14-23.

Karunaratne, S.B., Bishop, T.F.A., Odeh, I.O.A., Baldock, J.A., Marchant, B.P., 2014b. Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. Soil Research 52(4), 349-365.

Karunaratne, S.B., Bishop, T.F.A., Lessels, J.S. Baldock, J.A., Odeh, I.O.A., (2015). A space-time observation system for soil organic carbon. Soil Research, 53(6), 647-661.

Kidd, D., Malone, B., McBratney, A., Minasny, B. and Webb, M., 2015. Operational sampling challenges to digital soil mapping in Tasmania, Australia. Geoderma Regional, 4, 1-10.

Kölbl, A., Schad, P., Jahn, R., Amelung, W., Bannert, A., Cao, Z., Fiedler, S., Kalbitz, K., Lehndorff, E., Müller-Niggemann, C., 2014. Accelerated soil formation due to paddy management on marshlands (Zhejiang Province, China). Geoderma 228, 67-89.

Lal, R., 2002. Soil carbon sequestration in China through agricultural intensification, and restoration of degraded and desertified ecosystems. Land Degradation & Development 13(6), 469-478.

Lark, R.M., 2002. Robust estimation of the pseudo cross-variogram for cokriging soil properties. European Journal of Soil Science 53(2), 253-270.

Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. European Journal of Soil Science 57(6), 787-799.

Lin, L.I.K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45(1), 255-268.

Malone, B.P., McBratney, A., Minasny, B., Laslett, G., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154(1), 138-152.

Mattsson, E., Ostwald, M., Nissanka, S., Holmer, B., Palm, M., 2009. Recovery and protection of coastal ecosystems after tsunami event and potential for participatory forestry CDM– Examples from Sri Lanka. Ocean & Coastal Management 52(1), 1-9.

Mattsson, E., Ostwald, M., Nissanka, S., Marambe, B., 2013. Homegardens as a multi-functional land-use strategy in Sri Lanka with focus on carbon sequestration. Ambio 42(7), 892-902.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117(1–2), 3-52.

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers and Geosciences, 32:1378-1388.

McBratney, A.B., Pringle, M.J., 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture, 1(2), 125-152.

Minasny, B., McBratney, A.B., Hong, S.Y., Sulaeman, Y., Kim, M.S., Zhang, Y.S., Kim, Y.H., Han, K.H., 2012. Continuous rice cropping has been sequestering carbon in soils in Java and South Korea for the past 30 years. Global Biogeochemical Cycles 26(3), n/a-n/a.

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Adv. Agron 118, 1-47.

Nayak, A., Gangwar, B., Shukla, A.K., Mazumdar, S.P., Kumar, A., Raja, R., Kumar, A., Kumar, V., Rai, P., Mohan, U., 2012. Long-term effect of different integrated nutrient management on soil organic carbon and its fractions and sustainability of rice–wheat system in Indo Gangetic Plains of India. Field Crops Research 127, 129-139.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67(3–4), 215-226.

Pan, G., Li, L., Wu, L., Zhang, X., 2004. Storage and sequestration potential of topsoil organic carbon in China's paddy soils. Global Change Biology 10(1), 79-92.

Panabokke, C., 1996. Soils and agro-ecological environments of Sri Lanka.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30, 683-691.

Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation—An example from Scotland. Geoderma 232, 284-299.

Ranasinghe, C., Thimothias, K., 2012. Estimation of carbon sequestration potential in coconut plantations under different agro-ecological regions and land suitability classes. Journal of the National Science Foundation of Sri Lanka 40(1), 77-93.

Ratnayake, R., Kugendren, T., Gnanavelrajah, N., 2014. Changes in soil carbon stocks under different agricultural management practices in North Sri Lanka. Journal of the National Science Foundation of Sri Lanka 42(1), 37-44.

Ribeiro Jr, P.J., Diggle, P.J., 2001. geoR: A package for geostatistical analysis. R news 1(2), 14-18.

Richards, G.P., 2001. The FullCAM carbon accounting model: development, calibration and implementation for the national carbon accounting system. Australian Greenhouse Office, National Carbon Accounting System Technical Report No. 28

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorportaing operational constraints. In: Digital Soil Assessments and Beyond. Proceedings of the 5th Golobal Workshop on Digital Soil Mapping, Sydney, Australia.

