Interactive comment on our manuscript entitled « Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm by Ellili-Bargaoui et al

Yosra Ellili-Bargaoui, yousraellili91@gmail.com

Thank you for taking the time to review our manuscript. We will address the comments and revise the paper accordingly. Below reviewer's comments are provided in blue text and our responses are marked in black text.

The manuscript is relevant as it tackles two very practical problems in completing missing spatial soil information in general: 1) how to fully exploit partly heavily aggregated legacy soil maps and 2) how to include otherwise available knowledge into this process. Two types of knowledge were separately tested (but not combined): soil legacy data and local expert knowledge of the study region. The latter seems a very relevant endeavor as it can reduce reconnaissance survey efforts and drop the costs of creating more accurate maps significantly. The manuscript is mostly well assembled, logically structured and mostly written in adequate language. However, I would like the editor and authors to consider the following remarks:

1 Novelty

Three methods are applied to the same study region and their performance to predict soil units (STU) are compared in the manuscript. The first method it is the DSMART default algorithm published by Odgers et al. (2014). The second includes actual soil observations. This is new to my knowledge, but also quite straightforward. The innovative part of including expert knowledge stored in DoneSol in a structured way was, however, already published in Vincent et al. (2018, Geoderma). Comparing three methods and evaluating their performance justifies an additional article as long as the approaches are applied in a very sound statistical framework. Here, improvements are recommended (see below).

2 Introduction

The introduction should revised. First, it relies on few publications only. Then, it splits the approaches in two groups (L83-34) of which the first group is not advised for the presented study region extent. The actual opposed groups here are not approaches using no covariates (e. g. ordinary kriging – which is an obsolete approach for digital soil mapping as with the spatial coordinates present universal kriging should at least be applied) and approaches using covariates (as e. g. DSMART). For the large study area presented here I would never advise for kriging without covariates. The difference might be made between approaches that use actual observations as response (e. g. DSM in Nussbaum et al. 2018 and many others) while other approaches generate artificial observations from available covariates (this would theoretically not be limited to legacy soil maps).

<u>**RESPONSE**</u>: We thank Dr M. Nussbaum for the constructive feedback. As detailed below we have tried to address the reviewer's concerns about the introduction.

In the introduction, we tried to present at the beginning the main needs and challenges for improving soil information resolution and scale. These needs deal with solving environmental issues and improving the consideration of soils in management and planning strategies at various spatial scales. Moreover, we presented possible approaches that can be used to characterize the spatial distribution of soil information as regard to existing soil data and available environmental covariates. The general approach synthesizes the decision tree for digital soil mapping based on legacy soil data as proposed by Minasny and McBratney 2010 (Figure 1). Hempel et al (2014) also recommend using this workflow to create GlobalSoilMap.net soil property information and generate digital soil maps at high spatial resolution.

According to Minasny and McBratney, 2010 "The methods used for digital soil mapping depends on the availability of soil data. The possibilities in the order from the richest to the poorest soil information are:

1. Detailed soil maps with legends and soil point data This is the richest information that can give the best prediction of soil properties. Soil properties can be derived from both soil maps and soil point data. The available methods are: extracting soil properties from soil map using a spatially weighted measure of central tendency, e.g. the mean, spatial disaggregation of soil maps, scorpan kriging and combination of these. An example of such an application is Henderson et al. (2001, 2005) in Australia.

2. Soil point data When soil point data are available, soil properties can be interpolated and extrapolated to the whole area by using a combination of empirical deterministic modelling and a stochastic spatial component. We have called this the scorpan kriging approach.

3. Detailed soil maps with legends When only soil maps are available, we need to extract soil properties from soil maps using some central and distributional concepts of soil mapping units.

4. No data When no data or soil maps exist in area, we will use an approach we call homosoil, which means that we need to estimate the likely soil properties under the observed soil-forming factors or scorpan factors".

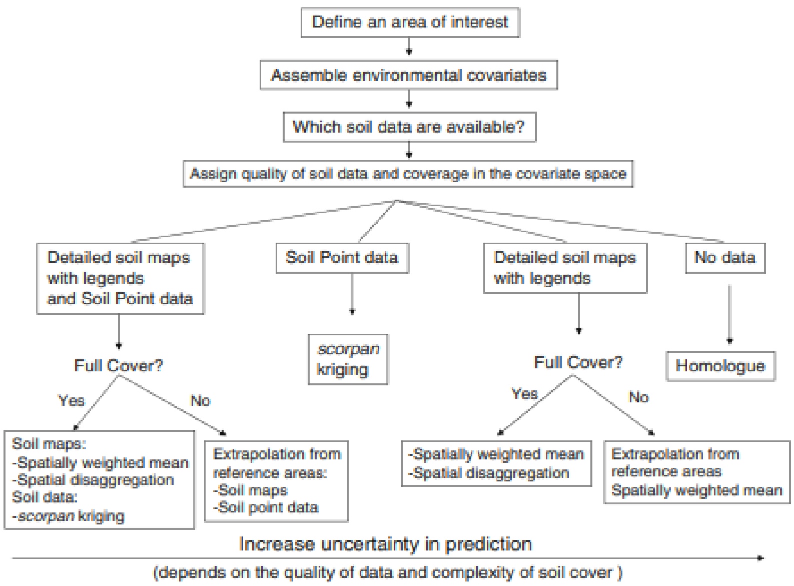

Figure 1: A decision tree for digital soil mapping based on legacy soil data (Minasny and Mcbratney, 2010, Hempel et al 2014)

On the other hand, in a recent study entitled "*Disaggregation of conventional soil maps by generating multi realizations of soil class distribution (case study: Saadat Shahr plain, Iran)"* (Jamshidi et al., 2019), the authors emphasize the need of using digital soil mapping approaches, particularly spatial disaggregation of legacy soil data, which considered as the most exhaustive soil information available over large areas. In other related DSMART studies like Odgers et al., 2014, Chaney et al., 2016, the researchers have focused on spatial disaggregation approaches of legacy maps and presented the main steps of the DSMART algorithm as well as the structure of the legacy soil data.

As suggested, we added more references to illustrate the use of observations and soil points data to calibrate soil prediction model (Malone et al., 2009, Nelson and Odeh, 2009, Abdel-Kader, 2011, Jafari et al., 2013, Kempen et al., 2012, Brungard et al., 2015, Mosleh et al., 2016, Viloria et al., 2016, Nussbaum et al. 2018, Padarian et al., 2019). However, in the literature, only few studies have used legacy soil maps and environmental covariates to generate virtual soil observations to disaggregate legacy maps as done by Odgers et al., 2014, Holmes el 2015, Chaney et al., 2016, Costa et al., 2019, Jamshidi et al., 2019; Moller et al., 2019, Zeraatpisheh et al., 2019.

3 Covariates not comprehensive

The authors state that for this landscape waterlogging is very characteristic. However, curvatures or TPI (see detailed comment on L185) representing terrain depressions were only used at one scale/resolution. Was there a reason for that? There are many publications showing the benefit of including a multitude of terrain attributes. Therefore, I suggest to also include other terrain attributes as e.g. MRVBF (multi resolution valley bottom flatness, see Nussbaum et al. 2018 for application and references).

RESPONSE: In our study, the TPI was used at a unique spatial resolution of 50 m for many raisons. Firstly, for running DSMART algorithm, all the environmental covariates must be expressed at the same spatial resolution. In our case, the selected resolution depends mostly on the resolution of the available DEM over the whole area and its accessibility as well. Secondly, in our context the selected resolution allowed to characterize and capture the main variation of topographic and geomorphologic features of our study area. The TPI is based on the upstream drainage network, and therefore it intrinsically integrates the variability of the environment over all of the watersheds and not only on neighboring pixels. Therefore, using multiple resolution of this integrative covariate does not markedly improve the prediction process. As demonstrated in a previous study by Lacoste et al., 2014, using multiple covariates resolution introduce some noise because of the high correlation existing between these variables. This could lead to mis-modelling the drainage network, and consequently the soil deposition areas.

In the other hand, the selection of covariates was based on a prior knowledge of the study area and its soil forming factors particularly the parent material and some topographic characteristics like the elevation. The choice of environmental covariates was also based on previous studies carried out over the same study area like Lacoste et al. 2011, Lacoste et al. 2014, Lemercier et al. 2012.

Following this aspect, it is not clear to me that plain DSMART algorithm would actually be outperformed by method 3 which includes expert-based rules. There were just not enough covariates included in the model to fully represent the soil forming factors in method 1.

RESPONSE: When comparing soil map depicting dominant soil type unit (STU) of each soil map unit (SMU) with the three disaggregated soil maps, we observed that disaggregated maps capture the main pattern of soil distribution over the study area. The visual inspection of these maps shows that the original DSMART approach promote the prediction of the dominant soil type unit (STU) with high proportion undependably from soil forming factors. However, local variations and clear internal disaggregation were located in the south part of the study area. The validation results using the three soil data (legacy soil profiles, independent soil profiles and accurate maps) highlight the absence of significant differences between disaggregated maps and almost the same performance of the three DSMART approaches. However, according to a prior pedological expertise and knowledge of the study area, we noticed that soil map derived from DSMART with soil/landscape rules gives more coherent soil type distribution and clear internal disaggregation of SMU with a well-developed hydrographic network using the same soil forming factors. Hence, the contribution of implemented soil /landscape rules were judged according to a prior expert knowledge of the study area and not proven by the validation results. The outperformance of the DSMART with expert based rules approach was not statically confirmed but it is clear that the data mining was able to detect the relationships between soil class and landscape over many realizations.

4 Weighing scheme for approach with legacy soil profiles (method 2)

The authors should maybe consider to apply a weighting scheme to the response during the model fit for method 2. The 755 actual observations are mixed with 14 000 artificial observations drawn from the legacy soil map polygons. The artificial observations largely outnumber the "more true" observations. I understand that the random assignment of STU (L212, step 2) in each iteration is only done for the artificial observations while the actual observations stay the same. However, the actual observations most likely "drown" in the abundance of the artificial ones during model fit. Giving higher weight to the actual observations might increase model performance. I suggest that the authors at least test a weighing scheme and evaluate its efficiency through e. g. cross-validation (the weighing scheme cannot be selected based on the validation soil data).

RESPONSE: It is a good suggestion to give high weight to legacy soil profiles, which represent a small percentage of the virtual observations drown from the legacy soil map polygons. However, the 755 extra soil profiles used to calibrate the model were already used to define the spatial boundaries of legacy polygons. Consequently, giving more weight to soil observations can bias predictions and overestimate the performance of this approach. Maybe the best way would be to use an independent soil dataset with extra soil profiles and giving more weight for the additional soil dataset.

5 Statistical approach

To train the models the C5.0 decision tree approach was used (CART with some simplification of the rules after tree growth). However, classification and regression trees (CART) are often outperformed by ensemble tree approaches (see e.g. Liess et al. 2012, Liess, M., Glaser, B., and Huwe, B.: Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models, Geoderma, 170, 70–79, doi: 10.1016/j.geoderma.2011.10.010, 2012) more complex methods often yield better results. Usage of ensemble tree methods (e. g. boosted classification trees, cubist with committees or random forest) or other models able to catch complexity (e. g. support vector machines) might improve model performance substantially. The models trained on artificially generated data are anyway not open to much pedological interpretation. Using a simple single tree approach does not result in any advantage. Ensemble tree methods also allow for covariate importance plots (and partial dependence plots for further interpretation).

RESPONSE: The objective of this study was not to select the best model that can be implemented in the DSMART algorithm to disaggregate legacy soil polygons as done by Moller et al., 2019 *"Improved disaggregation of conventional soil maps"*. Our study aimed to assess the contribution of soil/landscape rules in the disaggregation procedure of existing legacy soil maps. Most of studies like Odgers et al, 2014, Holmes el 2015, emphasize the need of implementing expert based rules in the original DSMART algorithm in order to improve the performance of prediction of soil types. However, as mentioned by Moller et al., 2019 no study has verified this hypothesis and assessed the real contribution of soil landscape rules in the disaggregation procedure nor how these rules can enhance the spatial characterization of soil distribution. To this end, we applied the same model as Vincent et al., 2018 at large spatial extend and we tried to characterize the differences between disaggregated soil maps generated by each DSMART based approach by using different validation approaches and pairwise comparison method. However, it worthwhile to investigate in futures studies the use of ensemble tree methods in the DSMART algorithm and optimizing the disaggregation process to improve the spatial characterization of soil distribution.

6 Evaluation of model performance

It remains unclear what is meant by the reported overall accuracy. Most likely the hit rate / percentage correctly allocated STUs was reported. Please specify in the methods This measure, however, might be hedged (Wilks, 2011, Chapt. 8). Scoring rules should be applied that evaluate the gain of prediction accuracy compared to a random assignment (e.g. pierce skill score, see Wilks, Chapter 8, Statistical methods in the atmospheric sciences, 2011, R Package verification). Brier skill score would by suitable for the probabilistic multi-category setting presented here. With a percentage correct of about 20–30 % it can be expected that a skill score would be as low as 0.1 (interpretation of a skill score: 0: predictions are completely random, 1: perfect predictions, -1: predictions are completely biased to predict the opposite). The properly evaluated model performance is expected to be very low and not much better than a random map generator (to await authors response). Therefore, all three approaches might not justify a map production nor a publication as a success. I am not against failure publications, but they should be discussed as such and possible reasons for the situation and improvements should be given.

RESPONSE: Many studies, which mobilized DSMART algorithm, like Odgers et al., 2014. Holmes et al, Chaney et al 2016, Vincent et al., 2018, Moller et al., 2019, Zeraatpisheh et al., 2019, Jamshidi et al., 2019 have used the term « the overall accuracy » to report the percentage of soil profiles where observations meet predictions. In this context, the overall accuracy corresponds to the number of correctly predicted classes to the total classes. For example, if we have 755 observations, and we well predict the STU of 200 profiles, the overall accuracy equals to (200/750)* 100= 26.7%.

In our study, the low overall accuracy values are explained by the complexity of the legacy soil data and the high number of STU that contained the soil database. Indeed, the Donesol database contains 171 STU, which are in most of case very similar and differ by some pedological criteria like the clay content or the thickness of some diagnostic horizons. These similarities affect the model performance, particularly where the differences between STU are not easily detected by learning rules.

Improving validation results and model performance were discussed in the manuscript, particularly in the sections 4.2 (legacy dataset) and 4.3 (taxonomy similarities). Here, we suggested to simplify the legacy soil data and to create a new soil typology by grouping similar soil types and we also suggest to use taxonomic distance to validate soil maps.

In a recent publication entitled "Validation of digital soil maps derived from spatial disaggregation of legacy soil maps" (Ellili-Bargaoui et al, 2019, https://doi.org/10.1016/j.geoderma.2019.113907), we developed a validation strategy to validate STU maps, single classification criterion maps (parent material, soil depth, soil natural drainage class, soil type) and continuous soil property maps using an independent validation dataset, selected by stratified random sampling design. Overall, our findings show that we correctly predict single classification criterion with good accuracy measures.

As recommended, we explored more validation treatments of produced digital soil maps. We computed the kappa index, which is based on the confusion matrix of soil types and characterizes the agreement degree between predictions and observations of soil types at validation sites.

7 Pairwise map comparison,

section 2.6 The authors spent a lot of words/formulas in the manuscript in defining measures to pairwise compare the predicted maps. However, all three maps remain one realization without a claim of being completely valid. The statement of one realization is a bit more similar to the second than the third does not confirm the validity of the predictions. Such a comparison is not meaningful without any further justification/goal. Moreover, one predicted map being more heterogeneous than the other does not mean it is more valid. I suggest to drop the entire sections or to explicitly justify why comparing the predictions is meaningful.

RESPONSE: Disaggregated soil maps were not generated from only one realization but from 100 realizations for both original DSMART and DSMART with extra soil observations approaches and from 50 realizations for DSMART with soil/landscape rules. All realizations were stacked together to compute the probability of occurrence of the 171 STU (Donesol database) at each pixel and then attribute the most probable STU to each elementary pixel.

