



HAL
open science

Evaluation biophysique des services écosystémiques des sols cultivés - Adaptation de l'information pédologique pour la modélisation dynamique du fonctionnement des sols

Yosra Ellili-Bargaoui

► To cite this version:

Yosra Ellili-Bargaoui. Evaluation biophysique des services écosystémiques des sols cultivés - Adaptation de l'information pédologique pour la modélisation dynamique du fonctionnement des sols. Sciences de la Terre. Agrocampus Ouest, 2019. Français. NNT : 2019NSARD087 . tel-03340240

HAL Id: tel-03340240

<https://theses.hal.science/tel-03340240>

Submitted on 10 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT

AGROCAMPUS OUEST
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 600
Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation
Spécialité : Sciences de la Terre et de l'Environnement

Par

Yosra Ellili Bargaoui

Evaluation biophysique des services écosystémiques des sols cultivés - Adaptation de l'information pédologique pour la modélisation dynamique du fonctionnement des sols

Thèse présentée et soutenue à AGROCAMPUS OUEST, Rennes, le 03/12/2019
Unité de recherche : UMR Sol Agro et hydrosystème Spatialisation INRA, AGROCAMPUS OUEST
Thèse N° : 2019-21_D-87

Rapporteurs avant soutenance :

Philippe Lagacherie : Ingénieur de recherche, INRA, UMR LISAH, Montpellier
Catherine Keller : Professeur, Cerege, Aix-Marseille Université

Composition du jury :

Examineurs : Patrice Cannavo	Professeur, AGROCAMPUS OUEST, unité EPHor, Angers
Isabelle Cousin	Directrice de Recherches, INRA, UR Sols, Orléans
Elies Hamza	Professeur, INA Tunis
Laura Poggio	Senior researcher, ISRIC Wageningen

Directeur de thèse : Christian Walter, Professeur, AGROCAMPUS OUEST, UMR SAS, Rennes
Co-encadrant de thèse : Didier Michot, Maître de conférence, AGROCAMPUS OUEST, UMR SAS, Rennes

Invité

Co-encadrante de thèse : Blandine Lemerrier, Ingénieure de recherche, AGROCAMPUS OUEST, UMR SAS, Rennes

Résumé

Les sols constituent la composante centrale de l'environnement et des ressources terrestres. Ils rendent de nombreux services écosystémiques essentiels au maintien de la vie planétaire et au développement socio-économique. Toutefois, les sols sont soumis à une pression croissante pour satisfaire les demandes d'une population croissante en matière de production alimentaire et énergétique. En vue d'une meilleure gestion, des outils d'aide à la décision intégrant la diversité des sols sont nécessaires pour évaluer l'impact des activités anthropiques sur l'évolution des sols.

Le sujet de thèse porte sur l'établissement et la validation d'une méthodologie de mise à disposition de l'information pédologique contenue dans les bases de données existantes en vue de l'évaluation biophysique de six services écosystémiques rendus par les sols cultivés. Les six services instruits sont la régulation du climat ; la régulation de la qualité de l'eau ; la recharge de la nappe souterraine ; la fourniture d'eau à la plante ; la fourniture d'azote à la plante ; et la production de biomasse agricole.

Le travail de thèse comporte trois volets principaux i) l'amélioration de l'information pédologique disponible en mobilisant des techniques numériques de désagrégation spatiale des cartes de sols existantes au 1/250,000 pour en dériver des cartes des types de sol et de propriétés des sols à haute résolution spatiale (50m) ; ii) la validation des cartes produites par désagrégation spatiale en mobilisant un jeu de données indépendant couvrant la variabilité géo-pédologique du département d'Ille et Vilaine en Bretagne (6 800 km²) et iii) l'évaluation de la sensibilité des services écosystémiques fournis par les sols cultivés à la source de l'information pédologique par une approche de modélisation dynamique (modèle INRA STICS) mise en œuvre sur une période de 30 ans.

La comparaison de trois méthodes de désagrégation spatiale montre l'intérêt de la prise en compte de règles décrivant les lois de distribution des sols pour restituer l'organisation des sols au sein des paysages et produire des cartes des types et propriétés de sol conformes aux spécifications du consortium international *Globalsoilmap*. La validation des cartes ainsi produites montre que leur qualité prédictive varie selon le critère pédologique considéré et le support géographique de validation retenu. Par ailleurs, les évaluations biophysiques des services écosystémiques des sols cultivés présentent une double composante de variabilité intra-annuelle d'origine climatique et de variabilité pédologique. Les évaluations de services écosystémiques réalisées à partir des données pédologiques initiales et des données pédologiques désagrégées

sont similaires et diffèrent significativement des évaluations réalisées à partir des observations de terrain.

La méthode de modélisation dynamique ainsi que la technique de désagrégation spatiale mises en œuvre permettent de promouvoir la prise en compte de l'information pédologique dans l'évaluation des services écosystémiques des sols cultivés. Cependant de nouvelles pistes de recherches sont envisagées pour améliorer la procédure de mise à disposition de l'information pédologique ainsi que la procédure d'évaluation des services écosystémiques des sols cultivés. Cela repose notamment sur l'exploration de nouvelles approches de cartographie numérique et l'adaptation d'une voie hybride d'évaluation des services écosystémiques combinant à la fois la modélisation et l'utilisation des proxys pertinents pour favoriser la considération formelle des sols dans des stratégies de gestion durable et de planification territoriale.

