Geoderma 209–210 (2013) 1–14

Contents lists available at SciVerse ScienceDirect

Geoderma

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

Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates

Laura Poggio [a,*], Alessandro Gimona [a], Mark J. Brewer [b]

[a] The James Hutton Institute, Craigiebuckler, AB15 8QH, Aberdeen, Scotland, UK
[b] BioSS, Craigiebuckler, AB15 8QH, Aberdeen, Scotland, UK

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form 27 March 2013
Accepted 26 May 2013
Available online xxxx

Keywords:
Digital Soil Mapping
Variables selection
Complex models
Stepwise
Variables tournament
Organic soils

ABSTRACT

Knowledge of soil properties with complete area coverage is needed for policy-making, land resource management, and monitoring environmental impacts. Remote sensing offers possibilities to support Digital Soil Mapping, especially in data-poor regions. The aim of this work was to test the potential of time-series of MODIS (Moderate Resolution Imaging Spectroradiometer) vegetation and drought indices to provide relevant information to model topsoil properties in a Boreal-Atlantic region (Scotland) focussing on differentiation between soils with high and soils with low organic matter contents. For each of the three considered years, 345 MODIS data sets were included in the exploratory analysis; 15 data products for 23 dates (bi-weekly) per year. Terrain parameters derived from Shuttle Radar Topography Mission were also included. A methodology was implemented to exploit fully the high number of covariates, to identify the band, index or product that best correlates with the soil property of interest. In particular the proposed approach i. relies on freely globally available data-sets; ii. uses statistical criteria to select the combination of covariates providing the highest predictive capability, among the data considered and available; iii. deals with both continuous (using Generalized Additive Models, GAMs) and multinomial categorical (using Random Trees) types of variables; iv. takes into account fully the spatial autocorrelation of the data; v. provides estimates of the spatial uncertainty for each pixel; and vi. is computationally efficient when compared with methods such as forward stepwise. The models fitted show a fairly good agreement with existing data sets, presenting a consistent spatial pattern. The use of MODIS data as covariates increased the predictive capabilities of GAMs using only terrain parameters. The misclassification error for organic matter classes was between 25 and 35%. The assessment provided of the spatial uncertainty of the modelled values can be used in further modelling and in the assessment of consequences of climate-change and trade-off in land use changes. This approach can contribute to improving our understanding and modelling of soil processes and function over large, and relatively sparsely sampled, areas of the world.

1. Introduction

Knowledge of soil properties is needed for policy-making, land resource management, and monitoring the environmental impact of development. For example, the water and carbon contents of soil have a major role in many environmental, pedogenic and geomorphological processes and they are fundamental for issues linked to climate change and food security (e.g. Lal, 2004). While soil survey data most often have point support, global and regional models that address climate change, land degradation and hydrological processes need soil input parameters with complete area coverage.

Remote sensing (RS) offers possibilities for derivation of soil properties using Digital Soil Mapping (DSM) to extend existing soil survey data sets (Mulder et al., 2011). RS data can be combined with soil, land use nd Digital Elevation Models (DEMs) to produce soil attributes maps following the scorpan approach presented in McBratney et al. (2003), especially in data-poor regions. RS has been used widely to model soils and their properties at different scales, from plot- to global-scale applications (see Mulder et al., 2011, for a recent review). The most commonly used sensors were Landsat, AVHRR (Advanced Very High Resolution Radiometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) and MODIS (Moderate Resolution Imaging Spectroradiometer). For example, soil moisture has been modelled using mainly radar-type sensors (Mulder et al., 2011) and the modelling has become easier with availability of the SMOS (Soil Ocean Salinity Mission) data; however its spatial resolution is too coarse and the algorithms used often require soil information (Expert Support Laboratories et al., 2011; Merlin et al., 2008). RS data have been also used to model soil organic carbon (e.g. Kheir et al., 2010), to detect peat areas (e.g. Harris and Bryant, 2009) and assess peat condition (e.g. Connolly et al., 2011; Dabrowska-Zielinska et al., 2009; Dimitrov et al., 2011), often with the support of extensive soil surveys. However, the potential of available time series of images from sensors such as MERIS (MEdium Resolution Imaging Spectrometer) and MODIS appears,

* Corresponding author. Tel.: +44 1224395270.
E-mail address: laura.poggio@hutton.ac.uk (L. Poggio).

0016-7061/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.05.029

Table 1
Summary of the modelled soil parameters and of the covariates included in the variables tournament (VT).

Abbreviation	Explanation
Soil	
OM	Organic matter (%)
ORG	Organic soil (presence/absence)
Elevation	
srtm_elev	Elevation
srtm_slp	Slope
srtm_asp	Aspect
MODIS	
EVI	Enhanced Vegetation Index
LAI	Leaf Area Index
LSTd	Land Surface Temperature during day
LSTn	Land Surface Temperature during night
NDDI	Normalised Difference Drought Index (at ρ_{SWIR} 1240, 1640, 2130)
NDWI	Normalised Difference Water Index (at ρ_{SWIR} 1240,1640,2130)
NMDI	Normalised Multi-band Drought Index
SAVI	Soil Adjusted Vegetation Index

Note: the suffix for date is yearDOY, where DOY is the Day Of the Year from the beginning of the year.

until now, to be under-exploited. This paper will concentrate on the use of MODIS time series.

