Digital soil mapping of soil classes at 1:50,000 scale in the Franche-Comté region, France

Sébastien Lehmann, Micheline Eimberck, Manuel Pascal Martin, Dominique D. Arrouays

DIGITAL SOIL MAPPING
OF SOIL Classes at 1:50,000 sCale
In the Franche-COMTE REGION, France

INRA INFOSOL
2163, Av de la Pomme de Pin
CS 40001 Ardon
45075 Orléans cedex 2

sebastien.lehmann@orleans.inra.fr

Authors: S. Lehmann, M. Eimberck, M. P. Martin, D. Arrouays

1 – Objectives

The aim of this work is to produce a predicted soil classes map, at a scale of 1 :50,000, validated by point data in the area of Vercel (Jura, France). On this area (Fig.1), the approach we detail here led to mapping a surface of which only ¼ was previously mapped. 2348 point data scattered over the whole territory were used to calibrate and validate the model.

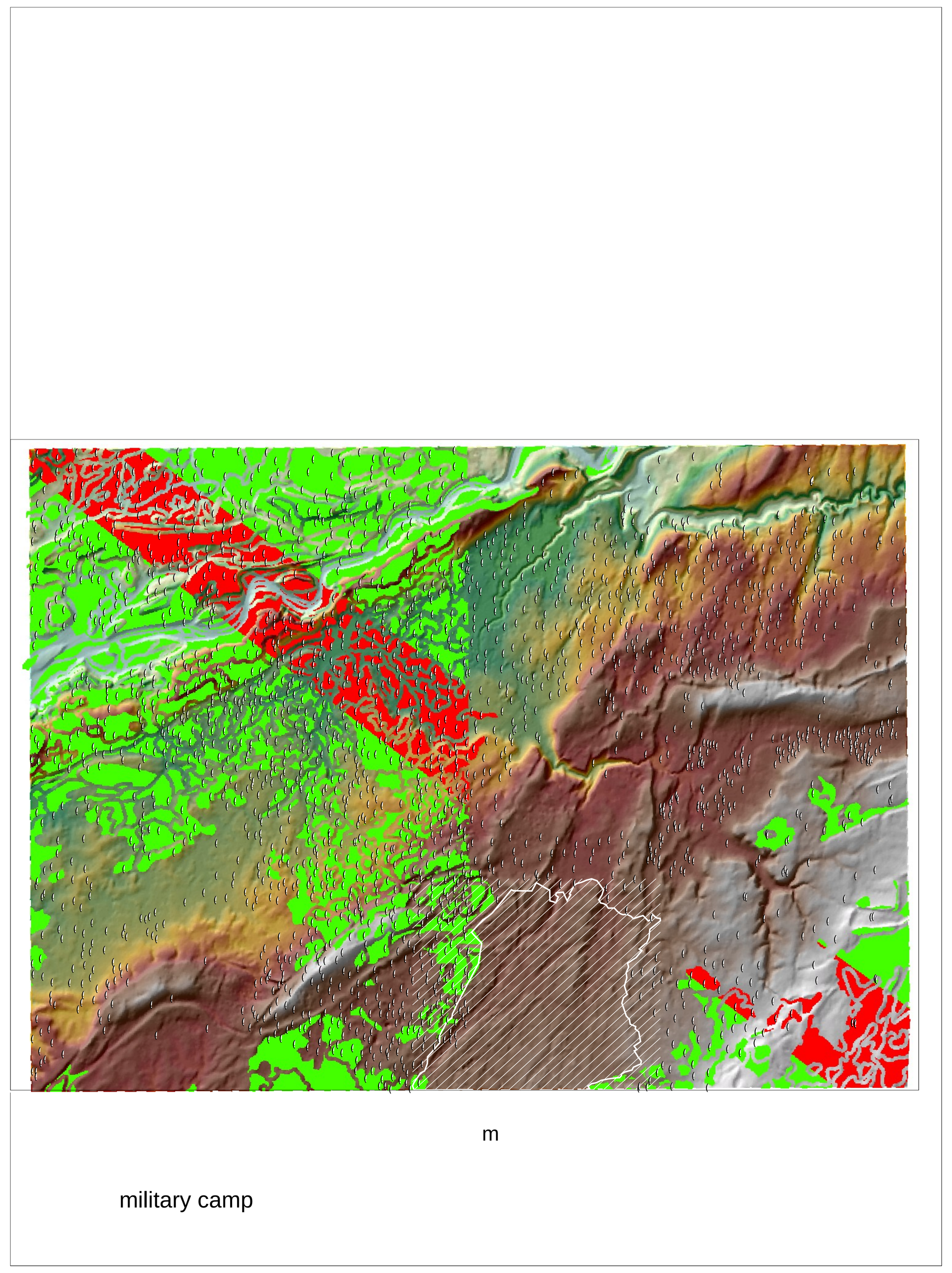

