Geoderma Regional 7 (2016) 137–152

Contents lists available at ScienceDirect

Geoderma Regional

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

# Evaluating large-extent spatial modeling approaches: A case study for soil depth for France

M. Lacoste [a,*], V.L. Mulder [b], A.C. Richer-de-Forges [b], M.P. Martin [b], D. Arrouays [b]

[a] INRA, UR0272 Science du sol, F-45075 Orléans, France
[b] INRA, US1106-Infosol, F-45075 Orléans, France

A R T I C L E   I N F O

*Article history:*
Received 5 August 2015
Received in revised form 19 February 2016
Accepted 22 February 2016
Available online 24 February 2016

*Keywords:*
Soil depth
*GlobalSoilMap*
High resolution
Mapping
France
Cambisols
Luvisols
Leptosols

A B S T R A C T

The objective of this work was to compare three spatially explicit modeling approaches for soil depth (SD) for France (540 K km$^2$), produced using: i) a straight forward digital soil mapping (DSM) approach, based on regression treemodeling (*RTM*), ii) gradient boosting modeling (*GBM*), and iii) multi-resolution kriging (*MrK*) for large datasets. SD was defined from the USDA Soil Survey Manual as the depth (in cm) to a lithic or paralithic contact. SD was determined for 2116 sites from the French Soil Monitoring network (RMQS) and 14,113 sites from French Soil Inventory (IGCS) program. The RMQS dataset had a better spatial coverage and was collected following a standardized procedure. The RMQS dataset was used for calibration; the IGCS dataset was used for validation. SD ranged from 0 to 300 cm for RMQS (mean = 102 cm), and from 3 to 295 cm for IGCS (mean = 93 cm). Exhaustive environmental data were used to characterize the climate, the organisms, the topography and the other known soil forming factors, following the Soil-Landscape paradigm. The final maps were compared using the same strategy; the maps' accuracy and difference were assessed by comparing the SD predictions to observed data (RMQS and IGCS datasets) and to an available soil map (1:1M). The three DSM approaches predicted the SD trend mainly from the covariates derived from the digital elevation model (DEM). Then, the most important covariates were soil properties, climate covariates, and finally land use. The three predictive maps showed similar accuracy, and were consistent with the 1:1M SD map. The four maps presented similar spatial pattern at the country scale, but the *RTM* and *GBM* map showed the higher spatial heterogeneity, while the *MrK* map was smoother. The three maps had poor performance to estimate the shallower and deeper SD values. This issue was discussed and three options were presented to solve it (implementing calibration dataset, addition of kriged residuals, quantile transformation). Finally, this work showed that the choice of the modeling approach should be done considering the users' goals.

## 1. Introduction

Soils are supporting major ecosystem services, defined as provisioning services, regulating services and cultural social services (Millennium Ecosystem Assessment, 2005). Soil depth (SD) is a controlling factor in numerous surface and subsurface soil processes. SD is closely linked to soil erosion and slope stability, and was proven to be a crucial soil indicator for soil erosion impact, landscape evolution (Heimsath et al., 2001) and landslide control (Dietrich et al., 1995). SD also influences vegetation growth (Meyer et al., 2007), and is a key variable when dealing with soil storage capacity. That is valid for soil water storage, as SD partly constrains the available water capacity and the general hydrological response of a landscape (Dietrich et al., 1995; Wang et al., 2006), but also for soil organic carbon (SOC) storage. As a result, SD is a crucial input parameter, among others, in hydro-ecological models (Tesfa et al., 2009). Therefore, SD knowledge is important and SD exhaustive mapping was set as a requirement of the *GlobalSoilMap* project (Arrouays et al., 2014a; Arrouays et al., 2014b). However, determining SD proved to be hampered by i) high SD spatial variability, in relation with different soil forming factor in natural or cultivated areas (Vanwalleghem et al., 2010), ii) cost of SD estimation (in terms of time, field work and money), especially for the deep soils (>1.5 m) (Dietrich et al., 1995; Tesfa et al., 2009) and iii) discordance in the definition of SD. According to the authors, different terms were used (Table 1), and these designations were not always defined or even equivalent. For these reasons, values of SD are often missing in legacy data, or data are sparse and of unequal meaning. In this work, as in *GlobalSoilMap*, SD was defined as the depth (in cm) from the soil surface to a lithic or paralithic contact (Soil Survey Division Staff, 1993).