Mots-clés : sol, services écosystémiques, agroécosystèmes, bases de données pédologiques, indicateurs, modélisation, gestion multifonctionnelle des sols, exploitation agricole

Abstract

Soils are the central component of the Earth's environment and resources. They provide many ecosystem services essential to sustain human life and socio-economic development. However, soils are subjected to increasing pressure from current activities, including intensive land use to satisfy demands of a growing population for food and energy. To improve soil management, decision-support tools that consider soil diversity are required to assess impacts of human activities on soil dynamics.

The objective of this study was to develop and validate a method to enhance existing pedological information in order to use it to estimate soil ecosystem services. Six soil ecosystem services were investigated: climate regulation, water quality regulation, groundwater recharge, water provision to plants, nitrogen provision to plants and biomass production. The study had three main components: i) improve legacy soil information, available at 250 000 spatial scale, using a spatial disaggregation approach according to *GlobalSoilMap* specifications to generate soil property maps at a 50 m spatial resolution; ii) assess the predictive quality of these disaggregated soil property maps using an independent dataset covering the geo-pedological variability of the Ille-et-Vilaine department of Brittany (68 000 km²), and iii) assess the sensitivity of estimated soil ecosystem services to the source of soil information over a baseline period of 30 years using the STICS crop model.

The main results highlight the importance of using expert rules that describe soil distribution to derive coherent soil property maps. The validation of disaggregated soil maps showed that the predictive quality varied among soil properties and validation supports. The biophysical assessment of ecosystem services from cultivated soils identified intra-annual variability due to variability in both climate and soil. Moreover, ecosystem services estimated from coarse soil data and disaggregated soil data were similar but differed significantly from those estimated from field observations.

The dynamic modelling and spatial disaggregation approach promote the use of pedological information when estimating soil ecosystem services. However, future research must be performed to improve the availability of soil information and the method for estimating soil ecosystem services. In particular, it can rely on exploring new approaches to digital mapping and using a hybrid ecosystem services approach, which combines dynamic modelling and proxies, to promote the consideration of soil in sustainable management strategies and territorial planning.

Keywords: soil, ecosystem services, agroecosystems, pedological databases, indicators, modelling, multifunctional soil management, field

Table des matières

Introduction générale	1
Chapitre 1 : Synthèse bibliographique sur les services écosystémiques fournis par les sols : concepts et liens avec l'information pédologique	10
1.1. Le concept de services écosystémiques : enjeux et définition.....	11
1.2. Origine et émergence du concept de service écosystémique	12
1.3. Un concept encore sujet à débat	14
1.4. Les typologies de classification des services écosystémiques	17
1.5. Les approches d'évaluation des services écosystémiques.....	17
1.6. Le sol est un capital naturel à l'origine de plusieurs services écosystémiques	19
1.6.1. Le sol est le centre d'intérêt de la recherche scientifique actuelle	19
1.6.2. La controverse entre SES et contribution des sols à la fourniture des services écosystémiques.....	21
1.6.3. Contribution des sols à la fourniture des services écosystémiques	22
1.6.4. Les services écosystémiques des sols.....	25
1.7. Le sol est un composant central des écosystèmes anthropisés	27
1.7.1. Définition d'un écosystème agricole vs agrosystème	27
1.7.2. L'intérêt porté au sol dans un contexte agricole	28
1.7.3. Exemples de cadres conceptuels d'évaluation des services écosystémiques spécifique aux systèmes agricoles.....	29
1.8. Les indicateurs d'évaluation biophysique des services écosystémiques liés aux sols ..	33
1.8.1. Les données pédologiques mobilisées dans l'évaluation des services écosystémiques liés aux sols	35
1.8.2. La cartographie des services écosystémiques.....	37
1.8.3. Les incertitudes associées à l'évaluation et la cartographie des services écosystémiques des sols	39
Conclusion du chapitre 1	40
Chapitre 2 : Matériels et Méthodes	42
2.1. Présentation du site d'étude.....	43
2.1.1. Contexte climatique.....	43
2.1.2. Contexte géologique.....	44
2.1.3. Contexte pédologique.....	47
2.1.4. Structure du paysage et occupation du sol	48
2.2. Les données pédologiques	50
2.2.1. Le RRP de Bretagne.....	51
2.2.2. Structuration de la base de données	51
2.2.3. Extraction des données pédologiques du RRP de Bretagne	53
2.2.4. Analyse du taux de complétude de la base de données	55