Fig. 1 - Location of the study area, training and validation area used with point and polygon data

2 – Data

The derivatives from the ASTER DEM (Fig.2) were averaged using focal mean circular windows of 8 varying sizes (radius from 30 to 1800 m). We have also used categorical data such as land use and geology rasterized to the same resolution.

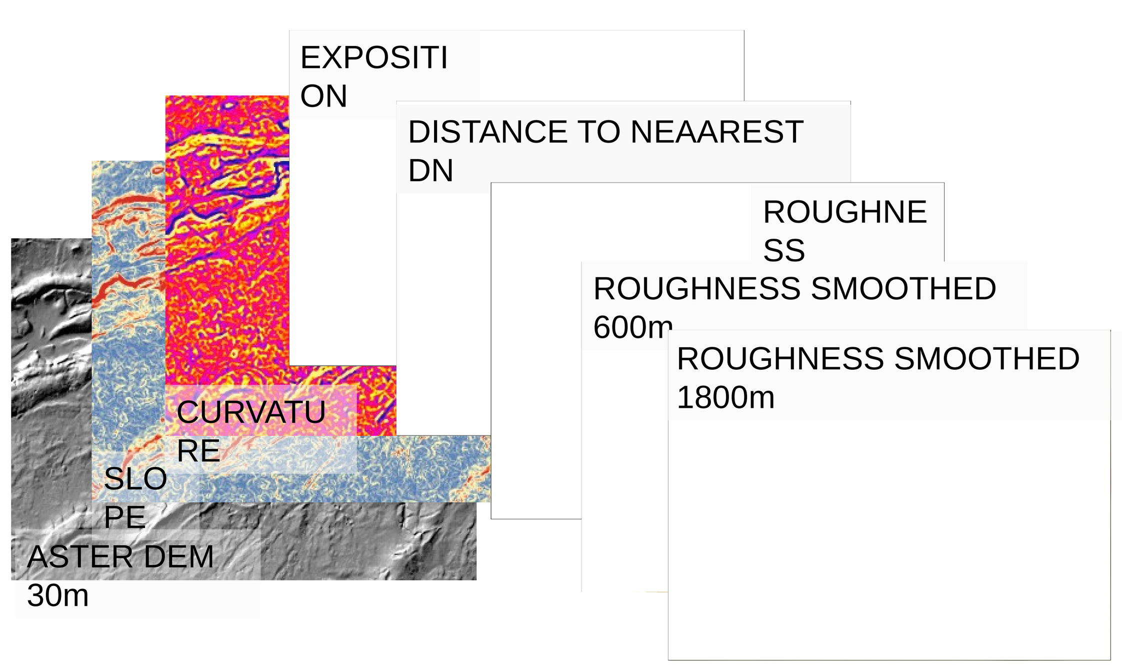

Fig. 2 – Main covariate predictors derived from DEM used to predict soil types

3 – Methods

First, we produced a predictive map from point data. Half of these points were used to calibrate a model using boosted regression trees. The remaining were used for validation. We tested 8 iterations (Fig. 3) of increasing size sliding window (focal mean). We then used the same approach but taking the existing soil units as calibration and validation data.

