Public document

Validation and performances report

FP7 – DIGISOIL Project Deliverable D3.3

N° FP7-DIGISOIL-D3.3
March 2011

…

The DIGISOIL project (FP7-ENV-2007-1 N°211523) is financed by the European Commission under the 7th Framework Programme for Research and Technological Development, Area "Environment", Activity 6.3 "Environmental Technologies".

Public document

Validation and performances report

FP7 – DIGISOIL Project Deliverable 3.3

N° FP7-DIGISOIL-D3.3
March 2011

G.Grandjean

With the collaboration of

K.Samyn (BRGM), I.Cousin, M.Seger A.Besson, (INRA) S.Lambot, B.V. Wasemael (UCL), F.Galfagnoli (UNIFI), L.Chiarantini, (Selex Galileo)

Checked by:	**Approved by**:
Name: B. Van Wasemael	Name: A.Vagner
☑	☑

BRGM's quality management system is certified ISO 9001:2000 by AFAQ.

Keywords: geophysics, sensors, soil properties

In bibliography, this report should be cited as follows:
Validation and performances report, L.Chiarantini, 2009. Validation and performances report. Report N° FP7-DIGISOIL-D3.3; 48 pages.

Synopsis

This deliverable present a validation of the geo-physical techniques, based on different sets of criteria: i) accuracy of the measurements generally estimated by the RMSE, ii) capacity to represent spatial patterns and spatial variability.

The geophysical data are interpreted in terms of different thematic maps such as: soil thickness, stone content, water content, clay content and C content. For each of these maps the performance are evaluated by measuring the gap between the estimated soil properties and those observed on the field or stored in existing soil database. Special attention is also given to the spatial patterns of errors and uncertainties inherent to the geophysical techniques.

The different maps obtained on the Luxembourg or Mugello test sites are presented. First order maps are obtained by a simple inversion of geophysical signals into soil properties; second order maps are computed by several first order maps that are combined to produce more accurate information or a soil property of a high level.

Contents

1. Introduction 9

1.1. MAIN OBJECTIVES OF THIS STUDY 9

1.2. GENERAL WORKFLOW : FROM INPUT DATA AND FIELD OBSERVATIONS TO THEMATIC MAPS 9

2. First order soil maps 11

2.1. SOIL THICKNESS MAP 11

2.1.1. Final workflow for the considered map 11

2.1.2. Resulting map description 12

2.1.3. Performances 14

- Spatial error 14
- RMSE estimation 15

Spatial resolution 15

2.2. EFFECTIVE CLAY CONTENT MAP 17

2.2.1. Mapping procedure 17

2.2.2. Resulting map description 18

2.2.3. Performances 19

- Validation and RMSE estimation 19

2.3. STONE CONTENT MAP 21

2.3.1. Final workflow for the considered map 21

2.4. WATER CONTENT MAP 23
2.4.1. Resulting map 25
2.4.2. Performances 25
- Spatial error 25
- RMSE estimation 25
2.4.3. Spatial resolution 28
- Technical or economical constraints 28
- Spatial structure and data interpolation 28
2.5. BULK DENSITY MAP 30
2.5.1. Final Workflow 30
3. Second order soil maps 35
3.1. CLAY CONTENT 35
3.1.1. Final workflow for the considered map 35
3.1.2. Resulting maps 36
3.1.3. Performances 38
- Validation and RMSE estimation 38
3.1.4. Spatial resolution 40
- Technical or economical constraints 40
- Interpolation aspects 40
3.2. C STOCK 41
4. Conclusions 49
5. References 51

List of illustrations

Figure 1 : Revised workflow of the Digisoil's system for estimating soil properties maps 10

Figure 2 : From acquisition to inversion of surface waves (MASW methodology)..... 12

Figure 3 : (a) Location map of seismic and penetrometer data points over the Luxembourg site; (b) Comparison between a penetrogram and trench observations. . 13

Figure 4 : Comparison between the soil depth obtained from MASW (left) and from the validation penetrometric data set (right). 13

Figure 5 : Map of the RW limit depth spatial error obtained calculating the difference between RW limit depth from MASW and from the penetrometric validation data set. 14

Figure 6 :Plot of the linear regression between the RW limit depth predicted using MASW and from the validation penetrometric data set..... 15

Figure 7 : (Left) Example of a variogram estimated in the direction 0° from geographic north. This variogram is fitted with a theoritical variogram model using a quadratic component with scale=0.0551, Length=185, anisotropy ratio=2 and anisotropy angle=23.08°, (Right) kriging standart deviation map. 16

Figure 8 : Complete workflow for the considered map..... 17

Figure 9 : From acquisition to inversion of hyperspectral data..... 18

Figure 10: Comparison between hyperspectral-derived clay map and IDW interpolation of laboratory values..... 19

Figure 11:Plot of the linear regression between clay content predicted using SIM-GA and from the validation sampling data set. 20

Figure 12: Correlation between clay content and absorption peak at 1414 nm, 1914 nm and 2210 nm..... 20

Figure 13 : Calibration relationship to estimate the stone content from electrical resistivity data and direct measurements in the field. 21

Figure 14 : Modelisation of the stone content..... 22

Figure 15 : Validation of the model. The orange dots, far from the 1:1 line, correspond to points located in the anthropogenic North-East part and the stone content at these locations can not be estimated by the model..... 23

Figure 16. Data acquisition from the field and principles of the STA method. 24

Figure 17 : Soil water content maps derived from off-ground (left) and on-ground (right) GPR data for dry (top) and wet (bottom) conditions at the Luxembourg site. 26

Figure 18 : Wet condition soil moisture maps derived by both off- and on-ground GPR techniques compared with ground truths in shallowest 10 cm depth. 27

Figure 19 : Plot of the linear regression for soil volumetric water content between the off-ground (left) and on-ground (right) GPR techniques and soil core sampling. 27

Figure 20 : Plot of the linear regression for soil volumetric water content between the off-ground (a) and on-ground (b) GPR techniques and soil core sampling. 29

Figure 21 :Calibration plot for the bulk density estimation from ART measurements... 31

Figure 22: a) real resistivity map of the cultivated horizon b) associated inversion error map 32

Figure 23 : Bulk density map..... 33

Figure 24: (a) GPR-EMI setup and (b) workflow for clay content estimation..... 36

Figure 25: Kriged maps of clay content estimates from EMI and GPR measurements 38

Figure 26: Validation relationships between measured and estimated clay content 40

Figure 27: Standard deviation map for kriged 0-10 cm clay content estimated from horizontal dipoles EMI measurements during the second measurement day (wet conditions) 41

Figure 28 : Map of the SOC stocks of the upper 20 cm soil layer. The values and residuals are expressed in Mg ha^{-1}. Point numbers refer to the validation points. 43

Figure 29 : Map of the SOC stock (Mg ha^{-1}) for the entire soil profile until the unweathered bedrock as detected by the seismic profiler. 44

Figure 30 : Predicted against observed SOC stocks (Mg ha^{-1}) for the upper 20 cm. The numbers refer to the validation points. 45

Figure 31 : Semi variogram of the SOC stock of the upper 20 cm..... 47

1. Introduction

1.1. MAIN OBJECTIVES OF THIS STUDY

Thiis deliverable present a validation of the geo-physical techniques, based on different sets of criteria: i) accuracy of the measurements generally estimated by the RMSE, ii) capacity to represent spatial patterns and spatial variability.

The geophysical data are interpreted in terms of different thematic maps such as: soil thickness, stone content, water content, clay content and C content. For each of these maps the performance are evaluated by measuring the gap between the estimated soil properties and those observed on the field or stored in existing soil database. Special attention is also given to the spatial patterns of errors and uncertainties inherent to the geophysical techniques.

In the following sections we present the different maps obtained on the Luxembourg or Mugello test sites. First order maps are obtained by a simple inversion of geophysical signals into soil properties; second order maps are computed by several first order maps that are combined to produce more accurate information or a soil property of a high level.

1.2. GENERAL WORKFLOW : FROM INPUT DATA AND FIELD OBSERVATIONS TO THEMATIC MAPS

The principle of estimating soil properties maps from geophysical signals was presented within a general workflow in the deliverables of workpackage 2. Due to different constraints coming from local field measuring contexts, i.e., soil heterogeneity, surface conditions such as roughness, particular physical properties unfavourable to one or several geophysical methods, the theoretical workflow has been modified so that it corresponds now to what it can be really obtained from our methodologies. The following figure (Figure 1) shows the revised approach of the Digisoil's system, including the input geophysical data and the contributions of ancillary ones in the different processes.

This new workflow starts with the different techniques (seismic, geoelectric, GPR, EMI, hyperspectral) validated during the project and describes the main processing levels (inversion, calibration, interpolation). The results are presented as two kinds of maps: the first ones derive directly from the data processing (soil thickness, water content, different clay content maps, surface C content), the second ones are a combination of the previous ones and propose a highest level of information (clay content map, C stock map)

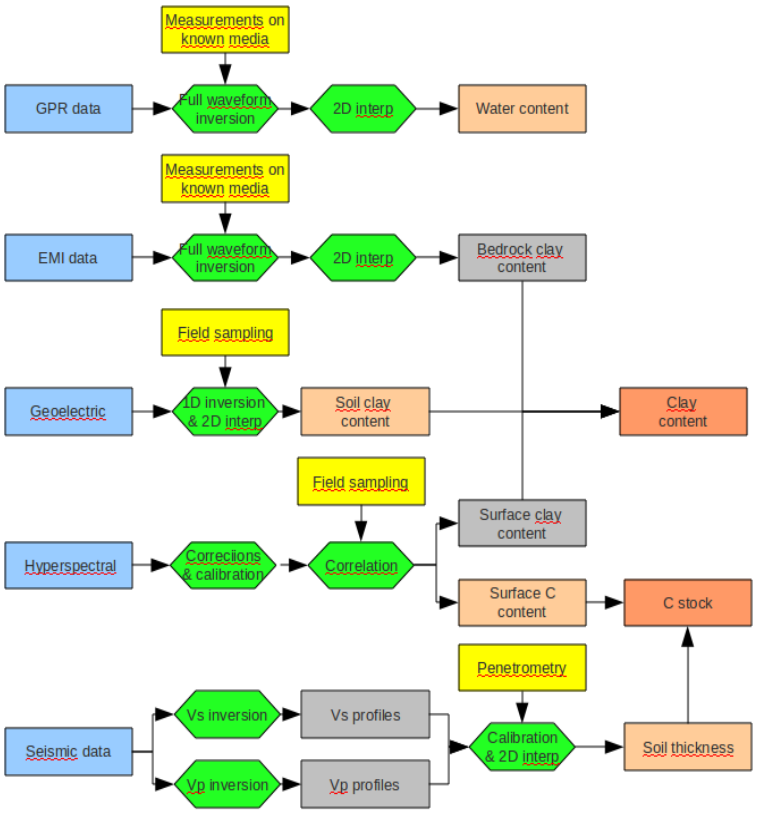


Figure 1 : Revised workflow of the Digisoil's system for estimating soil properties maps

2. First order soil maps

2.1. SOIL THICKNESS MAP

2.1.1. Final workflow for the considered map

The MASW seismic method allows obtaining a velocity model at a given point because of the dispersive character of surface waves (Figure 2). The dispersion diagram reveals the dependence of phase velocity with frequency. After this stage, the inversion process of dispersion curves aims to find the vertical S-waves velocity model. A linearized inversion method is then used to obtain the S-waves velocity model and layers thicknesses. For each iteration of the inversion process, the algorithm tends to fit a computed dispersion curve to observed real data. In some case, there is a contraindication for the use of the MASW methodology. Hard soils, non-tabular media particularly don't allow to compute well resolved dispersion diagrams. In such cases, a second processing workflow based on the inversion of P-waves first-time arrivals inversion was used (the example of Luxembourg test site). This method allows to retrieve 2D P-waves velocity (Vp) models. The processing workflow presented hereafter is a summary of the methodology developed in the report D1.3 and based on the different tests included in D1.1-2 and 2.1-2.3.

	Luxembourg	Mugello
Input data	Seismic records	Seismic records
Processing	P-wave tomography	MASW
1.	Amplitude correction (AGC)	Dispersion diagram computation
2.	First breaks picking	Picking of dispersion curves
3.	First-time arrivals inversion	Dispersion curves inversion
Data inversion	Production of 2D Vp models	Production of 1D Vs models
Data interpolation		Production of 2D Vs models
Calibration	Penetrometric data set	Penetrometric data set
Soil/bedrock limit	Seismic horizon picking using penetrometric calibration	Seismic horizon picking using penetrometric calibration
Mapping	Interpolation of the data points using kriging technique	Interpolation of the data points using kriging technique

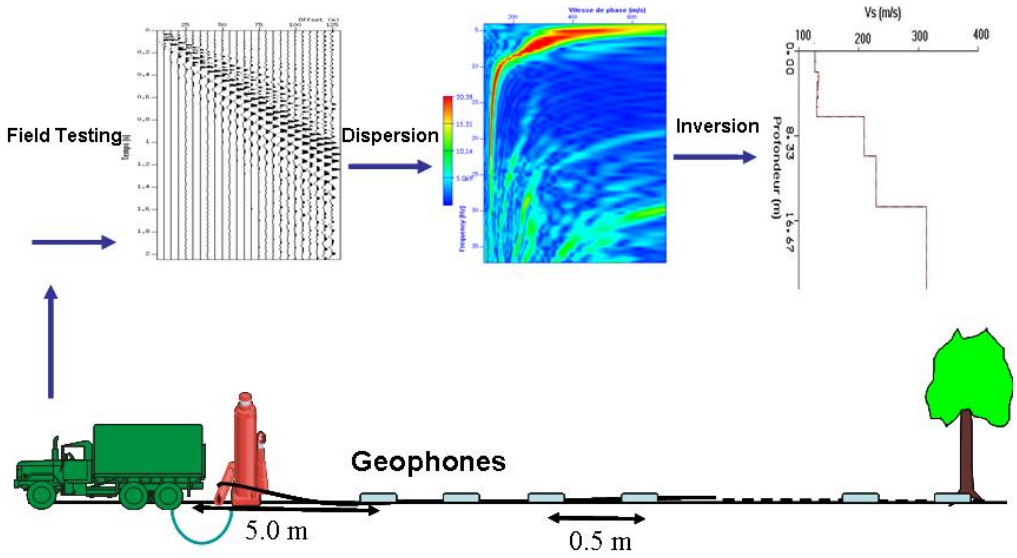


Figure 2 : From acquisition to inversion of surface waves (MASW methodology).

2.1.2. Resulting map description

Seismic experiments on Luxembourg site led to realize 5 profiles covering around 5 ha for almost 200 seismic shots. In addition, 30 penetrometer soudings were performed on the same area (Figure 3a). The soundings were divided into 2 groups with the purpose of (i) calibrating the Vp isovalue for the soil/bedrock limit horizonation (15 soundings) and (ii) validating the obtain soil depth map using this methodology (15 soundings). In this case, the soil/bedrock limit was defined as the boundary between subhorizontal schistosity red material (~50 to 90 cm) and horizontal schistosity white material (from ~1m) (RW limit). On a matter of fact, the change is the schistosity constitute a great mechanical contrast which strongly influence surface-waves. The RW limit is identified as a Qd step around 90 cm on the whole penetrometric data set (Figure 3b).

When interpolating the soil depth data points obtained from the MASW method, a map of the soil depth can be derived. This map can afterwards be compared to the validation dataset, i.e., RW limit depth obtained from the 2nd group of penetrometric data, to estimate the a posteriori uncertainty related to the methodology (Figure 4).