Various drivers influencing SD have been identified. At a given time *t*, SD obviously results from the history of soil production and erosion (Heimsath et al., 1997, 1999), involving physical, chemical and biological processes. Soil production by parent material weathering partly depends on the interactions between parent material and climate. From the results compiled by Minasny and McBratney (1999), soil formation

---

* Corresponding author.
*E-mail address:* marine.lacoste@orleans.inra.fr (M. Lacoste).

http://dx.doi.org/10.1016/j.geodrs.2016.02.006
2352-0094/

**Table 1**
Term used in studies dealing with soil depth.

| Term used | Reference | Definition |
|---|---|---|
| Depth of entire soil profile | Gessler et al. (2000) | A + B horizons thickness |
| Depth of solum | Odeh et al. (1994) | Not defined |
| Depth to bedrock | Zebarth et al. (2002) | Depth to bedrock |
| Depth to bedrock | Odeh et al. (1994) | Not defined |
| Soil depth | Ziadat (2010) | Depth to the maximum depth of the auger or to an impeding layer (rock or large stones) |
| Soil depth | Tesfa et al. (2009) | Depth to bedrock |
| Soil depth | Dahlke et al. (2009) | Depth from the ground surface to the surface of the bedrock or an impermeable layer |
| Soil thickness | Meyer et al. (2007) | Soil depth […] included A + C and Cr horizon thickness |
| Soil thickness | Catani et al. (2010) | Soil thickness […] intended […] as depth to bedrock […], or depth to the first marked change in hydrologic properties |
| Solum depth | Gessler et al. (1995) | A + E + B horizons depth |
| Solum depth | Martin and Timmer (2006) | Depth of soil from the bottom of the litter layer to the top of the parent material, or C horizon |
| Solum thickness | Rahman et al. (1996) | Not defined |

can range from 0.0003 mm $yr^{-1}$ for basalt in North Queensland to 7.78 mm $yr^{-1}$ for volcanic materials in Indonesia. Soil erosion could cause SD decrease, but also SD increase when erosion is counterbalanced by sedimentation processes. All these processes depend on climate (temperature, precipitations), land use, human practices, topography, etc. Especially, topography has long been identified as an important soil forming factor (Jenny, 1941). This was reinforced by many studies, which specified those variables linked to topography involved in the soil formation vs. degradation processes: slope, curvature, upslope contributing area, topographic wetness index (Florinsky et al., 2002; Gessler et al., 2000; Vanwalleghem et al., 2010). Indeed, topography influences soil redistribution, and Dietrich et al. (1995) underlined that soils were thinner or even absent on sharply defined ridges and thickest in unchannelled valleys. Strong relations between SD and landscape position (or landform) were also detailed in later studies (e.g. Martin and Timmer, 2006; Vanwalleghem et al., 2010; Zebarth et al., 2002).

SD, as other soil properties, can thus be derived from the covariates established in the soil-Landscape paradigm defined by Jenny (1941) and later reviewed by McBratney et al. (2003) as the SCORPAN model, where a soil property can be derived from the relations between other soil properties (S), climate (C), organisms (O), topography (R), parent material (P), past history (A) and spatial location (N). The influence of topographic attributes onto SD depends mainly on: i) the spatial scale of topographic variation in the area, ii) the nature of the processes involved in SD spatial variation, and iii) the degree to which terrain-soil relationships have been disturbed by human activities (Boer et al., 1996; Kuriakose et al., 2009; Vanwalleghem et al., 2010; Ziadat, 2010).