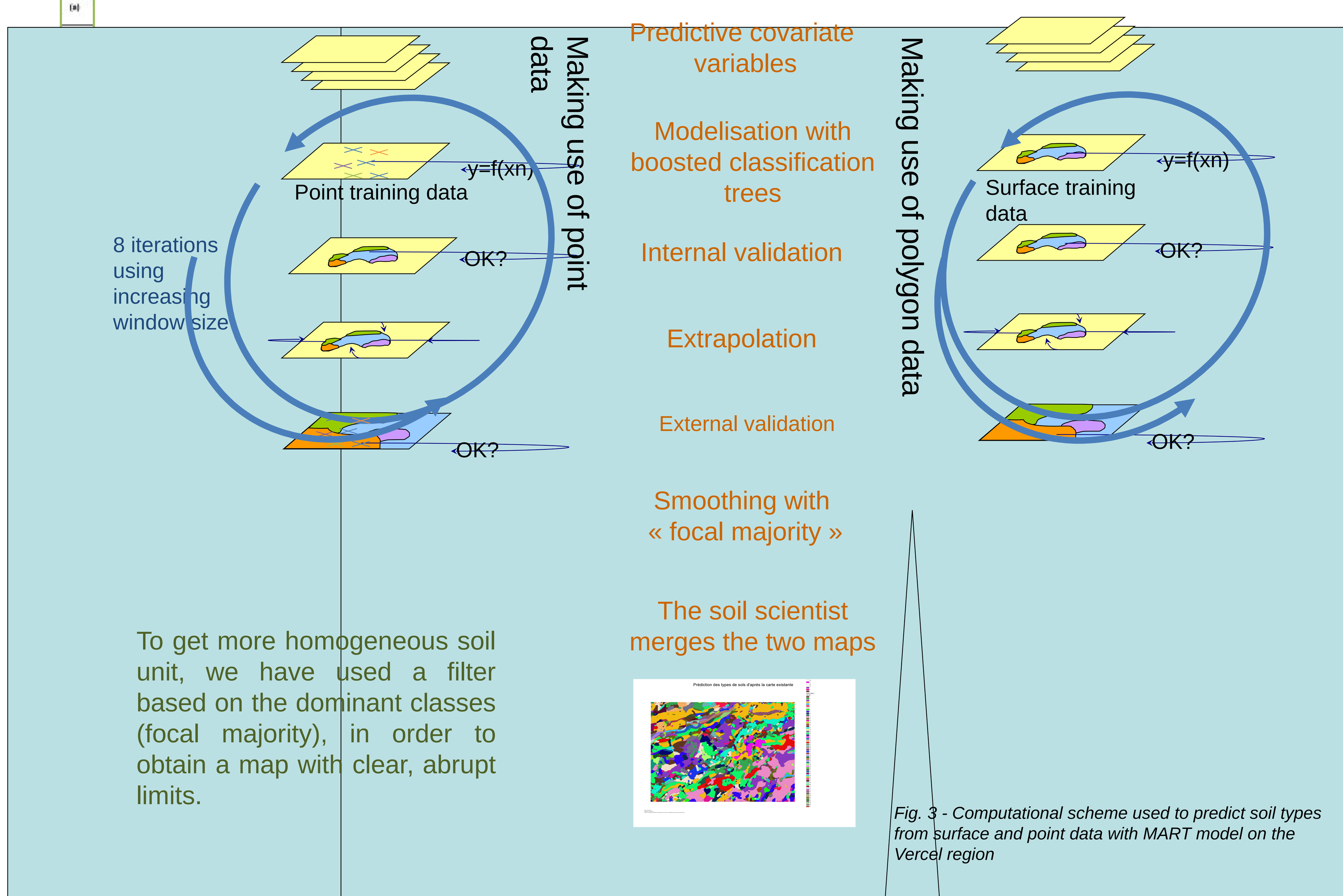

Fig. 3 - Computational scheme used to predict soil types from surface and point data with MART model on the Vercel region

4 – Results

A – Global perfomances of predictions

The relative contribution of covariates predictors was calculated from the data used to fit the model with the point data. Considering all soil classes, the most influent variables are the geology and the aspect. Both accounting for more than 90%. Some terrain attributes with a small focal mean radius were of little importance, usually accounting for less than 50% in the prediction, regardless of soil class. However, elevation (mne), curvature (curv) and slope (pente) with a large focal mean radius (60 pixels) have a moderate influence in predicting the main soil class.

