Geoderma xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Geoderma

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

Combining measured sites, soilscapes map and soil sensing for mapping soil properties of a region

Emily Walker[a], Pascal Monestiez[a], Cécile Gomez[c], Philippe Lagacherie[b]

[a] BioSP, INRA, Avignon 84000, France
[b] INRA Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), Campus de la Gaillarde, 2 place Viala, Montpellier 34060, France
[c] IRD Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), Campus de la Gaillarde, 2 place Viala, Montpellier 34060, France

A R T I C L E I N F O

Article history:
Received 18 February 2016
Received in revised form 8 December 2016
Accepted 13 December 2016
Available online xxxx

Keywords:
Digital soil mapping
Remote sensing
Hyperspectral data
Kriging
Cross validation
Soil map
Soil properties

A B S T R A C T

The limited availability of soil information has been recognized as a main limiting factor in digital soil mapping (DSM) studies. It is therefore important to optimize the joint use of the three sources of soil data that can be used as inputs of DSM models, namely spatial sets of measured sites, soil maps and soil sensing products.

In this paper, we propose to combine these three inputs, through a cokriging with a categorical external drift (CKCED). This new interpolation technique was applied for mapping seven soil properties over a 24.6 km^2 area located in the vineyard plain of Languedoc (Southern France), using an hyperspectral imagery product as example of a soil sensing data. Cross-validation results of CKCED were compared with those of five spatial and non-spatial techniques using one of these inputs or a combination of two of them.

The results obtained in the La Peyne Catchment showed i) the utility of soil map and hyperspectral imagery products as auxiliary data for improving soil property predictions ii) the greater added-value of the latter against the former in most situations and iii) the feasibility and the interest of CKCED in a limited number of soil properties and data configurations. Testing CKCED in case study with soil maps of better quality and soil sensing techniques covering more area and depths should be necessary to better evaluate the benefits of this new technique.

1. Introduction

Given the relative lack of, and the huge demand for, quantitative spatial soil information to be used in environmental managing and modelling, digital soil mapping (DSM) has been proposed as an alternative to the classical soil surveys for the quantitative mapping of soil properties over regions at intermediate (20–200 m) spatial resolutions (McBratney et al., 2003). McBratney et al. (2003) proposed the equation S = f(s,c,o,r,p,a,n) for summarizing the general principle of DSM. According to this equation, a soil property (S) can be predicted by a spatial inference function (f) using, as input, the existing soil information (s), the spatial covariates that map the different factors of soil formation early defined by Jenny (1941) (c,o,r,p,a,) and the geographical location (n) that can highlight any spatial trends missed by the other covariates.

It has been early stressed that the limited availability of the soil information (the s component) was a severe limiting factor in DSM applications (Lagacherie, 2008). Up to now, most of the soil information used as input in DSM for mapping soil properties has been either soil maps or spatial sampling of sites with measured soil properties. When available under the form of soil databases (Rossiter, 2004), the former may provide estimates of soil properties over larger areas with however limited spatial resolutions and accuracy (Marsman and Gruijter, 1986; Leenhardt et al., 1994; Odgers et al., 2012). Pedometricians have developed a large range of algorithms for exploiting spatial sampling of sites for mapping soil properties, using sites with measured soil properties combined with spatial covariates (Oliver and Webster, 1989). Recent operational applications of DSM are converging toward the use of regression kriging (Malone et al., 2011; Hengl et al., 2014) in which the two sources of soil data are used together, soil map as a soil covariate among others and spatial sampling with measured soil properties as input data for calibration of the regression model and for spatial interpolation of the regression residuals. However, in situations of sparse spatial sampling that often occurs in operational DSM, the performances of the regression kriging remain severely limited (Vaysse and Lagacherie, 2015).

The spatial estimations of soil properties produced by Soil Sensing are a third type of soil information that may be considered also as a DSM input that may mitigate the dearth in soil data. A growing number of sensors is now available for producing very high resolution

http://dx.doi.org/10.1016/j.geoderma.2016.12.011
0016-7061/© 2016 Elsevier B.V. All rights reserved.

(<5 m) images of estimated soil properties, either by field-based (or proximal) soil sensing techniques (Adamchuk and Viscarra Rossel, 2010; Mouazen et al., 2007) or by airborne sensing techniques (Selige et al., 2006; Stevens et al., 2010; Gomez et al., 2008). However, these soil sensing products are most often available over uncompleted and scattered areas because of their high costs and of their limited conditions of application. This prevents from using them as soil covariates in a classical regression kriging approach. As an alternative for mapping soil properties over a region with soil sensing products, we proposed a co-kriging approach (Lagacherie et al., 2012) that combined such input with a spatial sampling of measured sites. By taking hyperspectral-based estimations of clay content over a limited set of fields with bare surfaces as an example of soil sensing input, we showed that soil sensing could bring a significant increase of accuracy of clay content predictions over a whole region.

In this paper, we went a step further by developing and testing a new kriging approach, namely cokriging with a categorical external drift (CKCED), which combines the three possible soil inputs - soil map, spatial sampling of measured sites and soil sensing products. This approach was compared with spatial and non-spatial techniques using one of these inputs or a combination of two of them. The comparisons were performed for seven soil properties (Clay, silt, sand, Calcium Carbonate, pH, Total Iron and CEC) mapped over a 24.6 km^2 area located in the vineyard plain of Languedoc (Southern France).