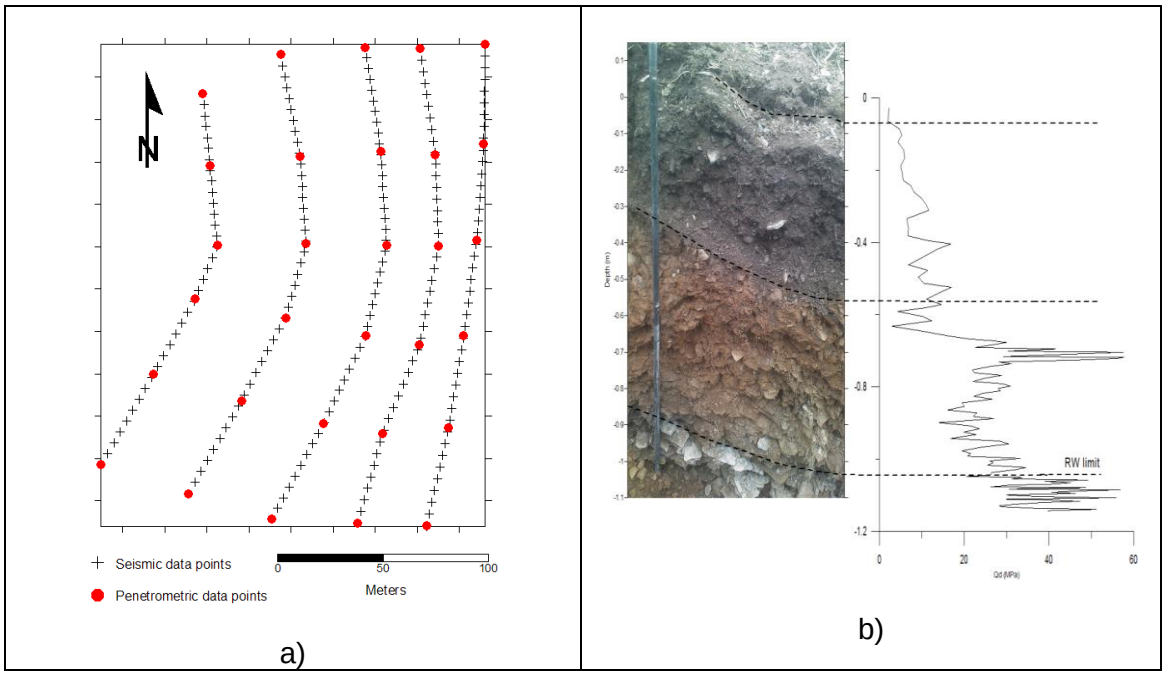


Figure 3 : (a) Location map of seismic and penetrometer data points over the Luxembourg site; (b) Comparison between a penetrogram and trench observations.

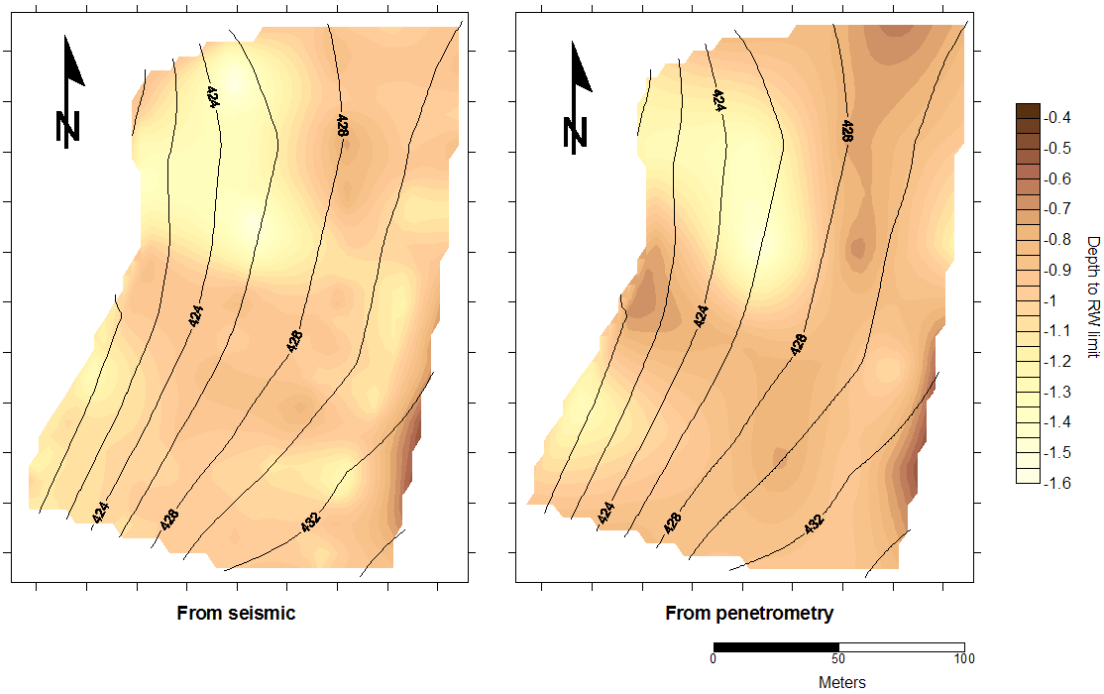


Figure 4 : Comparison between the soil depth obtained from MASW (left) and from the validation penetrometric data set (right).

2.1.3. Performances

- Spatial error

The map of the difference between RW limit depth from MASW and from the validation penetrometric data set allows analyzing the spatial variations of the misfit between the prediction and the validation data set (Figure 5). The prediction error vary between ~-20 cm and ~20 cm. Isolated positive error spots (brown colors on the Figure 5) can be considered as singular values, probably due to a bad picking of the RW limit horizon on the seismic section. For those points, it would be judicious to reiterate one part of the processing workflow. On the other side, we observe that the negative error cloud (yellow colors on the Figure 5) is concentrated in the area of lowest RW limit depth (eastern area), according to the results of Figure 4. On a matter of fact, very low soil depth (lower than ~50 to 70 cm) constitute a limitation in terms of resolution for the seismic method.

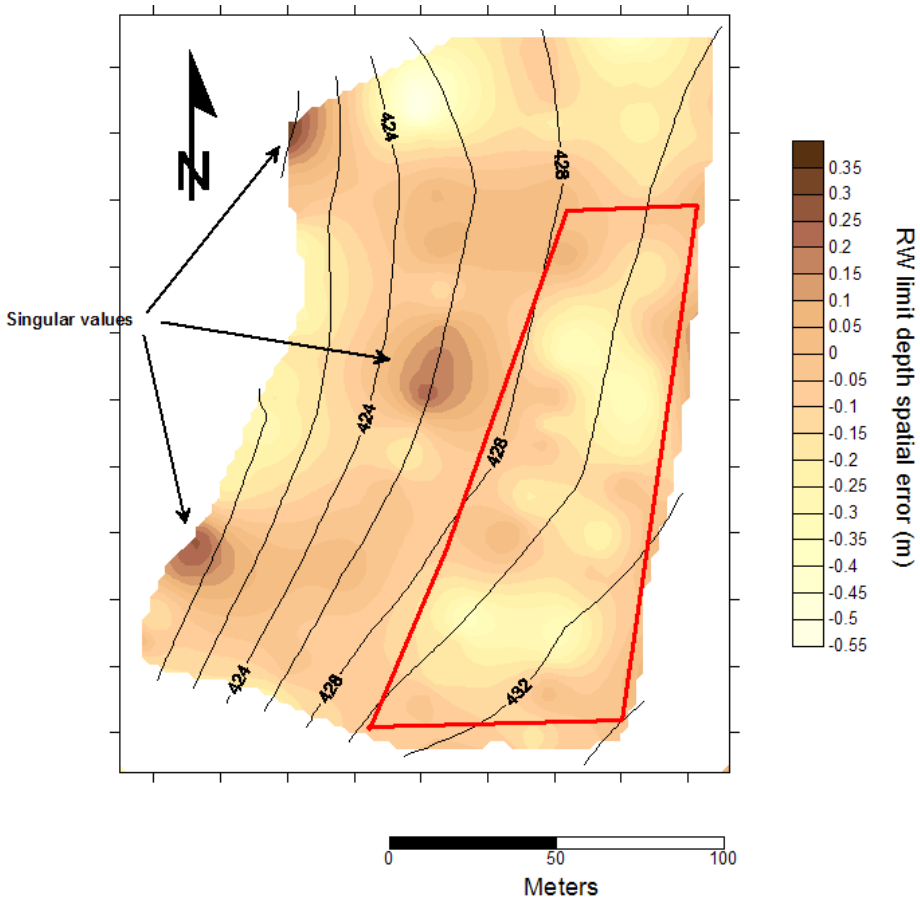


Figure 5 : Map of the RW limit depth spatial error obtained calculating the difference between RW limit depth from MASW and from the penetrometric validation data set.

- RMSE estimation

Linear regression analyzes the relationship between two variables, X and Y. For each subject (or experimental unit), knowing both X and Y, the best straight line through the data has to be found. The goal of linear regression is to adjust the values of slope and intercept to find the line that best predicts Y from X. More precisely, the goal of regression is to minimize the sum of the squares of the vertical distances of the points from the line. The Figure 6 shows the linear regression between the RW limit depth predicted using the MASW methodology and from the validation penetrometric data set at the location of the penetrometric soundings. A significant correlation is observed between these two variables (R^2=0.6255). This means that ~62% of the real RW limit depth should be explained using this linear regression obtained using the MASW methodology. This constitutes a consistency check of the method.

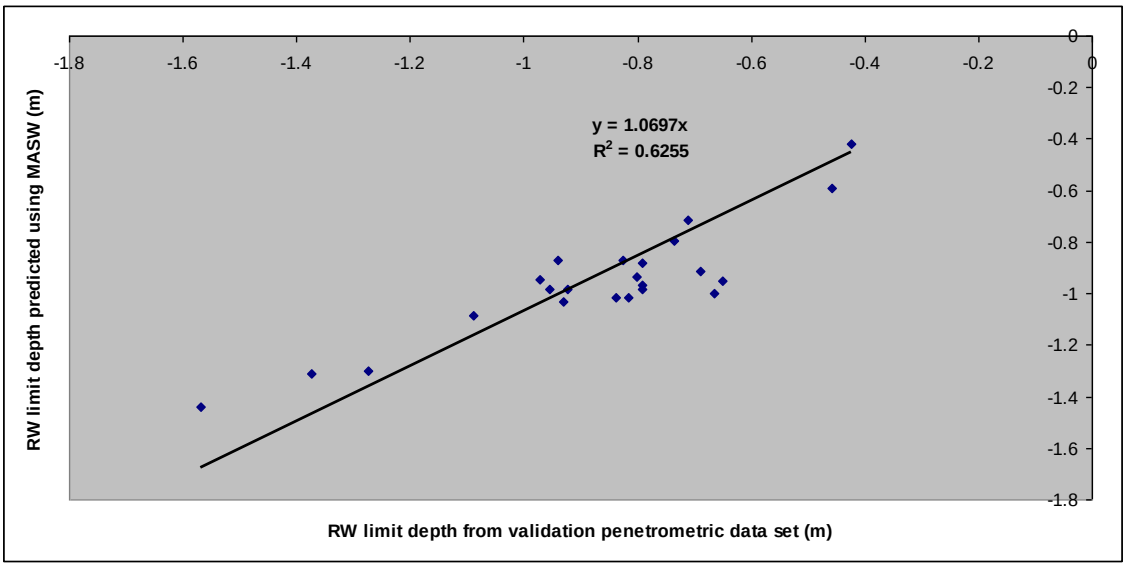


Figure 6 :Plot of the linear regression between the RW limit depth predicted using MASW and from the validation penetrometric data set.

Spatial resolution

- Technical or economical constraints

Timetable and human effort: For this study, the Digisoil seismic system was run on the field by two operators during two days; another day was dedicated to penetrometric measurements and trench observations. This implies that a consequent part of the spatial resolution in predicting soil depth is due to acquisition parameters. On a matter of fact, for example, it would be too expensive and too long to realize a seismic survey with an excessively reduced shot interval. In the current state, the performance of the system is about 0.75 Ha per day with a 4 m shot interval and 20 m between seismic lines.

- Spatial structure and data interpolation

The spatial structures of the soil depth was assessed from variograms estimated along four directions: 0°, 35°, 90° and 135° from geographic north. The variograms were generated from all possible sample pairs in a given direction grouped into classes (lags) of approximately equal distance (Matheron, 1965). The variance (one-half of the mean squared difference) of the paired sample measurements were then plotted as a function of the distance between the samples to provide a means of quantifying the spatial structure of the data. The soil depth obtained using seismic method was then interpolated by ordinary kriging - a geostatistical method that takes into account both the distance and the degree of variation between known data points and relies on the data's spatial correlation structure to determine the weighting values. Ordinary kriging has been shown to perform better for soil parameters than other available methods (e.g. Burgess et al., 1981; Myers, 1994). The interpolations were accomplished by fitting each of the various theoretical variogram models (Quadratic components with scale=0.0551, Length=185, anisotropy ratio=2 and anisotropy angle=23.08°) to the empirical isotropic variogram via the least-square method (Figure 7). The best fit model was used for the interpolation. Data points were then interpolated to a regular 5×5m grid using a full second-order polynomial drift function, as is common practice. With the interpolation process, we then create an output grid of kriging standart deviations which brings informations about the interpolation error. Figure 7 shows that the minimum interpolation error (between 0.01 and 0.05m) is situated at the location of data points and the maximum interpolation error (between 0.05 and 0.1m) is located between data points.

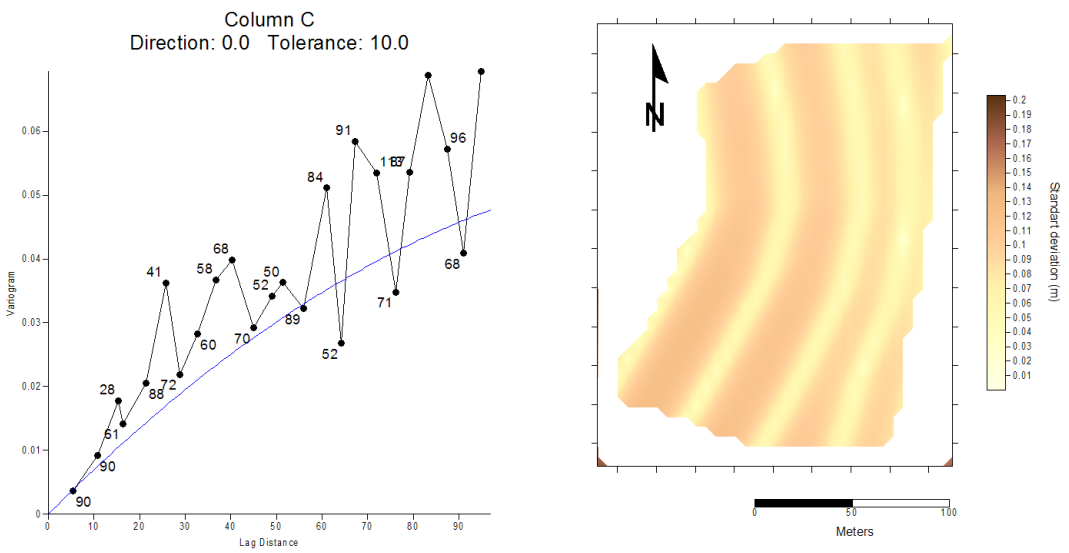


Figure 7 : (Left) Example of a variogram estimated in the direction 0° from geographic north. This variogram is fitted with a theoritical variogram model using a quadratic component with scale=0.0551, Length=185, anisotropy ratio=2 and anisotropy angle=23.08°, (Right) kriging standart deviation map.

