Populating soil maps with legacy data from a soil testing databases

Jean-Baptiste Paroissien, Nicolas Saby, Benjamin Louis, Dominique Arrouays

Populating soil maps with legacy data from a soil testing databases

J-B. Paroissien[1], N.P.A. Saby[1], B. Louis[2], A.C. Richer de Forges[1] & D. Arrouays[1]
[1]*INRA, US1106 InfoSol, F-45075 Orléans, France*
[2]*INRA, UMR1069, Sol Agro et hydrosystème Spatialisation, F-35000 Rennes, France*

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 the soil properties (pH, CaCO3 content, cation-exchange capacity, clay content and silt content) of the NSTD and soil map units mean values of the same parameters
- Results showed a consistent spatial distribution of SOC content

Context

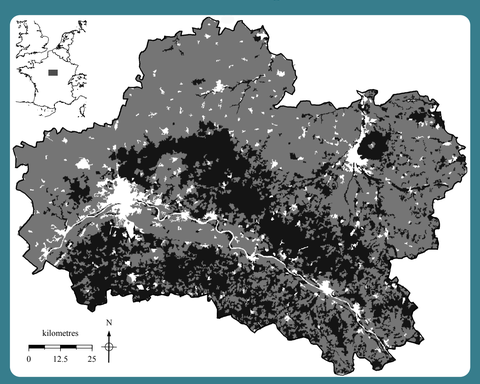
NSTD offers the advantage to gather a very large number of analytical results since 1990 but these analytical results were not originally intended for the purposes of soil mapping (Schvartz et al., 1997). The legacy soil data do not always include information about the statistical distribution within 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.

Aim

- Re-allocate soil organic carbon content analysis from soil testing database to soil map units of the Regional Soil Survey (RSS) (Richer de Forges 2008) of the French department "Loiret"
- Compute a pedological distance (Carré et al., 2009) between soil properties (pH, CaCO3 content, clay content, silt content and sand content) of the NSTD analyses and soil map units mean values

Methods

1. "Loiret", the study area



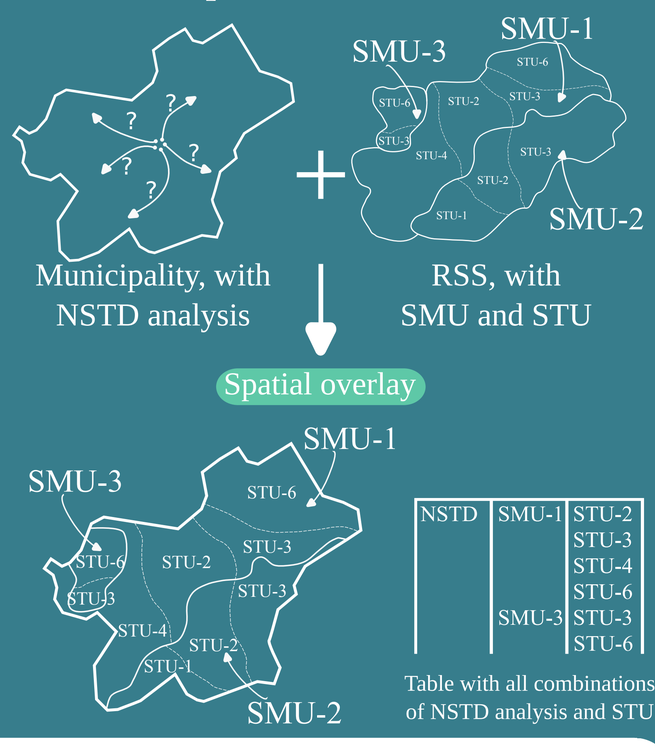
Urban & water area | Crops | Forest

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



NSTD	SMU-1	STU-2
		STU-3
		STU-4
		STU-6
	SMU-3	STU-3
		STU-6

Table with all combinations of NSTD analysis 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.

Pedological criteria used are : soil texture, pH and $CaCO_3$.

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 :

$$d_{i,j} = \sqrt{\sum_k w_k (x_{i,k} - x_{j,k})^2}$$

$x_{i,k}$ 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

The pedological distance is then used as selection criteria. Analyzes with a pedological distance lower than the 10th quantile calculated using the distribution of all the distances are assigned to the selected STU.

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 calculate the mean of SOC by SMU. This dataset is named *SOC dataset validation*.