The visual inspection of the three-disaggregated maps shows high similarities and local differences. As validation results do not allowed selecting the best disaggregation approach, we have based on the expert pedological knowledge to choose the best disaggregated map which will be used later to derive soil property maps. These maps are required to calibrate decision support and diagnostic tools needed for sustainable soil-landscape management. Using pairwise comparison of disaggregated maps allowed simultaneously visualizing and locating the main differences between the reference map chosen by the expert (DSMART with expert based rules) and the two other maps. Disaggregated soil maps differ mainly by the numbers of regions, which correspond to the spatial delimitation of STU in each complex SMU and the predicted STU at each pixel. Consequently, the pairwise comparison gives a visual support to compare maps and highlights the contribution of expert based rules. For example, we observe that soil landscape rules promote the prediction of hydromorphic soils in the bottom valley area. Almost, similar trend characterizes DSMART with extra soil observations map, particularly in the north part of the study area where extra observations have been collected. Moreover, pairwise comparison method is a new approach, which never has been used before in soil sciences field despite its potentialities. To this end, we decided to keep this section and showing the results of the pairwise comparison of soil maps to illustrate how *V-measure* method can be used in soil sciences field and help to interpret soil maps differences derived from different methods.

8 Unbalanced response

It seems the response STU categories do not have equal probability distribution. Hence, the nominal response is unbalanced. According to the manuscript (L348) the less frequent STU were rarely or not predicted. Tree-based methods especially tend to overpredict the majority categories. The prediction is calculated by majority vote in the final tree leaf and minority classes will in most tree leaves be outvoted and not predicted although the tree splits were meaningfully done. The authors should consider to test a sampling scheme that balances the response. Or in case this was used, please specify and put this aspect explicitly in the text.

RESPONSE: It is a good suggestion to test a sampling scheme that promote the prediction of less frequent STU. In our study, we do not test this approach but we discussed guiding sampling scheme in the section 4.5) (Improvement and future work). It may be a relevant way to improve the disaggregation process and promote the prediction of less frequent and particular STU.

9 Detailed comments

(L: line in the discussion manuscript):

P1L52-53, Abstract: What accuracy measure did you use? Hit rate/percentage correct? Please specify?

RESPONSE: The accuracy measure corresponds to the percentage of soil profiles where predictions meet observations. For example, if we have 755 observations, and we well predict the STU of 200 profiles then the overall accuracy equals to (200/750) * 100= 26.7%

As requested, this was clarified and pointed out in the abstract.

L91: Please replace "developed" by "formalized". The approach was already used before (what this publication widely shows).

Revised as suggested developed was replaced by formalized

L119: It is not relevant that the authors used a HPC (it would be, if your article would focus on HPC and DSM). Please consider dropping.

Revised as suggested

L167-169: As long as this publication is not accessible: Please consider at least adding the stratification criteria and weights between strata

RESPONSE: This publication is accessible online and entitled "Validation of digital soil maps derived from spatial disaggregation of legacy soil maps" (Ellili-Bargaoui et al, 2019, https://doi.org/10.1016/j.geoderma.2019.113907).

L170: Was this "purposive sampling" by expert knowledge of soil surveyors? Please specify.

RESPONSE: The validation dataset contains 755 legacy soil profiles. These profiles were sampled based on expert knowledge to characterize pedological diversity. This sentence was revised as suggested to point out the purposive sampling strategy followed to collect these profiles.

L173: Incomplete sentence.

RESPONSE: Sentence checked and completed.

L177: A thought on a detail: How exactly did you convert the point data (e. g. point shapefile) to a raster of 50 m resolution? Where there never 2 profiles in the same pixel? Which could be technically possible and asks for resolution of the conflict.

RESPONSE: We used Arc Toolbox from ArcGIS software to create a raster layer from punctual soil observations and we select the assignment type "Most Frequent", and a cell size of 50 m. In our case, we never have 2 profiles in the same pixel.

L179, Section 2.3: Original pixel resolution is not given for every dataset. Please consider reporting it here.

RESPONSE: The original spatial resolution of soil and environmental covariates are as following:

Soil parent material and waterlogging index covariates were predicted in previous studies using machine learning and point dataset at a spatial resolution of 50m. These studies were done before the achievement of the 250,000-soil map of Brittany. For more details, please refer to Lacoste et al (2011) and Lemercier et al (2012).

Gamma-ray spectrometry data was obtained from an airborne geophysical survey in which flying lines were spaced 250–1000 m apart, and measurements were interpolated by kriging to achieve a final data resolution of 250m (Bonijoly et al., 1999).

Land use is a 250 m-pixel size landscape classification resulting from a supervised classification of MODIS (MODerate resolution Imaging Spectroradiometer) imagery (Le Du-Blayo et al., 2008).

The rest of terrain attributes: elevation, slope, Compound Topographic Index (CTI) were directly derived from a DEM at a 50 m-resolution (IGN, 2008).

As requested, we added a supplement information about the original covariate resolution in table 1.

L185: Please give a direct citation of the TPI algorithm instead of an application paper. Was it: Jenness, J.: Topographic Position Index (TPI) v. 1.2, http://www.jennessent.com, 2006 ? Moreover, according to Vincent et al. 2018 you did not use the TPI itself, but a TPI based landscape classification (according to Weiss ca. 2001?). A TPI is zero-centered continuous covariate similar to curvature not a categoric covariate.

RESPONSE: As suggested, the reference Jenness, 2006 was added. Like Vincent et al, 2018, we have used a TPI based landscape classification, which classifies the landscape into 5 classes: ridges, upper slopes, steep slopes, gentle slopes, lower slopes and valleys.

L236: Please try to avoid "extrapolate" without further specification (you mean spatial extrapolation here). Extrapolation outside of the given data value ranges should only be done exceptionally. Better wording would be something like: "From this fitted model we computed predictions for each node of the 50m-grid throughout the study area".

RESPONSE: Revised as suggested

L243: Please explain UTS. (or did you mean STU?)

RESPONSE: It was a mistake. It was checked and fixed.

L243: Please specify what you mean with "This approach...". Method 3 or the work of Vincent et al.? L254: Please give more details on "a fixed number". How was it determined?

RESPONSE: Method 3 is the work of Vincent et al., 2018. For DSMART with expert based rules we used Vincent et al's., 2018 findings, extracted at the Ille-et-Vilaine department. The fixed number drown from each polygon was determined based on the literature (Odgers et al., 2014). For more details, please refer to the article of Vincent et al., 2018.

L256: Please specify proportion of what, occurrence count, area?

RESPONSE: Area proportion. We added area to clarify the random sampling procedure followed.

L256: How many samples from the expert rules and the random set? Please specify.

RESPONSE: The number of samples from the expert rules can be easily deducted. In the line 266 of the manuscript, we specified that for each realization 18,320 samples were generated, where 14,370 virtual points are randomly selected (line 214). Therefore 18,320 – 14,370 = 3950 points were derived from expert knowledge.

As requested, this was clarified and pointed out in the manuscript (Line 266-268).

L258: What do you mean by "a unique". Please consider removing.

RESPONSE: Revised as suggested

L299: What is the difference of regions and zones? Are these e. g. predictions calculated by method 1 and method 2? Please specify.

RESPONSE: Exactly, it means that prediction calculated by method 1 are called regions and predictions calculated by method 2 are called zones.

L345: For method 2 172 STU were predicted. Is this number correct as the maximum STU is 171?

For method 2 (DSMART with 755 supplement soil profiles) we predicted 172 STU because to calibrate the model (C5.0) we merge two sources of data:

- Virtual soil samples derived from random sampling of legacy polygons to be then assigned to 171 STU (STU contained in the legacy soil data)

-755 legacy soil profiles which have been assigned to 172 different STU. Hence, there is an extra STU which not exists in the legacy soil database.

L380: Please consider replacing "quality" by "uncertainty".

Revised as suggested.

L383-387: Please always report in the same order. Consider using labels as "method 1", "method 2" to ease readability.

Revised as suggested

L391: Please consider reformulation, e. g. replace "recorded" by "suggested".

Revised as suggested.

All figures: some text is too small. Figure 1: In the map legend please specify the scale for "Accurate soil maps". Moreover, please change a different color or shape (e. g. triangles) for the red and green dots. Having the same color saturation, they are not visible for about 10 Figure 2: Please slightly enlarge the smallest fonts and explain the abbreviations in the figure caption for readers only checking this figure. Figure 3: Please replace numbers in legend with soil type unit names or at least indicate the general meaning of the numbers in the figure caption. Figure 4 and 8: One legend is enough (if they contain the same color scheme). Figure 6: x-axis labels are missing. Please add.

Revised as suggested.

Many thanks to accept my comments, Best regards, M. Nussbaum, BFH-HAFL

Many thanks for your suggestions that allowed us to improve our paper.

Interactive comment on "Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm" by Yosra Ellili et al.

Caroline Chartin (Referee)

caroline.chartin@uclouvain.be Received and published: 24 September 2019

Thank you for taking the time to review our manuscript. We will address the comments and revise the paper accordingly. Below reviewer's comments are provided in blue text and our response are marked in black text.

General comments

This paper focuses on testing if, and in which way, disaggregating legacy soil map is improved by adding supportive data in the procedure, i.e. soil legacy data and soil-landscape relationships deduced from local expert knowledge. The purpose of this study is important given the lack of accurate soil information in many regions of the world. Those are particularly needed to better face the current process threatening/degrading soils. Moreover, some methods tested here could considerably help diminishing the time and cost for producing new accurate soil maps by reducing field work efforts. In my opinion, the manuscript is mostly well structured, logical, and the language correct. However, I have some concerns about the approach and the methodology.

Specific comments

My concerns about this study join those highlighted earlier in the discussion by the referee Madlene Nussbaum. Indeed, the authors proposed to compare three methods of disaggregation, each based on the DSMART algorithm, and to test them on the Ille-et-Vilaine department. As far as I understand, the method 3 was proposed by Vincent et al. in 2018 who applied it to the entire Brittany (which includes the Ille-et-Vilaine department) using the same covariates (at the same resolution) and validation databases used here, but obviously at a bigger extent. Although Vincent et al. (2018) do not detail the results obtained by using the classical version of DSMART (i.e., the Method 1 here), they already visually compared the maps resulting from Methods 1 and 3 on a reduced area of Ille-et-Vilaine. The maps obtained here and in Vincent et al. (2018) showed that only ~ 20 % of the validation data had been correctly predicted. The authors of the latter study already highlighted that adding the soil-landscape relationships (Method 3) did not substantially improve the results accuracy but tend to produce a more pedologically coherent map. Hence, the authors of Vincent et al. (2018) proposed different coherent ways to optimize the disaggregation procedure and improve its performance (through improving soil data, covariates, and predictive models). Here, the authors proposed to improve input data by combining DSMART algorithm to legacy soil data. Unfortunately, the legacy soil data are very largely outnumbered by the observations created artificially by the algorithm, limiting greatly their potential effect on the model performance. In this context, applying Method 2 is almost the same as applying Method 1. A weighing procedure should be implemented in the procedure for the Method 2. Considering the low performance of the different methods, I suggest the authors to dig in more in improving the procedure as suggested by Vincent et al. (2018) before applying a pairwise map comparative study. For example, the use of the legacy soil data could be optimized in Method 2 as proposed above and more complex and efficient ensemble tree methods could be tested (e.g., random Forest, cforest. . .) which have many advantages as, among others, integrated validation procedure and clear estimation of the respective variable importance in the model.

We thank Dr C. Chartin for the constructive feedback. As detailed below we have tried to address the reviewer's concerns about the methodology followed.

RESPONSE: The suggestion to give high weight to legacy soil profiles, which represent a small percentage of the virtual observations drown from the legacy soil map polygons (Method 2) is very relevant. However, the 755 extra soil profiles used to calibrate the model were already used to define the spatial boundaries of legacy polygons. Consequently, giving more weight to soil observations can bias predictions and overestimate the performance of this approach. May be the best way could be to use an independent soil dataset with extra soil profiles and giving more weight for the additional soil dataset.

The objective of this study is not to select the best model that can be implemented in the DSMART algorithm to disaggregate legacy soil polygons as done by Moller et al., 2019 *"Improved disaggregation of conventional soil maps"*. Our study aimed to assess the contribution of soil/landscape rules in the disaggregation procedure of existing legacy soil maps. Most of studies like Odgers et al, 2014, Holmes el 2015, emphasize the need of implementing expert-based rules in the original DSMART algorithm in order to improve the performance of prediction of soil types. However, as mentioned by Moller et al., 2019 no study has verified this hypothesis and assessed the real contribution of soil landscape rules in the disaggregation procedure nor how these rules can enhance the spatial characterization of soil distribution. To this end, we applied the same model as Vincent et al., 2018 at large spatial extend and we tried to characterize the differences between disaggregated soil maps generated by each DSMART based approach by using different validation approaches and pairwise comparison method.

Technical corrections

l.143: Please, replace the underscore '_' by the dash '-' in "0_20 m" and "20_50 m".

Revised as suggested

l. 165-178: §2.2.2. 'Soil validation data'

- As the existing detail maps define one of the three validation datasets, l. 173-174 should be aligned with l.167-172.

Revised as suggested

- The dataset extracted from 'Sols de Bretagne' is used for validation of M1 and M3 but also used as calibration datasets in M2: it has to be clear somewhere in the text.

RESPONSE: As suggested, we added the following sentence to clarify this point in line 297

"In this study, the validation dataset with 755 observations was used to assess the accuracy of digital maps derived from method 1 and method 3 and it was used as additional calibration dataset for method 2". This was also pointed out in the Figure 2, which presents the schematic of DSMART based approaches investigated in our study.

- Could you please precise what are the main characteristics considered by an expert to define a STU and how you converted legacy data points and vector maps to raster (l. 176-178)?

RESPONSE: The STU nomenclature respects the French soil classification system (Baize and Girard, 2008). It reflects different information at the same time like the weathering degree of soil parent material, the redoximorphic conditions, the soil type (referring to the identification of diagnostic horizons depicting pedogenetic processes), and the soil depth. This was clarified in the manuscript (Line 163-167).

We used Arc Toolbox from ArcGIS software to create a raster layer from punctual soil observations using the tool points to raster conversion. In our case, we never have 2 profiles in the same pixel of $50m^2$.

We also used Arc Toolbox from ArcGIS software to create raster maps from accurate maps using the tool polygon to raster conversion and we selected the assignment type "Most Frequent", and a cell size of 50 m.

l. 277-291: The validation procedure should be more explicit and maybe improved by computing one or two more parameters in order to better apprehend the performance of the models.

RESPONSE: Most of studies that applied DSMART algorithm like Odgers et al., 2014, Holmes el 2015, Chaney et al., 2016, Jamshidi et al., 2019; Moller et al., 2019, Zeraatpisheh et al., 2019 have computed the same validation measures "the overall accuracy". Moreover, only few studies like Chaney et al., 2016, Vincent et al., 2018 have considered pixel neighbored, as we done in our study, to compute validation measures with some flexibility.

In a recent publication entitled "Validation of digital soil maps derived from spatial disaggregation of legacy soil maps" (Ellili-Bargaoui et al, 2019, https://doi.org/10.1016/j.geoderma.2019.113907), we developed a validation strategy to validate STU maps, single classification criterion maps (parent material, soil depth, soil drainage class, soil type) and continuous soil property maps using an independent dataset, selected by stratified random sampling design.

As recommended, we explored more validation treatments of produced digital soil maps. We computed the kappa index, which is based on the confusion matrix of soil types and characterizes the agreement degree between predictions and observations of soil types at validation sites.

l. 179-200: §2.3 'Soil covariates' Please, could you quickly justify the choice of the covariates used in this procedure, and maybe make a parallel with the characteristics considered for defining STU?

RESPONSE: The selection of covariates was based on a prior knowledge of the study area and its soil forming factors particularly the parent material and some topographic characteristics like the elevation. This choice was also based on previous studies carried out over the same study area like Lacoste et al. 2011, Lacoste et al. 2014, Lemercier et al. 2012.

Moreover, some soil covariates particularly soil parent material and soil drainage characteristics are also used to define STU.