2. Case study

2.1. Study area

The study was carried out in the La Peyne catchment (Fig. 1) in the South of France $43^o9'0''N$ and $3^o2'0''$ E. Vineyards form the primary land use in the area. Marl, limestone and calcareous sandstones from Miocene marine and lacustrine sediments formed the parent material of several soil types observed in this area, including Lithic Leptosols, Calcaric Regosols and Calcaric Cambisols (WRB soil classification, ISSS-ISRIC-FAO, 1998). These sediments were partly covered by successive alluvial deposits ranging from the Pliocene to Holocene and differed in their initial nature and in the duration of weathering conditions. These sediments have produced an intricate soil pattern that includes a large range of soil types, such as Calcaric, Chromic and Eutric Cambisols, Chromic and Eutric Luvisols and Eutric Fluvisols (Coulouma et al., 2008). The local transport of colluvial material along the slopes has added to the complexity of the soil patterns. An earlier ground sampling made in the study region (Lagacherie, 2008) showed that these complex soil patterns correspond to a great variability of clay content at the soil surface (from 65 $g.kg^{-1}$ to 452 $g.kg^{-1}$). A study area of 24.6 km^2(Fig. 1) was defined by intersecting this region of interest with the hyperspectral image used in this study.

2.2. Data

2.2.1. Spatial sampling of measured sites

143 sites (average sampling density of 1 site / 17 ha) were sampled in the study area for measurements of soil properties. All of these samples were composed of five sub-samples collected to a depth of 5 cm for representing a 5 m × 5 m square. The geographical position at the centre of this square was recorded by a decimetric GPS instrument. After homogenization of the sample, and removal of plant debris and stones, sieving and air drying, about 20 g was devoted to soil properties laboratory analysis. Seven soil properties for which previous estimations from hyperspectral data were attempted (Gomez et al., 2012a) were determined using classical physico-chemical soil analysis (Baize, 1988): calcium carbonate content (CaCO3), clay content (granulometric fraction $\prec 2\ \mu m$), silt content (granulometric fraction between 2 to 50 μm), sand content (granulometric fraction between 0,05 and 2 mm), free iron content, cation-exchange capacity (CEC) and pH.

Two subsets of sites can be distinguished among the set of 143 sites. 95 sampled sites were located in the bare soil fields. Both soil properties measurements and hyperspectral data suitable for estimation of soil properties were available for these 95 sites (Fig. 1 left). The remaining 48 sites had soil content measurements but unsuitable hyperspectral data because they were located in vineyard fields covered by vegetation. Both subsets were sampled for obtaining an even spatial distribution of sites while respecting the relative importance of the soil mapping units delineated by Coulouma et al. (2008). It must be noted that the criteria of selection of the two subsets of sites (bare soil vs vegetated fields) was totally independent from the spatial distribution of soils, which therefore did not generate any sampling bias.

2.2.2. Soil map

The soil map was derived from a very detailed soil map of the study area (Coulouma et al., 2008) by an expert-based grouping of the initial soil units into seven soilscapes as homogeneous as possible regarding the topsoil properties focused in this study. These soilscapes were described in details in Gomez et al. (2012a). The grouping into soilscapes was necessary for obtaining soil mapping units that included a number of sites large enough for applying the tested geostatistical procedures.

2.2.3. Airborne HYMAP image and its derivative

The HYMAP airborne imaging spectrometer measured reflected radiance in 126 non-contiguous bands covering the 400–2500 nm spectral range with around 19 nm bandwidths and average sampling intervals of 17 nm in the 400–2500 nm domain (http://www.intspec.com/). The HYMAP image was acquired on 13 July 2003 from a 3000 m altitude, providing a 5 × 5 m spatial resolution. Radiometric calibration was performed inflight (Richter, 1996) using nadir ground measurements (Beisl, 2001). The ATCOR4 code for airborne sensors was used for atmospheric corrections (Richter and Schlapfer, 2000). Topographic corrections were performed with a high-resolution digital elevation model from the Institut Géographique National (www.ign.fr) and DGPS ground control points.

The image was masked by using NDVI to remove living vegetation (essentially vineyards). The cellulose absorption band (2010 nm) was used to remove dry vegetation. Small areas of bare soils located at the parcel margins or along roads and pathway were also removed since they were not judged as representative of the neighbouring soil surfaces. Finally, the image provided usable data over 33, 690 pixels covering 3.5% of the total area only, that is the 192 bare soil fields that were randomly scattered over the region at the date of measurement.

3. Methods

3.1. Experimental set-up

We present hereafter the general workflow of our testing (Fig. 2). The details on methods are presented further.

The new algorithm combining the three possible types of soil information (CKCED) was compared with five non spatial and spatial methods that involved less types of soil information (Fig. 2). Ordinary Kriging (OK) and Partial-Least-square-Regression (PLSR) were applied for providing estimations of soil properties (denoted products in Fig. 2) from the spatial sampling of measured sites and from hyperspectral data respectively. Soil Map and spatial sampling of measured sites were combined twice, first by a baseline method that consists in computing a mean per soil mapping units (SMM), second by a more sophisticated Kriging with Categorical Drift (KCED, Monestiez et al., 2001). Finally the product derived

Please cite this article as: E. Walker et al., Combining measured sites, soilscapes map and soil sensing for mapping soil properties of a region, Geoderma (2016), http://dx.doi.org/10.1016/j.geoderma.2016.12.011