Sahrawat, K.L., 2003. Organic matter accumulation in submerged soils. Advances in Agronomy 81, 169-201.

Samuel-Rosa, A., Heuvelink, G., Vasques, G., Anjos, L., 2015. Do more detailed environmental covariates deliver more accurate soil maps? Geoderma 243, 214-227.

Saunders, A., Boettinger, J., 2006. Incorporating classification trees into a pedogenic understanding raster classification methodology, Green River Basin, Wyoming, USA. In:

P. Lagacherie, A.B. McBratney, M. Voltz (Eds.), Developments in Soil Science, Amsterdam, pp. 389-620.

Sumfleth, K., Duttmann, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecological Indicators 8(5), 485-501.

Survey Department of Sri Lanka, 2007. National Atlas for Sri Lanka, Colombo, Sri Lanka.

R Development Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

Wang, L., Liu, H., 2006. An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. International Journal of Geographical Information Science 20(2), 193-213.

Webster, R., Oliver, M.A., 2007. Geostatistics for environmental scientists. John Wiley & Sons.

Wu, J., 2011. Carbon accumulation in paddy ecosystems in subtropical China: evidence from landscape studies. European Journal of Soil Science 62(1), 29-34.

Zhao, M.S., Rossiter, D.G., Li, D.C., Zhao, Y.G., Liu, F., Zhang, G.L., 2014. Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index. Ecological Indicators,*39*, 120-133.