2.2.5.	Résultats de caractérisation de la base de données Donesol 3 du RRP de Bretagne ..	56
2.2.6.	Jeu de validation	58
2.2.7.	Description des sols et analyses	60
2.2.8.	La caractérisation pédologique du jeu de validation	63
	Conclusion partielle	66
2.3.	Méthodologie d'évaluation des services écosystémiques des sols	67
2.3.1.	Présentation du modèle de culture STICS	67
2.3.2.	Le développement phénologique	68
2.3.3.	La croissance aérienne	69
2.3.4.	L'élaboration du rendement	69
2.3.5.	La croissance racinaire	69
2.3.6.	Le fonctionnement hydrique	70
2.3.7.	La dynamique de l'azote	70
2.3.8.	Transferts de chaleur, d'eau et de nitrates	70
2.3.9.	Gestion des pratiques culturales	73
2.3.10.	Microclimat	73
2.3.11.	Les services écosystémiques instruits et le choix des indicateurs	73
	Conclusion partielle	75
Chapitre 3 :	<i>Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm</i>	77
	Introduction du chapitre 3	78
	Comparing three approaches of spatial disaggregation of legacy soil maps based on DSMART algorithm.....	80
	Abstract	80
	3.1. Introduction.....	82
	3.2. Materials and methods	84
3.2.1.	Study area	84
3.2.2.	Soil data	85
3.2.2.1.	Regional soil database at 1:250 000 scale	85
3.2.2.2.	Soil validation data	86
3.2.3.	Environmental covariates	86
3.2.4.	Disaggregation procedure: DSMART algorithm.....	88
3.2.4.1.	Original DSMART algorithm (Method 1)	88
3.2.4.2.	Original DSMART algorithm + soil observations (method 2)	89
3.2.4.3.	Original DSMART algorithm + expert rules (Method 3)	89
3.2.4.4.	Prediction of the most probable STUs	90
3.2.5.	Validation of disaggregated soil maps	90
3.2.6.	Pair wise comparisons of disaggregated soil maps	93
	3.3. Results	94
3.3.1.	Disaggregated soil maps	94
3.3.2.	Covariates importance in the decision trees	102
3.3.3.	Validation of disaggregated soil maps	104
3.3.4.	Comparing disaggregated maps.....	107
	3.4. Discussion.....	110
3.4.1.	Performance of the disaggregation procedures.....	110

3.4.2.	Legacy soil data	111
3.4.3.	Taxonomic similarities	112
3.4.4.	Mapping comparison	113
3.4.5.	Improvements and future work.....	113
3.5.	Conclusion	114
	Conclusion du chapitre 3	116
	Chapitre 4 : Mapping soil properties at multiple depths from disaggregated legacy soil maps	119
	Introduction du chapitre 4	120
	Mapping soil properties at multiple depths from disaggregated legacy soil maps	121
	Abstract	121
	4.1. Introduction.....	123
	4.2. Materials and methods	125
4.2.1.	The study area.....	125
4.2.2.	Regional soil database at 1:250 000 scale	126
4.2.3.	Assessing and mapping soil properties at the regional scale	127
4.2.3.1.	Disaggregation of complex mapping units.....	127
4.2.3.2.	Modelling depth variation using spline functions	128
4.2.3.3.	Correlation analysis among soil properties	130
4.2.3.4.	Prediction of a complex soil variable (ECEC)	130
4.2.3.5.	Digital soil mapping of soil properties	130
4.2.3.6.	Spatial correlation of soil properties	131
	4.3. Results	132
4.3.1.	Spline-function outputs	132
4.3.2.	Correlation between soil properties before and after fitting smoothing spline functions	133
4.3.3.	Multicollinearity among soil properties	133
4.3.4.	Stepwise multiple linear regression	135
4.3.5.	Soil-property mapping at multiple depths.....	136
	4.4. Discussion.....	143
4.4.1.	Equal-area spline method: advantages and limitations	143
4.4.2.	Spatial structure of disaggregated soil-property maps	143
4.4.3.	Legacy soil data and DSMART limitations	144
	Conclusion and perspective	145
	Conclusion du chapitre 4	146
	Chapitre 5 : Validation of digital maps derived from spatial disaggregation of legacy soil maps.....	148
	Introduction du chapitre 5	149
	Validation of digital maps derived from spatial disaggregation of legacy soil maps.....	150
	Abstract	150
	5.1. Introduction.....	152

5.2. Materials and methods	155
5.2.1. Study area	155
5.2.2. Disaggregated soil map	157
5.2.3. Soil property estimation	158
5.2.4. Development of soil dataset for map validation	159
5.2.5. Laboratory analysis.....	161
5.2.6. Measures of map quality.....	162
5.2.6.1. Quality measures for continuous soil properties	162
5.2.6.2. Quality measures for categorical soil properties	163
5.2.6.2.1. Validation supports	163
5.2.6.2.2. Quality measures.....	164
5.3. Results	165
5.3.1. Soil spatial variability over short distances	165
5.3.2. Purity measures.....	165
5.3.3. Descriptive statistics of continuous soil properties for the validation dataset	171
5.3.4. Validation of continuous soil property predictions	174
5.4. Discussion.....	175
5.4.1. Soil map quality measures	175
5.4.2. Performance of categorical soil attributes prediction by spatial disaggregation	176
5.4.3. Accuracy of quantitative soil property predictions by spatial disaggregation	178
5.4.4. Improvement and future study	179
5.5. Conclusion	179
Conclusion du chapitre 5	181
<i>Chapitre 6 : Assessment of six soil ecosystem services by coupling simulation modelling and field measurement of soil properties</i>	<i>183</i>
Introduction du chapitre 6	184
Assessment of six soil ecosystem services by coupling simulation modelling and field measurement of soil properties	185
Abstract	185
6.1. Introduction.....	187
6.2. Materials and methods	190
6.2.1. The study area.....	190
6.2.2. The STICS soil-crop model	191
6.2.3. Datasets.....	191
6.2.3.1. Climate dataset	191
6.2.3.2. Soil dataset	192
6.2.3.3. Crop management	195
6.2.4. Simulation procedure	196
6.2.5. Assessment of SES indicators	196
6.2.6. Statistical analysis	199
6.3. Results	201
6.3.1. Inter-annual means of SES over the study area	201
6.3.2. Inter-annual temporal variability in SES indicators	201
6.3.3. Spatial variability of the annual estimates of SES indicators	204