MODIS global products provide valuable information for many environmental applications, especially in ecological and hydrological fields. The large number of available bands allows the calculation of indices to characterise different spectral responses. The use of MODIS was suggested for supporting DSM in areas with limited availability of data (Minasny et al., 2008). Various combinations of MODIS bands have been used to model different scorpan factors, such as definition of landscape (Le Du-Blayo et al., 2008), landforms (Ballantine et al., 2005), energy balance (Kalma et al., 2008) and, above all, vegetation status and patterns (Xie et al., 2008). Some MODIS products were also used to model soil parameters, such as carbon content (e.g. Takata et al., 2007), and peat condition and properties (Connolly et al., 2011; Dimitrov et al., 2011; Harris and Bryant, 2009).

The use of MODIS data has been limited, in most cases, to a few products and bands. The amount of information provided by the

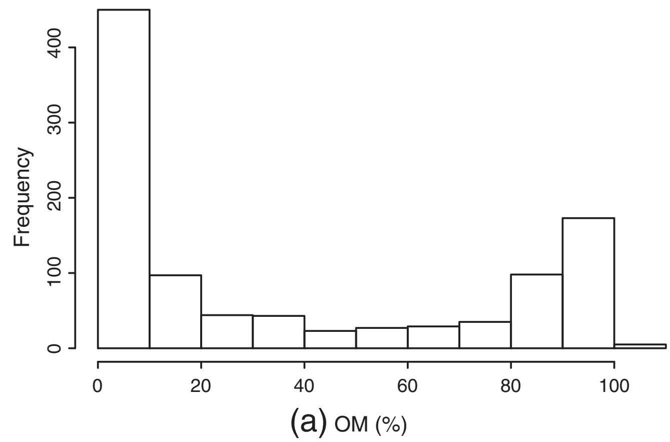

Fig. 2. Histograms of soil property values at 955 profile locations: 573 for mineral soils and 382 for organic soils.

sensor is therefore under-utilised, while there is scope for the use of the multiple bands, dates and indices available. Given the relatively high temporal resolution of data the choice of band, index or product that best correlates with the properties of interest is not straightforward. In this paper we propose a statistical approach to deal with such choice relying on existing well-established methods.

Different statistical methods have been proposed for fitting the empirical quantitative function linking the soil information to the scorpan approach factors, while taking into account the spatial structure of the data (McBratney et al., 2003). Regression kriging (Hengl et al., 2004) extends the methods of kriging and co-kriging and overcomes some limitations of the latter, namely by use of multiple covariates at the same time and having reduced influence of the second order stationarity assumption. Regression kriging has been further extended by the use of Generalized Additive Models (GAMs; Hastie and Tibshirani, 1990; Wood, 2004) in McBratney et al. (2000) and Poggio et al. (2010b), where estimation of uncertainty was also included.

When multinomial data are modelled, advanced non-parametric methods, such as CART (Classification and Regression Tree), can be used. CARTs have been used widely to estimate soil properties and,

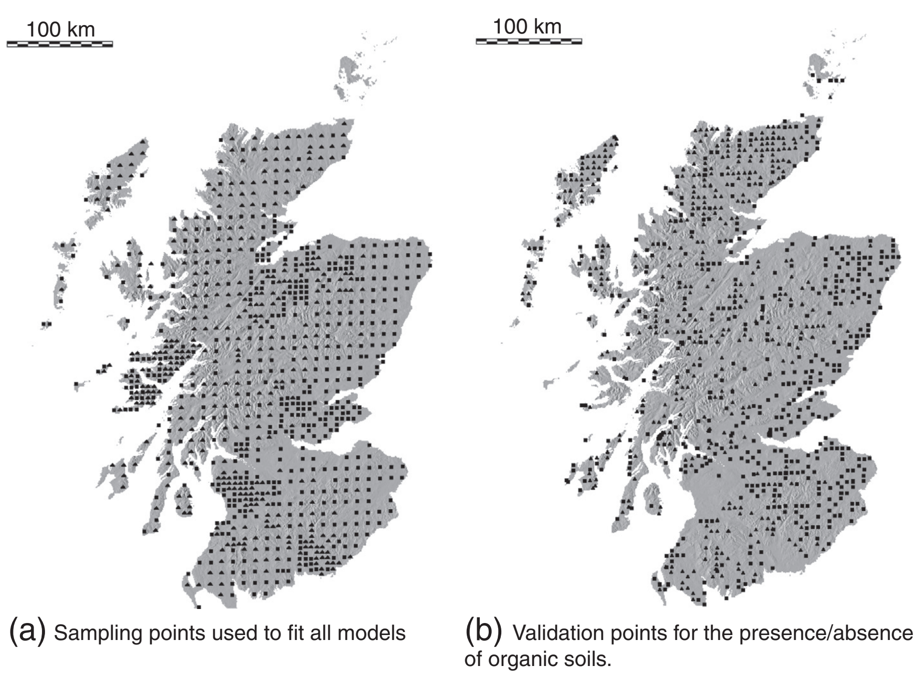

Fig. 1. Sampling and validation points from NSIS: triangle for organic soils, square for mineral soils.