Geoderma 213 (2014) 133–143

Contents lists available at ScienceDirect

Geoderma

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

Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks

B.P. Louis *, N.P.A. Saby, T.G. Orton, E. Lacarce, L. Boulonne, C. Jolivet, C. Ratié, D. Arrouays

INRA Orléans, Infosol Unit, US 1106, CS 40001, Ardon, 45075, Orléans CEDEX 2, France

ARTICLE INFO

Article history:
Received 17 July 2012
Received in revised form 3 July 2013
Accepted 21 July 2013
Available online 7 September 2013

Keywords:
Data harmonization
Random sampling
Latin Hypercube Sampling
Soil quality monitoring networks
Predictive performance

ABSTRACT

Regulations about soil quality are normally imposed at international level while many countries have set up monitoring networks at national scale. Since these networks use different sampling strategies, there is a strong need to harmonize a posteriori the collected data from the national networks in order to answer questions raised by the global regulations. For that purpose, calibration sites where different sampling strategies are carried out are necessary in order to construct harmonization functions between measurements from different sampling protocols. A case study is available for French forest soils that have been sampled twice simultaneously on the same sampling grid but with different sampling and analytical strategies: a first sampling for the French soil quality monitoring network (RMQS) and a second one for the European forest monitoring network (ICP Forests level I second survey i.e. Biosoil). However, the way to define the number and the position of these calibration sites remains a key issue. In this work, we compare both RMQS and Biosoil strategies for a set of measured variables of interest (carbon, potassium and lead contents and pH) and aim to define the minimum number of sites and their best location to establish reliable harmonization functions. Three statistical methods for construction of sampling designs are tested: random sampling, conditioned Latin Hypercube Sampling (cLHS, Minasny and McBratney, 2006) and D-Latin Hypercube Sampling (DLHS, Minasny and McBratney, 2010). With each method, we investigate the effects of the number of calibration data on the predictive quality of the harmonization functions. First, we show that both cLHS and DLHS are better than simple random sampling. Then, the difference between cLHS and DLHS performance depends mainly on the size of the samples, the nature of the soil property and the form of the pedotransfer functions.

1. Introduction

As a result of our increasing awareness of the important functions the soil fulfills in our ecosystems, soil quality has become a key issue in political agendas (e.g. European committee, 2002, 2006). In order to assess this quality, more and more surveys and monitoring tools have been implemented. The ENVASSO project (ENVironmental ASsessment of SOil for monitoring, Kibblewhite et al., 2008; Morvan et al., 2008) has identified most networks already implemented on the European continent, whether national or transnational. It shows a global lack of standardization due to heterogeneities in sampling strategies (e.g. sampling design, number of sampling sites, sampling depths) or in analytical protocols (e.g. measured soil property, measuring methods). Such differences bring out the question of compatibility between the data from these networks which can lead to a lack of comparability of the measures from different sources (Köhl et al., 1999).

The lack of comparability between measured soil properties from different networks complicates significantly the establishment and application of international regulations and makes data harmonization deeply necessary. Köhl et al. (1999) define the harmonization of data as the bringing together of existing concepts in such a way as to make them easier to compare. This means that existing data measured by different networks with different strategies should be transformed in a way that data put side by side can be meaningfully compared (i.e. values compared refer to the same property in the same conditions of sampling and analysis). Though "horizontal standardization" (i.e. standardization of the monitoring protocols on all the monitoring networks) would "avoid unnecessary duplication of work" (Gawlik et al., 2004), most countries already have established national monitoring protocols and switching them to harmonized protocols could make it impossible to compare data from previous surveys (Morvan et al., 2007). In light of this observation, the INSPIRE directive (European Parliament and European Union council,

Abbreviations: LHS, Latin Hypercube Sampling; cLHS, conditioned Latin Hypercube Sampling; DLHS, D-Latin Hypercube Sampling; RMQS, French soil quality monitoring network; RMSEP, Root Mean Squared Error of Prediction; HF, harmonization function; RPD, ratio of standard deviation to RMSEP; RPIQ, ratio of inter-quartile range to RMSEP.

* Corresponding author. Tel.: +33 2 23 48 52 31; fax: +33 2 23 48 54 30.
E-mail addresses: benjaminlouis.pro@gmail.com (B.P. Louis), nicolas.saby@orleans.inra.fr (N.P.A. Saby), Thomas.orton@orleans.inra.fr (T.G. Orton), eva.lacarce@laposte.net (E. Lacarce), line.boulonne@orleans.inra.fr (L. Boulonne), claudy.jolivet@orleans.inra.fr (C. Jolivet), celine.ratie@orleans.inra.fr (C. Ratié), dominique.arrouays@orleans.inra.fr (D. Arrouays).

0016-7061/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.07.018

2007) places the total coverage of the European territory for 2015 in interoperable datasets on soil quality. Interoperability of datasets relies on i) having documentation in a metadata form, ii) having flexible data structures and standardized exchange formats (Baritz and Eberhardt, 2008; Lacarce et al., 2009) and iii) developing tools for data interpolation and harmonization.