6.3.4.	Interrelation between SES.....	207
6.3.5.	Interrelation between SES and soil properties	210
6.3.6.	Multivariate analysis of interrelations between SES and soil properties	211
6.4.	Discussion.....	214
6.4.1.	Comparing SES evaluations with a focus on the modelling approach limitations.....	214
6.4.2.	Explaining and interpreting SES interrelations	216
6.4.3.	Explaining and interpreting interrelations between SES and soil properties	217
6.5.	Conclusion	219
	Conclusion du chapitre 6	220
	Chapitre 7 : Sensibilité de l'évaluation des services écosystémiques rendus par les sols à la source de l'information pédologique	222
	Abstract.....	223
7.1.	Introduction.....	224
7.2.	Matériel et méthodes	225
7.2.1.	Données d'entrée pour la modélisation mise en œuvre.....	225
7.2.2.	La procédure de simulation STICS.....	226
7.2.3.	Evaluation des services écosystémiques des sols	226
7.3.	Résultats	227
7.3.1.	Distribution des SES moyens sur 30 ans selon la source d'information pédologique ...	227
7.3.2.	Comparaison des SES évalués à partir des données de terrain et à partir des données cartographiques	230
7.3.3.	Analyse en composantes principales (ACP) des services écosystémiques	232
7.4.	Discussion et conclusion.....	235
7.4.1.	La qualité de la base de données Donesol du RRP de Bretagne.....	235
7.4.2.	Les limites de la méthode de désagrégation des cartes	236
7.4.3.	L'approche d'évaluation des services écosystémiques mise en œuvre.	238
	Chapitre 8 : Discussion et conclusion générale.....	240
8.1.	L'amélioration de la prédiction spatiale des types de sol et de leurs propriétés	242
8.1.1.	La méthodologie de désagrégation spatiale	242
8.1.2.	Perspectives d'amélioration de l'approche de désagrégation spatiale.....	245
8.1.3.	La cartographie prédictive des propriétés des sols à l'échelle régionale.....	246
8.1.4.	La validation des cartes désagrégées	248
8.1.5.	Les incertitudes associées aux prédictions spatialisées sur les sols et la qualité de la base de données du RRP de Bretagne	251
8.2.	L'évaluation biophysique des services écosystémiques des sols cultivés.....	251
8.2.1.	Avantages et limites d'une approche d'évaluation des SES par modélisation dynamique.	253
8.2.2.	Vers une approche hybride d'évaluation des services écosystémiques des sols	253
	Références bibliographiques	257

Liste des tableaux

Tableau 2.1 : Synthèse des variables extraites de la base de données Donesol 3 du RRP de Bretagne.	54
Tableau 2.2 : Taux de renseignement des strates du RRP de Bretagne pour neuf propriétés de sol vis-à-vis leurs modalités de renseignement (valeurs modales et les bornes minimales et maximales) (SOC=teneur en carbone organique, CEC=capacité d'échange cationique, A=teneur en argiles, S=teneur en sables, L=teneur en limons, LG=teneur en limns grossiers, LF=teneur en limons fins et EG=charge en éléments grossiers).	57
Tableau 2.3 : Taux de complétude des UTS pour les différentes propriétés des sols en fonction des intervalles de profondeurs du programme <i>GlobalSoilMap</i> (SOC=teneur en carbone organique, CEC=capacité d'échange cationique, A=teneur en argiles, S=teneur en sables, L=teneur en limons, LG=teneur en limns grossiers, LF=teneur en limons fins et EG=charge en éléments grossiers).	58
Tableau 2.4 : Paramètres statistiques estimés de différentes propriétés analytiques de trois intervalles de profondeur (5-15 cm (n= 135) ; 30-60 cm (n=125) et 60-100 cm (n=103)).	65
Table 3.1 : Description of the environmental covariates selected.	87
Table 3.2 : Ten most extended STUs according to the regional soil database and their respective rank by area using three DSMART based disaggregation procedures.	101
Table 3.3 : Overall accuracies (%) obtained using various external validation approaches for the three most probable STU.	105
Table 3.4 : 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.	111
Table 4.1 : Pearson's correlation matrices between soil properties before (1020 strata) and after fitting spline functions (1710 standard horizons). SOC = soil organic carbon, CEC = cation exchange capacity. * P < 0.05; ** P < 0.01; *** P < 0.001.	135
Table 4.2 : Regression coefficients of effective cation exchange capacity (ECEC) regression models before and after fitting spline functions. SOC = soil organic carbon. * P < 0.05; ** P < 0.01; *** P < 0.001	136
Table 4.3 : Descriptive statistics of the disaggregated soil maps of the soil properties studied.	140
Table 5.1 : Characteristics of the 11 strata used for stratified simple random sampling to produce the validation dataset, dominant parent material, percentage of the study area and number of transects sampled.	160
Table 5.2 : Estimated partial, strict and mean purities (%) for the three soil maps over the entire study area.	166
Table 5.3 : Design-based estimates of purities of soil classification criteria, strict purity and mean purity for selected strata for the 1st most probable STU map.	168
Table 5.4 : Results of the χ^2 test performed to compare proportions of good prediction of classification criteria along hillslope positions for the three soil typological unit (STU) maps ($\alpha=5\%$).	170
Table 5.5 : Descriptive statistics of soil properties for the validation dataset at 5-15 and 30-60 cm (n = 135 and 125, respectively).	172