Finally, various methods taking advantages of these relations have been applied to produce digital SD maps. As classified by Kuriakose et al. (2009), three main methodological groups can be distinguished: i) physically-based methods, where SD is mainly predicted from weathering, erosion and accumulation rates and from physical properties of the regolith or underlying consolidated rock (Catani et al., 2010; Minasny and McBratney, 1999), ii) empirico-statistical methods, based on inferential and environmental correlations (Moore et al., 1993; Ziadat, 2010), and iii) interpolations from point samples using geostatistical methods (such as regression kriging in Vanwalleghem et al. (2010)).

The objective of this work was to compare three SD maps produced for mainland France following the specifications of the *GlobalSoilMap* project, which requires spatially explicit modeling at 3 arc-seconds resolution (equivalent to 90 m for France) and the estimation of the map uncertainties, and to assess to which extent their results were consistent. For France, regional work has already been done (e.g. Vaysse and Lagacherie, 2015), but, to our knowledge, the present work is the first presentation of a national SD product. Considering the total surface area of mainland France (about 540 K $km^2$), this implies that the employed methods need to be able to deal with large datasets. This is challenging, especially for geostatistical modeling. Thus, this study proposes to compare the results of three modeling approaches used to produce a SD map for mainland France, respectively based on i) a regression treemodeling (*RTM*) approach, ii) gradient boosting modeling (*GBM*), and iii) multi-resolution kriging (*MrK*) for large datasets.

## 2. Material and methods

### 2.1. Study area

This study was carried out in mainland France, without considering islands (such as Corsica) and overseas departments. The study area covers about 540 K $km^2$ (centered at approximately 47.0°N, 2.0°E) and is characterized by high landscape diversity. The altitude ranges from sea level in flat coastal areas up to more than 4000 m in the Alps, in the south-eastern part of France (mean altitude for France around 340 m). The actual geomorphology is driven by past orogenesis, past and still ongoing processes of erosion and sediment redistribution, which leads to a large variety of soil parent material (loose material, sedimentary, igneous and metamorphic rocks). Due to its location at the western part of Europe, France has mainly a temperate climate. However, different climate zones are identified: the western part of the area is mostly dominated by oceanic influences, the southern part by Mediterranean influences, and the eastern part by continental influences. These large trends are locally modified by the relief, which plays an important part in the spatial distribution of climate types. Mainland France is mainly covered by agricultural crops (annual and permanent cropping), permanent grasslands and forest. Intensive cropping, dairy systems and livestock farming are mainly present in the northern, western and central flat areas of the country, while forest areas are predominant in north-eastern and mountain areas or in the Landes coastal region. Even though orchards and vineyards are encountered in most of France regions, a large part of the production comes from the southern part of the country. In relations with these diverse features, soils also present a large diversity. However, at the highest taxonomic level, about 70% of the study area is covered by Cambisols, Luvisols and Leptosols (IUSS Working Group WRB, 2015).

### 2.2. Soil point datasets

Two available soil datasets with SD data were used (Fig. 1). The first dataset was a subset of 2116 sites from the French soil survey network (RMQS). This dataset encompasses a broad spectrum of climatic, soil and agricultural parameters and regularly covers the entire mainland France. The RMQS network is based on a 16 km × 16 km square grid and the sites are selected at the centre of each grid cell (Arrouays et al., 2002). For each site a soil pit description was made. The site surrounding (e.g. land use and geomorphology) and a detailed soil profile description were recorded, including soil horizon depth and SD. If the soil pit was not deep enough to determine the SD, an auger boring was done to complete the soil profile. Among the 2116 sites used in this study, SD has been explicitly recorded for 940 sites. For 1167 sites, only the soil horizon boundaries were described, and SD was estimated by adding 30 cm to the bottom limit of the deeper soil horizon. Finally,