2.2. EFFECTIVE CLAY CONTENT MAP

2.2.1. Mapping procedure

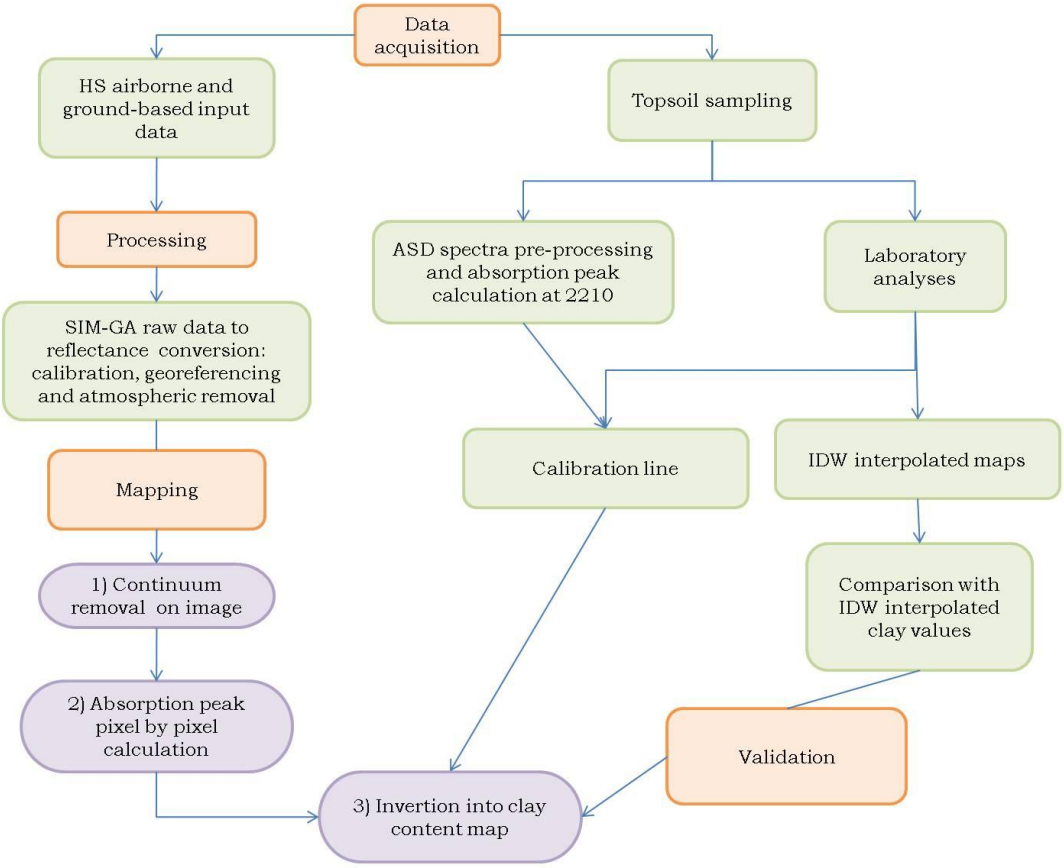
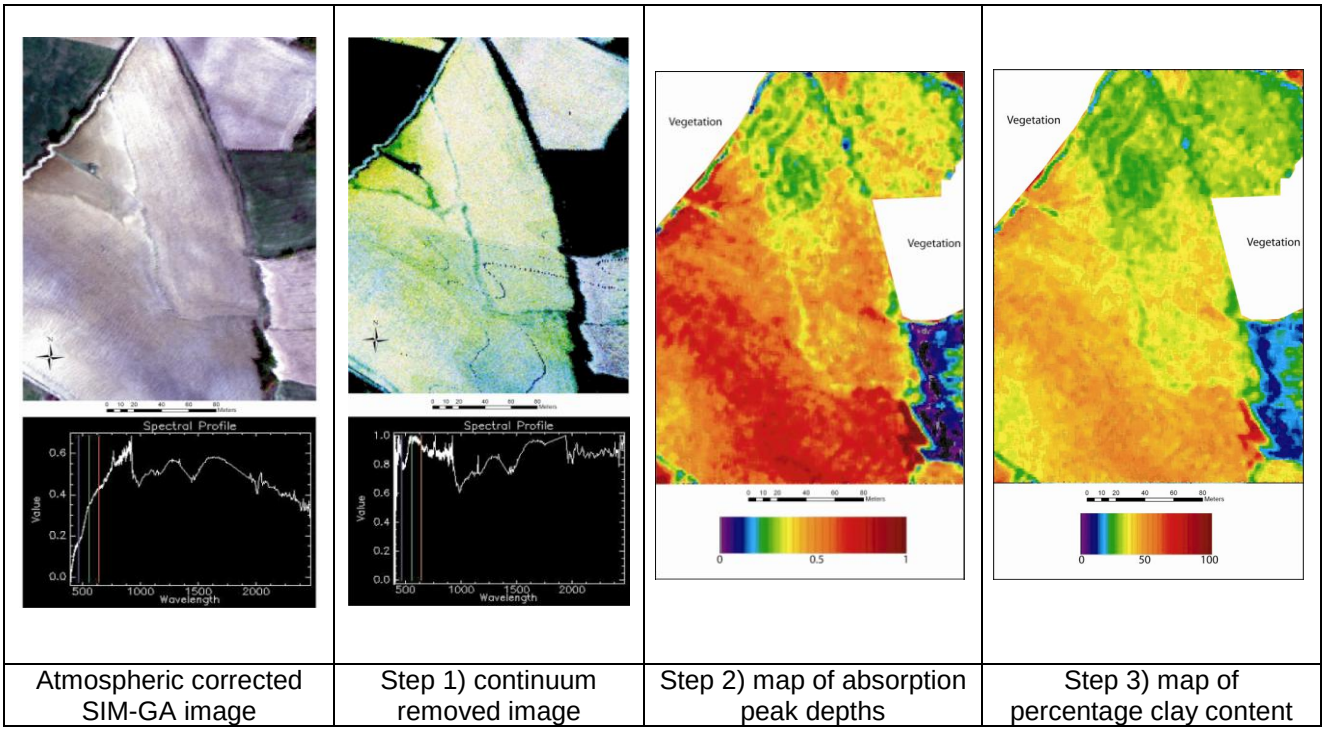


Figure 8 : Complete workflow for the considered map.

The procedure described hereafter allows obtaining a map of clay content of bare topsoils from airborne hyperspectral images, using the continuum removal technique (Clark and Roush, 1984). This procedure is based on the fact that the depth of an absorption feature is strongly related to the abundance of the absorbing material. Continuum removal normalizes reflectance spectra with the aim of allowing a direct comparison among absorption features from a common baseline, minimizing the effect of different scales or observation conditions and assuming that no other no other material has strong absorption features around that specific wavelength.

After CR the absorption peak depth at a certain wavelength is calculated and related to the atomic group responsible for the spectral feature. Clay content and mineralogy influence the short wave infrared portion of the spectrum (1300-2500 nm), but only the peak at 2210 nm can be detected in airborne sensors spectra, which are affected by atmospheric absorption bands.

Data acquired with Hyper SIM-GA sensor form Selex Galileo on September 23, 2009 were processed in order to obtain geo-referenced, calibrated and atmospherically corrected SWIR cubes. Clay absorption peak depth at 2210 nm from ASD indoor spectra, was correlated with total clay content obtained from laboratory analysis, thus obtaining a calibration line. After continuum removal, the absorption peak depth was calculated at 2210 nm, for every pixel of the image, using a dedicated IDL routine. Then, the laboratory relation was used to invert the hyperspectral images into clay content maps. As a matter of fact, as demonstrated by Lagacherie et al. (2008), laboratory-calibrated functions can be applied at map scale to hyperspectral images to map soil properties. This processing workflow represents a summary of the methodology developed in the report D1.3 and based on the different experiments included in D1.1-2 and 2.1-2.3 and is further explained in Figure 8 and Figure 9. Elaborations on images were performed using ENVI software (ITT VIS, Boulder, CO).



Atmospheric corrected SIM-GA image	Step 1) continuum removed image	Step 2) map of absorption peak depths	Step 3) map of percentage clay content

Figure 9 : From acquisition to inversion of hyperspectral data.

2.2.2. Resulting map description

The map of Figure 10 shows in false colors the spatial distribution of clay mineral content, resulting from manipulation on hyperspectral images, with orange-red corresponding to higher percentages and green-blue corresponding to lower percentages. A mask was applied to the neighbouring grassland. Lower values are concentrated in the northern part of the parcel, which is topographically more elevated

while an increasing trend towards the south (i.e.: parallel to the flow direction, towards the bottom) can be observed.

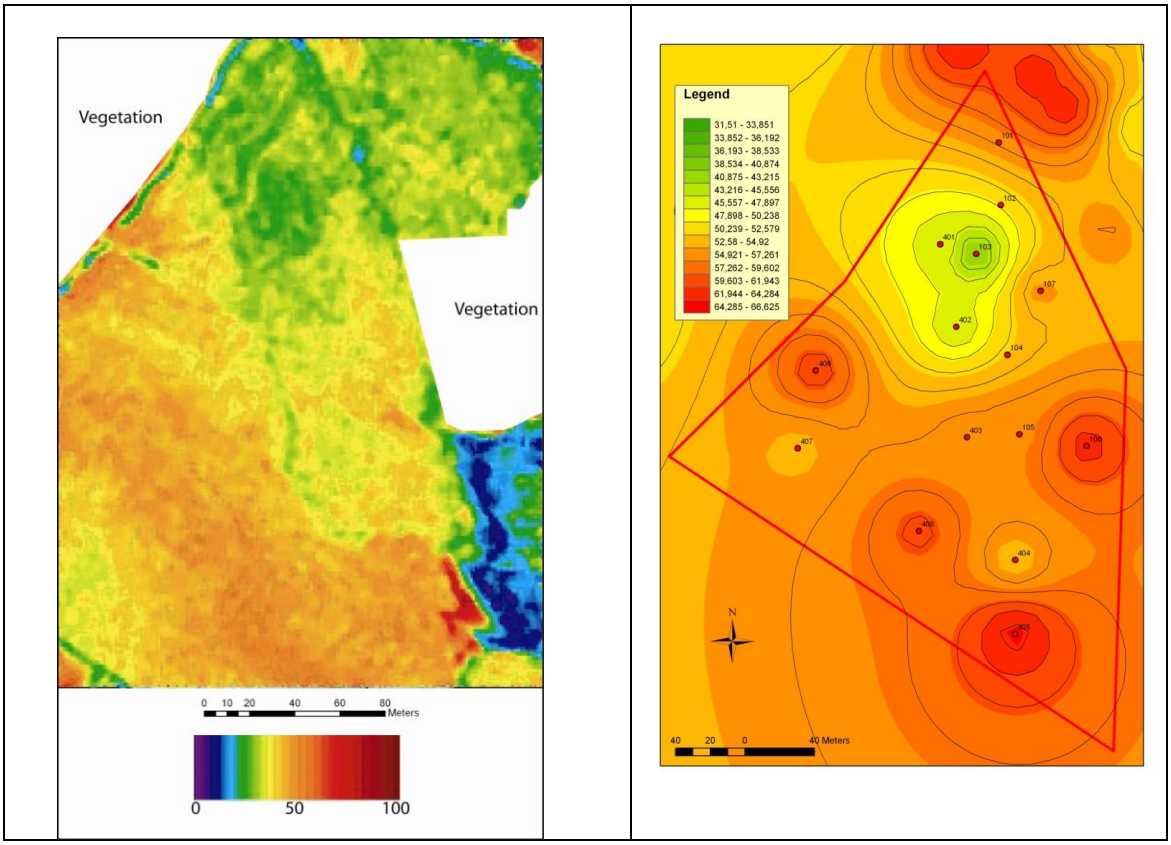


Figure 10: Comparison between hyperspectral-derived clay map and IDW interpolation of laboratory values.

2.2.3. Performances

- Validation and RMSE estimation

The clay content map deduced from hyperspectral data was compared with the interpolated laboratory values map of the total clay mineral concentration, obtained using the Inverse Distance Weighting algorithm, for validation (Figure 10). The general trend, described in the previous paragraph, can be retrieved in this second map, showing a good agreement between the predicted and the observed clay distribution tendency.

A reliability test of the described procedure can be performed using linear regression, which analyzes the relationship between two variables, X and Y, finding the best straight line through the data, with the goal of minimizing the sum of the squares of the vertical distances of the points from the line. The linear regression was calculated

between hyperspectral-predicted clay values and their correspondent clay mineral content, observed in the sampling data set (Figure 11). Correlation between the two variables is satisfactory, with a determination coefficient of R^2=0.599; as a consequence, our model can explain about 60% of actual clay content in the top level of soils in the study area.

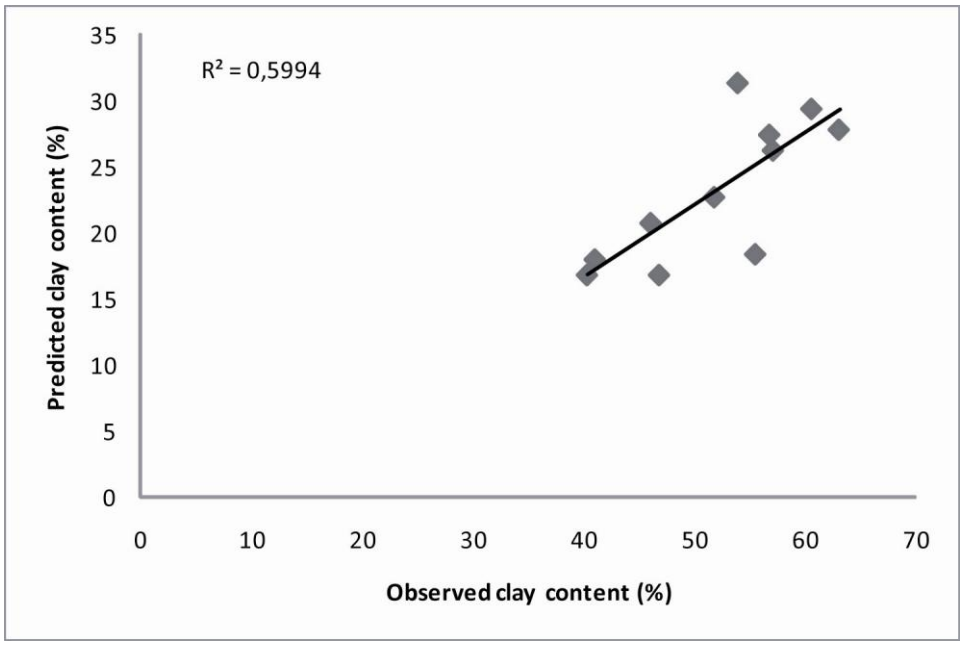


Figure 11:Plot of the linear regression between clay content predicted using SIM-GA and from the validation sampling data set.

The accuracy of clay content estimation and clay maps could be improved through the use of 1414 and 1914 nm clay absorption peaks, which show better correlations with the laboratory dataset, with respect to the used one (Figure 12).

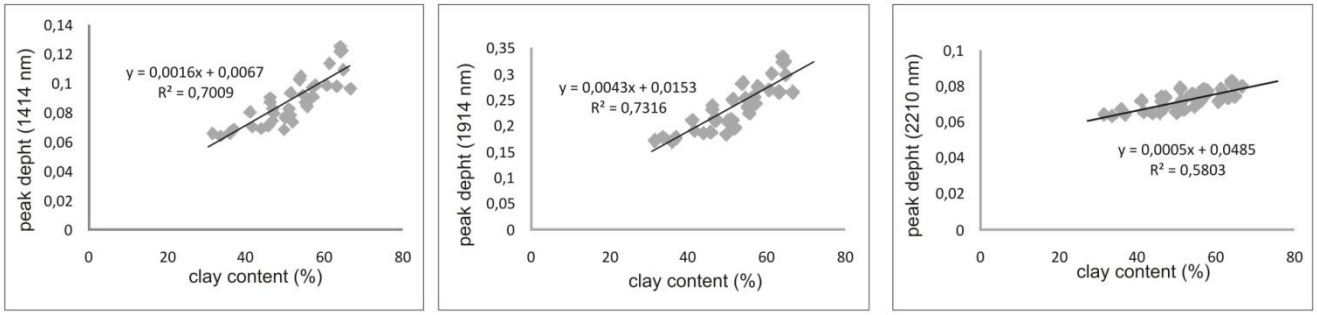


Figure 12: Correlation between clay content and absorption peak at 1414 nm, 1914 nm and 2210 nm.