Table 5.6 : Accuracy indicators of soil property prediction at 5-15 and 30-60 cm (n = 135 and 125, respectively, root-mean-squared error (RMSE), relative-root-mean-squared error (RRMSE), mean error (ME) and coefficient of determination (R^2)).	175
Table 6.1 : Characteristics of six soil profiles in the dataset of 64 sampling points based on soil analysis and estimates used as input parameters for the STICS model.	194
Table 6.2 : Descriptive statistics of mean values of soil ecosystem services (SES) over the baseline period estimated at the 64 sites (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision).	201
Table 6.3 : Mean (and standard deviation) of soil ecosystem service values over the baseline period for 16 sites. These sites were selected by ranking the 64 sites by increasing yield then retaining every fourth site (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision).	205
Table 6.4 : Pearson correlation coefficients (r) between soil ecosystem service (SES) indicators and soil properties. (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision; AW: available water capacity; Obs: maximum rooting depth; pH: soil pH of the topsoil; CC: soil organic carbon content of the topsoil; Cly: clay content of the topsoil)	209
Table 6.5 : Statistics of stepwise multiple linear regression models predicting means soil ecosystem service (SES) values over the baseline period as a function of soil properties. (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision; AW: available water capacity; Obs: maximum rooting depth; pH: soil pH of the topsoil; CC: soil organic carbon content of the topsoil; Cly: clay content of the topsoil).	213
Table 6.6 : Results of a Multivariate Analysis of Variance (MANOVA) model comparing clay, soil pH, parent material, soil type, soil depth and soil drainage effects on mean values of soil ecosystem services over the baseline period. The F-statistic, approximated from Wilks' lambda, denotes the importance of model terms in explaining SES.	213

Liste des figures

Figure 0.1: Schéma conceptuel positionnant le sol comme un capital naturel générant des services écosystémiques au profit des besoins humains (adapté de Dominati et al., 2013) ..	1
Figure 0.2 : Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.....	9
Figure 1.1 : Relations entretenues entre les services écosystémiques et les constituants du bien-être humain (adapté du MEA, 2005)	12
Figure 1.2 : Nombre de publications sur les services écosystémiques entre 1990 et 2014 avec l'expression "ecosystem services" dans les corps des articles (adapté de Adhikari et Hartemink., 2016).....	13
Figure 1.3 : Modèle conceptuel des liens existant entre les processus biophysiques, les fonctions écologiques, les services écosystémiques et le bien-être humain. (Adapté de Greiner et al., 2018).	16
Figure 1.4 : Principaux objectifs, outils et approches de l'évaluation de la qualité des sols à travers l'histoire (adapté de Bünemann et al., 2018).....	21
Figure 1.5 : Cadre conceptuel de la fourniture de services écosystémiques spécifique au capital naturel du sol (adapté de Dominati et al., 2010).	23
Figure 1.6 : Le modèle conceptuel des services écosystémiques développé dans le cadre du projet européen RECARE qui vise à atténuer les menaces de dégradations des sols européens (Adapté de Schwilch et al., 2016).....	25
Figure 1.7 : Les fonctions des sols et ses interrelations (adapté de Pardo et al., 2016).	27
Figure 1.8 : Cadre conceptuel du groupe de travail MAES (2013) qui reconnaît à la biodiversité une place centrale (adapté de Maes et al. (2013).	30
Figure 1.9 : Représentation des concepts clefs manipulés dans le cadre de l'EFESE-EA appliqués à la fourniture de deux SE : l'un rendu directement à la société, l'autre rendu directement au gestionnaire de l'agroécosystème et dont la société dérive un avantage indirect. (EFESE-EA, 2017).	32
Figure 2.1 : Diagramme ombrothermique (Base de données CLIMATIK de l'INRA, station de Saint Jacques de la Lande située à 48 m d'altitude)	44
Figure 2.2 : Répartition cartographique des terrains briovériens dans l'ouest de la France (seuil du Poitou, Bretagne et Normandie) (BRGM, 2009).....	46
Figure 2.3 : Cartographie prédictive du matériau parental en Bretagne (A=Argile, Altérite, B=Cuirasse ferrugineuse, C=Calcaire, D=Dune sableuse d'origine marine, E=Eboulis de pente, F=Micaschistes, G=Granite I=Gneiss, L=Limons, M=Marais, N=Schiste tendre, O=Schiste moyen, P= Schiste dur, R= Schiste gréseux, S= Sable, T= Terrasse caillouteuse, U= Matériau d'apport colluvial, V= Matériau d'apport alluvial, X= Quartz et Poudingue, Z= Matériau remanié.	46
Figure 2.4 : Cartes de types des sols les plus fréquents en Bretagne.....	48
Figure 2.5 : Les unités paysagères de la Bretagne déterminées par télédétection (Le Du-Blayo et al., 2008).....	50