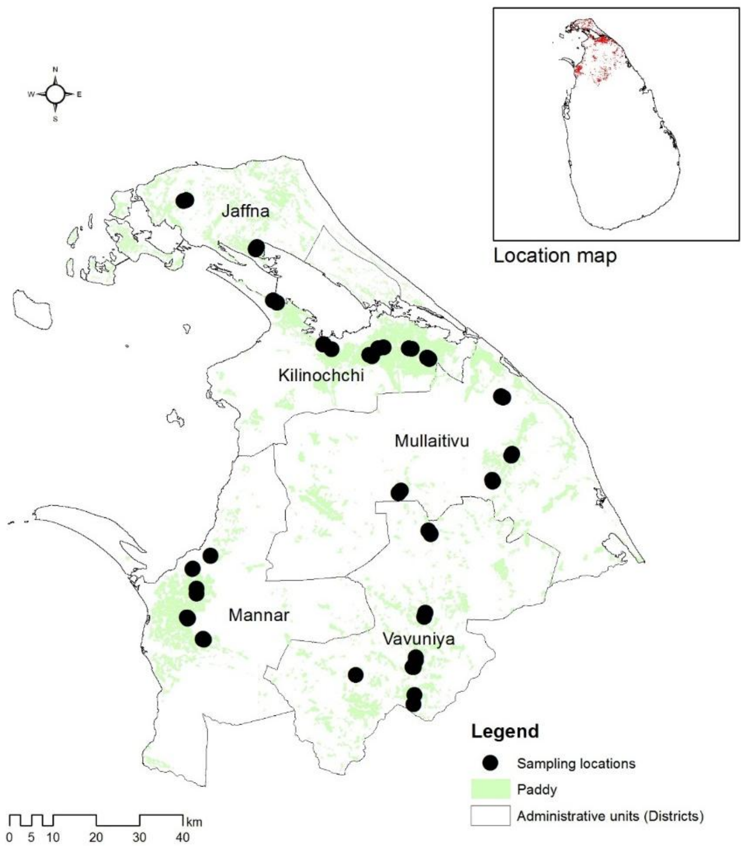


Figure 1. Study area and sampling locations

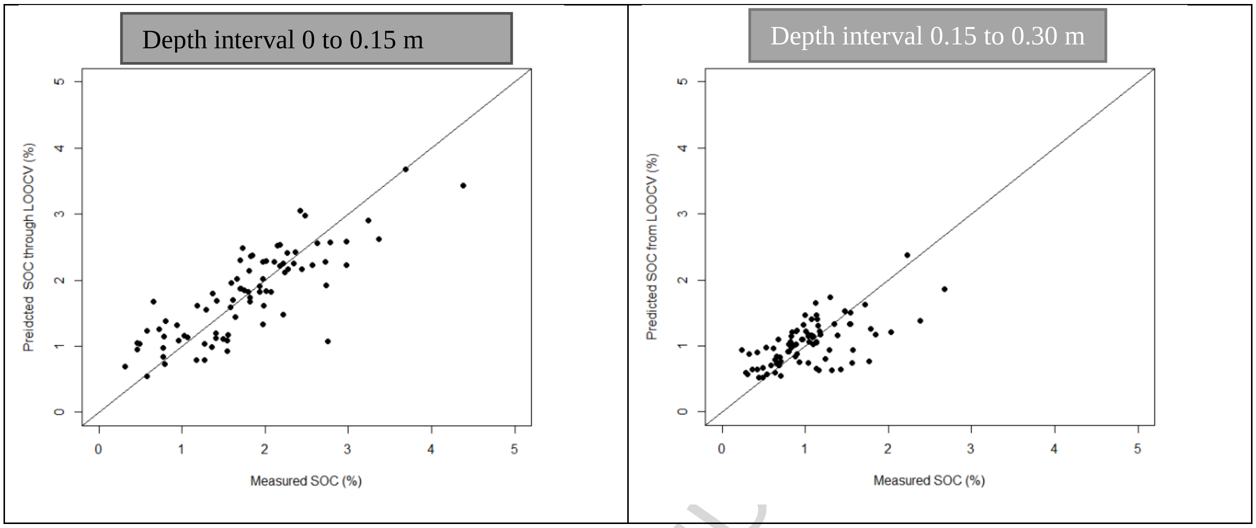


Figure 2. Relationship between measured SOC content *vs*. predicted SOC content *via* LOOCV