2.3. STONE CONTENT MAP

2.3.1. Final workflow for the considered map

- Input data

Input data are electrical resistivity measurements from the three arrays of the ARP. From the first array data, a standard deviation map has been calculated by the following method: each pixel is of 50 x 50 cm in size, and its value is the standard deviation of the electrical resistivity values calculated in a circle centred on the pixel with a radius of 5 m (see paragraph 2.2.1, page 29 in D3.2).

- Calibration data

The stone content has been measured on soil cores sampled by a driller over the first 30 cm (see the locations of the sampling in paragraph 2.1, page 17 in D3.2.). The dataset was splitted in two parts: four samples were selected randomly (except in the North-East anthropogenic part) and used as a calibration set.

A linear relationship was determined between the standard deviation of the electrical resistivity and the stone content (Figure 13). As expected, a higher stone content leads to the higher variability in electrical resistivity. To check the stability of the relationship, several random datasets of 4 samples were selected and analysed with the standard deviation of the electrical resistivity. The slope of the relationship was quite stable, equal to about 0.65. The quality of the relationship could be improved by taking into account more numerous samples. We have here decided to use the maximum data for the validation, and, as a consequence, the minimum data for the calibration.

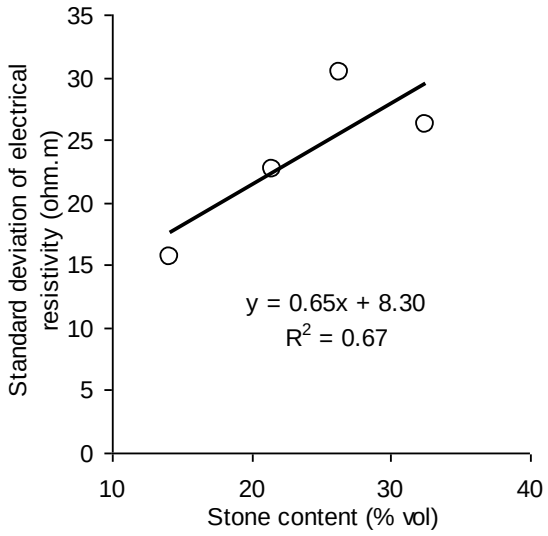


Figure 13 : Calibration relationship to estimate the stone content from electrical resistivity data and direct measurements in the field.

- Validation

The remaining 26 samples of the field dataset were used for the validation.

- Spatial resolution

Due to the high spatial resolution of the ARP (one point each 50 cm on a line, and lines separated by 2m), the initial geophysical measurements can be considered as an exhaustive dataset. During the data processing a 5x5 m grid was build. The final map results from a weighted inverse distance interpolation.

- Resulting map description

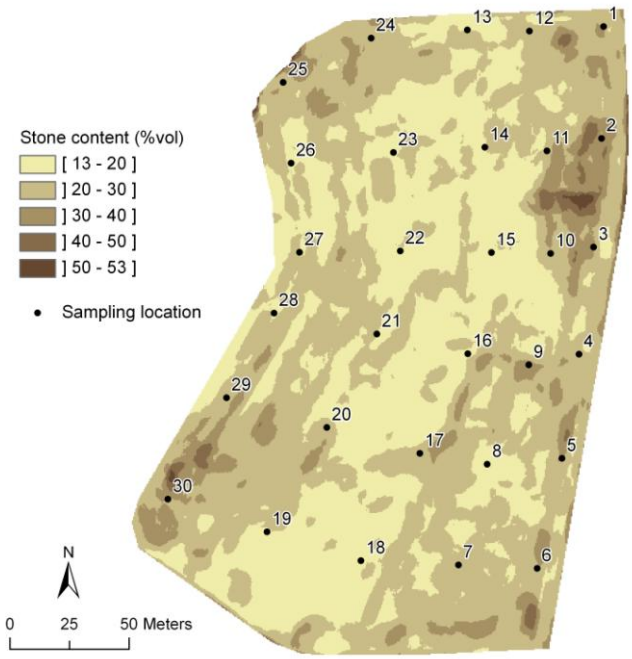


Figure 14 : Modelisation of the stone content

Figure 14 presents the stone content in the studied area. It varies between 13 and 53% and is higher in the South-East part and the Western border of the studied area.

- RMSE Estimation

The stone content measured on soil cores (validation dataset) was compared to the stone content measured by the model (Figure 15). Except at some locations in the anthropogenic part, the estimation is rather satisfying, but the model slightly overestimates the real stone content. The Root Mean Square Error has been calculated for the whole calibration dataset, except data from the anthropogenic part: it was equal to 9.7%. An estimation of the stone content at about 10% can be considered

as a good estimation, with a precision of the same order of magnitude as visual estimations in the field by a pedologist.

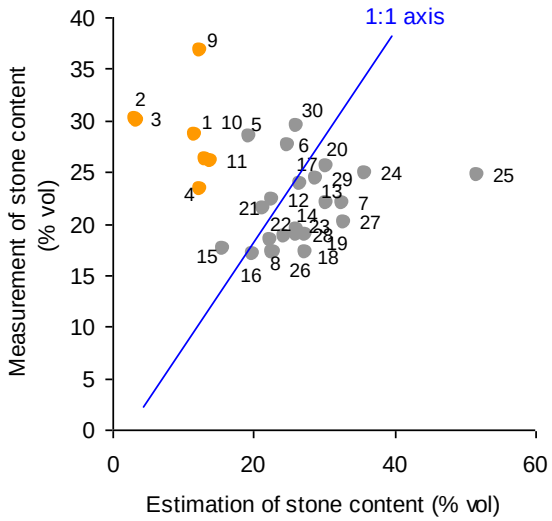


Figure 15 : Validation of the model. The orange dots, far from the 1:1 line, correspond to points located in the anthropogenic North-East part and the stone content at these locations can not be estimated by the model.

- Possible improvements of the method

The methodology can then be used extensively to determine the stone content at the scale of a parcel. Improvements would consist in better measurements of the stone content on the calibration points, by taking into account larger volumes of soil that would be more representative of the real stone content.

2.4. WATER CONTENT MAP

GPR surveys were carried out using both far-field (off-ground) and near-field (on-ground) systems (Figure 16). Depending on the configuration, different processing procedures were used to retrieve soil water content from the radar data. We used the Lambot et al. (2004, 2006) full-wave inversion method for far-field radar measurements. In that case, the GPR antenna effects are filtered out by complex linear transfer functions determined by antenna calibration in the laboratory and the filtered frequency-domain signal is converted to the time-domain using the Fourier transform. Then, focusing on the surface wave reflection, the full-wave radar model is inverted to retrieve the soil surface dielectric permittivity. Finally, the soil dielectric permittivity is converted to volumetric water content using Topp's (1980) equation. For on-ground GPR, the direct ground wave (DGW) method is used. The single trace analysis (STA) for DGW allows retrieving the soil dielectric permittivity in the shallow soil layer (less than 0.5 m), depending on the antenna center frequency and soil moisture (Figure 16). In order to retrieve the soil dielectric permittivity, the propagation time from the transmitting antenna (Tx) to the receiving one (Rx) should be calculated. In this case,

the Tx-Rx offset is fixed and the ground propagation time is directly related to soil dielectric permittivity and, consequently, to soil water content:

$$\varepsilon_r = \left(c.\frac{\Delta t}{\Delta x} \right)^2$$

where ε_r is soil relative dielectric permittivity (-), c is the speed of light in vacuum ($m.s^{-1}$), Δt is propagation time (s), and Δx is the Tx-Rx offset (m).

	Luxembourg	Mugello
Input data	Off-ground frequency-domain GPR/ On-ground time-domain GPR	Off-ground frequency-domain GPR/ On-ground time-domain GPR
Processing	Surface-reflection waveform inversion/ Single Trace Analysis (DGW)	Surface-reflection waveform inversion/ Single Trace Analysis (DGW)
Calibration	Antenna calibration/ none	Antenna calibration/ none
Soil/bedrock limit	Roughness	Roughness
Mapping	Interpolation of the data points using kriging technique	Interpolation of the data points using kriging technique

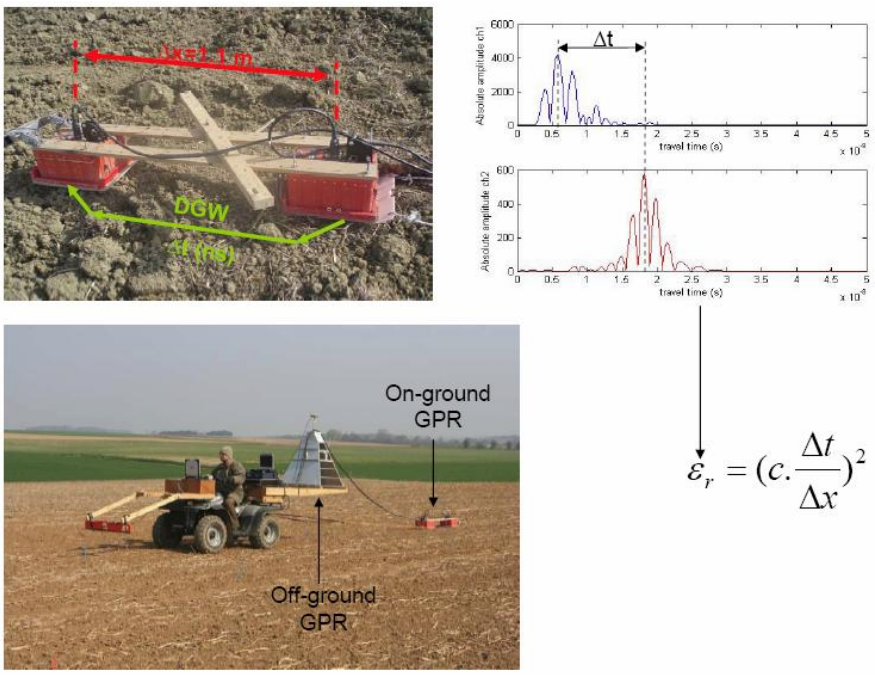


Figure 16. Data acquisition from the field and principles of the STA method.

2.4.1. Resulting map

Off- and on-ground GPR surveys were carried out at the Luxembourg site during different weather conditions: dry and wet. In the dry condition, 11 transects were performed with about 1300 traces for off-ground and 14000 traces for on-ground over ~5 ha field. In the wet conditions 13 transects were performed with about 1700 traces for off-ground and about 24500 traces for on-ground. Soil core sampling was carried out at 30 locations over the entire field close to the GPR transects for ground-truth volumetric water content measurements at different depths (0-10, 10-20, 20-30, 30-40, and 40-50 cm). Three of these ground-truth points were sampled during the dry day while sampling at the other 27 locations was performed under wet soil conditions. Applying the STA method to on-ground measurements and surface-reflection inversion to off-ground measurements allowed to retrieved the soil dielectric permittivity from each trace. Then, the soil volumetric water content estimates from Topp's equation were interpolated over the entire field area using kriging. Figure 17 shows the soil water content maps from both off- and on-ground GPR data in dry and wet conditions. It is worth noting that the off-ground derived map represents only the surface moisture (top ~2-3 cm). The characterization depth for the on-ground map is difficult to specify accurately as it depends on both the operating center frequency and soil moisture. In our case, the characterization depth is expected to be around 30 cm for the dry soil and around 10 cm for the wet conditions.

2.4.2. Performances

- Spatial error

The off-ground GPR allows retrieving surface soil water content and the on-ground GPR (DGW method) allows retrieving the shallowest less-lossy layer soil water content depending on moisture conditions. Figure 10 shows the two water content maps derived from off- and on-ground GPR surveys under wet conditions, corresponding to the collection of the major part (27 out of 30) of the ground-truth samples. Although both GPR techniques provided consistent soil moisture maps, comparison with the ground-truths is complicated by the different characterization scales and depths as well as by the great local variability of soil moisture (Figure 18).

- RMSE estimation

We have 30 ground truth points in different depths to validate the GPR derived soil water content. Only 3 points of them correspond to the dry conditions and the remaining points were sampled during the wet conditions. Therefore, we are only able to validate the wet derived water content estimates. To quantify the error between the 0-10 cm depth ground-truth soil moisture and both GPR derived water content, we used the nearest neighbor of GPR data to the location of ground-truth data. Then linear regressions were used to analyze the relationships between the two variables. Figure 19 shows the plots of data validation for both GPR techniques. The RMSE is about 3.6% for off-ground GPR and 4.9% for on-ground related to the best fit line. Discrepancies between ground-truth measurements and GPR soil water content estimates may arise for several reasons. First, a part of the differences would result

from the contrasted characterization depths and scales of GPR compared with ground-truth measurements (100 cm³ cylinders).

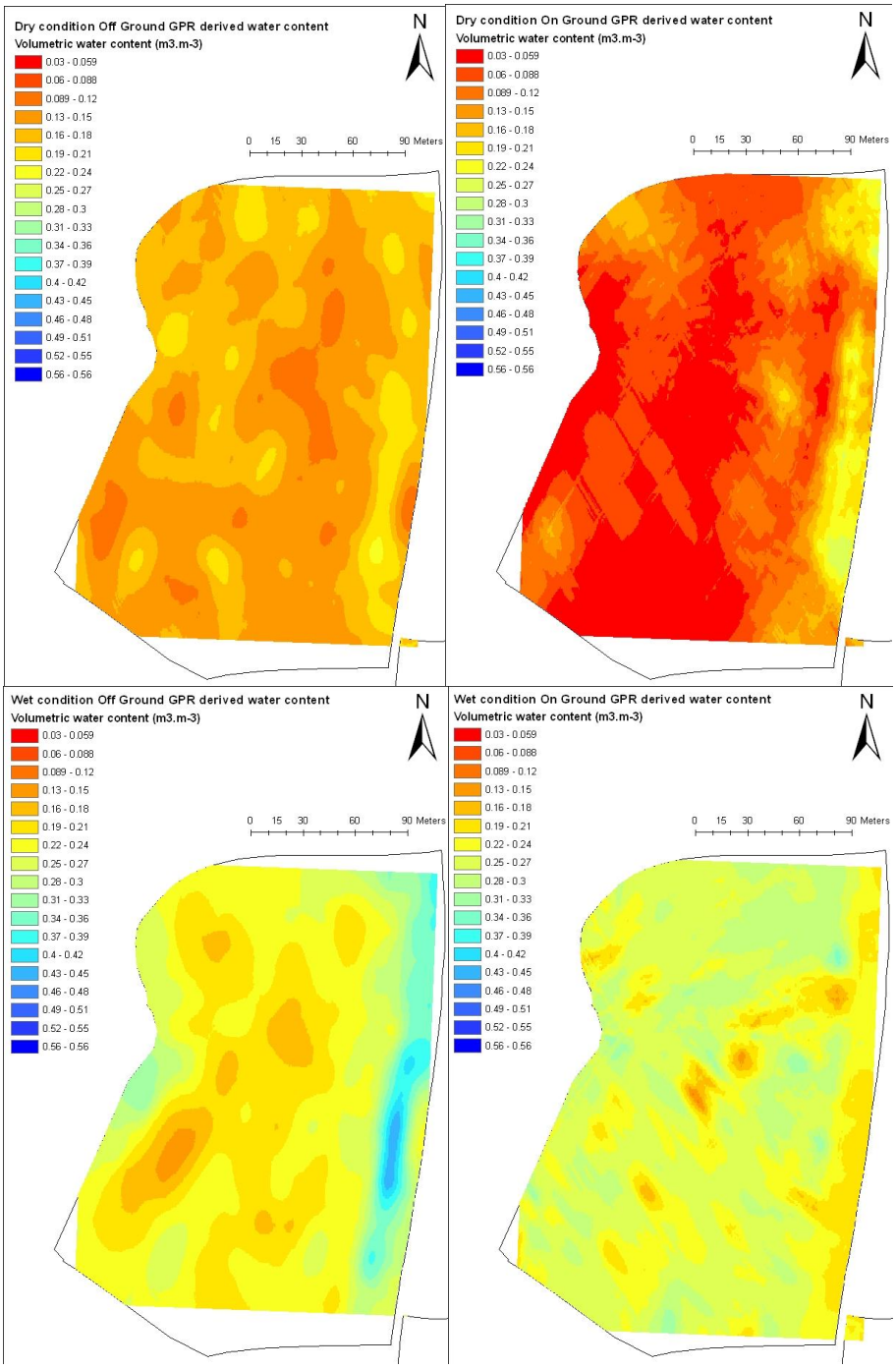


Figure 17 : Soil water content maps derived from off-ground (left) and on-ground (right) GPR data for dry (top) and wet (bottom) conditions at the Luxembourg site.