- The TPI and waterlogging parameters are categorized here. I understand that it facilitates the computation of the soil landscape relationships, but have you try to input the continuous versions of these parameters in the models?

RESPONSE: Like Vincent et al, 2018, we have used a TPI based landscape classification, which classifies the landscape into 5 classes: ridges, upper slopes, steep slopes, gentle slopes, lower slopes and valleys. In our study, we do not use the continuous version of the TPI to calibrate the model, but we expect found similar results.

- The landscape unit parameter is an aggregation of vegetation, land use and relief attributes. Why did you prefer to use one aggregated layer instead of more accurate maps about land use, vegetation and relief attributes? Is there a significant correlation between all of these parameters? Is it in order to take into account the landscape morphology at different scales, i.e. main features with the Landscape units and then local features within thanks to more accurate relief attributes layers?

RESPONSE: The landscape classification resulting from a supervised classification of MODIS (MODerate resolution Imaging Spectroradiometer) imagery (Le Du-Blayo et al., 2008). This landscape classification particularly focuses on agricultural land use and spatial organization, considering not only land cover and relief, but also elements of the landscape as the network of hedges. It allowed considering the landscape morphology and capturing the main landscape and local feature of our study area.

Until now, there is no more accurate exhaustive information on landscape units taking into account spatial organization of the agricultural land.

l. 256: Could you precise which proportions of the 18,320 samples used in the Method 3 are derived from expert knowledge and from the random selection implemented in the DSMART algorithm?

RESPONSE: The number of samples from the expert rules can be easily deducted. In the line 266 of the manuscript, we specified that for each realization 18, 320 samples were generated, where 14,370 virtual points are randomly selected (line 220). Therefore 18,320 – 14,370 = 3950 points were derived from expert knowledge. As requested, this was clarified and pointed out in the manuscript (Line 266-268).

Figure 1: Please, reduce the size of the dots or change to triangles. Precise the scale of the detail maps.

Revised as suggested.

Figure 3: Please, could you precise the names of the STU in the legend or in the caption.

Revised as suggested.

Figure 6: Please, add the x-axis labels.

The x-axis labels correspond to STU ordered by increasing cumulative area as mentioned on the Figure.

Figure 7: Please, harmonize the covariate names with main text.

Revised as suggested.

I wish my comments would be considered and helpful. I stay available for any discussion. Best regards, Caroline Chartin

Many thanks for your suggestions that allowed us

Dear Editor,

We are pleased to submit the revised version of our article entitled “Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm”. We tried to take into account as far as possible the reviewers recommendations and comments along the revised text.

As suggested by reviewers1, we added more references in the introduction to illustrate the use of different digital soil mapping approaches around the world at different spatial scales and in different pedo climatic conditions. We detailed the environmental covariates mobilized in our study and as suggested, we added a table with the original spatial resolution of all the covariates. However, we pointed out that we used only the covariates which are judged relevant in our context and which allowed characterizing our study area features according to a prior well known of our study area and the pedological expertise. Some details were added along the text, to clarify some steps of the implemented approaches and some technical corrections were also made.

In this paper, we pointed out that the objective of our study is not to select the best model that can be implemented in the DSMART algorithm to disaggregate legacy soil polygons as done by Moller et al., 2019 “Improved disaggregation of conventional soil maps”. Our study aimed to assess the contribution of soil/landscape rules in the disaggregation procedure of existing legacy soil maps. Most of studies like Odgers et al, 2014, Holmes el 2015, emphasize the need of implementing expert-based rules in the original DSMART algorithm in order to improve the performance of prediction of soil types. However, as mentioned by Moller et al., 2019 no study has verified this hypothesis and assessed the real contribution of soil landscape rules in the disaggregation procedure nor how these rules can enhance the spatial characterization of soil distribution.

Furthermore, as mentioned in our responses, all the digital soil maps were derived from 100 realizations and not from only one realization as claimed by the reviewer 1.

Moreover, as recommended by both reviewers, we explored more validation treatments of produced digital soil maps. We computed the kappa index, which is based on the confusion matrix of soil types and characterizes the agreement degree between predictions and observations of soil types at validation sites. The obtained results confirm the results derived by analyzing the overall accuracy using pixel-to-pixel validation, a window of 3x3 and the semantic validation.

Finally, all figures and tables were also revised according to the reviewer’s recommendations and comments. We look forward to seeing this manuscript in print in SOIL journal.

Title: Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm

Authors:

Yosra Ellili-Bargaoui[1,2], Brendan Philip Malone[3], Didier Michot[4], Budiman Minasny[5], Sébastien Vincent[1], Christian Walter[4] and Blandine Lemercier[4]

[1]UMR SAS, INRA, AGROCAMPUS OUEST 35000 Rennes, France

[2]Interact, Unilasalle, 60 000 Beauvais, France

[3]CSIRO, Agriculture and Food, Canberra, ACT, Australia

[4]UMR SAS, AGROCAMPUS OUEST, INRA 35000 Rennes, France

[5]Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University of Sydney, NSW, Australia

Corresponding Author: Yosra Ellili

Corresponding Author's Institution: UMR SAS, INRA, AGROCAMPUS OUEST 35000 Rennes, France

Corresponding Author's contact (email) yousraellili91@gmail.com

Abstract:

Enhancing the spatial resolution of pedological information is a great challenge in the field of Digital Soil Mapping (DSM). Several techniques have emerged to disaggregate conventional soil maps initially available at coarser spatial resolution than required for solving environmental and agricultural issues. At the regional level, polygon maps represent soil cover as a tessellation of polygons defining Soil Map Units (SMU), where each SMU can include one or several Soil Type Units (STU) with given proportions derived from expert knowledge. Such polygon maps can be disaggregated at finer spatial resolution by machine learning algorithms using the Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART) algorithm. This study aimed to compare three approaches of spatial disaggregation of legacy soil maps based on DSMART decision trees to test the hypothesis that the disaggregation of soil landscape distribution rules may improve the accuracy of the resulting soil maps. Overall, two modified DSMART algorithm (DSMART with extra soil profiles, DSMART with soil landscape relationships) and the original DSMART algorithm were tested. The quality of disaggregated soil maps at 50 m resolution was assessed over a large study area (6,775 km²) using an external validation based on independent 135 soil profiles selected by probability sampling, 755 legacy soil profiles and existing detailed 1:25,000 soil maps. Pairwise comparisons were also performed, using Shannon entropy measure, to spatially locate differences between disaggregated maps. The main results show that adding soil landscape relationships in the disaggregation process enhances the performance of prediction of soil type distribution. Considering the three most probable STU and using 135 independent soil profiles, the overall accuracy measures (the percentage of soil profiles where predictions meet observations) are: 19.8 % for DSMART with expert rules against 18.1 % for the original DSMART and 16.9 % for DSMART with extra soil profiles. These measures were almost twofold higher when validated using 3x3 windows. They achieved 28.5% for DSMART with soil landscape relationships, 25.3% and 21% for original DSMART and DSMART with extra soil observations, respectively. In general, adding soil landscape relationships as well as extra soil observations constraints the model to predict a specific STU that can occur in specific environmental conditions. Thus, including global soil landscape expert rules in the DSMART algorithm is crucial to obtain consistent soil maps with clear internal disaggregation of SMU across the landscape.

Key words: digital soil mapping, soil landscape relationships, spatial disaggregation, DSMART

1) Introduction

Characterizing soil variability especially over large areas, remains a crucial challenge to foster sustainable management of agronomic and environmental issues and help stakeholders to design regional projects (Chaney et al., 2016). At the regional as well as country level, soil maps are often available at coarse spatial resolution (Bui and Moran, 2001) which limits their ability to depict accurate soil information. For instance, the finest soils maps covering France were elaborated by administrative region at 1:250,000 scale, via a set of polygons, called Soil Map Units (SMU) with crisp boundaries. The delineation of SMU is based on soil survey programmes involving pedologists' expertise. In a coarse scale map, each polygon includes one or several Soil Type Unit (STU), which are not explicitly mapped, but their proportions and their environmental conditions, as well as soil characteristics, are provided in a detailed database (Le Bris et al., 2013).

To improve soil variability knowledge and overcome the limitation of a coarse mapping scale, several methods have emerged in the field of Digital Soil Mapping (DSM). These methods offer useful tools to predict soil spatial pattern from scarce or limited soil datasets by exploiting the availability of model based methods and an extensive array of spatialise (and more often than not gridded) environmental variables. In recent decades, DSM techniques have been increasingly used to downscale soil information and improve their spatial resolution. Depending on the quality of data and the complexity of soil cover, Minasny and McBratney (2010) supply a workflow that outlines different models that can be explored. In general, two main pathways can be distinguished: point based DSM approaches and map disaggregation approaches (Odgers et al., 2014; Holmes et al., 2015). Point DSM approaches used legacy soil profiles, which are irregularly distributed and collected according to specific objectives rather than to optimise a statistical criterion (Holmes et al., 2015). The spatial distribution of soil properties can be estimated by fitting geostatistical models such as ordinary kriging (Odgers et al., 2014; Holmes et al., 2015; Chaney et al., 2016; Santra et al., 2017; Vincent et al., 2018; Chen et al., 2018) or cokriging, which takes into account the spatial interrelations among several soil properties (Webster and Oliver, 2007). Additionally, McBratney et al. (2003) formalized the SCORPAN soil landscape model. It is an empirical quantitative function of environmental covariates, allowing predicting soil attributes (soil type or soil property) based on correlative and statistical relationships with predictor variables.

The second approach, known as spatial disaggregation, attempts to downscale the soil map unit information to delineate unmapped STUs (Bui and Moran, 2001; Odgers et al., 2014; Holmes et al., 2015). Alternatively, it can be defined as the process that allows estimating soil properties at a finer scale than the initial soil map. Several techniques have been demonstrated through soil science literature and tested in different case studies around the world. For instance, Kempen et al. (2009) have explored the use of multinomial logistic regression (MLR) for digital soil mapping. Other techniques have also been applied as decision trees using rule based induction (Bui and Moran, 2001), Bayesian techniques (Bui et al., 1999) and an area to point kriging method (Kerry et al., 2012).

In the DSM field, machine learning techniques are increasingly used to elucidate the spatial distribution of both soil type and soil properties across a large range of scale (Bui and Moran., 2001; Scull et al., 2005; Malone et al., 2009; Nelson and Odeh, 2009; Abdel-Kader, 2011; Lacoste et al., 2011; Lemercier et al., 2012; Kempen et al., 2012; Jafari et al., 2013; Nauman and Thompson, 2014; Brungard et al., 2015; Mosleh et al., 2016; Viloria et al., 2016; Nussbaum et al. 2018; Vaysse and Lagacherie, 2015; Ellili et al., 2019; Padarian et al., 2019).They were also applied to disaggregate superficial geology maps available at 1: 250 000 scale in Australia (Bui and Moran, 2001). The main advantage of these approaches is they allow handling both quantitative and categorical (ordinal or nominal) soil and environmental variables, as explanatory covariates (Bui and Moran, 2001).

Odgers et al. (2014) have developed a machine learning algorithm entitled Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART) to predict STU as a function of the high resolution environmental data supplied over different study areas in Australia. The DSMART algorithm is based on a calibration dataset derived from a random selection of a fixed number of sampling points within each soil polygon. Each sampling point is then assigned to one soil type following a weighted random allocation procedure based on the proportions informed in the soil map database. The same procedure was applied by Chaney et al. (2016) to spatially disaggregate the soil map of the contiguous United States at a 30m spatial resolution. Because integration of pedological knowledge has been recognized as an effective way to improve digital soil mapping approaches (Cook et al., 1996; Walter et al., 2006; Stoorvogel et al., 2017; Machado et al., 2018; Møller et al., 2019), Vincent et al. (2018) have applied the

DSMART algorithm with additional expert soil landscape rules describing soil distribution in the local context of the Brittany region (France). By adding supplement sampling points to the calibration dataset selected according to soil parent material, soil redoximorphic conditions and topographic features, and by integrating soil landscape relationships in the DSMART sample allocation scheme, the authors obtained a coherent soil spatial distribution observing soil organisation along hillslopes and occurrence of intensely waterlogged soils in the stream neighbourhood, as observed in Brittany.

This study aimed to test the hypothesis that adding soil landscape relationships in the disaggregation procedure improved the accuracy of produced disaggregated soil maps. This involves assessing the contribution of soil landscape relationships implemented in the DSMART algorithm by Vincent et al. (2018). To achieve this objective, we compared disaggregated soil maps either derived from the original DSMART algorithm, the DSMART algorithm with extra soil observations and the DSMART algorithm fed by soil landscape relationships over an area of 6,775 km² in the eastern part of Brittany, France.

2) Materials and methods

2.1) Study area

The Ille et Vilaine department covers an area of 6,775 km² and is located at the eastern part of Brittany, France (48°N, 2° W) (Fig 1). It is drained by the rivers Ille and Vilaine and their tributaries. Its climate is oceanic, with a mean annual rainfall of 669 mm and mean annual temperature of 11.3° (Source: Climate Data EU). Main land uses comprise arable land, temporary and permanent grasslands, woodland, and urban areas. In the present study, anthropogenic areas were not considered. Elevation ranges between 0_20 m in the coastal zone and 20_150 m almost everywhere expect in the western part of the department where it tills 256 m. The topography is generally gentle with maximum slopes not exceeding 16%. The Ille et Vilaine department is part of the Armorican Massif with complex geology (BRGM, 2009): intrusive rocks (granite, gneiss and micaschist) in northern and north western zones, sedimentary rocks (sandstone) and metamorphic rocks (Brioverian schist) in the central and southern zones, and superficial deposits (Aeolian loam with decreasing thickness from north to south overlaying bedrock, alluvial and colluvium deposits). According to the World Reference Base of Soil Resources, soils occurring in

Ille et Vilaine include Cambisols, Luvisols Stagnic Fluvisols, Histosols, Podzols, and Leptosols (IUSS Working Group WRB, 2014).

2.2) Soil data

2.2.1) Regional soil database at 1:250 000 scale

In Brittany, soils are represented through a regional geographic database called "Référentiel Régional Pédologique (RRP)" available at 1:250,000 scale (INRA Infosol, 2014).This regional database identifies soils within Soil Map Units (SMUs), each containing one to several soil types called Soil Type Units (STUs). STUs are defined as areas with homogeneous soil forming factors, such as morphology, geology, and climate. In the study area, 96 SMU and 171 STU have been distinguished and represented by a spatial coverage of 479 polygons.

In the regional database, SMUs were spatially delimited with crisp boundaries, while STUs were not explicitly mapped, but their proportion in each SMU as well as associated environmental and soil characteristics were accurately described in a semantic database (Le Bris et al., 2013; INRA Infosol, 2014).

2.2.2) Soil validation data

To assess the quality of disaggregated soil maps, three validation datasets were used (Fig. 1):

- 135 soil profiles chosen following a stratified random sampling design and specifically described and sampled from March to May 2017 for independent validation purposes in the framework of the Soilserv research project (Ellili-Bargaoui et al., 2019).
- 755 legacy soil profiles collected between 2005 and 2008 during the "Sols de Bretagne" programme (INRA Infosol, 2014).These profiles were sampled following a purposive sampling design by expert knowledge of soil surveyors to characterize hydromorphic soil conditions and soil landscape heterogeneity.
- Existing detailed soil maps (1:25,000) covering 87,150 ha, were surveyed according to Rivière et al. (1992) and revised later to adapt to the STU typologies developed in the RRP (Le Bris et al., 2013).

All soil profiles were allocated after description and analysis by an expert to a suitable STU. Both legacy soil profiles and detailed maps were converted to raster format to perfectly meet the prediction raster at 50m spatial resolution.

