Evaluating the quality of soil legacy data used as input of Digital Soil Mapping models

Philippe Lagacherie, Maider Arregui, David Fages, Malithi Weerasekara

Evaluating the quality of soil legacy data used as input of Digital Soil Mapping models

Philippe Lagacherie[1], Maider Arregui[2] , David Fages[1] & M. Weerasekara[1]

1. LISAH, University Montpellier, IRD, INRA, Montpellier SupAgro (France) 2. BRL Exploitation, Nîmes (France)

Objective

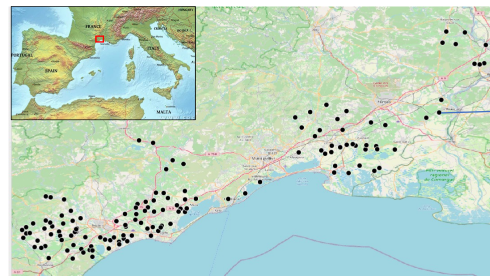
Most of the Digital Soil Mapping products now available across the globe have been developed from legacy measured soil profiles inherited from several decades of soil survey activity. To our knowledge, no reference on the quality of these soil legacy data have been established yet. Our objectives was then i) to evaluate the differences between values of soil properties provided by legacy measured soil profiles and actual certified soil laboratory measurements , ii) to depict the different causes of these differences and iii) to investigate how to reduce some of them.

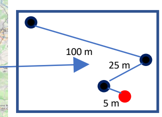
Legacy measured soil profiles

8200 soil profiles with laboratory analysis collected between 1955 and 1963 by BRL soil surveyors

Integrated in the French Soil database using text recognition (Chemchem et al, this workshop)

Control soil sampling





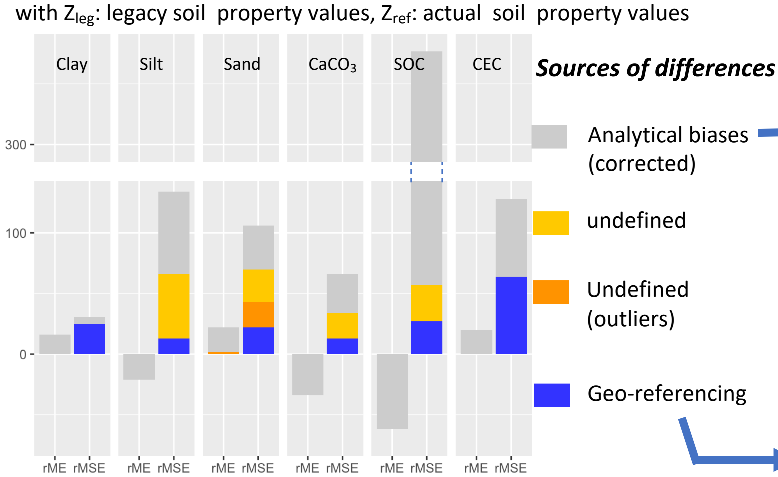
- 137 sites over the BRL soil survey area (6 636 km^2)
- Four samples per sites (local variations of soil properties)
- Analysis of topsoil (0-20 cm) properties

Differences between legacy and actual soil property values

rME: relative Mean error : rME = $\text{mean}(Z_{leg} - Z_{ref})/ \text{Mean}(Z_{ref})$

rMSE: relative Mean Square Error = $\text{mean}(Z^2_{leg} - Z^2_{ref})/ \text{variance}(Z_{ref})$

with Z_{leg}: legacy soil property values, Z_{ref}: actual soil property values



- High relative Mean errors and Mean Square errors revealing great differences between legacy and actual soil property values
- Important contributions of analytical biases that could be corrected
- Variable and substantial contributions of the geo-referencing errors related with the relative importance of the local variability
- For some properties, a large amount of the difference remained undefined (time evolution, analytical errors, recording errors,…)

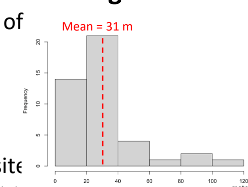
Correcting analytical biases

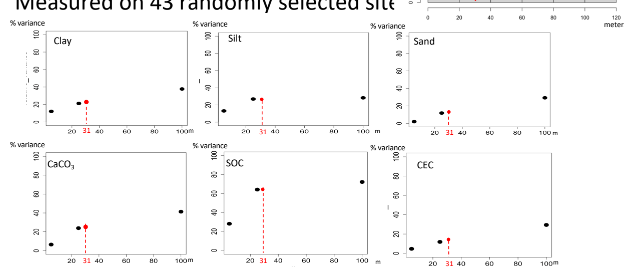
$Z_{corrected} = a + b\, Z_{leg} + c\, CaCO3_{leg}$

Property	a	b	c
Clay	9.3980	0.7531	-
Silt	17.5961	0.4246	- 0.4723
Sand	1.2228	0.7133	-
CaCO3	5.2143	0.9089	-
SOC	0.7111	0.4728	-
CEC	4.577	0.522	-

Assessing the impact of geo-referencing errors

- Geo-referencing errors = differences of relative positions from remarkable points given by the geographical coordinates of soil profiles *vs* by the localisation plan
- Measured on 43 randomly selected site





Local variabilities of soil properties (% of total variance for 5m, 25m and 100m spacing) and impacts of a 31 m geo-referencing error on the differences of soil property values.

Take home messages

- The legacy measured soil profiles usually used as input of our DSM models may have substantial biases and uncertainties that could be a major limitation of DSM prediction performances.
- Biases and a part of these uncertainties can however be corrected by using control sampling. Such control step should become systematic in DSM applications using legacy data
- Geopositionning is an important source of uncertainty for old soil datasets (< 2010) collected with manual geo-referencing
- Future DSM models should handle such input uncertainties and biases