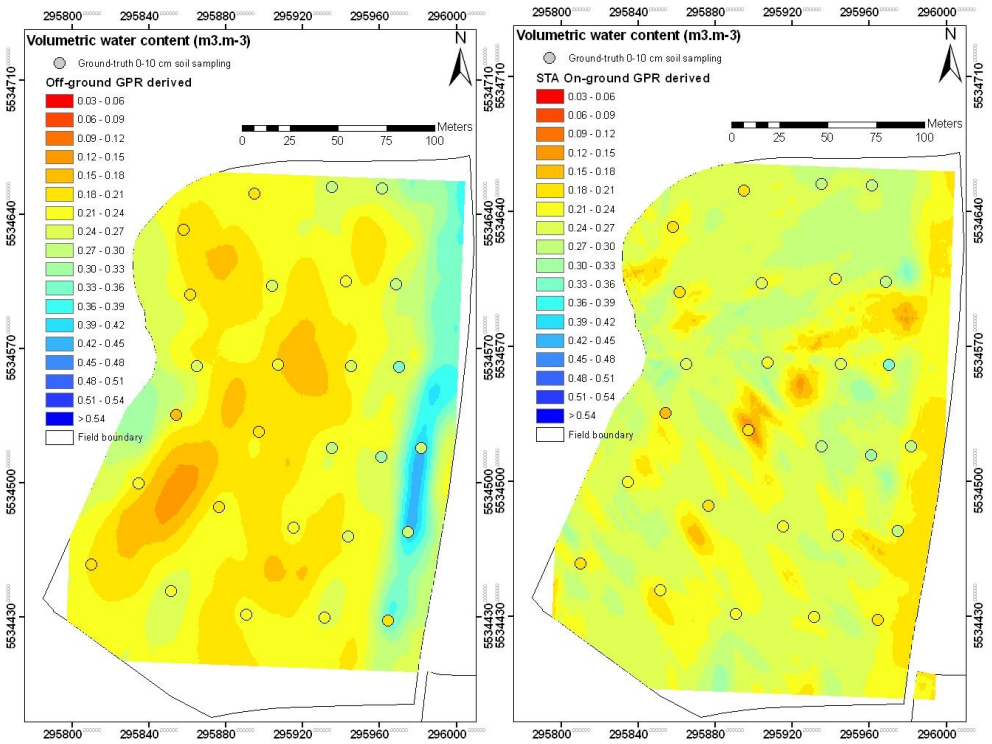


Figure 18 : Wet condition soil moisture maps derived by both off- and on-ground GPR techniques compared with ground truths in shallowest 10 cm depth.

Furthermore, while GPR surveys were completed within about one hour and a half, soil core sampling over the entire field took one complete day and soil water content is likely to vary during that time as a result of infiltration, especially within the hours following a rain event as it is the case for this field campaign. Moreover, infiltration is likely to evolve spatially as a result of spatial variation of topography and clay content over the field area.

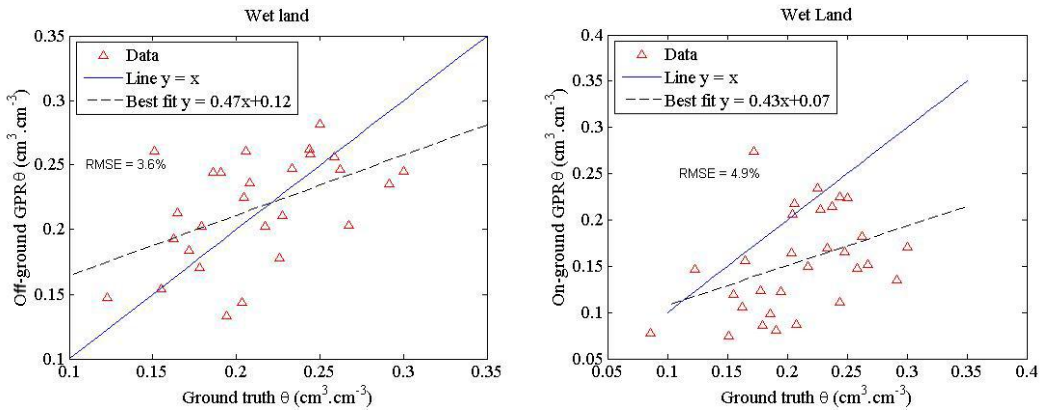


Figure 19 : Plot of the linear regression for soil volumetric water content between the off-ground (left) and on-ground (right) GPR techniques and soil core sampling.

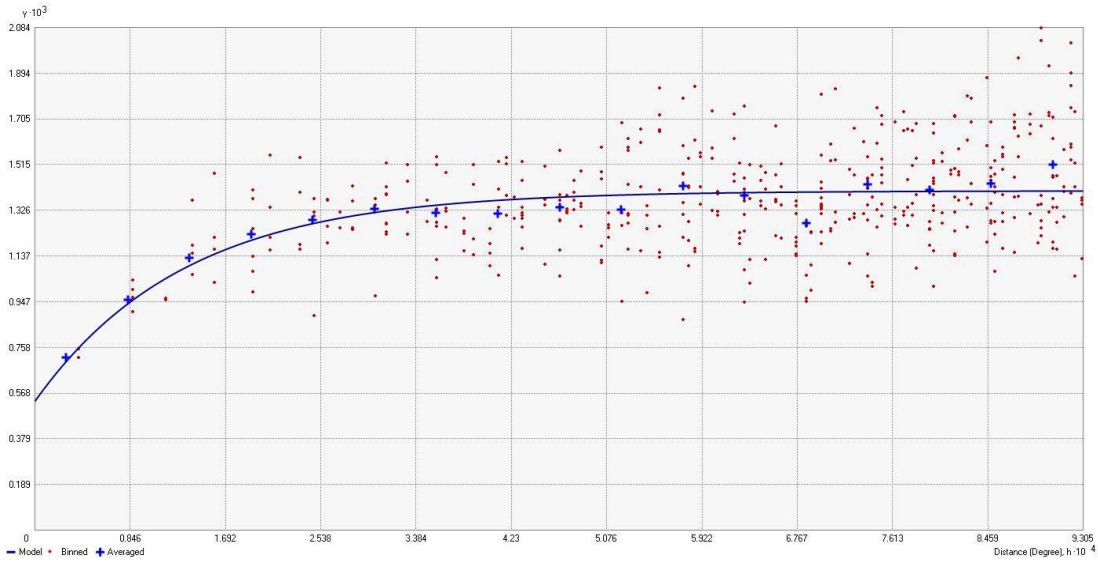
2.4.3. Spatial resolution

- Technical or economical constraints

Timetable and human effort: For proper high-resolution (1 m) mapping of soil properties, GPR measurements should be carried out at a velocity lower than 2 m/s. On the Luxembourg site, the average velocity of the platform during the measurements was 1.9 m/s Therefore, each coverage of the total field area took between half an hour and one hour for one person.

- Spatial structure and data interpolation

The derived soil moisture by both GPR methods was interpolated by ordinary kriging. The interpolations were accomplished by fitting each of the various theoretical variogram models to the empirical isotropic variogram via the least-square method. Figure 20 shows an example of variogram and the spatial error based on standard deviation of soil water content entire the field.



a)

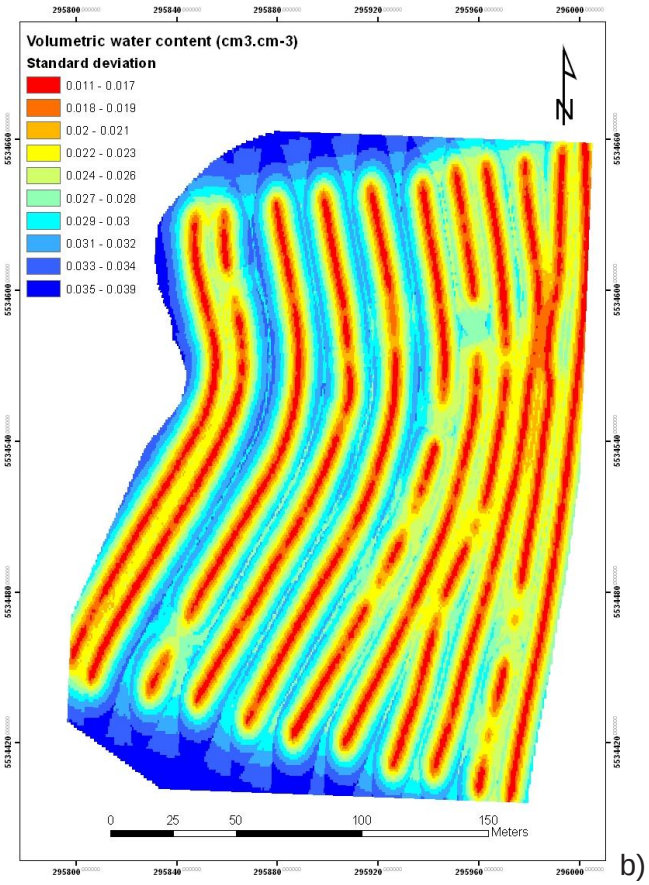
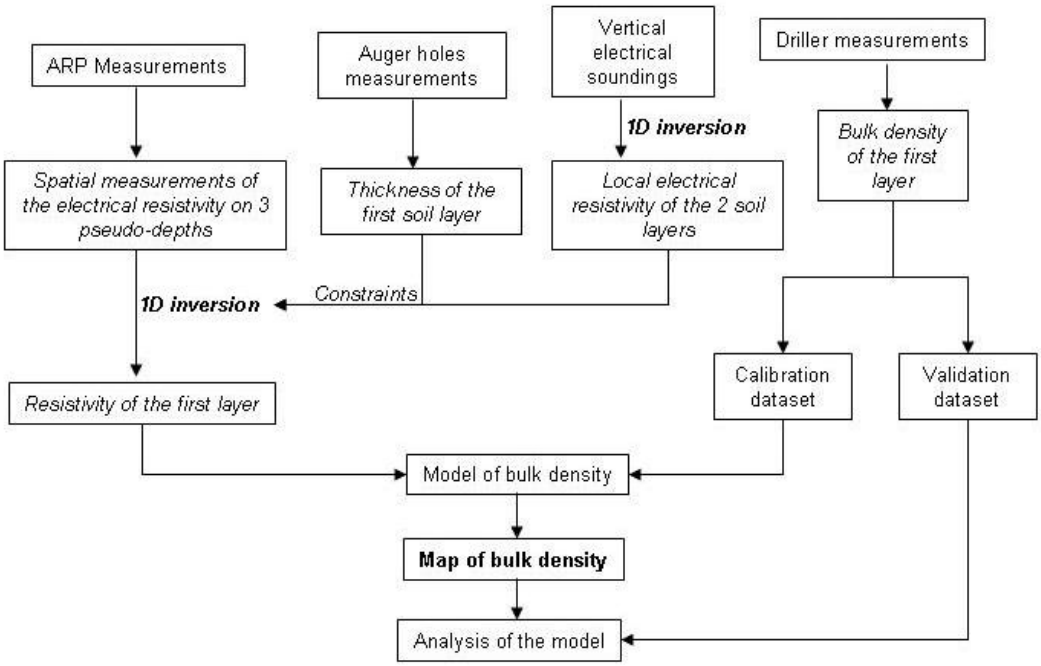


Figure 20 : Plot of the linear regression for soil volumetric water content between the off-ground (a) and on-ground (b) GPR techniques and soil core sampling.

2.5. BULK DENSITY MAP

2.5.1. Final Workflow



A local 1D inversion was calculated for all the ARP data, by using the QWin1D model from J. Tabbagh, to estimate the real resistivity of the cultivated soil horizon. A double-layer initial model was used. According to the auger holes measurements, the thickness of the first layer was taken equal to about 30 cm and the thickness of the second layer was considered as infinite. Outside the anthropogenic part, the resistivity was taken equal to 150 ohm.m in the first layer and to 1000 ohm.m in the second layer. These values come from the 1D inversion of some local vertical soundings.

In the anthropogenic part, the thickness of the first layer was taken equal to 60 cm and the resistivity was equal to 300 ohm.m, according to the apparent resistivity data from the ARP measurements. The resistivity of the second layer (thickness: infinite) was taken equal to 150 ohm.m.

- Calibration and validation datasets.

The bulk density was measured on soil cores sampled by a driller over the first 30 cm (see the locations of the sampling in paragraph 2.1, page 17 in D3.2.). The dataset was splitted in two parts: 15 samples were selected randomly for the calibration (Figure 21), and 15 were used as a validation dataset.

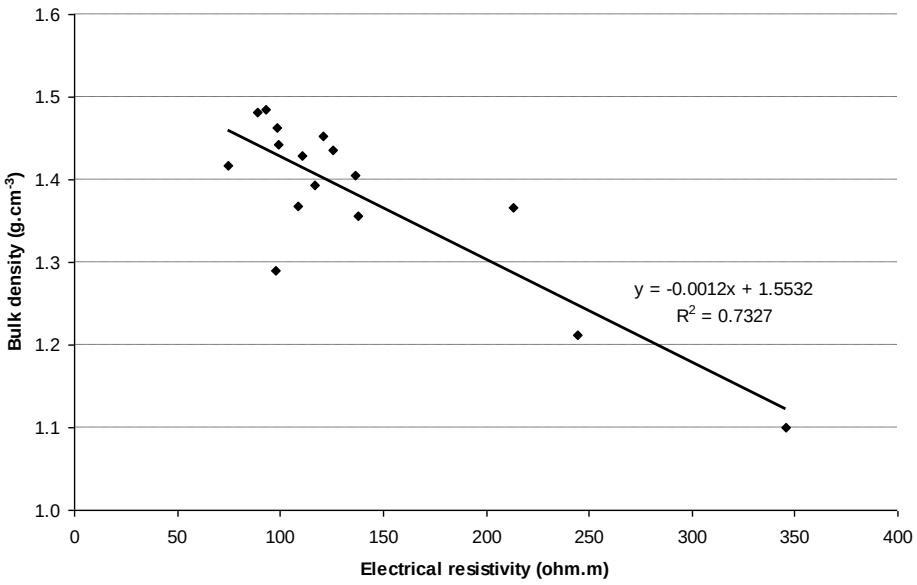


Figure 21 :Calibration plot for the bulk density estimation from ART measurements

A relationship between the real electrical resistivity of the first layer and the bulk density of the cultivated horizon was determined and used as a model (Figure XX). The determination coefficient of the model was equal to 0.73 but we have to consider that the model was strongly influenced by 3 points of low bulk density and high electrical resistivity. The model was then applied over the whole studied area. A weighted inverse distance interpolation was used to create the final map. Finally, the modelled bulk density values were compared to the validation data.

- Resulting map description and RMSE estimation

Figure 22 presents the real resistivity map of the cultivated horizon and the associated error map. As expected, the real resistivity was higher in the anthropogenic part, and in the south-East part of the studied area. The lowest resistivity was observed in the central part. The error associated to the inversion ranged from 0.5 to 8 %, with a mean value of 4%, which means that the inversion can be considered as satisfying.