2.3) Environmental covariates

The SCORPAN concept (McBratney et al., 2013) allows one to predict STU as a function of a set of covariates describing seven soil forming factors, namely soil properties (s), climate (c), organisms (o), relief (r), parent material (p), age (a) and geographic position (n). In this study, ten environmental variables (Table 1) were considered as covariates in the disaggregation process at a 50m spatial resolution. Terrain attributes included elevation, slope, Compound Topographic Index (CTI) (Beven and Kirkby, 1979, Merot et al., 1995) and Topographic Position Index (TPI) (Jenness, 2006, Vincent et al., 2018) that together were derived from a 50m resolution Digital Elevation Model (IGN, 2008). These attributes were computed using ArcGIS 10.1 (ESRI, 2002) and MNT surf software (Squividant, 1994).

Environmental attributes describing soil parent material (Lacoste et al., 2011) and hydromorphic soil conditions via waterlogging index (Lemercier et al., 2012) were obtained using decision tree methods. Waterlogging index derives from a natural soil drainage prediction. Four classes were distinguished: well drained, moderately drained, poorly drained and very poorly drained. Aeolian silt deposits and Soil Map Units boundaries are environmental covariates also obtained via expert knowledge from soil scientists.

Landscape units reflecting vegetation, land use, and relief attributes were derived from a MODIS imagery by supervised classification (Le Du Bayo et al., 2008). The Airborne gamma ray spectrometry variable (K:Th ratio) (Messner, 2008), characterizing the degree of weathering of the geological material, was also taken into account.

All soil environmental covariates were converted to raster format at 50 m spatial resolution.

2.4) Disaggregation procedure: DSMART algorithm

2.4.1) Original DSMART algorithm (Method 1)

The open source DSMART algorithm (Odgers et al., 2014) was applied to spatially disaggregate the existing legacy soil map at 1:250,000 scale. DSMART algorithm uses machine learning classification trees implemented in C5.0 (Quinlan, 1993) to build a decision tree from a target variable (STU) and the environmental covariates supplied. The DSMART algorithm was written in the Python programming language by Odgers et al. (2014) and was recently translated in the R programming language.

Running DSMART algorithm requires four main steps (Fig. 2):

1) Polygon sampling by a random selection of a fixed number of sampling points (n=30) within each polygon. This procedure allowed to select a total of 14,370 sampling points, per iteration, covering the study area and ensured that all polygons were sampled.
2) Soil Type Unit (STU) assignment to each sampling point following a weighted random allocation method. This step was based on the proportion of each STU informed in the RRP database.
3) Decision tree generation: the full set of sampling points were spatially intersected with the selected environmental covariates. This georeferenced dataset was then used as a calibration dataset to build the decision tree allowing the prediction of an STU as a function of environmental covariates. C5.0 created explicit models, which were applied to the covariates rasters to generate a realisation of STU distribution over the study area at 50 m resolution.

These three steps were repeated 100 times to generate 100 realisations of the potential soil type distribution over the study area at 50 m of resolution.

4) Computing the probabilities of occurrence: the 100 realisations were stacked to calculate the probability of occurrence of each predicted STU by counting the frequency of each STU at each pixel. This procedure led to a set of 171 rasters depicting the probability of occurrence of 171 STU.

2.4.2) Original DSMART algorithm + soil observations (method 2)

This disaggregation approach is similar to the original DSMART algorithm. However, the main difference is that 755 additional soil profiles, spatially collocated, were added to the calibration dataset to build decision trees. These soil profiles make it possible to incorporate real field observations with established soil landscape relationships. For each realisation, a calibration dataset (15, 125 samples) including virtual samples randomly selected from polygon units, as well as soil observations were used to model soil type with environmental covariates. From this fitted model we computed predictions for each node of the 50m-grid throughout the study area.

2.4.3) Original DSMART algorithm + expert rules (Method 3)

Including soil landscape relationships in the disaggregation process was explored by Vincent et al. (2018) in a specific regional pedoclimatic context in Brittany (France). Expert soil landscape relationships were used to assign STU to sampling points. These relationships were based on expert pedological knowledge, which takes into account soil parental material as well as topography and waterlogging in the STU allocation procedure. This approach combines two sources of the dataset to calibrate the model. The first one was derived from semantic information for each SMU/STU combination. It consists in attributing a barcode to each SMU/STU combination, derived from a concatenation of four features contained in the RRP database (parent material, SMU identifier, TPI and waterlogging index), and to compare these barcodes to a stack of regional covariates representing the same four features, to assign each pixel of the study area to a suitable STU. This procedure allowed matching soil exhibiting specific features with their potential spatial distribution. For instance, hydromophic soils occur with slope sequences and valley positions, while well drained soils occur in upslope or middle slope positions. Using a random sampling stratified by SMU's area, a set of sampling points was selected with a proportion of one sample for every 5 hectares and a minimum of five samples per polygon unit (3950 virtual samples).

The second dataset was derived from a random sampling of a fixed number of sampling points in each polygon unit. This procedure ensured that all polygons had been sampled. STU allocation was based on the soil map unit area proportions. The full set of each realisation (18, 320 samples) combining expert calibration dataset as well as dataset derived from random sampling procedure was spatially intersected with existing environmental covariates and used as a calibration dataset to build decision trees.

2.4.4) Prediction of the most probable STUs

From all soil type probability rasters obtained, only the three most probable STUs (with the highest probability of occurrence) were considered: for each pixel, the final prediction was the combination of the three most probable predicted STUs (1^{st} STU, 2^{sd} STU, and 3^{rd} STU) and their associated probability of occurrence.

The classification confusion index (CI) between the first most probable STU and the second most probable STU was calculated following Eq.1:

$$CI = 1- (P_{1^{st}\,STU} - P_{2^{sd}\,STU}) \qquad [1]$$

Where $P_{1^{st}\,STU}$ and $P_{2^{sd}\,STU}$ denote respectively the highest probability of occurrence for 1^{st} STU and the second highest probability of occurrence for 2^{sd} STU, calculated at each pixel (Burrough et al., 1997; Odgers et al., 2014).

This index was considered as an indicator of certainty assessment about the most probable predicted soil class and is ranging between 0 and 1. It tends to 1. When the 1^{st} STU and 2^{sd} STU are predicted with similar probability of occurrence and zero when the probability of occurrence of the 2^{sd} STU is close to zero.

2.5) Validation of disaggregated soil maps

The quality of soil maps resulting from the three DSMART algorithm based approaches was assessed by combining both spatial and semantical validation methods. Spatial validation is divided into 2 sub approaches (“pixel to pixel” and “window of 3x3 pixels”). For detailed soil maps and accurate soil profiles, “pixel to pixel” validation consists in checking, at each pixel, if the predicted STU respects the observed STU value (Heung et al., 2014; Nauman et al., 2014; Chaney et al., 2016; Møller et al., 2019). The “window of 3x3 pixels” validation assumes that, for each pixel, the predicted STU respects the observed STU value if it matches at least one of its 9 surrounding neighbours (Heung et al., 2014; Chaney et al., 2016). This method provides some flexibility by compensating spatial referencing error of soil maps and avoids the impact of fine scale spatial noise.

The semantical validation was also performed considering either each STU or a group of STUs sorted by expert on the basis of similar pedogenesis factors and similar diagnostic horizons (Vincent et al., 2018; Møller et al., 2019). From the initial 171 STUs described in the soil database, the sorting procedure led to 78 groups and 11 STU remained single.

Moreover, to assess the performance of the three DSMART based approaches, the confusion matrix was used to derive the Kappa index. This Kappa index corresponds to a chance-corrected index of the agreement between observed and predicted soil types (Cohen, 1960; Elith et al., 2008). It assumes values between −1 and 1. The higher the value, the better the prediction (Bergeri et al., 2002).

In this study, the validation dataset with 755 observations was used to assess the accuracy of digital maps derived from method 1 and method 3 but it was used as additional calibration dataset for method 2.

2.6) Pair wise comparisons of disaggregated soil maps

To compare the soil type rasters derived from the three DSMART based approaches, pairwise comparisons were performed using *Vmeasure* method implemented as open source software in an R package called Spatial Association Between REgionalisations (SABRE) (Rosenberg and Hirschberg, 2007). This is a spatial method developed to compare maps in the form of vector objects and it was commonly used in computer science to compare (non spatial) clustering.

We divide the entire study area into 2 different sets of regions, referred to as regionalizations R and Z. The first regionalization R divides the domain into n regions r_i (i=1 to n) and the second regionalization Z divides the domain into m zones z_j (j=1 to m). Superposition of the 2 regionalization R and Z divides the domain into n x m segments having a_{ij} area. The total area of a region r_i is $A_i = \sum_{j,1}^{m} a_{ij}$, the total area of a zone z_j is $Aj = \sum_{i,1}^{n} a_{ij}$ and the total of the domain is $A = \sum_{j=1}^{m} \sum_{i=1}^{n} a_{ij}$.

The SABRE package calculates a degree of spatial agreement between two regionalizations using an information theoretical measure called the *V measure*. *V measure* provides two intermediate metrics: *homogeneity* and *completeness*. *Homogeneity* is a measure of how well regions from the first map fit inside zones from the second map (Eq 2). *Completeness* measures how well zones from the second map fit inside regions from the first map (Eq 5). The final value of *V measure* is calculated as the weighted harmonic mean of homogeneity and completeness (Eq 8). All metrics range between 0 and 1, where larger values indicate better spatial agreement. *V measure*, homogeneity, and completeness are global measures of association between the two regionalizations.

Additional indicators of disaggregation quality were calculated using Shannon entropy index of regions and zones (Shannon 1948; Nowosad and Stepinskie, 2018). These indicators qualify local associations by highlighting the region’s inhomogeneities (Eq 3, Eq 4), or zone’s inhomogeneities

(Eq 6, Eq 7). Two normalized Shannon entropy was also computed using the ratios (S_j^R/S^R) and (S_i^Z/S^Z) to derive maps of local spatial agreement between the two regionalizations R and Z. These measures have a range between 0 and 1.

When S_j^R (Eq3) is close to zero, this denotes that the zone j is homogenous in terms of regions (each zone is within a single region). However, when S_j^R value increases the zone is increasingly inhomogeneous in terms of regions (it overlays an increasing number of regions). Therefore, S_j^R (Eq 3) assesses the degree of this inhomogeneity or a variance of region in zone *j*. A global indicator that measures a homogeneity of a given zone in terms of regions is given via Eq 2.

Analogous to homogeneity but with the roles of regions and zones reversed, the dispersion of zones over the entire area is also computed using Shannon entropy (Eq 4 and Eq 7), and a global indicator *C* (Eq 5) measures a homogeneity of a given region in terms of zones.

$$h = 1 - \sum_{j=1}^{m}\left(\frac{A_j}{A}\right)\left(\frac{Variance\ of\ regions\ in\ zone_j = S_j^R}{Variance\ of\ regions\ in\ the\ domain = S^R}\right) \quad [2]$$

$$S_j^R = -\sum_{i=1}^{n}\left(\frac{a_{i,j}}{A_j}\right)\log\left(\frac{a_{i,j}}{A_j}\right) \quad [3]$$

$$S^R = -\sum_{i=1}^{n}\left(\frac{A_i}{A}\right)\log\left(\frac{A_i}{A}\right) \quad [4]$$

$$c = 1 - \sum_{i=1}^{n}\left(\frac{A_i}{A}\right)\left(\frac{Variance\ of\ zones\ in\ region_i = S_i^Z}{Variance\ of\ zones\ in\ the\ domain = S^Z}\right) \quad [5]$$

$$S_i^Z = -\sum_{j=1}^{m}\left(\frac{a_{i,j}}{A_i}\right)\log\left(\frac{a_{i,j}}{A_i}\right) \quad [6]$$

$$S^Z = -\sum_{j=1}^{m}\left(\frac{A_j}{A}\right)\log\left(\frac{A_j}{A}\right) \quad [7]$$

$$V_\beta = \frac{(1+\beta)hc}{(\beta h)+c} \quad [8]$$

β is a coefficient that allows promoting the first or the second regionalization, and by default, β equals 1. V_β has a range between 0 and 1. It equals 0 in case of no spatial association and 1 in case of perfect association.

The *V measure* method was applied in two main situations (DSMART+expert rules, Original DSMART) and (DSMART + expert rules, DSMART+extra soil observations). The reference map is always the map derived from DSMART algorithm with expert soil landscape relationships.

3) Results

3.1) Disaggregated soil maps

Applying DSMART based approaches yielded a set of soil maps and associated probability of occurrence rasters. The original DSMART approach allowed to disaggregate the 96 SMUs into 108 STUs while DSMART with expert rules approach yielded 158 STUs and DSMART with extra soil observations approach yielded 172 STUs with respect to the first most probable STU map. A total of 171 STUs were identified in the Ille et Vilaine department within the RRP database. Unpredicted STUs correspond mainly to rare STUs with low proportions ranging between 2 and 10% within the SMUs containing them.

Figure 3 shows the three maps of the 1st most probable STU derived from each approach as well as the original soil map. Overall, the three most probable STUs maps captured the main pattern of soil distribution of the coarse soil map. As one could expect according to the geological parent material map (Lacoste et al., 2011), extensive areas of deep silty soils are developed in Aeolian loam deposits encountered in the north east as well as in the north central parts of the study area. Colluvial and alluvial soils were mainly predicted in the north coast part and large valleys zones.

The visual comparison of disaggregated soil maps highlighted global similarities in the soil spatial distribution markedly affected by SMU boundaries. The three approaches distinguished very well soils developed in marsh parent material in the coastal part (north) of the study area. However, DSMART with soil landscape expert rules map as well as DSMART with extra soil observations map remained more detailed and underlined a clear internal disaggregation of SMUs especially in the north and the central parts of the Ille et Vilaine department. Visual inspection of the obtained DSMART with extra soil observations map as well as DSMART with expert rules map showed an increase in soil heterogeneity when compared to Original DSMART map. More importantly, legacy soil profiles made it possible to take into account some rare soil types with low probability to be predicted. Therefore, adding supplement sampling points via the expert calibration dataset and the 755 extra soil profiles allowed to predict STUs characterized in the soil database with a low spatial extent. Nevertheless, the three DSMART based approaches spatially disaggregated the most frequent components disregarding the less frequent ones.

Figure 4 shows maps of the global probability of redoximorphic soils across the study area. STU probability rasters, depicting hydromorphic soils, were added together to produce continuous maps of hydromorphic soil probability. Visual inspection of three maps highlighted global similarities, but local differences were recorded along the hydrographic network and in the southern part of the study area. As could be expected, DSMART with expert rules well predicted hydromorphic soils in valleys and coastal areas, with a probability of occurrence exceeding 80%. Adding soil landscape relationships in the allocation process constrained hydromorphic soil predictions in specific landscape positions. The same trend characterized DSMART with extra soil observations map, particularly in the central part of the study area. Therefore, including 755 soil profiles had an important role in the disaggregation process in the northern and the central parts where these profiles were located.

The uncertainty of maps resulting from DSMART based approaches was quantified via the probabilities of occurrence of each STU predicted and the confusion index maps (Fig. 5). The latter measure indicated areas where the probability of occurrence of the two most probable soil types was close. Over the study area, the average probability of occurrence of the most probable soil type achieved respectively 0.41 for DMSART map (method 1), 0.28 for DSMART with extra soil observations maps (method 2) and 0.68 for DMSART with expert rules (method 3). Meanwhile, the average confusion index reached 0.8 for the original DSMART approach (method 1) while DSMART with extra soil observations (method 2) and DSMART with expert rules (method 3) achieved 0.9 and 0.43, respectively. Although the most probable soil classes provide plausible maps of soil distribution, there is a significant prediction uncertainty as depicted by these measures.

In regions where disaggregated soil maps showed low confusion index, particularly in northwest and north coast areas of Ille et Vilaine department, high confidence in predictions was suggested. These areas were predominantly deep loamy soils or developed in alluvial and colluvium deposits.

Figure 6 compares the cumulative area of the STUs estimated from the three disaggregated maps and that derived from the regional soil database. For each STU, its relative predicted area was estimated by counting the number of pixels where it was predicted. For the regional soil database, each STU area was computed from total SMU area multiplied by the proportion of the STU. This comparison shows that some STUs were overestimated by the disaggregation approaches when comparing to the soil database. DSMART with extra soil observations and original approaches showed similar cumulative STU areas under the curve whereas DSMART with expert rules had a shape similar to the regional soil database.