Figure 2.6 : Organisation des principaux objets pédologiques (UCS, UTS, strates, profils et horizons) dans le système d'information pédologique (SIG+DoneSol) (https://sols-et-territoires.org).....	52
Figure 2.7 : Les unités cartographiques des sols (UCS) en Bretagne (http://sols-de-bretagne.fr/donnees-et-outils/CARTES/PEDOPAYSAGES/).....	53
Figure 2.8 : Liens entre les différentes tables de DoneSol qui caractérisent les données surfaciques (https://sols-et-territoires.org/produits-du-reseau/publications-marquantes/)...	54
Figure 2.9 : Topographie du département d'Ille-et-Vilaine et localisation des sites de validation sélectionnés par un échantillonnage aléatoire stratifié selon le matériau parental dominant.....	60
Figure 2.10 : Exemples de profils de validation sur des schistes briovériens. Pour chacune de ces 4 photos, le profil de sol au milieu représente le profil situé sur le transect et les deux autres représentent les profils à courtes distances (à 20 m).	62
Figure 2.11 : Situation dans le triangle du GEPPA de la composition granulométrique des 1020 strates du RRP de Bretagne (A) et des 367 analyses de terre réalisées en Ille-et-Vilaine (B).....	64
Figure 2.12 : Schéma de l'organisation modulaire du modèle de culture STICS (adapté de GARCIA DE CORTAZAR ATAURI,2006).	68
Figure 2.13 : Conception des différents compartiments du sol selon le module de transfert de STICS. a) Description de la circulation de l'eau en sol profond et filtrant ; b) description de la circulation de l'eau en sol profond et tassé ; c) description de l'eau en sol superficiel sur substrat poreux. HCCF est l'humidité à la capacité au champ de chaque couche; HMINF l'humidité au point de flétrissement permanent de chaque couche ; et H est l'humidité pondérale de l'horizon. (adapté de GARCIA DE CORTAZAR ATAURI,2006).	72
Figure 3.1 : Location of the study area and the validation datasets.....	85
Figure 3.2 : 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.	92
Figure 3.3 : 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) DSM.....	96
Figure 3.4 : 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.	98
Figure 3.5 : Confusion index maps for a) Classic DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations.....	99
Figure 3.6 : Cumulative area of the 171 STUs estimated from the regional soil database and predicted by different DSMART based approaches.	100
Figure 3.7 : 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.....	103
Figure 3.8 : 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.	109
Figure 4.1 : Location of the Brittany study area, showing boundaries of soil map units (SMU) and elevation.	126
Figure 4.2 : The workflow used to derive disaggregated soil-property maps at the regional scale. STU = soil typology unit.	128
Figure 4.3 : The classification tree used to characterise lithic and paralithic horizons. STU = soil typological unit (adapted from Vaysse,2015).	129
Figure 4.4 : Strata silt content (solid grey rectangles), equal-area spline curves (red), and estimated mean silt content for <i>GlobalSoilMap</i> regular soil-depth intervals (open black rectangles) for four soil typological units.	132
Figure 4.5 : Principal component analysis based on physico-chemical and texture characteristics of soil typological unit strata (a) before and (b) after fitting spline functions. SOC: soil organic carbon content, ECEC: effective cation exchange capacity, sand: sand content, CF: coarse fragments, F-silt: fine silt, C: silt: coarse silt, and silt: silt content. ...	134
Figure 4.6 : Disaggregated maps of soil organic carbon (SOC) content in Brittany at the six soil-depth intervals.	138
Figure 4.7 : Disaggregated maps of silt content in Brittany for the six soil-depth increments.	139
Figure 4.8 : Experimental semivariograms derived from disaggregated soil maps for seven soil properties for six regular soil-depth intervals.	142
Figure 5.1 : Locations of the 45 transects by stratified simple random sampling based on dominant soil parent material strata.	156
Figure 5.2 : The 50 m resolution disaggregated soil map showing the most probable soil typological unit and associated probability of occurrence.	158
Figure 5.3 : Diagram of the transect sampling strategy along the hillslope.	161
Figure 5.4 : Diagram of the validation strategy to compute quality measures of digital soil maps and their derived soil properties maps.	162
Figure 5.5 : Percentage of validation plots in which 1, 2 or 3 categories of a given soil classification criterion (soil depth, drainage class, parent material, soil type, soil typological unit) were observed within a 20 m radius.	166
Figure 5.6 : Relation over the entire study area between strict and mean purities (%) and probability of occurrence of the three most probable soil typological units (STUs).	167
Figure 5.7 : Distribution functions of soil type purity and their mean values (red dots) by stratum for the (left) 1st and (right) 2nd soil typological unit (STU) predicted.	169
Figure 5.8 : Mean percentage of good soil type prediction and their 95% confidence intervals for the three most probable soil typological units (STUs) according to hillslope position. Dashed lines denote the overall soil type purity for the entire study area.	170
Figure 5.9 : Distribution of SOC content, pH and clay content according to observed parent material for the validation dataset. G: granite, I: gneiss, L: aeolian loam, N: soft schist, O: medium schist, P: hard schist, Q: sandstone, R: gritty schist, T: alluvial terrace, U: colluvium deposits, V: alluvial deposits, Vm: marsh. Whiskers represent 1.5 times the interquartile range. The number in brackets is the number of soil samples collected from	