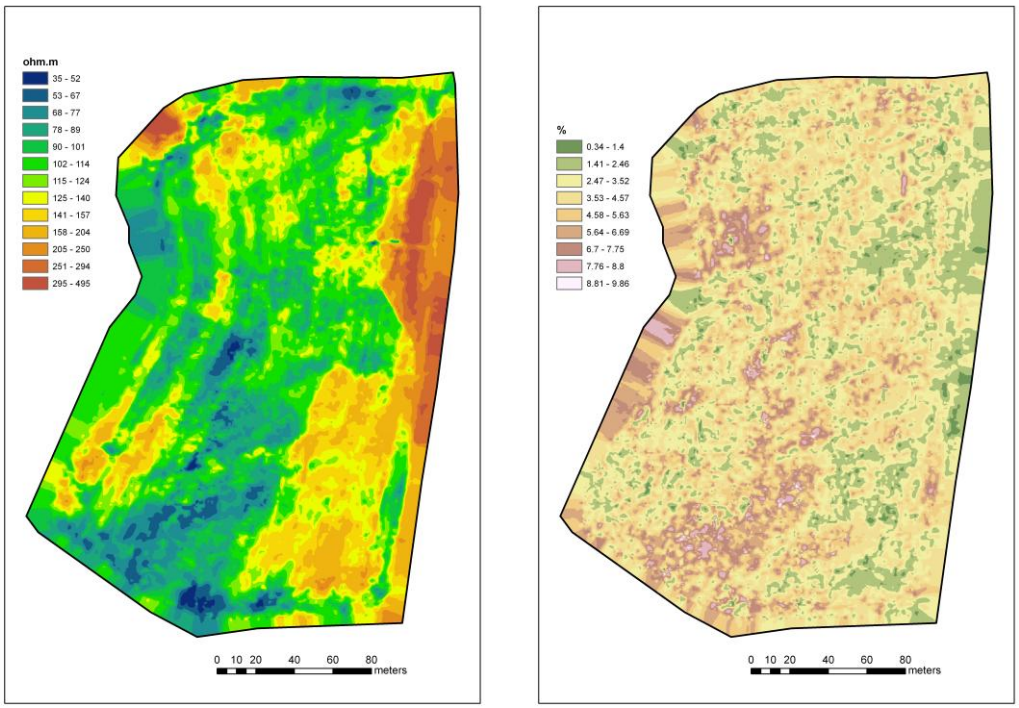


Figure 22: a) real resistivity map of the cultivated horizon b) associated inversion error map

The associated bulk density map is presented in Figure 23. As expected, the lowest values of bulk density (about 1-1.2) were observed in the anthropogenic part, whereas the bulk density was higher elsewhere. The mean value of the bulk density was equal to 1.41 for the estimation and 1.44 for the measurements. The RMSE associated to this map was equal to 0.17 only. Nevertheless, this map has to be analysed with caution insofar as the measurements of bulk density in the field, i.e. the calibration and validation sets, were difficult due to the high stone content in the studied area.

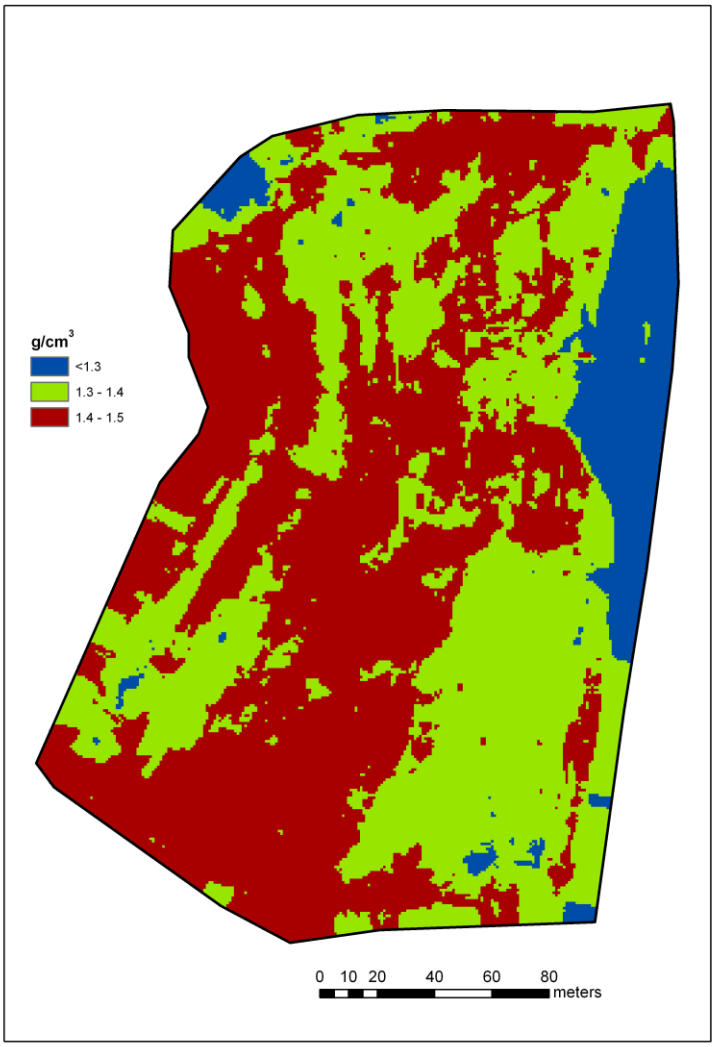


Figure 23 : Bulk density map

3. Second order soil maps

3.1. CLAY CONTENT

3.1.1. Final workflow for the considered map

Soil electrical conductivity of low-salinity soils is mainly affected by soil water content and clay content. Rhoades et al. (1976) proposed the following empirical model to relate soil electrical conductivity to soil physico-chemical properties:

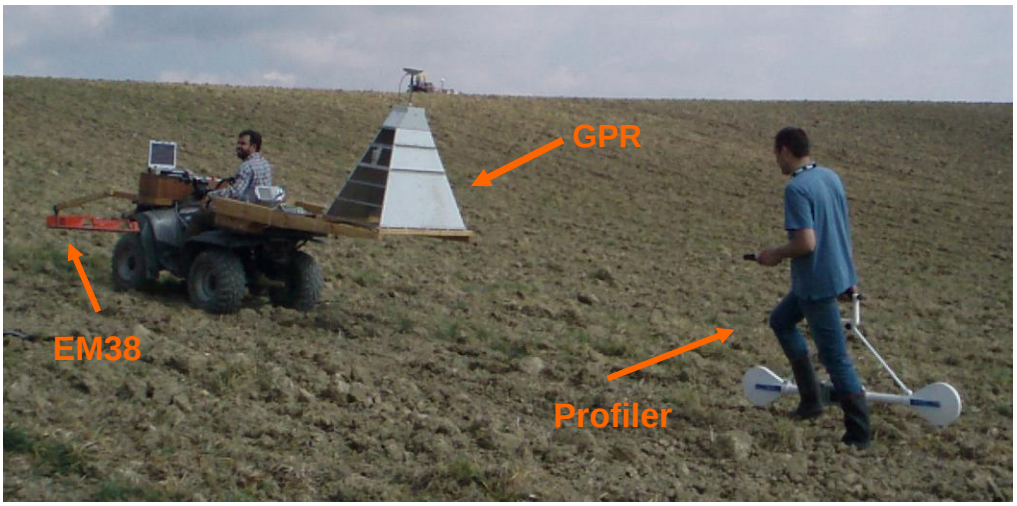
$$\sigma = (a\theta^2 + b\theta)\sigma_w + \sigma_s$$

where σ is the bulk soil electrical conductivity ($S.m^{-1}$), θ is the soil water content ($m^3.m^{-3}$), σ_w is the soil solution electrical conductivity ($S.m^{-1}$), σ_s is the electrical conductivity of dry soil ($S.m^{-1}$), and *a* and *b* are soil specific empirical parameters. In this equation, σ_s may be expressed as a function of the soil clay content. We used this model to estimate clay content from EMI measurements considering the empirical parameter values (*a*=1.382, *b*=-0.093) found by Rhoades et al. (1976) for a soil comparable to that of the main soil unit observed at the study site and assuming σ_w= 0.05 $S.m^{-1}$. Water content estimates from GPR measurements were used to remove the effect of θ on measured soil electrical conductivity and provide estimations of σ_s. For a subsample of the ground truth points (0-10 cm and 20-30 cm clay content measured from soil samples collected along the EMI and GPR transects), relationships were then established between σ_s estimates and clay content measurements. Finally, these calibration relationships were applied to estimate soil clay content at each EMI measurement point and these estimations were compared with the complete ground truth measurement data set for validation.

	Luxembourg	Mugello
Input data	EMI-GPR	EMI-GPR
Processing	GPR: full-wave inversion → ε_r → θ EMI: standardization of measured electrical conductivity to reference temperature of 25°C (Sheets and Hendrickx, 1995)	GPR: full-wave inversion → ε_r → θ EMI: standardization of measured electrical conductivity to reference temperature of 25°C (Sheets and Hendrickx, 1995)
Calibration	Ground truth clay content	Ground truth clay content
Mapping	Interpolation of the data points using kriging	Interpolation of the data points using kriging

The GPR-EMI setup and the workflow are presented at Figure 24.

(a)



(b)

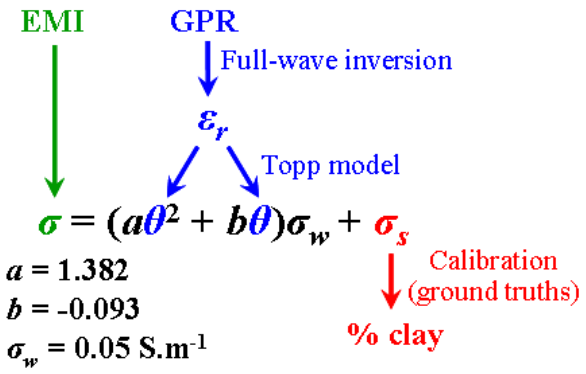


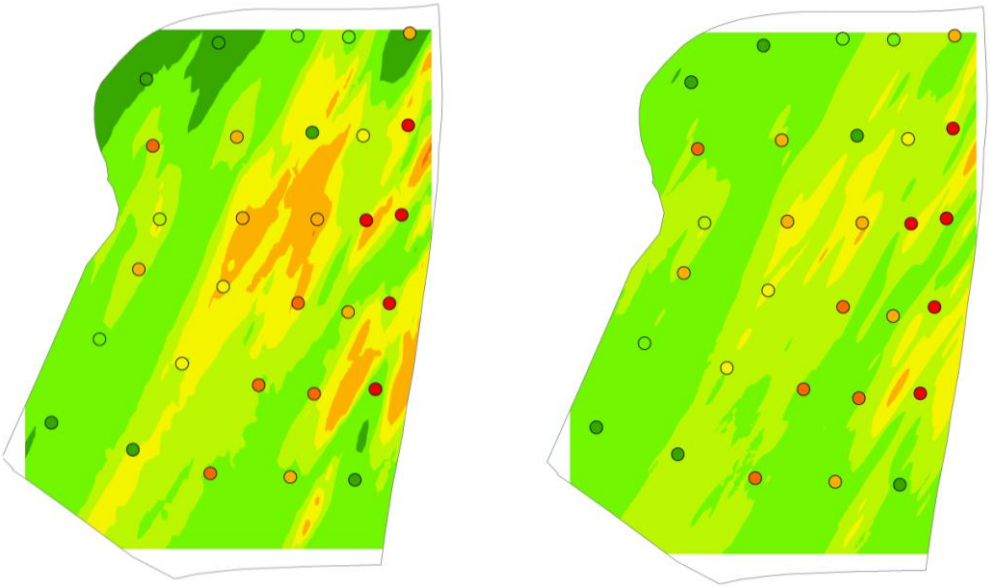
Figure 24: (a) GPR-EMI setup and (b) workflow for clay content estimation

3.1.2. Resulting maps

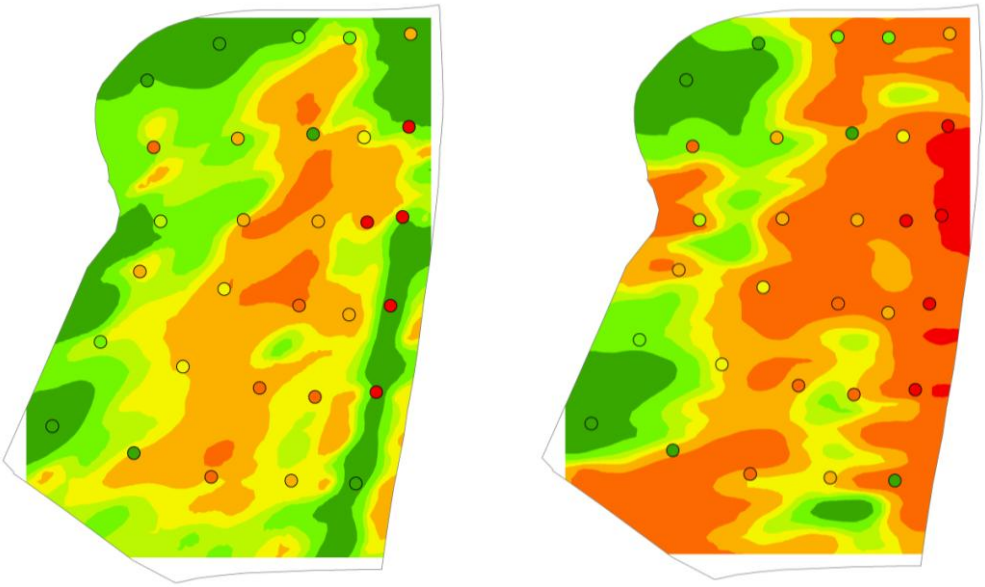
For clay content within the 0-10 cm soil layer, these analyses were performed using horizontal dipoles measurements as they present high sensitivity to the soil surface properties. Data sets from both the first (dry conditions) and second (wet conditions) measurement days were considered in order to compare clay content estimates using data from contrasted soil water content conditions. For clay content below the plough layer (20-30 cm), vertical dipole measurements were used as this configuration presents higher sensitivity to the deeper soil layers, only one set of EMI data (second day of measurements) is available in this case. For each data set, clay content was determined with and without correcting electrical conductivity values for water content (using the Rhoades model), in order to investigate the effect of this correction on clay content estimates. The resulting maps are presented in Figure 25.