The most abundant STU in the database (431: Stagnic Fluvisol developed from alluvial and colluvium deposits) was predicted as the most frequent STU by DSMART with extra soil observations and DSMART with expert rules, and it was predicted as the second most abundant STU by the original DSMART algorithm. The 10 most abundant STUs in the soil database covers almost 43% of the study area. Of them, 7 belong to the 10 STU most predicted by the three disaggregation approaches (Table 2).

3.2) Covariates importance in the decision trees

Figure 7 gives the relative importance of the covariates used in DSMART based approaches. Soil parent material and SMU boundaries were used systematically in condition rules regardless of the disaggregation method. This was consistent with the contrasting pattern of geology and the dependence relationship between SMU and its soil components. Considering the original DSMART approach (Fig. 7.a), distribution functions of Aeolian silt deposits, airborne gamma ray spectrometry variable (K:Th ratio) and elevation contributions were more dispersed according to the STU considered than those of other covariates. For instance, Aeolian silt deposits contribution varied between 20 and 80% with a median value of 42%, whereas slope contribution ranged between 20 and 40 % with a median value of 28%. Aeolian silt deposits have an important weight in STU predictions, due to its ability to represent soils inherited from this superficial parent material, which is poorly represented in lithological maps.

DSMART with soil landscape relationships (Fig. 7.b) showed almost the same distribution function of all covariates except for elevation where its distribution function was more dispersed. Since a part of training samples was chosen with expert knowledge based on three environmental covariates: TPI, a waterlogging index and soil parent material, we would expect the prominent role of waterlogging index and TPI to constrain hydromorphic soils predictions and to achieve STU distribution in the appropriate order along the toposequence. This most likely explains the dominance of Fluvisol Stagnic in valleys areas followed by a transition to Cambisols commonly found at upslope and midslope positions along the toposequences.

Analogous to the original DSMART algorithm, DSMART with extra soil observations (Fig. 7.c) highlighted almost the same distribution of use of soil environmental covariates in the decision trees, except for aeolian silt deposits, K:Th ratio and elevation. The latter covariates contributions remained less dispersed compared to the original DSMART approach.

3.3) Validation of disaggregated soil maps

The validation procedure was performed for each DSMART based approach applied, considering the three most probable soil types and using both semantic objects (STU or soil group) and spatial neighbourhood (per pixel or 3x3 window of pixels).

Considering 755 legacy soil profiles prospected in the framework of “Sols de Bretagne” project, per pixel validation accuracy reached 27%, for original DSMART maps and 34 % for DSMART with expert rules (Table 3). A similar comparison using 135 validation sites derived from Soilserv project showed that 18.1 % of soil profiles match DSMART maps, 19.8 % match DSMART with expert rules maps and only 16.9 % match DSMART with extra soil observations maps (Table 3). Using a 3 x 3 window of pixels markedly improves the global accuracies, which increased for the two validation datasets (Table 3). DSMART with soil landscape relationships remained the best performing method.

When compared to accurate soil maps (1:25,000), the validation procedure showed that DSMART with extra soil observations as well as DSMART with soil landscape expert rules had almost the same performance (37% and 38%) while best accuracy (44%) was observed for original DSMART maps (44%) (Table 3). These scores were clearly improved by considering soil groups and 3x3 pixels neighbourhood. For instance, the accuracy of DSMART with expert rules maps using soil group reached 45.9% and increased to 62.1 % when considering 3x3 pixels windows (Table 3).

Moreover, disaggregated soil maps were compared to soil type maps extracted from existing 1:25,000-scale soil maps using the kappa index, which was computed based on the confusion matrix of the first most probable soil type of each soil mapping approach (method1, method2, and method3). Overall, the Kappa index ranging from 0.43 to 0.49, which can be considered moderate. Method 3 showed better performance, with a higher kappa index (0.49). The most accurately predicted soil types were cambisol and fluvisol. The Kappa index of the method 1 reached 0.45 meanwhile method 1 (original DSMART algorithm) showed the worst Kappa index (0.43).

3.4) Comparing disaggregated maps

Figure 8 shows inhomogeneity maps measured by Shannon entropy. The map derived from DSMART with soil landscape relationships was chosen as a reference map. This map deeply disaggregates the initial SMUs into 120,653 regions with irregular shapes. By contrast, Original DSMART map remained very similar to the original map and delineated the study into 40,459 regions. Both disaggregated maps reflect the main pattern of soil distribution over the study area despite the difference in the disaggregation process. Visual inspection of maps DSMART with soil landscape rules map and Original DSMART map revealed an overall similarity between disaggregated maps, but local differences between them were depicted.

We calculated h_1= 0.49, c_1 = 0.58 and V_1 = 0.53 as global measures of spatial agreement between the two maps (DSMART+expert rules and Original DSMART). The average homogeneity of the DSMART with soil landscape rules map with respect to the Original DSMART map was qualified via h homogeneity index. Similarly, the average homogeneity of the Original DSMART map with respect to the DSMART with soil landscape rules map was qualified via c completeness index. Visually, the Fig. 8.b map seemed to be more homogeneous than the map Fig. 8.a in agreement with the statistical assessment $c > h$. The large number of DSMART with soil landscape rules map regions, which was three times higher than Original DSMART map zones, might explain this difference. It is more likely that DSMART with soil landscape rules map regions cross through multiple Original DSMART map zones than vice versa. However, two disaggregated maps remained spatially associated according to the high V_1 score. The two inhomogeneity maps (Figs. 8a and 8b) highlighted the locations of greatest differences between two maps, mainly along the hydrographic network.

When comparing disaggregated soil maps derived from modified DSMART algorithm (DSMART with soil landscape rules and DSMART with supplement soil observations), we note that the DSMART with extra soil observations map delineated the study area into 132,942 regions. For both maps, internal disaggregation was well pronounced expect for DSMART with extra soil observations map in the southern part of the study area. Visual inspection of selected maps showed high spatial agreement and highlighted some locations of greatest differences, particularly in the southern part of the Ille et Vilaine department. Even if the hydrographic network was well detailed in both maps, it appeared more developed in DSMART with extra soil observations soil map.

Applying *V measure* method for assessing the spatial similarity between DSMART with soil landscape rules map and DSMART with supplement soil observations map provided similar information theoretical measures h_2 = 0.47, c_2 =0.48, and V_2 = 0.47. Visual comparison of soil inhomogeneity maps revealed constant variance measured by normalized Shannon entropy. This was in agreement with the quantitative assessment $c = h$. Overall, the two disaggregated maps were spatially correlated, as indicated by the global spatial agreement measure V_2.

4) Discussion

4.1) Performance of the disaggregation procedures

Produced disaggregated soil maps closely resemble the abundant soils in the original soil map (Holmes et al., 2015; Fig.3). The 1st most probable STU map derived from DSMART based approaches captured the main spatial pattern of soil distribution across the study area. More internal variation within SMUs was found when using DSMART with added point observations and DSMART with soil landscape relationships. Local soil heterogeneity reflecting inherent pedological complexity was depicted by the 1st STU maps which deliver a deterministic soil landscape distribution, continuously varying with landscape features.

External validation was performed to assess the quality of disaggregated soil maps. Using 135 independent soil profiles and a per pixel validation approach, the overall accuracy reached 18.1% for DSMART algorithm 1st STU map, 19.8% for DSMART with expert rules 1st STU map and 16.9% for DSMART with extra soil profiles 1st STU map. In the DSM literature, researchers who applied classification tree decision methods founded similar validation results. For instance, by applying DSMART algorithm in eastern Australia and using 285 legacy soil profiles, Odgers et al. (2014) achieved an overall accuracy of 23%. Similarly, Nauman and Thompson (2014) explored the use of expert rules for soil landscape relationships in the United States and achieved global accuracy ranged between 22% and 24%. Similar disaggregation performance was recorded by Holmes et al. (2015) in Western Australia (20%), Chaney et al. (2016) in the United States (17%) and Møller et al. (2019) in Denmark (18%) using DSMART algorithm (Table 4). In contrast to the latter studies, a large number of STU (171 STU) compose our soil dataset. This could certainly decrease the chance of predicting the right STU, even through mobilizing relevant geographic dataset to implement soil landscape relationships.

When considering a window of 3x3 pixels, the overall accuracy increased considerably for the three DSMART based approaches maps, but DSMART with expert soil landscape relationships achieved the highest accuracy scores. Chaney et al. (2016) highlighted a high degree of spatial noise in the predictions by including pixel validation neighbours. Overall, prediction accuracy increased twofold with a 3x3 pixel validation window and when grouping soils to a coarser level of soil classification (171 versus 89 soil group). This was recorded for all disaggregated maps regardless of the disaggregation procedure and suggests that fine soil taxonomic dissimilarities can not be accurately mapped by disaggregation processes.

4.2) Legacy soil data

Legacy soil data used in this study provide an overall representation of soil over large areas (1: 250,000 scale). This database was derived from several soil surveys and pedological expert knowledge. SMUs were spatially delineated, and their spatial organisation, as well as STUs features, were described according to available soil data and pedological expertise. STUs and their associated landscape characteristics were identified as accurately as possible using legacy soil profiles collected according to a not probabilistic sampling design between 1968 and 2012. Hence, differences in survey methods covering a large area over a long sampling period could lead to errors in the STU definition or uncertainties in the estimation of their area in a given SMU.
Moreover, soil survey intensity was not uniform within SMUs. Thus, SMU components may be derived from the unequal representation of soil samples across SMUs.
Harmonising soil data to reduce the number of STU is a great challenge by itself. Grouping some STUs regarding their pedological similarities such as sharing comparable morphological criteria, having similar pedogenic horizons and occurring in analogous environmental conditions is worthwhile to be investigated. More importantly, unifying soil data according to more functional aspects such as soil agricultural potential allows also to generate a relevant regional soil database easily handled by soil users to satisfy their needs. Many countries around the world have already harmonized their soil databases such as Denmark and Australia, where high pedological complexity was captured with a reasonable STU number, with not exceeding 23 soil groups in Denmark (Møller et al., 2019) and 73 soil groups in Australia (Holmes et al., 2015).

4.3) Taxonomic similarities

In the recent DSM literature, DSMART approach is considered as an efficient tool to disaggregate existing coarse soil maps. In this study, we compared variants of the DSMART based approach, which differed by the training dataset used to calibrate the C5.0 model and the allocation procedure. Modified DSMART algorithms used additional calibration datasets derived from supplement soil observations and expert sampling of polygons. Hence, taxonomic similarities were not taken into account neither in the calibration process nor in the current component assignment scheme. Even if there is a large number of STUs addressing inherent soil landscape heterogeneity, there is most likely a short taxonomic distance between many of them. As a result, these STUs may have similar forming conditions, making it a challenge to suitably constrain the prediction probabilities using DSMART algorithm. This likely explains the high confusion index scores recorded in the present study, particularly for original DSMART and DSMART with extra soil profiles approaches. As demonstrated by Minasny and McBratney (2007), including taxonomic distance in decision trees using pedological knowledge is a relevant way to decrease the misclassification error. Therefore, future effort and improvements of the DSMART algorithm should take into account the taxonomic distance between STU in the disaggregation procedure.

4.4) Mapping comparison

A quantitative comparison between disaggregated soil maps was performed using a novel approach called *V measure* method. This method was commonly used to assess the spatial agreement between land cover maps and thematic biotic and abiotic factors maps, as done by Nowosad and Stepinski (2018) in the United States, but never before for soil maps.

In the present study, V_1 (0.53) was larger than V_2 (0.47) suggesting that DSMART with expert soil landscape relationships map is much more similar to Original DSMART map than DSMART with extra soil observations map. This might be explained by the allocation procedure for training samples. The original DSMART algorithm tends to promote most abundant STUs with high proportions of occurrence within polygons and penalized STUs with low proportions (comprise between 2 and 10%). Therefore, frequent STUs are more likely to be predicted rather than rare

STUs. Meanwhile, by adding supplement soil profiles, preliminarily assigned to a suitable STU to the training dataset, we constrain STUs with low proportions of occurrence predictions.

Major differences between DSMART with expert rules map and DMSART with soil observations were mainly observed in the southern part of the study area and valleys areas. In general, Fluvisol Stagnic soils were overestimated by DSMART with extra soil observations. This was likely due to the purposive sampling design followed to supplement soil observations. The 755 legacy soil profiles were selected to characterize hydromorphic soil conditions and to characterize inherent soil landscape variability supposed to be organized along the hillslope.

4.5) Improvements and future work

Even though this work emphasizes the contribution of pedological knowledge in the disaggregation process, other pathways can also be explored to improve map's accuracy. As recommended by Mulder et al. (2016), compensating the temporal changes and differences in laboratory analytics is a good option to improve the quality of legacy soil data. This suggests harmonising local soil database and regrouping some STUs with similar soil forming factors through statistical modelling. Moreover, additional environmental covariates with high spatial resolution should be used to capture micro landscape variability (Lacoste et al., 2014; Odgers et al., 2014; Chaney et al., 2016; Møller et al., 2019). For example, adding a more detailed Digital Elevation Model allowed to capture small terrain features, where may be particular, STUs occurs. Improving both polygon sampling procedure and current components assignment scheme turned out to be important to reduce uncertainty prediction. This suggests drawing virtual soil samples proportionally to polygons areas and using supplement STU characteristics based on surveyor observations (slope shape, hillslope position, soil texture …) to guide STU allocation procedure (Møller et al., 2019). Assuming that the decision tree can be built to relate STU descriptors to legacy soil data, this method can replace weighted random allocation procedure and should help minor STU prediction by constraining raster probabilities.

5) Conclusion

We applied three DSMART based approaches, including original DSMART algorithm, DSMART with extra soil observations and DSMART with soil landscape relationships, to disaggregate legacy soil polygons over a large area in Brittany (France). Regardless of the disaggregation approach, the produced soil maps at 50 m spatial resolution successfully address the main soil spatial pattern regarding prior pedological knowledge of our study area. Performance assessed against 135 independent soil profiles, 755 legacy soil profiles, and accurate 1:25,000 soil maps highlighted that DSMART with expert rules maps achieved highest validation measures. Overall, modified DSMART algorithms allowed minor STUs prediction, whereas original DSMART algorithm promoted abundant STUs prediction with poor spatial structure improvement. Adding pedological knowledge as well as extra soil observations in the prediction process constrained STU probabilities, even STUs with low proportions. However, some particular STUs reflecting hydromorphic soils or loamy soils were greatly overestimated for all the three DSMART based approaches.

Soil maps produced using the original DSMART and DSMART with expert rules have a high spatial agreement, but the latter map appeared more detailed and provided a spatially continuous and consistent STU's prediction. Therefore, generalizing soil landscape relationships taken to account several STU descriptors and landscape features should be implemented in the future version of DSMART algorithm to capture soil landscape heterogeneity and consequently guarantee coherent variability of soil properties.

Acknowledgments

The authors gratefully acknowledge all farmers at the Ille et Vilaine site involved in our research. We thank the technical staff who actively participated in field sampling and laboratory analysis. This research was performed in the framework of the INRA "Ecoserv" metaprogram. This work was also supported by Sols de Bretagne project and Soilserv program funded by ANR (Agence Nationale de la Recherche) (ANR-16- CE32-0005-01).