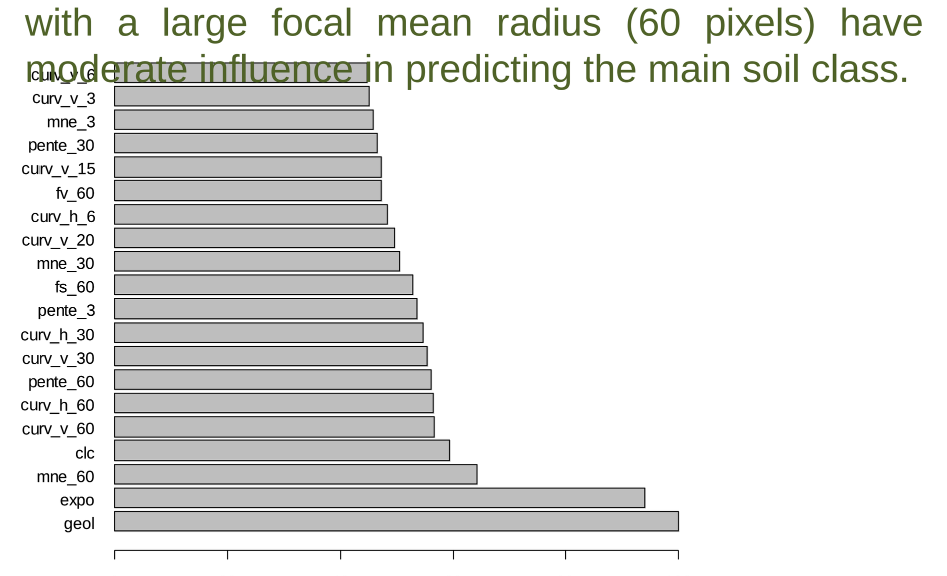

Fig. 4 - Relative contribution of the main variables for all classes with the point data

The evolution of the global classification index (G) and the kappa index (K) depending on the size of the smoothing window (focal mean function) applied to the topographic predictors shows a good correlation. The growth of the smoothing window impacts positively the G and the K. Many tests have been produced with the same dataset adding at each iteration predictors with a larger smoothing radius from 3 to 60 pixels. We had to stop after 8 iterations (Fig. 3) because the calculating

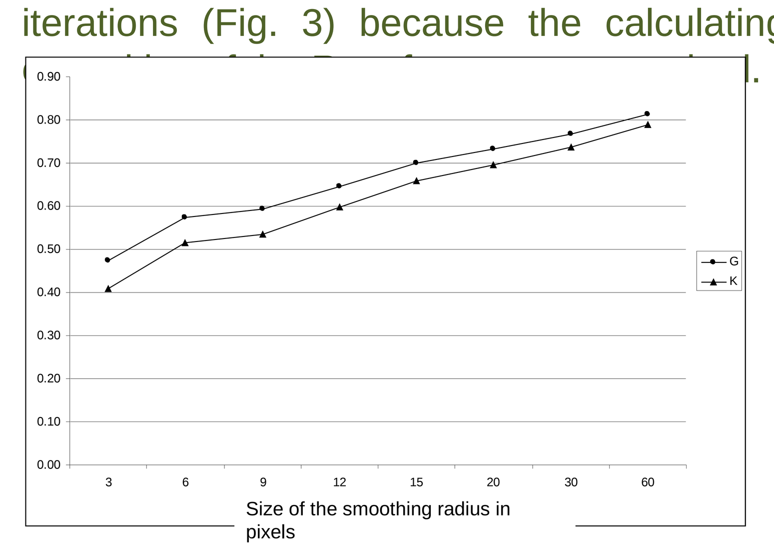

Fig. 5 - Global classification index (G) and Kappa index (K) evolution depending on the size of the focal mean window used on the topographic index in pixels.