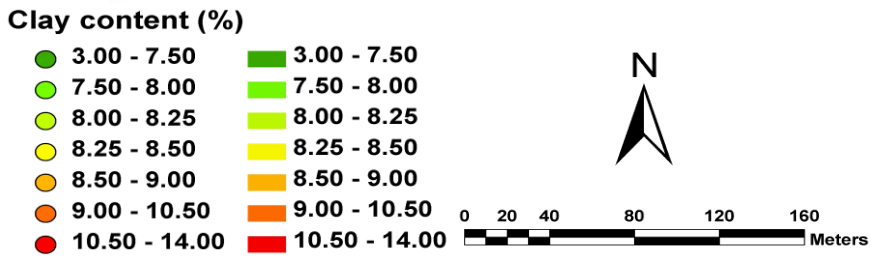
With correction for water content **Without correction for water content**

Clay content 0-10 cm, dry conditions



Clay content 0-10 cm, wet conditions





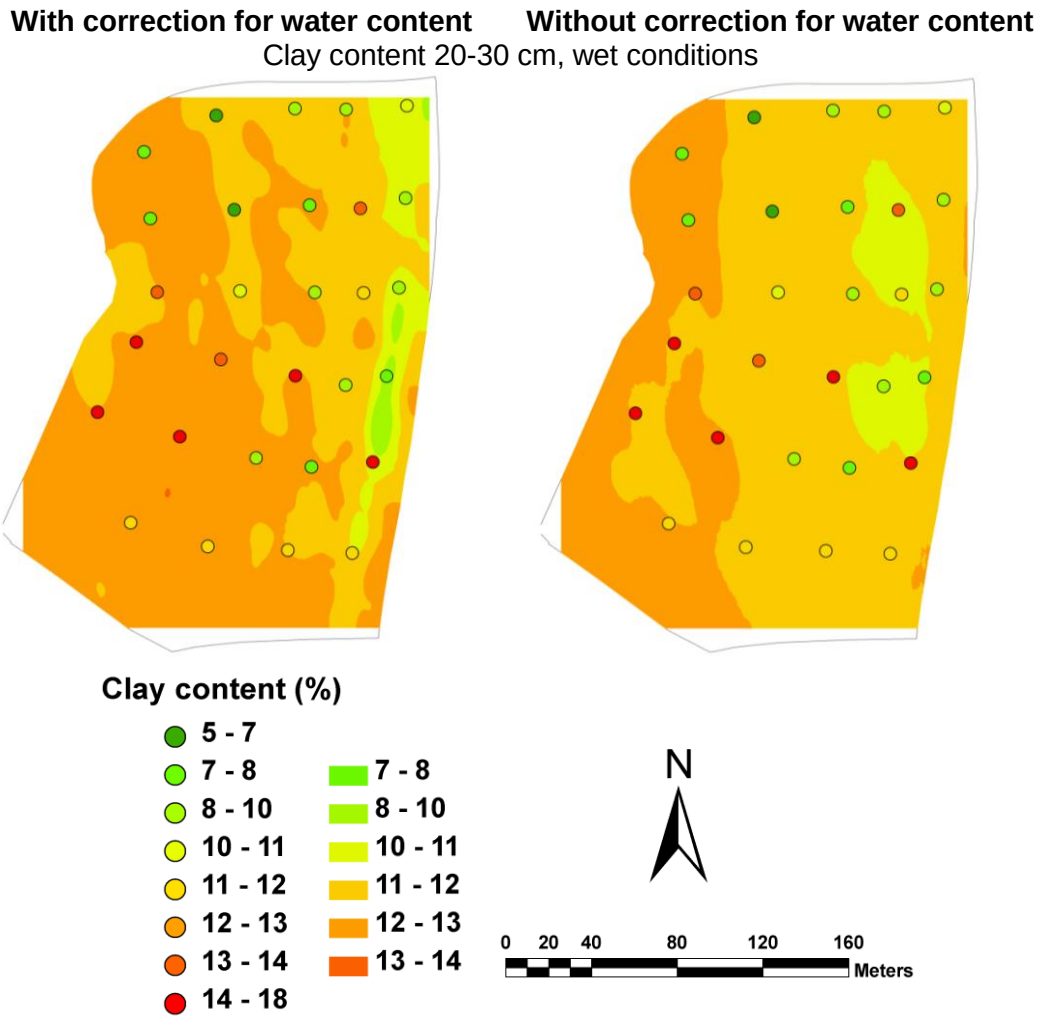


Figure 25: Kriged maps of clay content estimates from EMI and GPR measurements

3.1.3. Performances

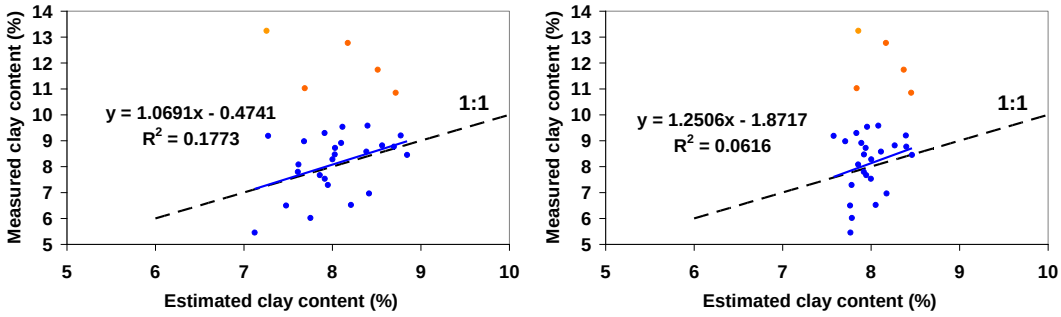
- Validation and RMSE estimation

Clay content estimates from EMI measurements may be compared with ground truth measurements for validation (Figure 26). Lower R^2 values of the relationships relating measured clay content to estimated clay content are generally observed when values of soil electrical conductivity are not corrected for soil water content (i.e., when the Rhoades model is not applied and clay content is directly related to measured soil electrical conductivity), indicating the importance of accounting for this correction. For the 0-10 cm layer, better agreement between clay content estimates and ground truth measurements is found for the wet soil conditions (R^2=0.27) compared with the dry conditions (R^2=0.18). This would be explained by the more contrasted patterns of soil electrical conductivity over the area under wet than under dry conditions, which would

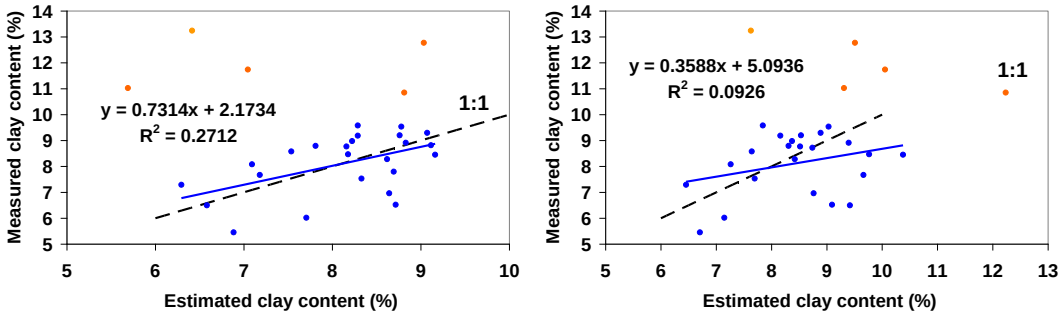
allow to better retrieving spatial variations of clay content. Nevertheless, in both cases, very poor agreement is observed between measurements and estimates of clay content at the location of the anthropogenic soil along the eastern limit of the field (represented by the orange dots in the graphs, not included in the establishment of the measured vs estimated clay content relationships). The very different nature of this soil compared with the rest of the studied field may explain these observations, the values of the model's empirical parameters being soil-specific. Regarding the 20-30 cm layer, poor agreement is found between clay content estimations and measurements, especially within the north-east part of the field where large overestimations of clay content are observed. Such low R^2 values and the sometimes large discrepancies found between ground truth and estimates may at least partly arise from the fact that EMI measurements integrate a large volume of soil while clay content was determined from relatively small samples characterising a rather thin (10 cm thick) layer of soil. Furthermore, the rather low clay content at the study site associated with its narrow range of spatial variation over the investigated area also limit the accuracy of the estimations of this soil property from soil electrical conductivity measurements.

With correction for water content **Without correction for water content**

Clay content 0-10 cm, dry conditions



Clay content 0-10 cm, wet conditions



Clay content 20-30 cm, wet conditions

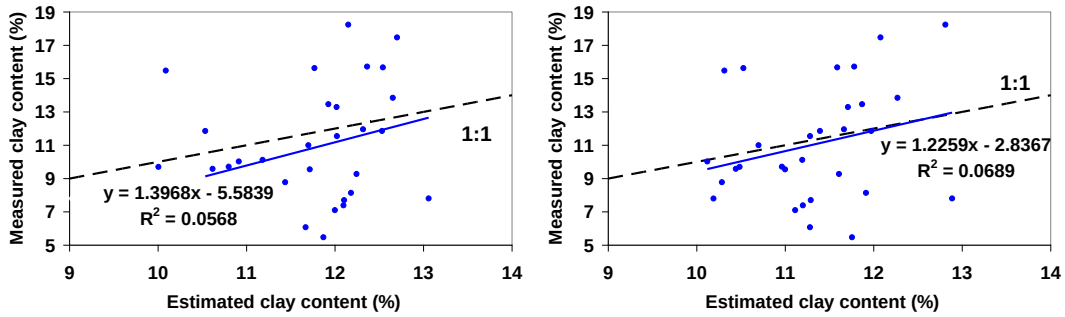


Figure 26: Validation relationships between measured and estimated clay content

3.1.4. Spatial resolution

- Technical or economical constraints

With the EM38, mounted of a quad together with the GPR setup, on average 0.75 hour was necessary for one person to collect each data set. For the Profiler, carried at walking speed over the area, the time requested to perform each measurement set amounted to around 1.25 hour. The total area of the study field is around 5 ha. The non-invasiveness of the techniques and the fast, almost instantaneous, data acquisition rate allow to cover large areas with fine spatial and temporal resolutions.

- Interpolation aspects

Ordinary kriging was used to interpolate clay content estimates from EMI and GPR measurements all over the study area and to produce the clay content maps presented above (see Figure 25). The exponential model associated with a nugget effect was found to provide better fitting results to the empirical variogram in each case. The map of kriging standard deviations for the 0-10 cm clay content estimates determined from horizontal dipole measurements during wet conditions (providing the best validation results, see Figure 26) indicates that the interpolation error is minimum and maximum along and between the measurement transects, respectively (Figure 27).

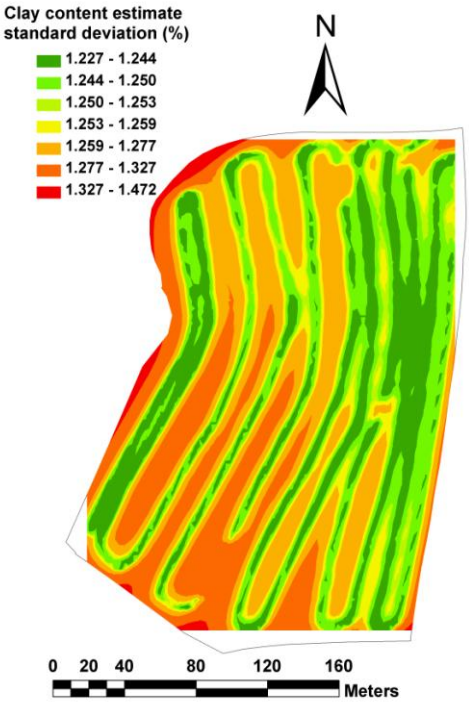


Figure 27: Standard deviation map for kriged 0-10 cm clay content estimated from horizontal dipoles EMI measurements during the second measurement day (wet conditions)

3.2. C STOCK

- Input data

The SOC stock (Mg ha^{-1}) is calculated from the SOC content (C, g kg^{-1}), the bulk density (BD, g cm^{-3}), the rock fragment content by mass (RM, dimensionless) and the thickness of the layer considered (d, m).

$$SOC = C * BD *(1-Rm)*d*10$$

The SOC stock requires a compilation of soil property maps in order to be able to calculate the SOC stock equation for each pixel. We have chosen to represent the SOC stock for the pixels (2.6 by 2.6 m) of the airborne hyperspectral image acquired by the AHS 160 sensor. The processing of the signals required to determine the individual soil properties have been explained elsewhere (C content : D3.2, section 2.2.3 ; bulk density : D3.3, section 2.5.1 ; stone content : D3.2, section 2.2.1 ; soil depth : D3.3, section 2.2.1).

	Luxembourg
Input data	Hyperspectral, Geo-electric, seismic
Processing	Hyperspectral, PLSR regression for C content of the AHS 160 image (D3.2, section 2.2.3 Geo-electric, BD, Rm: (2.3.1 and 2.5.1) Seismic, d (2.1.1)
Calibration	C content: independent calibration on previous campaign BD: 15 random points Rm: 4 random points D: independent penetrometer data
Mapping	BD, Rm, d maps interpolated to the 2*2 m pixels of the AHS 160, map calculator applied to eq. XX

- Calibration/validation

The maps of C content have been calibrated using 27 samples of the 0-20 cm top soil collected during a sample campaign in 2009 (D3.2, section 2.2.3), the bulk density on 15 of the 30 calibration/validation points (D3.3, section 2.5.1), the stone content on 4 points (D3.2, section 2.2.1) and finally the soil depth was calibrated against a set of 30 independent penetrometer values (D3.3, section 2.2.2). For the validation, we have calculated the SOC stock of the upper 20 cm of the soil profiles. Out of the 30 calibration/validation profiles, 15 have been selected that were not used in any form of calibration exercise. These profiles have been used to estimate the residuals between predicted and observed SOC stocks (Figure 28).

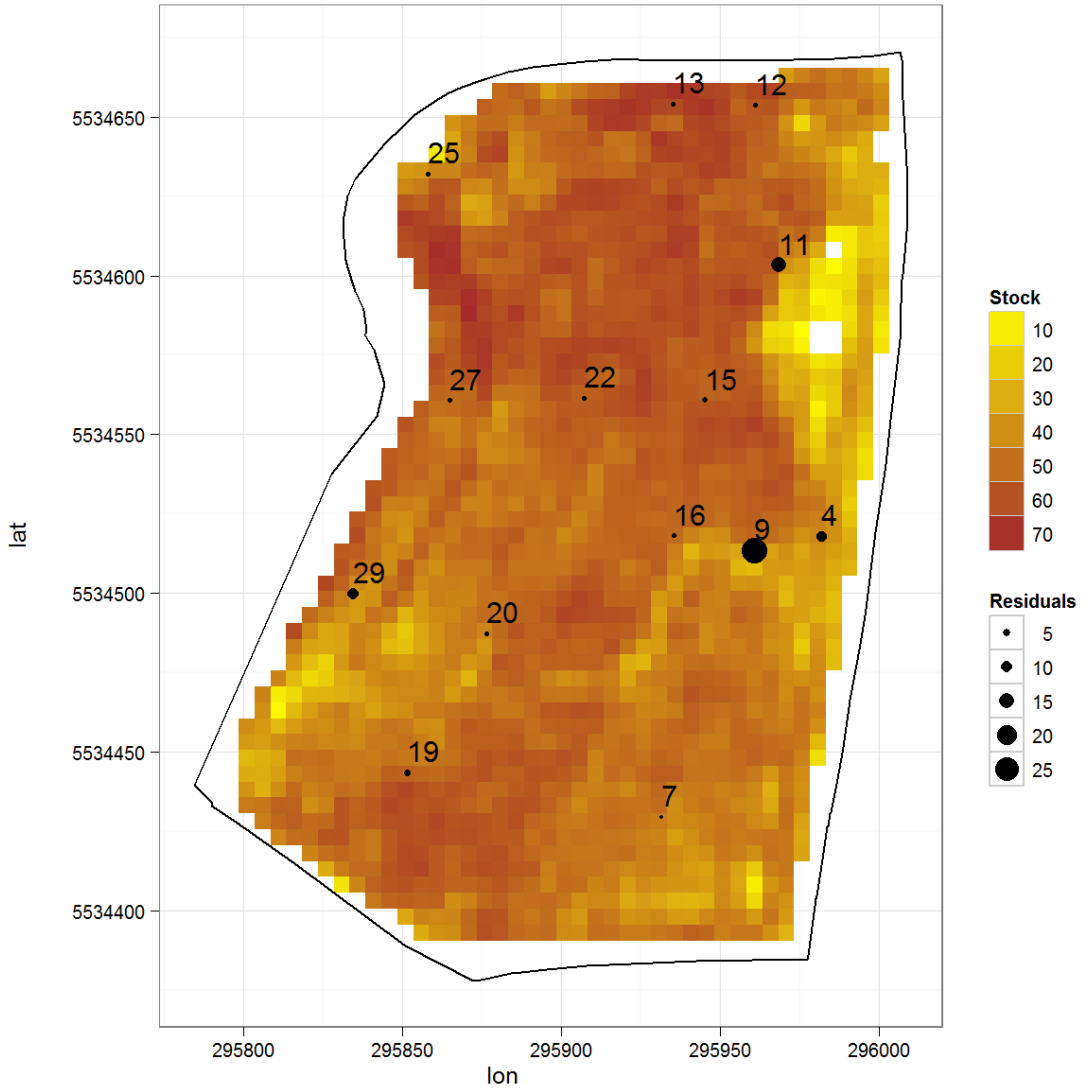


Figure 28 : Map of the SOC stocks of the upper 20 cm soil layer. The values and residuals are expressed in Mg ha^{-1}. Point numbers refer to the validation points.