Figure captions

Figure 1: Location of the study area and the validation datasets

Figure 2: Schematic of the DSMART based approaches algorithm. The steps in DSMART are: 1) construct the calibration dataset; 2) train C5.0 model; 3) estimate STU maps and their associated probabilities of occurrence

Figure 3: Digital soil map of the most probable STU and their associated probability of occurrence for the whole study area and for a focus zone, a) Legacy soil map: most probable STU for each SMU, b) original DSMART approach; c) DSMART with expert rules; d) DSMART with extra soil observations

Figure 4: Global probability of hydromorphic soils over the study area derived from a) original DSMART, b) DSMART with soil landscape relationships and c) DSMART with extra soil observations. The probabilities of the three STU with highest prediction occurrence are summed if they are hydromorphic

Figure 5: Confusion index maps for a) Original DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations

Figure 6: Cumulative area of the 171 STUs estimated from the regional soil database and predicted by different DSMART based approaches

Figure 7: Violin plots of the relative importance of each environmental covariate used in a) Original DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations

Figure 8: Spatial association between disaggregated maps of Ille et Vilaine department. a) map of inhomogeneity of DSMART with soil landscape relationships map in terms of original DSMART map b) map of inhomogeneity of original DSMART map in terms of DSMART with soil landscape relationships map c) map of inhomogeneity of DSMART with soil landscape relationships map in terms of DSMART with extra soil observations map d) map of inhomogeneity of DSMART with extra soil observations map in terms of DSMART with soil landscape relationships map. Inhomogeneity (variance) is measured by normalised Shannon entropy

Table headings

Table 1. Description of the environmental covariates selected. Summary of environmental covariates. P: parent material; S: soil properties; R: relief; O: Organisms; C: categorical; Q: quantitative.

Table 2. Ten most extended STUs according to the regional soil database and their respective rank by area using three DSMART based disaggregation procedures

Table 3. Overall accuracies (%) obtained using various external validation approaches for the three most probable STU

Table 4: Comparison between the size areas covered, number of soil map units, soil type units of the original legacy soil maps and the accuracy achieved in other studies using DSMART algorithm

References

Abdel-Kader, F.H.,: Digital soil mapping at pilot sites in the northwest coast of Egypt: a multinomial logistic regression approach. Egypt. J. Remote Sens. Space Sci. 14, 29–40. 2011.

Bui, E.N., Loughhead, A., Corner, R.: Extracting soil-landscape rules from previous soil surveys. Soil Research 37, 495, https://doi.org/10.1071/s98047, 1999.

Bui, E.N. and Moran, C.J.: Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103, 79–94. https://doi.org/10.1016/S0016-7061(01)00070-2, 2001.

Burrough, P.A., van Gaans, P.F.M., Hootsmans, R.: Continuous classification in soil survey: spatial correlation, confusion and boundaries. Geoderma 77, 115–135. https://doi.org/10.1016/S0016-7061(97)00018-9, 1997.

BRGM, 2009.http://sigesbre.brgm.fr/Histoire-geologique-de-la-Bretagne-59.html

Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C.W., Odgers, N.P.: POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. Geoderma 274, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025, 2016.

Chen, S., Richer-de-Forges, A.C., Saby, N.P.A., Martin, M.P., Walter, C., Arrouays, D.: Building a pedotransfer function for soil bulk density on regional dataset and testing its validity over a larger area. Geoderma 312, 52–63. https://doi.org/10.1016/j.geoderma.2017.10.009, 2018.

Climatedata.eu. https://www.climatedata.eu/

Cook, S., Corner, R., Groves, P., Grealish, G.: Use of airborne gamma radiometric data for soil mapping. Soil Research 34, 183. https://doi.org/10.1071/SR9960183, 1996.

Costa, J.J.F., Giasson, E., Silva, E.B. da., Campos, A.R., Machado, I.R., Bonfatti, B.R., Bacic, I.L.Z. Individualization of soil classes by disaggregation of physiographic map polygons. Pesquisa Agropecuária Brasileira 54, 00290. DOI: https://doi.org/10.1590/S1678-3921.pab2019.v54.00290, 2019.

Ellili, Y., Walter, C., Michot, D., Pichelin, P., Lemercier, B.: Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. Geoderma 351, 1–8. https://doi.org/10.1016/j.geoderma.2019.03.005, 2019.

Ellili-Bargaoui, Y., Walter, C., Michot, D., Saby, N.P.A., Vincent, S., Lemercier, B. Validation of digital maps derived from spatial disaggregation of legacy soil maps. Manuscript submitted to Geoderma, 2019.

ESRI, 2012. ArcMap 10.1. Environmental Systems Resource Institute, Redlands, California

Heung, B., Bulmer, C.E., Schmidt, M.G.: Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma 214–215, 141–154. https://doi.org/10.1016/j.geoderma.2013.09.016, 2014.

Holmes, K.W., Griffin, E.A., Odgers, N.P.: Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. Soil Research 53, 865. https://doi.org/10.1071/SR14270, 2015.

IGN, 2008. BD ALTI®. http://www.ign.fr.

INRA Infosol, 2014. Donesol Version 3.4.3. Dictionnaire de données.

IUSS Working Group WRB: World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome,116 pp., 2007.

Jafari, A., Ayoubi, S., Khademi, H., Finke, P., Toomanian, N.,: Selection of a taxonomic level for soil mapping using diversity and map purity indices: a case study from an Iranian arid region. Geomorphology 201, 86–97, 2013.

Jamshidi, M., Delavar, M.A., Taghizadehe-Mehrjerdi, R., Brungard, C: Disaggregation of conventional soil map by generatingmulti realizations of soil class distribution (case study:Saadat Shahr plain, Iran). Environ Monit Assess, 191-769, https://doi.org/10.1007/s10661-019-7942-x, 2019.

Jenness, J.: Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v. 1.3a. Jenness Enterprises (Available at: http://www.jennessent.com/arcview/tpi.htm), 2006.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J.: Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151, 311–326. https://doi.org/10.1016/j.geoderma.2009.04.023, 2009.

Kempen, B., Brus, D.J., Stoorvogel, J.J., Heuvelink, G., de Vries, F.,: Efficiency comparison of conventional and digital soilmapping for updating soil maps. Soil Sci. Soc. Am.J. 76, 2097–2115, 2012.

Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, B.P.: Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. Geoderma 170, 347–358. https://doi.org/10.1016/j.geoderma.2011.10.007, 2012.

Lacoste, M., Lemercier, B., Walter, C.: Regional mapping of soil parent material by machine learning based on point data. Geomorphology 133, 90–99. https://doi.org/10.1016/j.geomorph.2011.06.026, 2011.

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C.: High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311. https://doi.org/10.1016/j.geoderma.2013.07.002, 2014.

Le Bris, A.-L., Berthier, L., Lemercier, B., Walter, C. : Organisation des sols d'Ille-et-Vilaine. Version 1.1. Programme Sols de Bretagne, p. 266, 2013.

Le Du Blayo, L., Corpetti, T., Gouery, P., Bourget, E. : Esquisse cartographique des pédopaysages de Bretagne par télédétection. Rapport final du programme de recherche. CNRS : UMR6554 – Université de Bretagne Occidentale - Brest – Université de Caen – Université de Nantes – Université Rennes 2 - Haute Bretagne, p. 91, 2008.

Lemercier, B., Lacoste, M., Loum, M., Walter, C. : Extrapolation at regional scale of local soil knowledge using boosted classification trees: A two-step approach. Geoderma 171–172, 75–84. https://doi.org/10.1016/j.geoderma.2011.03.010, 2012.

Lemercier, B., Lacoste,M., Loum,M., Berthier, L., Le Bris, A.L.,Walter, C. : Apport de la cartographie numérique des sols pour prédire l'hydromorphie et l'extension des zones humides potentielles à l'échelle régionale. Etud. Gest. Sol 47–66, 2013.

Machado, I.R., Giasson, E., Campos, A.R., Costa, J.J.F., Silva, E.B. da, Bonfatti, B.R. : Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. Revista Brasileira de Ciência do Solo 42. https://doi.org/10.1590/18069657rbcs20170193, 2018.

Malone, B.P, McBratney, A.B, Minasny, B, Laslett, G.M. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma, 154,138-152, 10.1016/j.geoderma.2009.10.007, 2009.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B.: On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/s0016-7061(03)00223-4, 2003.

Messner, F. : Apport de la Spectrométrie Gamma Aéroportée pour la cartographie numérique des sols. Rapport de Master 2. Département des sciences de la terre et de l'environnement, Université d'Orléans, p. 52, 2008.

Merot, Ph., Ezzahar, B., Walter, C., Aurousseau, P.: Mapping waterlogging of soils using digital terrain models. Hydrological Processes 9, 27–34. https://doi.org/10.1002/hyp.3360090104, 1995.

Minasny, B., McBratney, A.B.: Methodologies for Global Soil Mapping, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Springer Netherlands, Dordrecht, pp. 429–436. https://doi.org/10.1007/978-90-481-8863-5_34, 2010.

Minasny, B., McBratney, A.B., 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140, 324–336. https://doi.org/10.1016/j.geoderma.2007.04.028

Møller, A.B., Malone, B., Odgers, N.P., Beucher, A., Iversen, B.V., Greve, M.H., Minasny, B.: Improved disaggregation of conventional soil maps. Geoderma 341, 148–160. https://doi.org/10.1016/j.geoderma.2019.01.038, 2019.

Mosleh, Z., Salehi, M.H., Jafari, A. The effectiveness of digital soil mapping to predict soil properties over low-relief areas. Environ. Monit. Assess. 188, 195, 2016. 10.1007/s10661-016-5204-8

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D.: National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34. https://doi.org/10.1016/j.geoderma.2015.08.035, 2016.

Nauman, T.W., Thompson, J.A.: Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213, 385–399. https://doi.org/10.1016/j.geoderma.2013.08.024, 2014.

Nauman, T.W., Thompson, J.A., Rasmussen, C.: Semi-Automated Disaggregation of a Conventional Soil Map Using Knowledge Driven Data Mining and Random Forests in the Sonoran Desert, USA. Photogrammetric Engineering & Remote Sensing 80, 353–366. https://doi.org/10.14358/PERS.80.4.353, 2014.

Nelson, M., Odeh, I.,: Digital soil class mapping using legacy soil profile data: a comparison of a genetic algorithm and classification tree approach. Soil Res. 47, 632–649. 2009

Nowosad, J., Stepinski, T.F.: Spatial association between regionalizations using the information-theoretical V –measure. https://doi.org/10.1080/13658816.2018.1511794, 2018.

Nussbaum, M., Spiess,K., Baltensweiler, A., Grob, U., Keller,A., Greiner,L., Schaepman,M.E., Papritz, A. Evaluation of digital soil mapping appraoches with large sets of environmental covariates. Soil, 4, 1-22. 10.5194/soil-4-1-2018, 2018.

Odgers, N., McBratney, A., Minasny, B., Sun, W., Clifford, D.: Dsmart: An algorithm to spatially disaggregate soil map units, in: GlobalSoilMap, edited by: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A., McBratney, Alex., CRC Press, 261–266. https://doi.org/10.1201/b16500-49, 2014.

Odgers, N.P., Holmes, K.W., Griffin, T., Liddicoat, C.: Derivation of soil-attribute estimations from legacy soil maps. Soil Research 53, 881. https://doi.org/10.1071/SR14274, 2015a.

Odgers, N.P., McBratney, A.B., Minasny, B.: Digital soil property mapping and uncertainty estimation using soil class probability rasters. Geoderma 237–238, 190–198. https://doi.org/10.1016/j.geoderma.2014.09.009, 2015b.

Padarian, J., Minasny, B., McBratney.,A.B: Using deep learning for digital soil mapping. SOIL, 5, 79–89, https://doi.org/10.5194/, 2019.

Quinlan, J.R.: C4.5: Programs for Machine Learning, 1.Morgan Kaufmann Publishers, 1993.

Rivière, J.M., Tico, S., Dupont, C. : Méthode Tarière Massif Armoricain. Caractérisation des sols, Rennes: INRA Editions, p. 20, 1992.

Rosenberg, A., Hirschberg, J.: V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure, in: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, June 2007, 410–420, 2007.

Santra, P., Kumar, M., Panwar, N: Digital soil mapping of sand content in arid western India through geostatistical approaches. Geoderma Reg., 9, 56-72, 2017.

Scull, P., Franklin, J., Chadwick, O.A.: The application of classification tree analysis to soil type prediction in a desert landscape. Ecological Modelling 181, 1–15. https://doi.org/10.1016/j.ecolmodel.2004.06.036, 2005.

Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal, 27, 379–423, 1948.

Squividant, H.: MNTSurf: Logiciel de traitement des modèles numériques de terrain. ENSAR, Rennes, France, p. 36, 1994.

Stoorvogel, J.J., Bakkenes, M., Temme, A.J.A.M., Batjes, N.H., ten Brink, B.J.E.: S-World: A Global Soil Map for Environmental Modelling. Land Degradation & Development 28, 22–33. https://doi.org/10.1002/ldr.2656, 2017.

Vaysse, K., Lagacherie, P. Evaluating Digital Soil Mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional 4, 20–30. https://doi.org/10.1016/j.geodrs.2014.11.003, 2015.

Viloria, J.A., Viloria-Botello, A., Pineda, M.C., Valera, A.,: Digital modelling of landscape and soil in a mountainous region: a neuro-fuzzy approach. Geomorphology 253, 199–207, 2016.

Vincent, S., Lemercier, B., Berthier, L., Walter, C.: Spatial disaggregation of complex Soil Map Units at the regional scale based on soil-landscape relationships. Geoderma 311, 130–142. https://doi.org/10.1016/j.geoderma.2016.06.006, 2018.

Walter, C., Lagacherie, P., Follain, S.: Integrating pedological knowledge into digital soil mapping. In: Lagacherie, P., McBratney, A., Voltz, M. (Eds.), Digital Soil Mapping. An Introductory Perspective. Development in Soil Science vol. 31. Elsevier, pp. 289–310 (ISBN-13: 978-0-444-52958-9), 2006.

Webster, R. and Oliver, M.: Geostatistics for Environmental Scientists. John Wiley & Sons, New York. http://dx.doi.org/10.1002/9780470517277, 2007.

Zeraatpisheha, M., Ayoubia S., , Brungardc , C. W. Finke, P.: Disaggregating and updating a legacy soil map using DSMART, fuzzy c-means and k-means clustering algorithms in Central Iran. Geoderma, 430, 249-258. DOI: 10.1016/j.geoderma.2019.01.005, 2019.