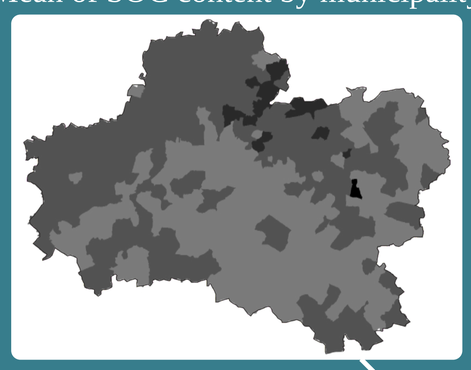
Open-source software used

R | PostGIS | GMT Generic Mapping Tools Graphics

Results

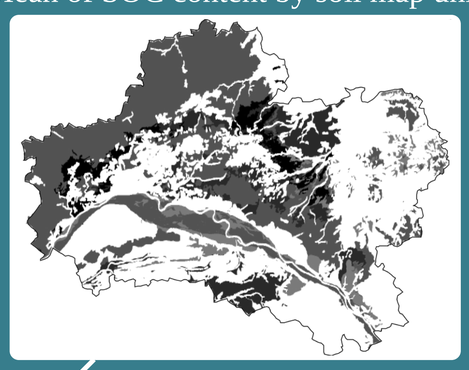
1. Maps of SOC content (no data areas are in color white)

Mean of SOC content by municipality



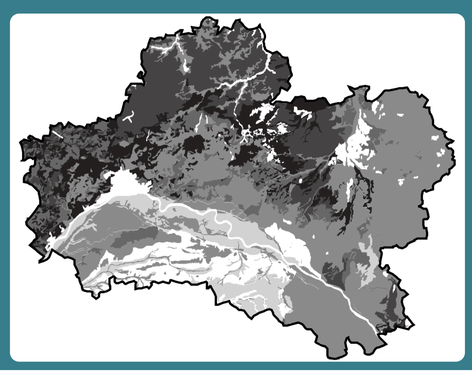
1-10
10-15
15-20
> 20
g.kg^{-1}

Mean of SOC content by soil map unit



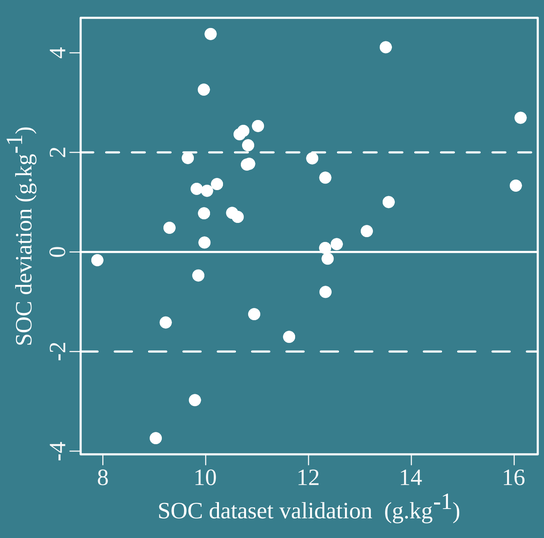
Re-allocated SOC content by SMU

6-8
8-10
10-11
12-13
> 13
g.kg^{-1}



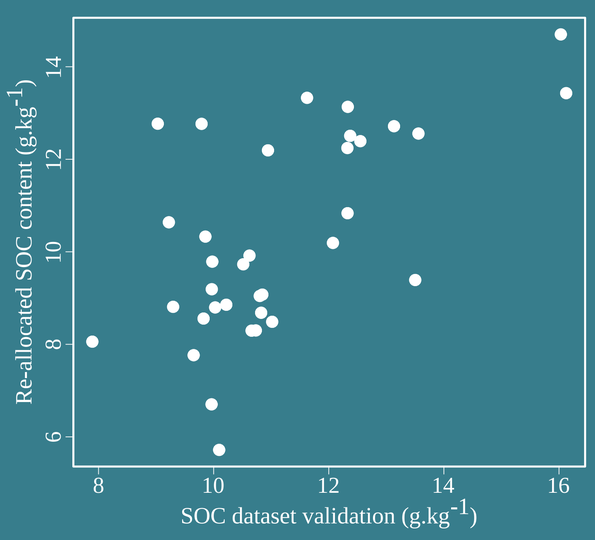
- Same pattern of SOC content across the study area
- Much lower SOC content in the south of the Loiret
- A complete SOC distribution for each SMU

2. Validation results



The deviation of the method is uniform along the range of NSTD values

- 71% of analyses are inside the envelope of acceptance



Significant correlation between dataset validation and re-allocated SOC content

- R^2 : 0.4
- RMSE : 2 g.kg^{-1}

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.

References

Carré, F. & Jacobson, M. 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148: 336-345

Richer de Forges, A.C., 2008. Référentiel régional pédologique de la région Centre, Notice explicative de la carte des pédo-paysages du Loiret à 1/250 000. Institut National de la Recherche Agronomique, France.

Schvartz, C., Walter, C., Claudot, B., Bouédot, T. & Aurousseau, P. 1997. Synthèse nationale des analyses de terre réalisées entre 1990 et 1994. 1 Constitution d'une banque de données cantonale. Etude et Gestion des Sols 4 : 191-204

l'Europe s'engage en région Centre | INRA SCIENCE & IMPACT