a) Resulting map description

The map of the SOC stocks of the 0-20 cm topsoil (Figure 28) corresponds to the first plough layer (AP1 : 0-10 cm) and half the thickness of the second plough layer (AP2 : 10-30 cm) as identified in the pedologic survey (D3.2, section 2.1). As this horizon is every year well mixed, there are no vertical gradients in SOC content, stone content or bulk density. Hence, mean values of SOC content (0-20 cm), stone content and bulk density (0-30 cm) were used to calibrate the geophysical signals. In general, The SOC stock of the plough layer decrease from 70 Mg ha^{-1} in the northern part of the field to 20

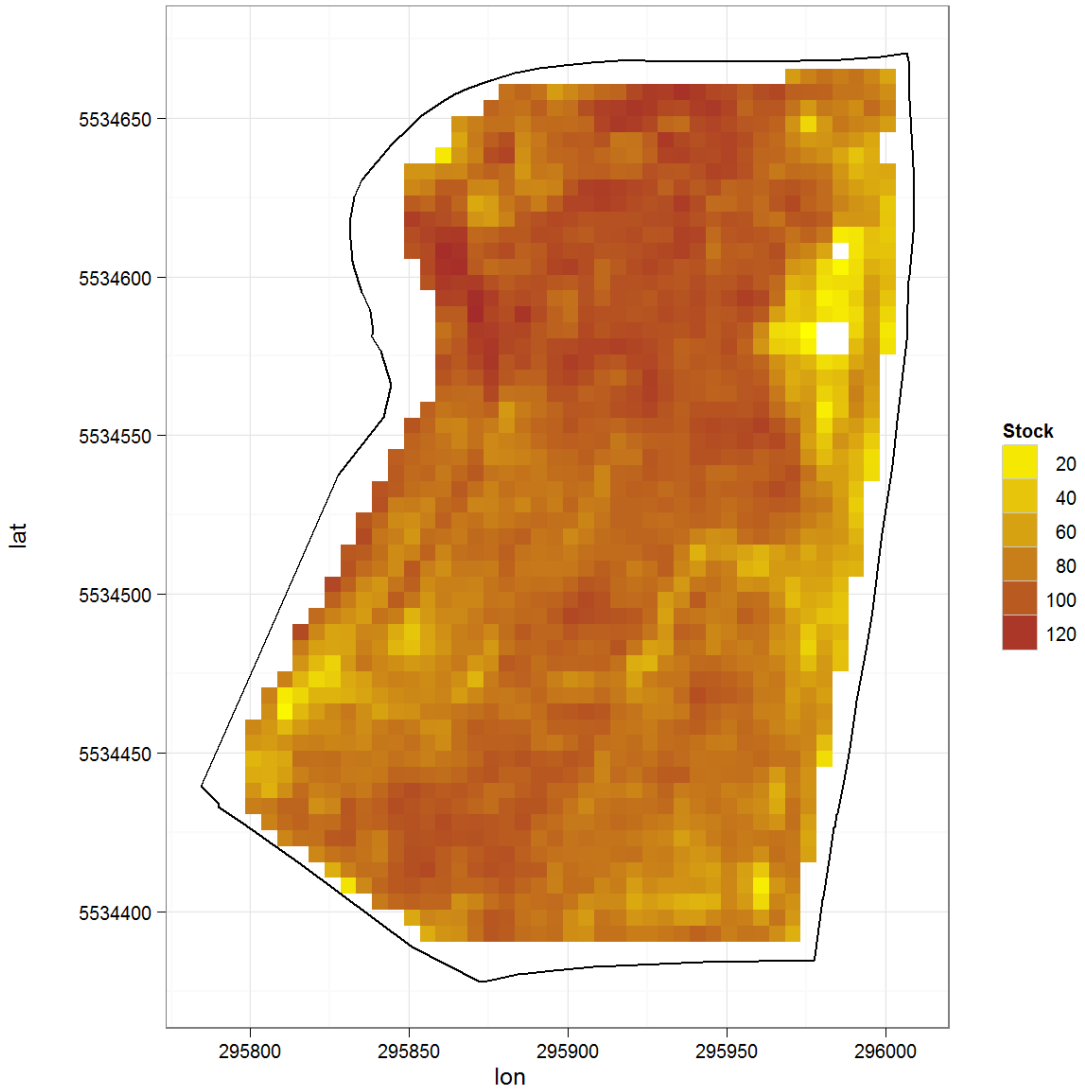


Figure 29 : Map of the SOC stock (Mg ha^{-1}) for the entire soil profile until the unweathered bedrock as detected by the seismic profiler.

Mg ha^{-1} on the southwestern margin. The low values on the northeastern field border are probably an artefact. According to the pedologic report, this is an anthropogenic area with a low stone content (D3.2, section 2.1). Unfortunately the stone content cannot be estimated for values below 20 % (2.3.1), and the resistivity erroneously indicated very high stone contents. The residuals for the validation points also indicate the largest errors in SOC stocks for this anthropogenic unit (points 4, 9 and 11, Figure 28).

Using the depth of the unweathered schist bedrock derived by the seismic signal (2.1.1) we were able to estimate the SOC stocks for the entire soil profile (Figure 29). For this map vertical extrapolation of the SOC content, stone content and bulk density for which only surface (0-30 cm) values were available proved to be necessary. For the SOC content we applied an exponential decrease with depth until it reaches zero at the unweathered bedrock detected by the seismic survey. The shape parameters for the curves were fitted on the 30 calibration/validation profiles). For the bulk density and stone content we used a linear increase until a value of 100 % for stone content and 2.65 g cm^{-3} for bulk density at the unweathered bedrock. Unfortunately these stocks cannot be validated as the soil was too stony for the corer in order to collect soil samples to the depth of the unweathered bedrock. The SOC stock for the entire profile broadly show the same pattern as the stock of the topsoil, and the values are roughly twice the topsoil values.

b) RMSE estimation

In general, the SOC stock (0-20 cm) was reasonable well predicted with an R^2 of 0.3 and an RMSE of 9.41 Mg ha^{-1} (Figure 30). As already discussed above, the prediction would considerably improve when the anthropogenic unit would be excluded from the SOC stock map (points 4, 9 and 11, Figure 28 and Figure 30).

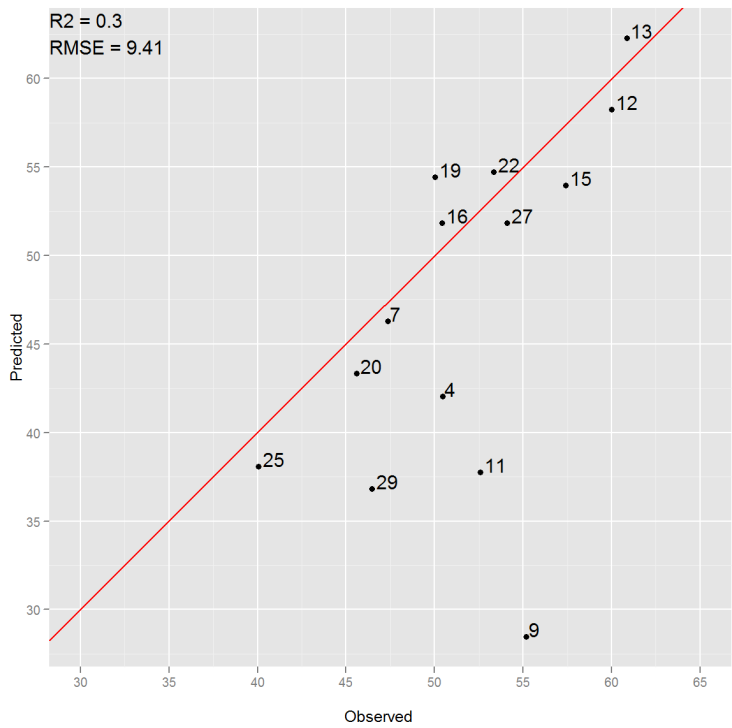


Figure 30 : Predicted against observed SOC stocks (Mg ha^{-1}) for the upper 20 cm. The numbers refer to the validation points.

- Spatial resolution

a) Technical and economical constraints

Hyperspectral remote sensing using the AHS160 airborne sensor provided the SOC signal for all bare cropland fields in a 60 by 10 km flight line taking 2 hours flight time. It is obvious that a large number of fields should be analysed for cost efficiency of such flight campaigns. The combination of hypespectral remote sensing with seismic and geo-electric signals was used to produce the SOC stock, and therefore the costs of the different techniques should be added up.

b) Interpolation aspects

An exponential function gives the best fit for the semivariogram of the predicted SOC stock of the first 20 cm (Figure 30). The spatial patterns occur mainly over 34 meters (i.e. the range of the semivariogram). Stevens et al (subm) have calculated the scale of spatial patterns in 14 cropland fields along a north-south transect in Luxembourg, and found that the range of semivariograms for C content varies between 27 and 76 m. Given the strong spatial dependence over short distances, the spatial correlation should be taken into account for the estimation of the mean SOC stock and its confidence limits, as the hypothesis of independence does not hold. Spatial dependence typically reduces the amount of information contained in a given set of observations. An option to take positive autocorrelation into account is to adjust the number of observations using the concept of effective sample size (Griffith, 2005). The effective sample size is defined as the number of independent observations *n** equal to the sample size *n* of a spatially autocorrelated dataset. Griffith (2005) proposed several model-informed solutions to estimate the effective sample size. Using a geostatistical model specification (i.e. based on semivariograms), the effective sample size can be approximated by :

$$n^* = 1 + \left(n - 1 / \left(1 + b \frac{r}{d_{max}}\right)^c\right)$$

Where *n* is the sample size (in our case the number of pixels), *b* and *c* are coefficients, *r* is the range and *dmax* the diagonal of the field. For an exponential semivariogram Griffith (2005) proposes to use a *b* of 51.4879 and a *c* of 1.7576. Given the number of pixels (1692), the effective sample size is reduced to 59. The mean value of the predicted SOC stock in the upper 20 cm is thus 48.62 ± 2.96 Mg C ha^{-1}, while the SOC stock (0-20 cm) calculated from the 30 calibration/validation points results in a value of 51.6 ± 2.37 Mg c ha^{-1}. The slightly lower predicted SOC stock is probably due to the poor prediction of the stone content in the anthropogenic unit, as illustrated by the high residuals in points 4, 9 and 11 (Figure 28).

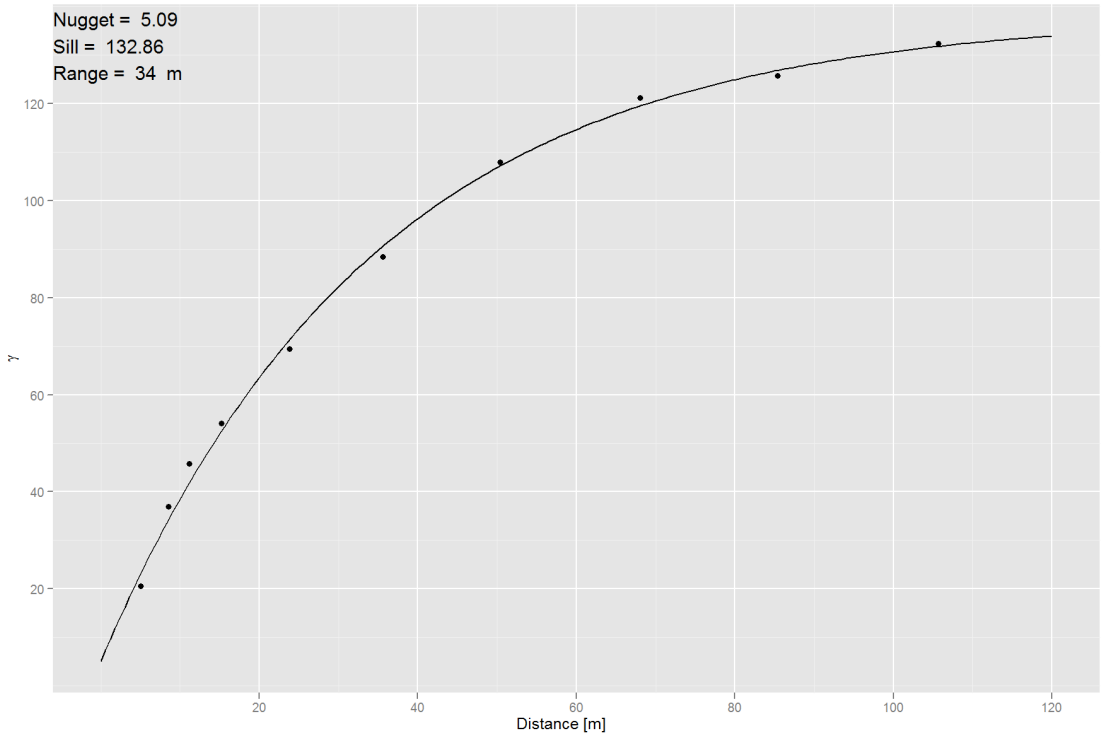


Figure 31 : Semi variogram of the SOC stock of the upper 20 cm.

4. Conclusions

This deliverable present a validation of the geo-physical techniques, based on different sets of criteria: i) accuracy of the measurements generally estimated by the RMSE, ii) capacity to represent spatial patterns and spatial variability.

The geophysical data are interpreted in terms of different thematic maps such as: soil thickness, stone content, water content, clay content and C content. For each of these maps the performance are evaluated by measuring the gap between the estimated soil properties and those observed on the field or stored in existing soil database. Special attention is also given to the spatial patterns of errors and uncertainties inherent to the geophysical techniques.

The different maps obtained on the Luxembourg or Mugello test sites are presented. First order maps are obtained by a simple inversion of geophysical signals into soil properties; second order maps are computed by several first order maps that are combined to produce more accurate information or a soil property of a high level.

5. References

Clark R.Nn, Roush T.L. (1984). Quantitative analysis techniques for remote sensing applications. Journal of Geophysical Research 89, B7: 6329–6340.

Griffith, D.A., 2005. Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. 95, 4, 740-760.

Lagacherie P., Baret F., Feret J.-B., Madeira Netto J., Robbez-Masson J.-M. (2008). Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements, Remote Sensing of Environment 112 (3): 825–835.

Lambot S., Slob E.C., van den Bosch I., Stockbroeckx B., Vanclooster M. (2004). Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties. Ieee Transactions on Geoscience and Remote Sensing 42: 2555-2568.

Lambot S., Weihermuller L., Huisman J. A., Vereecken H., Vanclooster M., Slob E. C. (2006). Analysis of air-launched ground-penetrating radar techniques to measure the soil surface water content. Water Resources Research, 42(11): W11403.

Matheron, G., 1965, Les Variables régionalisées et leur estimation, Masson, Paris.

Rhoades J.D., Raats P.A.C., Prather R.J. (1976). Effects of liquid-phase electrical conductivity, water content, and surface conductivity on bulk soil electrical conductivity. Soil Science Society of America Journal 40: 651-655.

Topp G.C., Davis J.L., Annan A.P. (1980). Electromagnetic determination of soil water content: measurements in coaxial transmission lines. Water Resources Research 16: 574-582.

Scientific and Technical Centre
RNSC Division
3, avenue Claude-Guillemin - BP 36009
45060 Orléans Cedex 2 – France – Tel.: +33 (0)2 38 64 34 34