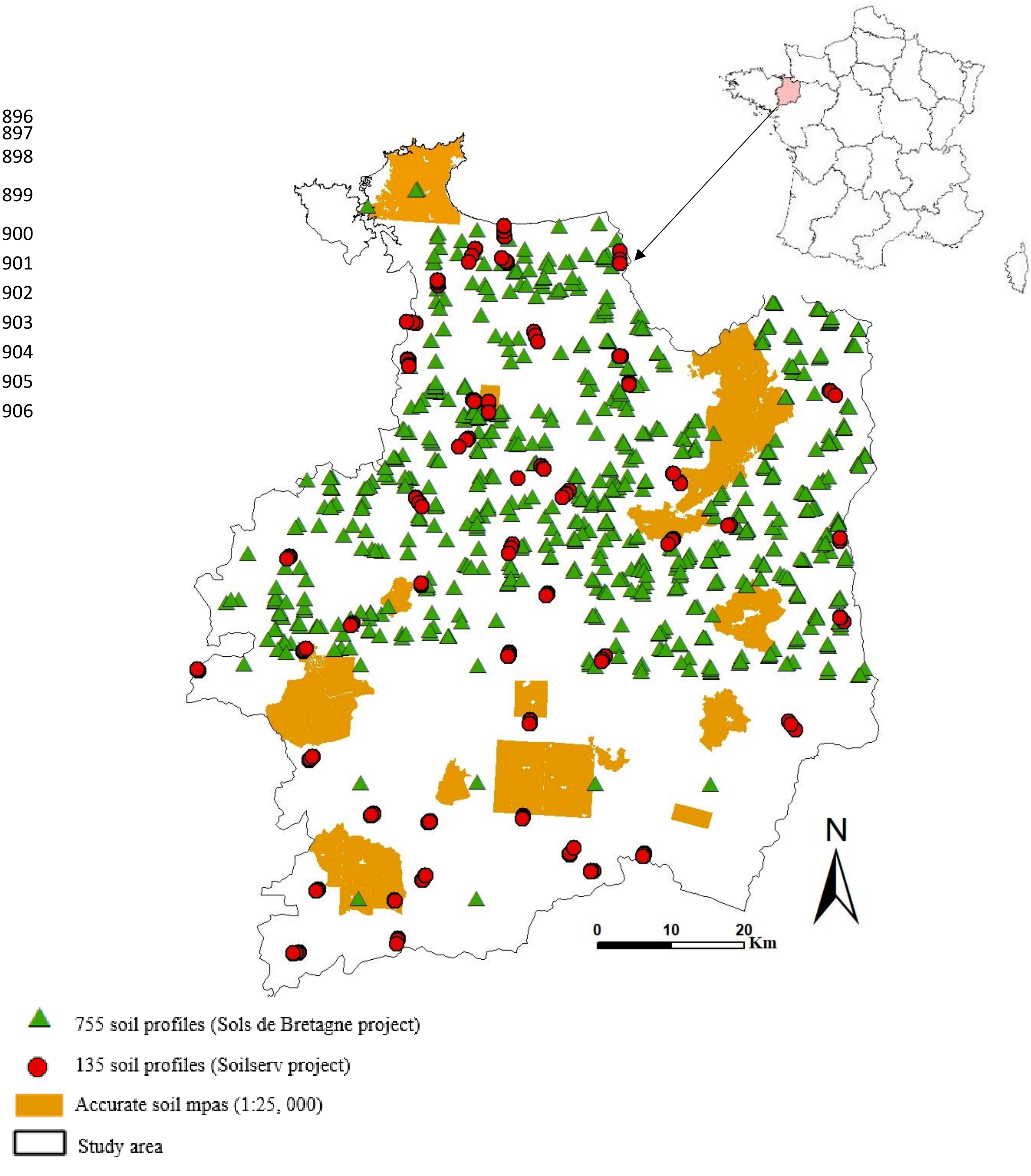

Figure 1: Location of the study area and the validation datasets

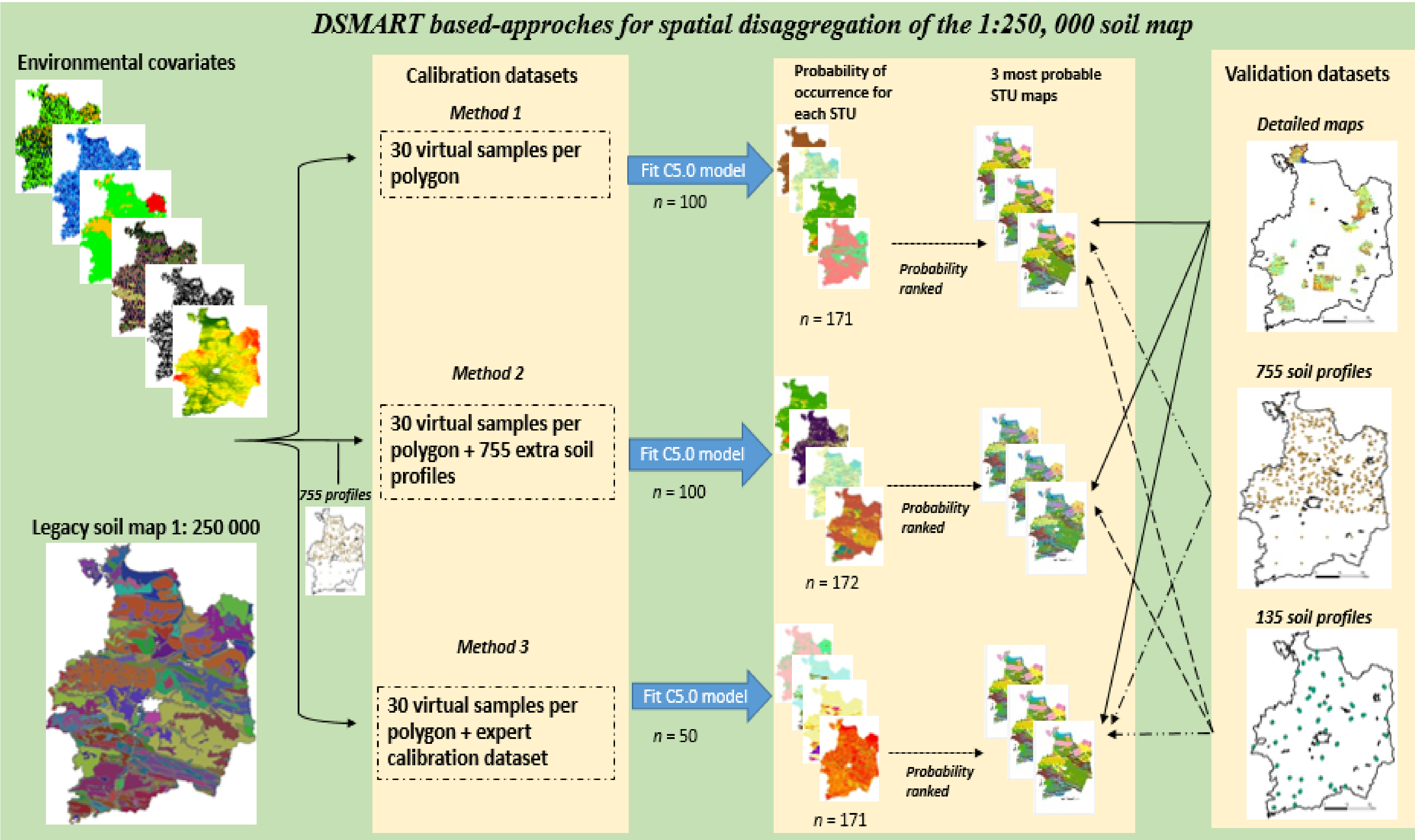

Figure 2: Schematic of the DSMART based approaches algorithm. The steps in DSMART are: 1) construct the calibration dataset; 2) train C5.0 model; 3) estimate Soil Type Unit (STU) maps and their associated probabilities of occurrence

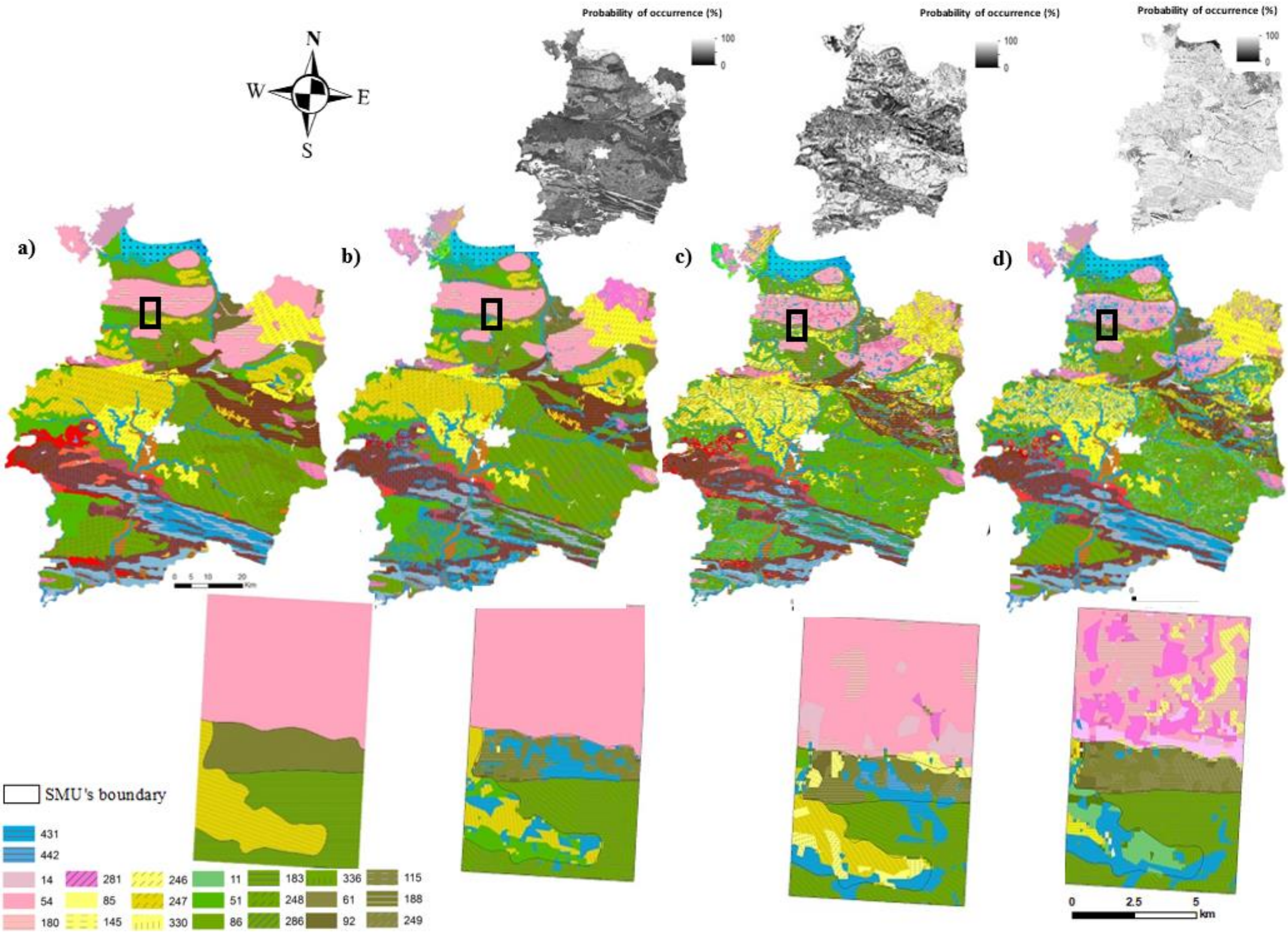

Figure 3: Digital soil map of the most probable STU and their associated probability of occurrence for the whole study area and for a focus zone, a) Legacy soil map: most probable STU for each SMU, b) original DSMART approach; c) DSMART with expert rules; d) DSMART with extra soil observations

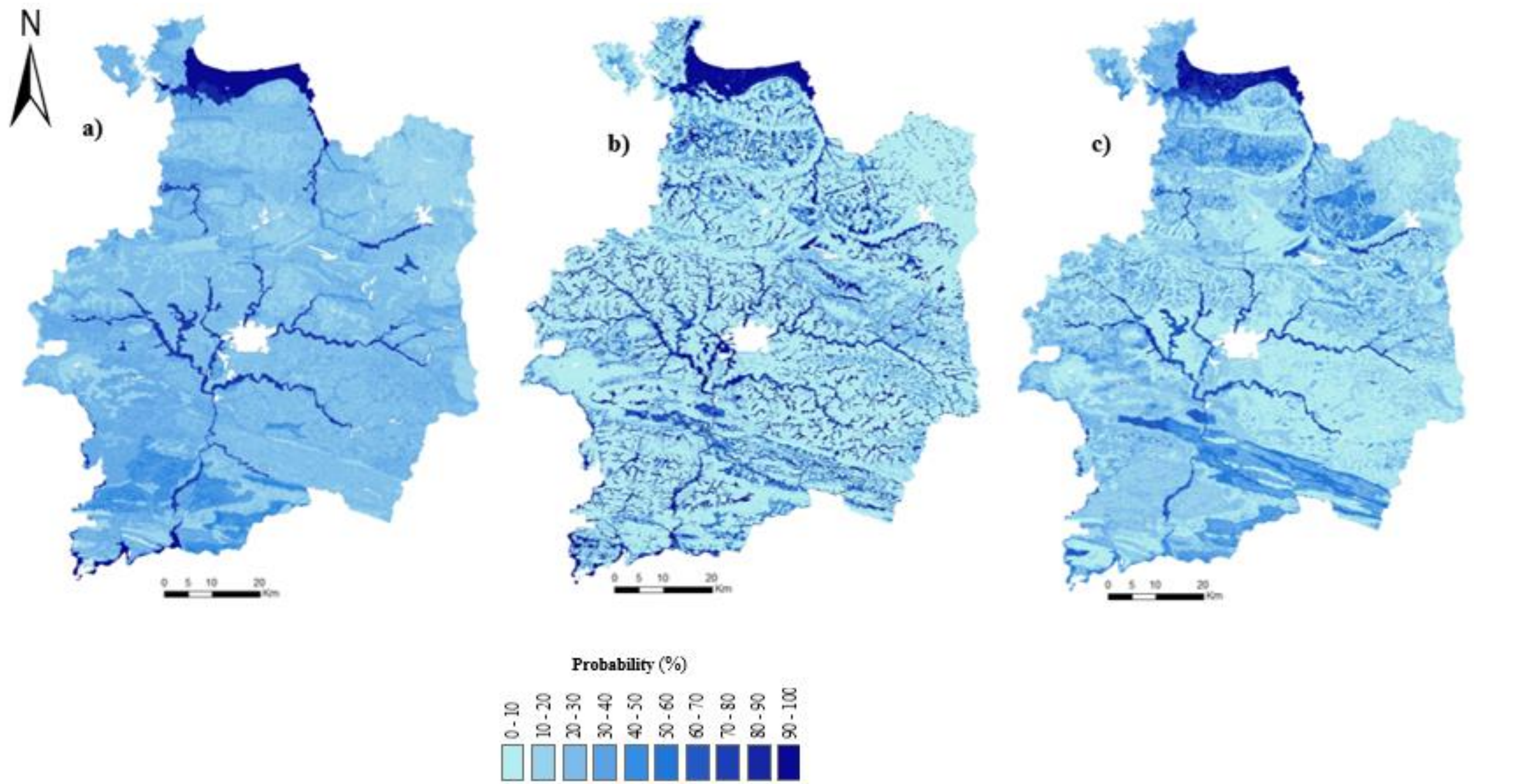

Figure 4: Global probability of hydromorphic soils over the study area derived from a) original DSMART, b) DSMART with soil landscape relationships and c) DSMART with extra soil observations. The probabilities of the three STU with highest prediction occurrence are summed if they are hydromorphic.