each parent material. Means of groups sharing a letter in the group label do not differ significantly at $\alpha=5\%$	173
Figure 6.1 : Mean annual rainfall (mm), global solar radiation (MJ m^{-2}) and mean annual temperature ($^{\circ}\text{C}$) of the study area.	192
Figure 6.2 : Location of the 64 soil sites selected by stratified simple random sampling based on dominant soil parent strata (Ellili et al., 2019).	193
Figure 6.3 : The crop rotation simulated over the baseline period (maize-wheat-catch crop).	196
Figure 6.4 : Annual soil ecosystem service values for maize and wheat crops for two contrasting soils over the baseline period (red: coarse-textured Leptosol overlying gritty schist with a depth of 30 cm; blue: Endostagnic Cambisol on deep loam with a depth of 150 cm.	203
Figure 6.5 : Mean soil ecosystem service indicators of the 64 sites studied, ranked by increasing plant biomass provision (YE). Error bars represent 1 standard deviation.	206
Figure 6.6 : Bagplots of pairs of soil ecosystem service indicators. Darker shading identifies the bag that contains 50% of observations, while lighter coloured shading identifies the boundary separating inliers from outliers (the “loop”). Asterisks identify the point with highest half space depth (median depth) with its 95% confidence region.	208
Figure 6.7 : Network chart showing interrelations among soil ecosystem service indicators based on their mean values over the baseline period. Only significant correlations are shown. Line width indicates correlation intensity.	209
Figure 6.8 : Network charts showing interrelations between the SES indicator means values over the baseline period (blue circles) and soil properties (blue circles). Only significant correlations are shown. Line width indicates correlation intensity.	210
Figure 6.9 : a) Plots of SES indicator means with their associated standard deviations and some soil properties as supplemental variables on the two first PCA principal components. b) Biplot of PCA variables and the 64 sites according to parent material ($\alpha=5\%$) (G: granite, L: Aeolian loam, N: soft schist, O: medium schist, R: gritty schist, T: alluvial terrace, U: colluvial deposits, V: alluvial deposits, Vm: marsh) c) Biplot of PCA variables and the 64 sites according to soil type ($\alpha=5\%$) (B: Cambisol, C: Endostagnic Cambisol, D: Stagnic Alluvisol, R: Leptosol, Th: Histosol, U: Colluvic Cambisol, V: Fluvisol).	212
Figure 7.1 : Statistiques non paramétriques en 64 sites d’Ille-et-Vilaine des services écosystémiques des sols moyens sur 30 ans, en fonction de la source de l’information pédologique : (i) Legacy correspond à l’utilisation de l’UTS dominante du RRP ; (ii) DSMART aux données issues des cartes désagrégées avec le modèle DSMART ; (iii) Obs aux données observées et analysées en chacun des sites. Les lettres représentent la significativité du Test de Tukey, la même lettre indique des moyennes non significativement différentes au seuil de 0.05.	229
Figure 7.2 : Représentation en chacun des 64 sites de l’évaluation des services écosystémiques moyens sur 30 ans à partir des données cartographiques (RRP : triangle noir ; cartes désagrégées : cercle bleu) en fonction de celle issue des données d’observation en abscisse.	231
Figure 7.3 : Représentation du premier plan factoriel d’une analyse en composantes principales regroupant les individus des trois jeux de données mobilisés pour l’évaluation des services écosystémiques : a) projection des variables et b) projection des individus (64 sites x3) selon l’origine des données : Legacy correspond à l’utilisation de l’UTS dominante	

du RRP ; DSMART aux données issues des cartes désagrégées avec le modèle DSMART ; Obs aux données observées et analysées en chacun des sites.	234
Figure 8.1 : Schéma général de différentes approches d'évaluation des services écosystémiques des sols	255

« Le sol est un des biens les plus précieux de l'humanité. Il permet la vie des végétaux, des animaux et des hommes à la surface de la terre. »

Article 1 de la charte européenne des sols

Introduction générale

Le sol est un élément clé de l'écosystème terrestre situé à l'interface de la lithosphère, de la biosphère, de l'hydrosphère et de l'atmosphère (Szabolcs, 1994) (Figure 0.1). Sa forte réactivité face aux facteurs biotiques et abiotiques lui confère un rôle régulateur des flux échangés entre les différents compartiments de l'environnement. Le sol est un milieu complexe marqué par une pédogenèse lente, mais néanmoins sensible aux pressions anthropiques et aux divers facteurs de dégradation qui altèrent ses composantes physiques, chimiques et biologiques, et ainsi sa capacité de fonctionnement. Il s'agit donc d'un capital naturel non renouvelable à l'échelle humaine, indispensable, à protéger contre les différentes menaces de dégradation (IPBES, 2018).