4 – Results

B – Comparison of the two approaches

1 - We have produced two prediction maps. The Fig. 6a was obtained with point data as training dataset and Fig. 7a with the existing soil map as training dataset. For the point training dataset, we can observe (Fig. 6b) that the probability levels of the prediction are well distributed on the map. This can be explained by the relative homogeneous spatial distribution of the point dataset over the study area.

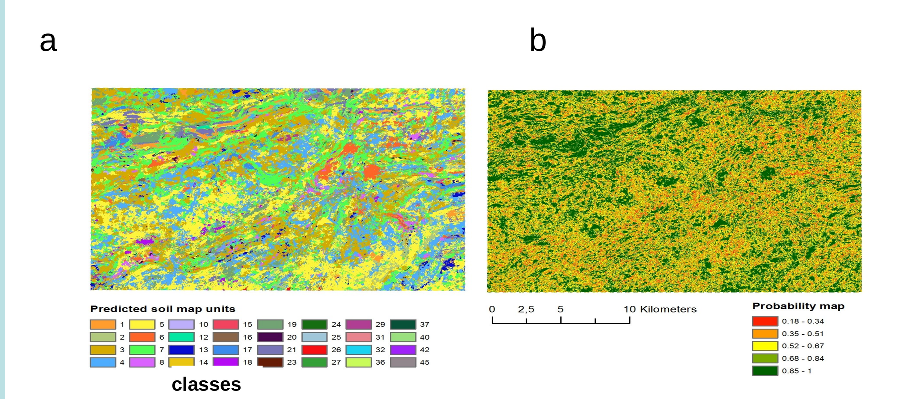

Fig. 6 - Soil classes predictions obtained using point data : predicted classes (a) and probability map (b)

2 - The soil classes obtained by using the existing soil map (Fig. 7a) shows a predictive probabilty with decreasing as the distance to the training area increases (Fig. 7b). The confusion matrix calculated for the internal validation (Tab. 1) shows that the model fit was good and that most of the soil classes are well predicted.

	Soil Unit													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**13**	**14**	**15**	**17**	**18**	**19**
Number of predicted pixels	108	254	1719	1208	1552	287	1351	152	125	69	79	109	122	502
Producer's accuracy (%)	69	70	85	77	86	76	80	78	59	68	70	79	67	70
User's accuracy (%)*	99	91	69	72	71	88	71	95	90	97	96	94	98	85
Overall accuracy (%)**	81													
Kappa Index	0,78													

* Producer's accuracy: probability that a known location is predicted well.
** User's accuracy: probability that a sample predicted as belonging to a certain class really represents that class.

*Tab. 1 - Accuracy of the main soil classes predictions in the **training** area (internal validation) with the point data*

3 - In the case of the external validation, when the training and validation surfaces belong to independent datasets, as shown in red and green in Fig. 1 , the confusion matrix (Tab. 2) gives lower results that stay around 46% accuracy. This quite important difference between internal and external validation could be explained by MART modeling which tends to be sensible to overfitting especially when using an existing map as training dataset.

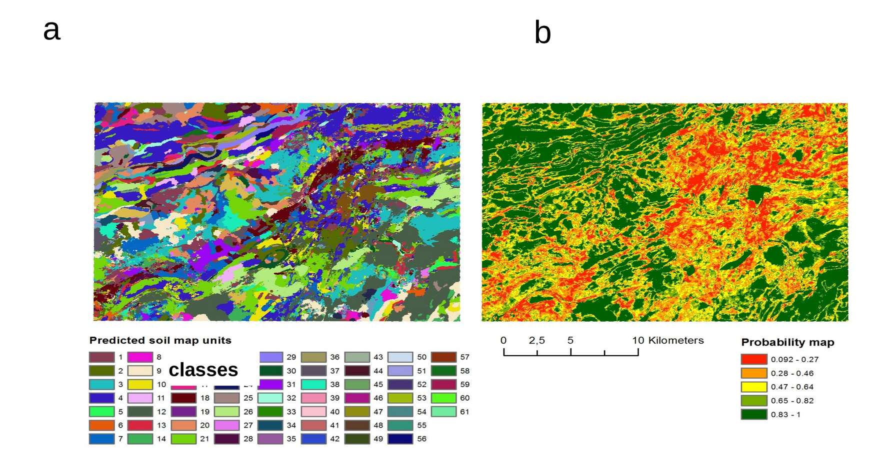

Fig. 7 - Soil classes predictions obtained using the existing soil map : predicted classes (a) and probability map (b)

4 - Finally, the soil surveyor expertise was used to produce a synthesis of these two predictions to obtain a chloropleth map of soil classes (Fig. 8). The main principle of this expertise was the following : when the results of the two predictions were conflicting, the point data were preferred in order to get a map corresponding, as much as possible, to the terrain observations.