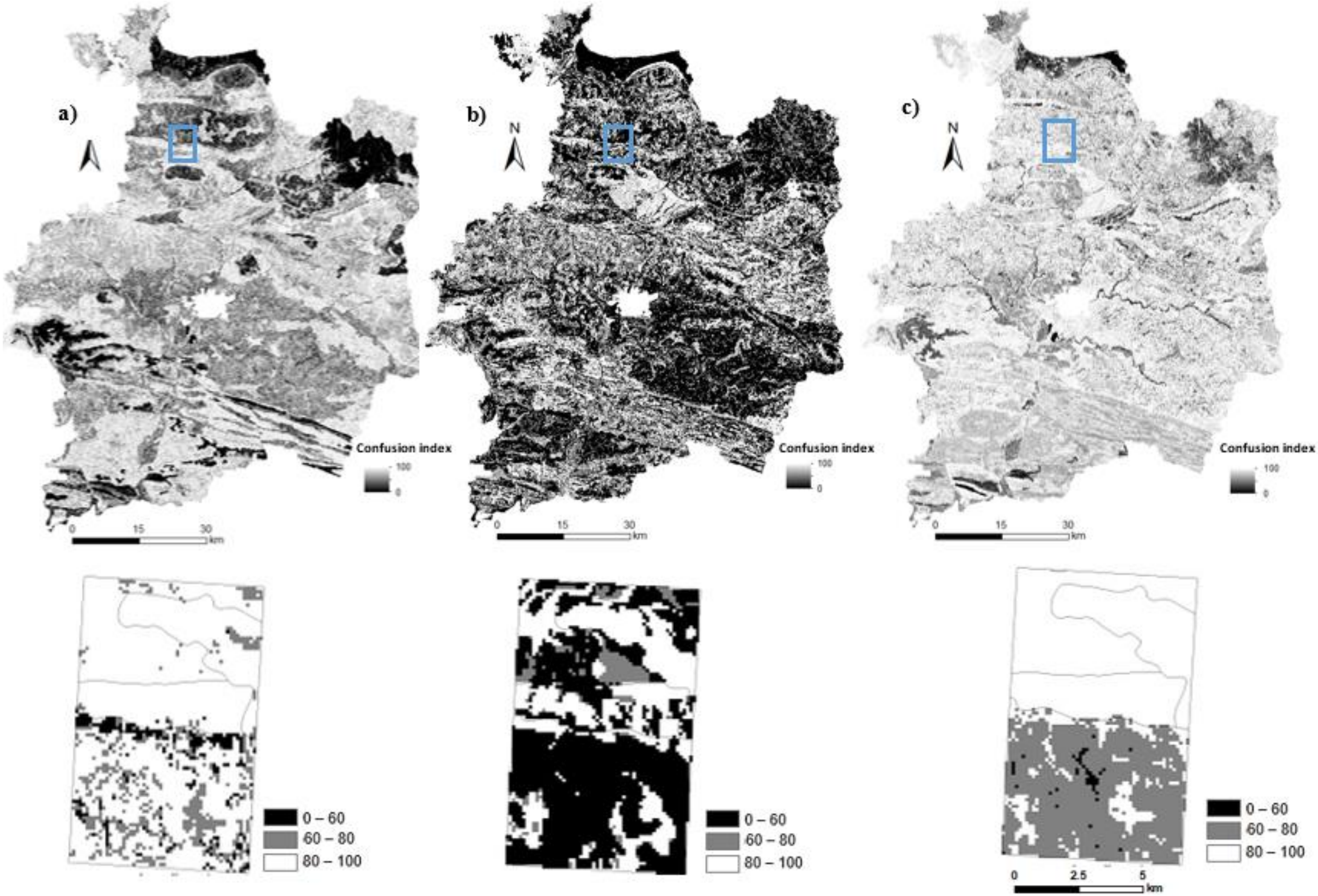

Figure 5 Confusion index maps for a) Classic DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations

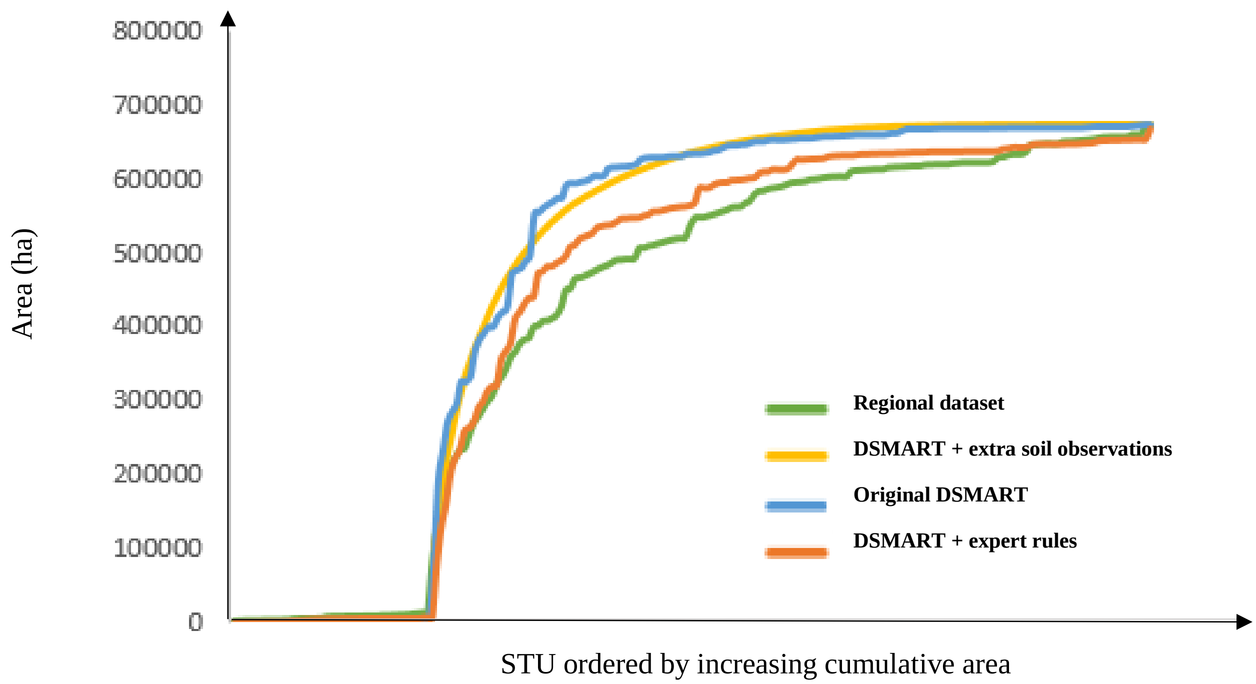

Figure 6: Cumulative area of the 171 STUs estimated from the regional soil database and predicted by different DSMART based approaches

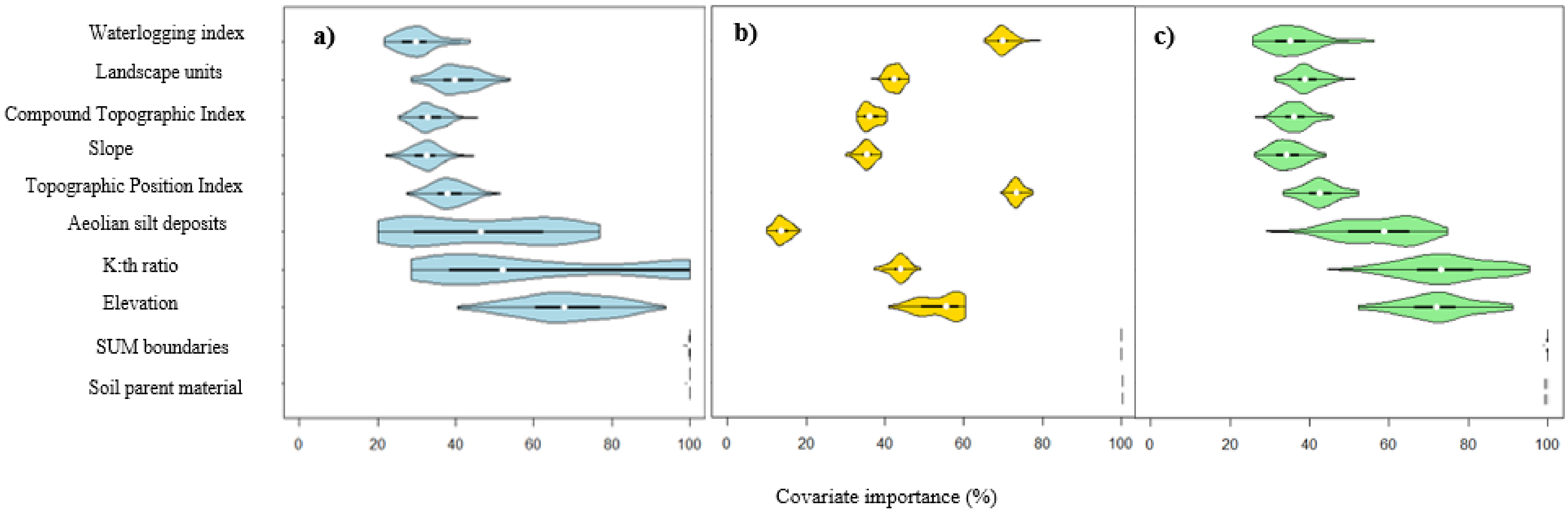

Figure 7: Violin plots of the relative importance of each environmental covariate used in a) Original DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations

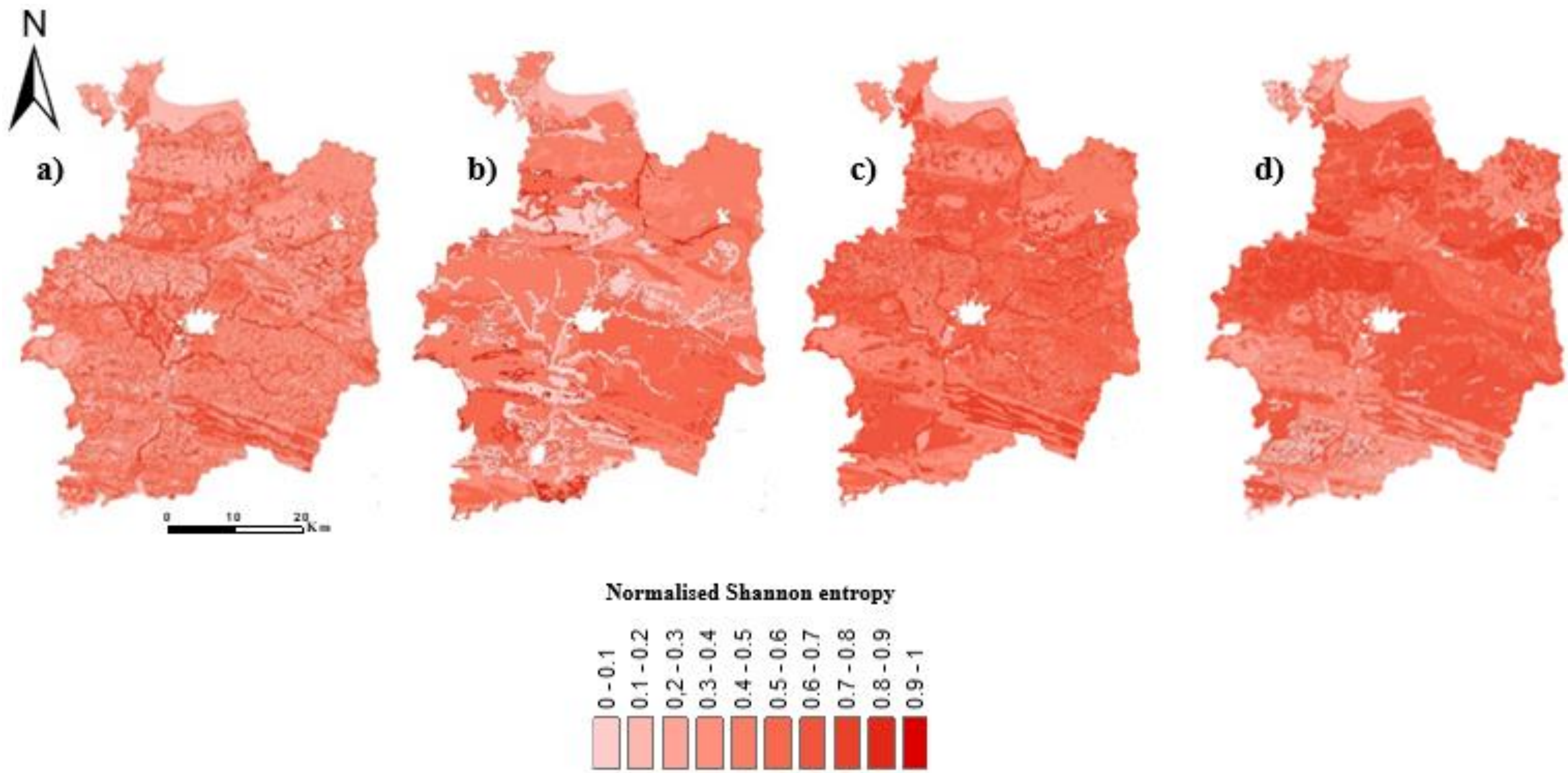

Figure 8: Spatial association between disaggregated maps of Ille et Vilaine department. a) map of inhomogeneity of DSMART with soil landscape relationships map in terms of original DSMART map b) map of inhomogeneity of original DSMART map in terms of DSMART with soil landscape relationships map c) map of inhomogeneity of DSMART with soil landscape relationships map in terms of DSMART with extra soil observations map d) map of inhomogeneity of DSMART with extra soil observations map in terms of DSMART with soil landscape relationships map. Inhomogeneity (variance) is measured by normalised Shannon entropy

Table 2 Ten most extended STUs according to the regional soil database and their respective rank by area using three DSMART based disaggregation procedures

STU			1 :250, 000 dataset		Original DSMART approach		DSMART with extra soil profiles		DSMART with expert rules	
Label	WRB classification	Parent material	Rank	Estimated area (km²)	Rank	Predicted area (km²)	Rank	Predicted area (km²)	Rank	Predicted area (km²)
431	Fluvisol Stagnic	Alluvial and colluvial deposits	1	688	2	757	1	983	1	740
248	Cambisol	Brioverian schists	2	480	1	1154	2	461	2	492
51	Cambisol	Brioverian schists	3	402	5	397	4	395	3	424
61	Cambisol	Gritty schists	4	227	9	177	30	53	14	128
183	Cambisol Stagnic	Sandstone	5	216	11	162	5	308	10	192
256	Cambisol	Aeolian loam	6	200	6	385	3	418	6	314
286	Cambisol Stagnic	Brioverian schists	7	179	23	62	9	187	24	80
86	Cambisol	Brioverian schists	8	169	12	126	15	124	4	358
340	Albeluvisol Stagnic	Granite and gneiss	9	168	7	347	10	177	11	189
54	Cambisol	Brioverian schists	10	167	4	451	18	98	5	324

Table 1. Description of the environmental covariates selected

Summary of environmental covariates. P: parent material; S: soil properties; R: relief; O: Organisms; C: categorical; Q: quantitative.

Environmental covariate	**SCORPAN factor**	**Type**	**Unit or number of classes**
Terrain attributes derived from the digital elevation model			
Elevation	R	Q	m
Slope	R	Q	%
Compound Topographic Index (TPI)	R	Q	Log (m^3)
Topographic Position Index	R	C	5 classes
Pedology and geology			
Soil parent material	P	C	22 classes
Soil Map Units	R	C	96 classes
Aeolian silt deposits	P	C	2 classes
Waterlogging index	S	C	4 classes
Organism			
Landscape units	O	C	19 classes
Gamma ray spectrometry from 250 m airborne geophysical survey interpolations			
K:Th ratio	P	Q	

Table 3. Overall accuracies (%) obtained using various external validation approaches for the three most probable STU

Pixel to pixel validation of STU					
	DSMART approach	Most probable STU	Second most probable STU	Third most probable STU	Total
Soil maps (87 150 ha)	Original DSMART	23	13	8	44
	DSMART with expert rules	19	11	7	37
	DSMART with extra soil observations	22	9	7	38
Independent soil profiles (n=135)	Original DSMART	11	5	3.8	18.1
	DSMART with expert rules	10	4.4	3.7	19.8
	DSMART with extra soil observations	8.2	6	2.7	16.9
Legacy soil profiles (n=755)	Original DSMART	14	7	6	27
	DSMART with expert rules	18	9	7	34
	DSMART with extra soil observations				

Pixel to pixel validation of STU group					
	DSMART approach	Most probable STU	Second most probable STU	Third most probable STU	Total
Soil maps (87 150 ha)	Original DSMART	26	13	9	48
	DSMART with expert rules	22.5	13.7	9.7	45.9
	DSMART with extra soil observations	25	10	7	42
Independent soil profiles (n=135)	Original DSMART	16	7	4.6	27.6
	DSMART with expert rules	18	8.4	5.2	31.6
	DSMART with extra soil observations	15	8	3.8	26.8
Legacy soil profiles (n=755)	Original DSMART	19	12	9	40
	DSMART with expert rules	23.4	15	11.8	50.2
	DSMART with extra soil observations				

Neighbourhood of 3 x 3 validation of STU					
	DSMART approach	Most probable STU	Second most probable STU	Third most probable STU	Total
Soil maps (87 150 ha)	Original DSMART	31	16	14	61
	DSMART with expert rules	29.6	19.4	13.1	62.1
	DSMART with extra soil observations	28	11	9	48
Independent soil profiles (n=135)	Original DSMART	15	6	4.3	25.3
	DSMART with expert rules	17	6.7	4.8	28.5
	DSMART with extra soil observations	11	7	3	21
Legacy soil profiles (n=755)	Original DSMART	19	10	7	36
	DSMART with expert rules	27.9	15	11.9	54.8
	DSMART with extra soil observations				

Table 4: Comparison between the size areas covered, number of soil map units, soil type units of the original legacy soil maps and the accuracy achieved in other studies using DSMART algorithm

Study	Area (km²)	Map units	Soil type unit	Accuracy
Odgers et al (2014)	68,000	1,110	72	23
Holmes et al. (2015)	2,500,000	5,069	73	20-22
Chaney et al. (2016)	-	-	-	17
Møller et al. (2019)	43,000	11-14	18-23	12-18