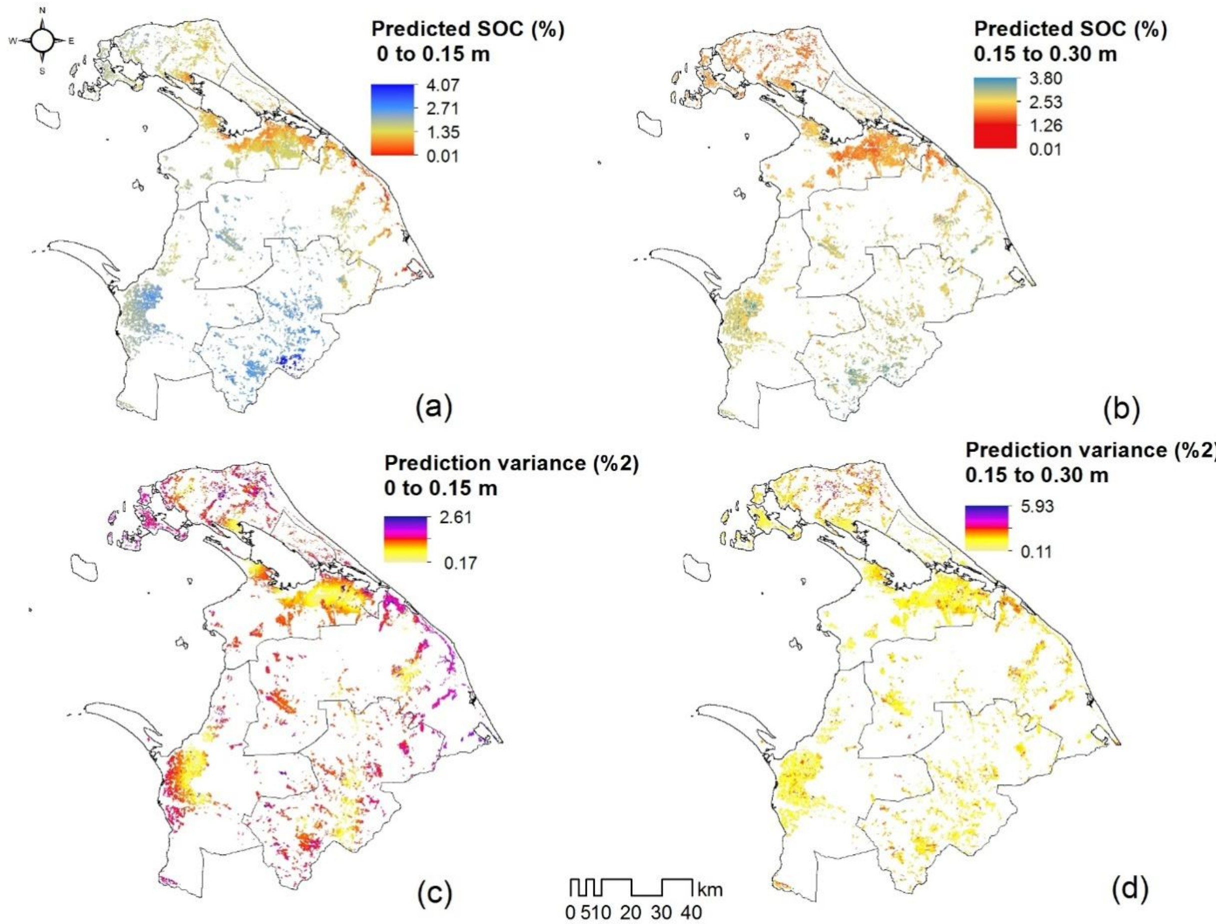


Figure 3. Predicted SOC contents and prediction variance maps for two depth intervals

Table 1.Distribution of main sampling sites and close range sampling sites

District	Number of main sampling sites	Number of samples at approximately 1 km distance	Number of samples at approximately 100m distance
Jaffna	2	4	8
Kilinochchi	6	12	24
Vavuniya	5	10	20
Mannar	4	8	15
Mullaitivu	4	8	16

Table 2. Summary of the environmental covariates used in the study: ranges of values in environmental covariates and coincided sampling locations

Environmental Covariate	Original spatial resolution (m)	*SCORPAN* factor	Value range: prediction grid	Value range: sample dataset
DEM	90	R	0 - 129	1 - 118
SAGA WI	90	R	0 – 22.26	8.34 – 19.68
Annual cumulative rainfall	1000	C	1050 -1544	1198 - 1447
Annual mean temperature	1000	C	27.2 – 28.1	27.3 – 27.9
Landsat band 2	30	O	0.09 – 0.70	0.10 – 0.35
Landsat band 3	30	O	0.06 – 0.96	0.09 – 0.34
Landsat band 4	30	O	0.04 – 0.75	0.06 – 0.35
Landsat band 5	30	O	0.06 – 0.87	0.17 – 0.52
Landsat band 6	30	O	0.03 – 0.76	0.12 – 0.41
Landsat band 7	30	O	0.02 – 0.54	0.07 – 0.30
Landsat band ratio 4/3	30	O	0.62 – 1.47	0.70 – 1.16
Landsat band ratio 4/7	30	O	0.51 - 4.86	0.64 – 1.35
Landsat band ratio 6/7	30	O	1.04 – 2.73	1.32 – 2.31
NDVI	30	O	-0.23 – 0.77	0.18 – 0.68
X coordinate	Unit less	S	89516 - 220216	115662 - 190704
Y coordinate	Unit less	S	371318 - 513218	387110- 503452

Table 3.Summary statistics for SOC content for two depth interval

Depth interval (m)	Min (%)	Q 1 (%)	Median (%)	Mean (%)	Q 3 (%)	Max (%)	SD	Skewness
0 to 0.15	0.31	1.28	1.80	1.78	2.21	4.38	0.78	0.48
0.15 to 0.30	0.24	0.69	1.00	1.03	1.21	2.68	0.47	0.96

Note: Min: Minimum; Max: Maximum, SD: Standard deviation; Q: Quantile

Table 4. Summary statistics of SOC with respect to soil types

Depth Interval (m)	Local Major Soil Type	USDA Soil taxonomy order	n	Min (%)	Q 1 (%)	Median (%)	Mean (%)	Q 3 (%)	Max (%)	SD	Skewness
0-0.15											
	Alluvials	Entisols	16	0.58	0.90	1.32	1.30	1.50	2.75	0.58	0.92
	Calcic Red Yellow Latosol	Oxisols	8	0.80	1.45	1.64	1.70	1.91	2.73	0.60	0.24
	Low Humic Gley Soils	Alfisols	32	1.36	1.74	1.97	2.18	2.38	4.38	0.69	1.51
	Solodized solonetz	Alfisols	27	0.31	0.79	1.93	1.61	2.24	2.97	0.83	-0.09
0.15-0.30											
	Alluvials	Entisols	16	0.29	0.44	0.57	0.73	0.74	1.76	0.46	1.16
	Calcic Red Yellow Latosol	Oxisols	8	0.36	0.63	0.74	0.77	0.88	1.32	0.28	0.51
	Low Humic Gley Soils	Alfisols	32	0.64	1.01	1.14	1.25	1.41	2.68	0.43	1.44
	Solodized solonetz	Alfisols	27	0.24	0.82	0.97	1.04	1.15	2.39	0.46	1.02

