Populating soil maps with legacy data from a soil testing databases
Jean-Baptiste Paroissien, Nicolas Saby, Benjamin Louis, Dominique Arrouays

To cite this version:
Jean-Baptiste Paroissien, Nicolas Saby, Benjamin Louis, Dominique Arrouays. Populating soil maps with legacy data from a soil testing databases. GlobalSoilMap, Oct 2013, Orléans, France. 2014. <hal-01190639>

HAL Id: hal-01190639
https://hal.archives-ouvertes.fr/hal-01190639
Submitted on 1 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Populating soil maps with legacy data from a soil testing databases

Abstract

In France, soil test results from samples of cultivated topsoil requested by farmers have been collected to constitute the National Soil Testing database (NSTD). Enriching soil maps with such data can be regarded as an important source of information to build GlobalSoilMap products when dense soil profile information does not exist.

We inferred the soil organic carbon content (SOC) distribution within the soil units of a soil map in a central region of France to re-allocated analyses from NSTD. The method is based on a pedological distance between soil properties (pH, CaCO3 content, cation-exchange capacity, clay content and silt content) of the NSTD analyses and soil map units mean values of the same parameters.

Results showed a consistent spatial distribution of SOC content.

Context

NSTD offers the advantage to gather a very large number of analytical results since 1990 but these analytical results were not originally intended for the purposes of soil mapping. The legacy data do not always include information about the statistical distribution within spatial unit of the soil properties (min, max and modal). Allocations of NSTD to soil map units may allow supplying a better statistical description of these distributions of their properties.

Methods

1. "Loiret", the study area

Databases used

NSTD : 12 594 topsoil samples spreaded into 327 municipalities
Map of Loiret : 291 Soil Typological Units (STU) & 95 Soil Maps Units (SMU)

2. Overlay between communes and the soil map

Table with all combinations of NSTD analyses and STU

3. Calculating pedological distance

The Euclidean distance is calculated using a Principal Component Analysis (PCA) built from the modal values of soil properties of the STU as active variables.

The values of the soil properties from NSTD analysis are introduced in the PCA as supplementary variables. A matrix containing the PCA coordinates of all combinations of NSTD analysis and STU is built and used to calculate a pedological distance following this formula:

\[ d_{i,j} = \sqrt{\sum_{k} w_k (x_{i,k} - x_{j,k})^2} \]

where \( x_{i,k} \) is the PCA coordinate of NSTD analysis \( i \) for the principal component \( k \), \( x_{j,k} \) is the PCA coordinates of STU \( j \) for the principal component \( k \), \( w_k \) is the percentage of variance explained by principal component \( k \).

4. Aggregating values by SMU

To map SOC, we first computed the mean of NSTD SOC analyses for each STU and then calculated the weighted mean of each SMU

5. Validation procedure

A part of NSTD (80% randomly selected) in each commune is used to populate the regional soil survey. The NSTD SOC analyses kept apart were aggregated by commune to compute the mean of SOC by SMU. This dataset is named SOC dataset validation.

Open-source software used

References


Conclusion

The principle of pedological distance has already shown good results in the prediction of soil classes (Carré et al., 2009). In this study, the results are consistent with the reference map. In the framework of GlobalSoilMap the main interest of this method is to populate existing soil maps to provide a better estimate of the range of values of a given soil property.