	Soil Unit													
	1	**2**	**3**	**4**	**5**	**7**	**8**	**9**	**10**	**13**	**14**	**15**	**17**	**19**
Number of predicted pixels	141	643	1107	1708	405	2860	36	1940	5428	1	841	16	162	163
Producer's accuracy (%)	20	53	36	23	13	42	4	48	91	0	5	0	84	12
User's accuracy (%)*	30	37	31	21	23	25	75	41	66	0	8	0	42	31
Overall accuracy (%)**	46													
Kappa Index	0,31													

* Producer's accuracy: probability that a known location is predicted well.
** User's accuracy: probability that a sample predicted as belonging to a certain class really represents that class.

*Tab. 2 - Accuracy of the main classes soil types predictions in the **validation** area (external validation) with the existing soil map as training data*

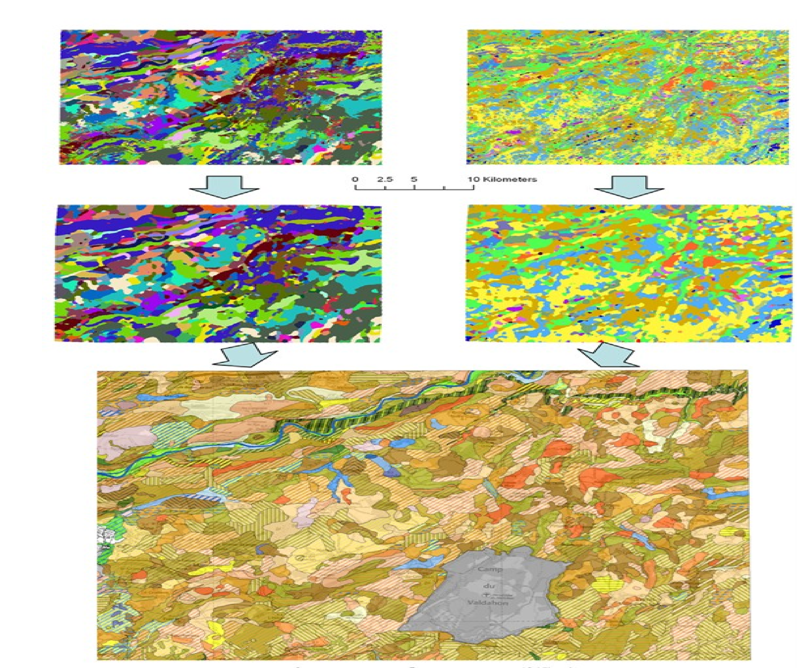

Fig. 8 - Synthesis of the two predicted maps made by the soil scientist as published for the CPF program (French Soil Map at 1/100 000 scale)

5 – Conclusion

The aim of this work was to make use of DSM tools to assist a soil surveyor to complete a soil map of which ¼ was previously conventionally mapped. We extrapolated soil observations using covariate predictors such as geology, terrain attributes, landuse etc. by using boosted regression trees. Our approach helps saving time of investigation. However it must be considered as a help for the surveyor and can not replace his knowledge of the soil landscape. The time saved with those tools, in comparison to classical soil mapping approaches, is estimated to be about 70-80 days for 36,000 ha.

DSM alone is not sufficient, knowledge of the terrain and external validation remain essential.

Aknowledgements
We thank the survey team : Line Boulonne, Eugénie Tientcheu, Laurent Richard, the people who kindly reviewed this poster: Joel Daroussin, Catherine Pasquier and
Sacha Desbourdes for the DTP.