Note: Min: Minimum; Max: Maximum, SD: Standard deviation; Q: Quantile

Table 5. Estimated fixed effect terms for two depth intervals

Model Name	Estimate coefficient	Std. error	t value	P value	
SOC (0 to 0.15 m)					
(Intercept)	$1.01 * 10^{+01}$	$2.67 * 10^{+00}$	3.801	< 0.001	***
DEM	$1.60 * 10^{-02}$	$3.47 * 10^{-03}$	4.611	< 0.001	***
Landsat band 2	$-2.59 * 10^{+01}$	$1.38 * 10^{+01}$	-1.876	0.064	.
Landsat band 4	$3.75 * 10^{+01}$	$1.53 * 10^{+01}$	2.448	0.017	*
Landsat band 7	$-1.47 * 10^{+01}$	$6.60 * 10^{+00}$	-2.229	0.029	*
X coordinate	$-1.40 * 10^{-05}$	$5.26 * 10^{-06}$	-2.656	0.009	**
Landsat band ratio 4/ 3	$-3.81 * 10^{+00}$	$2.18 * 10^{+00}$	-1.748	0.084	.
Landsat band ratio 4/ 7	$-2.47 * 10^{+00}$	$1.17 * 10^{+00}$	-2.115	0.037	*
SOC (0.15 to 0.30 m)					
(Intercept)	$1.40 * 10^{+01}$	$3.47 * 10^{+00}$	4.018	< 0.001	***
Landsat band 2	$-2.18 * 10^{+01}$	$5.86 * 10^{+00}$	-3.716	< 0.001	***
Landsat band 5	$1.71 * 10^{+01}$	$3.97 * 10^{+00}$	4.315	< 0.001	***
Landsat band 6	$-6.82 * 10^{+00}$	$3.64 * 10^{+00}$	-1.872	0.065	.
Y coordinate	$-6.72 * 10^{-06}$	$2.78 * 10^{-06}$	-2.417	0.018	*
Landsat band ratio 4/ 3	$-4.77 * 10^{+00}$	$1.92 * 10^{+00}$	-2.489	0.015	*
Landsat band ratio 4/ 7	$-1.54 * 10^{+00}$	$8.00 * 10^{-01}$	-1.929	0.057	.
NDVI	$-1.09 * 10^{+01}$	$2.63 * 10^{+00}$	-4.127	< 0.001	***

Level of significance: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Table 6. Estimated random effect terms (variogram parameters) for spatially correlated residuals for two depth intervals

Depth interval (m)	Model	Nugget (%^{2})	Sill (%^{2})	Range (m)	Nugget/Sill*100
0 to 0.15	Spherical	0.136	0.330	10000	41.21
0.15 to 0.30	Spherical	0.097	0.159	5000	61.00

Table 7. Model quality assessment through leave one out cross validation

Depth interval (m)	ME	RMSE	CCC
0 to 0.15	-0.006	0.441	0.81 (0.72 to 0.87)
0.15 to 0.30	-0.004	0.367	0.60 (0.45 to 0.71)

Note: ME: Mean error; RMSE: Root mean squared error; CCC: Lin's concordance correlation

Table 8: Calculated confidence interval for MSDR (mean and median) for two depth intervals

Theta statistics	Confidence interval (quantile)	Depth 0 to 0.15 m	Depth 0.15 to 0.30 m
Median	0.05	0.291	0.291
	0.50	0.458	0.455
	0.95	0.699	0.689
Mean	0.05	0.742	0.739
	0.50	0.989	0.994
	0.95	1.312	1.293

Highlights of the study

- Derived detailed digital soil organic carbon (SOC) map covering paddy growing soils
- Assessed the accuracy of the fitted spatial prediction models
- Provided firsthand information on baseline SOC content
- First ever, detailed large extent digital soil map produce in Sri Lanka