For this last point, some methodologies can be found in the literature. For instance, Skøien et al. (2010) use a geostatistical approach adding a particular bias for each network in the spatial model of the measured values and showed a good assessment of these biases. The limit of this method lies in the fact that it works only for overlapping or neighboring networks. Unfortunately, the issue of data harmonization cannot be reduced to these two cases: lots of networks belong to different regions that do not share boundaries. Projects which studied the question of data harmonization (e.g. Arrouays et al., 2012; Kibblewhite et al., 2008; van Beek et al., 2010) agreed on the lack of methodologies for harmonization. They underlined that, to elaborate such methodologies, sites sampled at the same time according to the strategies of different networks and measured for the same soil property are necessary to harmonize the data from these networks. These sites would allow the construction and calibration of harmonization functions (HFs, i.e. a function or model which links the data from one network to the data from another) for transforming data from a network so that they are comparable to data from a second network. Ideally, the function would be linear with an intercept of 0 and a regression coefficient of 1, which would mean that the values from the two networks are identical. This case is highly unlikely due to several sources of errors (Rawlins et al., 2009): i) the location of the sites: even if the locations of the two networks are very close, it is impossible to overlap them precisely, ii) the sampling errors due to differences between sampling strategies (e.g. sampling depths), iii) the subsampling errors in partitioning soil for analysis and iv) analytical errors or differences (laboratory measurements). The goal of the HF — which is calibrated using the collocated data from the two schemes being compared — is therefore to give the best fit between the data from two monitoring schemes. However, it is unfeasible to sample every location in every network being compared using each of the different sampling strategies. Then, how should we define the number and the relative position of these calibration sites?

We use a case study from France to investigate these issues. Forest soils across the country have been sampled on the same 16×16 km^2 systematic square grid by two monitoring schemes that use different sampling and analytical protocols. Soil samples were collected for the French soil quality monitoring network (RMQS, Arrouays et al., 2002; Jolivet et al., 2006), and also for the second campaign of soil sampling of the European program ICP-Forests level I in the framework of the project called Biosoil (Hiederer et al., 2011; Lacarce et al., 2009). Having two monitoring schemes with the same locations but with different sampling strategies presents the opportunity of studying the questions of the number and location of calibration sites. The comparability of measurements of the same soil property from these two monitoring schemes can be studied and HFs can be calibrated. The aim of this paper is to use this case study to determine the impact of the number and the relative location of calibration sites on the predictive quality of the HFs.

There are several methods that can be applied for designing sampling networks (e.g. simple random sampling, stratified random sampling or systematic random sampling; de Gruijter et al., 2006). In this work, we investigate three such approaches: a) simple random sampling (RS) and two stratified methods: b) conditioned Latin Hypercube Sampling (cLHS, Minasny and McBratney, 2006) and c) D-Latin Hypercube Sampling (DLHS, Minasny and McBratney, 2010). With each approach we investigate the effect of the number of calibration sample locations selected. Simple random sampling (i.e. selecting the calibration sites at random with no restrictions on their locations) is chosen as a reference method. Only the sample size is selected a priori with this type of design and all sites have the same probability to be sampled. Its advantage is its simplicity. The main disadvantage is that some domains of the distribution and modalities of parameters may be missing or large empty areas may occur in the sample. The two methods based on Latin Hypercube Sampling (LHS) are stratified methods, first proposed by McKay et al. (1979), which generate a set of samples that more precisely reflect the shape of a multidimensional distribution for a collection of chosen ancillary variables (i.e. the co-ordinates or important soil properties). Thus, the sites selected for sampling are more representative of all the sites (according to the chosen ancillary variables). Comparing the RS and LHS methods allows us to assess the impact of accounting for the shape of the multidimensional distribution for the ancillary variables in the construction of HFs.

2. Materials and methods

2.1. Data source

Our data come from the common sites between the French soil quality monitoring network (RMQS, Arrouays et al., 2002; Jolivet et al., 2006) and the European network of the International Cooperative Program on Assessment and Monitoring of Air Pollution Effects on Forests (ICP-Forests level I). The RMQS network contains about 2200 sampling sites spread over the entire French territory and is based on a 16 km $\times$ 16 km square grid. The sites were selected at the center of each grid cell on condition that the soil was accessible. If not, an alternative location was selected as close as possible. From the ICP-Forests (International Cooperative Program on Assessment and Monitoring of Air Pollution Effects on Forests) level I program network, Biosoil (Hiederer et al., 2011; Lacarce et al., 2009) was a test project for the development of a large scale harmonized soil monitoring network. It is based on the same square grid as RMQS but sampling is only carried out on forest soils. The two networks share 467 common sites, useable for our study, sampled simultaneously and spread over metropolitan France, separated by between 11 and 202 m (mean of the distances between pairs of sites = 21 m; median of the distances between pairs of sites = 19 m). However, there were differences between the sampling strategies used in the two projects. In particular, in RMQS the top layer was sampled from 0 (the limit between organic and organo-mineral horizons) to 30 cm depth. The equivalent sampled depth in Biosoil was divided in three sampled layers: 0–10 cm, 10–20 cm and 20–40 cm.

All analyses for both projects were performed in the same laboratory of INRA (Soil Analysis Laboratory, Arras, France). Organic carbon content, potassium content, and lead content and pH were measured, from the topsoil in RMQS (0–30 cm) and for the first three layers in Biosoil (except for lead which was measured only for the top layer in Biosoil; 0–10 cm). The analyses for SOC were carried out by dried combustion in both RMQS and Biosoil projects. Potassium content was measured using hydrofluoric acid dissolution, which enables the measurement of the total potassium content. Lead content was also measured after hydrofluoric acid extraction in the RMQS project, giving the total lead content. The lead content for the top-layer in Biosoil was extracted with aqua regia, which enables a strong but incomplete extraction of the total lead in the soil. Soil pH was measured similarly for the two projects by suspension in 1/5 of water. Table 1 summarizes the differences that we have described between the sampling strategies and analyses used in the two projects.

Table 2 shows the summary statistics for the data from the 467 common locations from the two networks for the four soil properties. Organic carbon values from the 0 to 10 cm layer of Biosoil are higher than the values from the RMQS-sampled layer, values from the 10 to 20 cm layer are similar to those from RMQS and values from the 20 to 40 cm layer are lower. This observation is due to the dynamic in soil organic carbon which stays generally in the topsoil (except in Podzols). Also, the variability of soil organic carbon in the topsoil decreases with