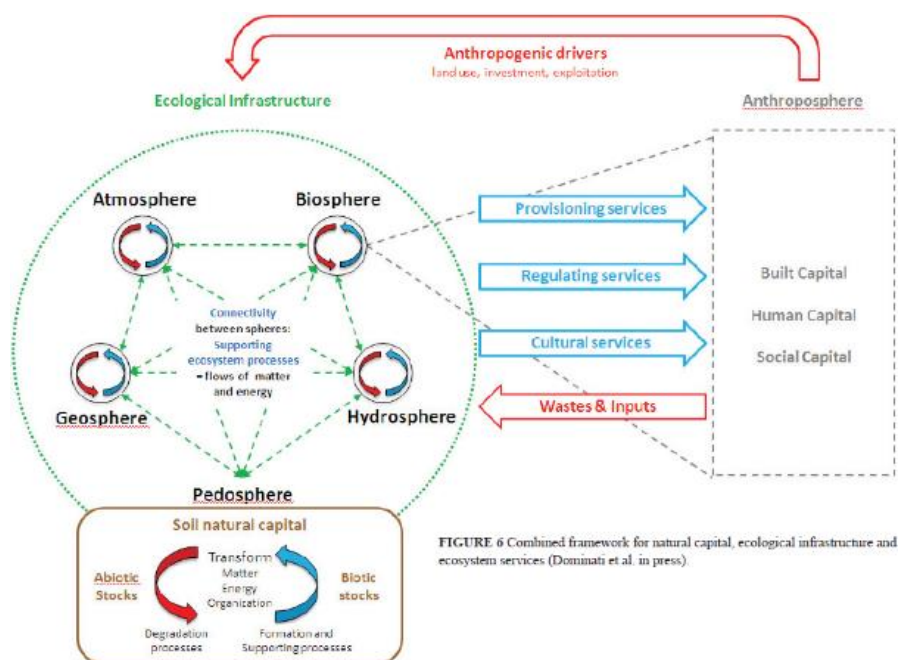


FIGURE 6 Combined framework for natural capital, ecological infrastructure and ecosystem services (Dominati et al. in press).

Figure 0.1: Schéma conceptuel positionnant le sol comme un capital naturel générant des services écosystémiques au profit des besoins humains (De Dominati et al., 2013)

Dans un contexte de changement climatique et de questionnement autour de la sécurité alimentaire, les sols constituent un levier important pour lutter contre les émissions des gaz à effet de serre et assurer la production alimentaire nécessaire à la population humaine. Plus encore, les sols constituent un réservoir d'eau et d'éléments nutritifs indispensables à la production agricole permettant de satisfaire les besoins primaires des êtres humains et des animaux.

A la lumière de la position centrale que le sol occupe, ce dernier représente une composante clé de l'agriculture, dont elle devra impérativement tenir compte pour limiter au maximum sa dégradation et optimiser ses usages. En effet, moyennant des pratiques agricoles non destructives et respectueuses de l'environnement, les espaces agricoles peuvent contribuer à la séquestration du carbone atmosphérique et à la lutte contre le réchauffement climatique tout en fournissant une nourriture abondante et de qualité. Plus encore, l'agriculture est un levier important pour que la France et plusieurs autres pays contribuent à la réussite de l'accord climat de la COP21. De ce fait, l'impérieuse nécessité de gérer durablement et efficacement les ressources en sol est devenue un enjeu crucial de la génération actuelle. Approfondir la connaissance du fonctionnement des sols demeure donc une nécessité incontournable pour répondre aux préoccupations liées à la durabilité des agrosystèmes. Cela suppose d'avoir une connaissance conjointe des propriétés du sol, de leur distribution spatiale, ainsi que de leurs déterminants agro-environnementaux.

Pour répondre aux enjeux planétaires tels que la conservation de la biodiversité, le maintien et l'amélioration du potentiel agronomique des terres agricoles, et la limitation des impacts environnementaux, les Nations Unies ont lancé en 2001 le programme du Millenium ecosystem assessment (MEA). Cette initiative a eu pour ambition d'évaluer l'état des écosystèmes naturels (éventuellement anthropisés), mais aussi de quantifier les avantages que l'humanité en retire. Elle a rassemblé de nombreux experts qualifiés qui ont mobilisé le concept de « service écosystémique » pour promouvoir le rôle de la biodiversité dans le maintien de l'équilibre écologique des écosystèmes. Selon le MEA (2005), les services écosystémiques sont définis comme les bienfaits que les êtres humains tirent du fonctionnement des écosystèmes naturels garantissant ainsi leur bien-être et leur développement socio-économique.

Depuis la publication du rapport du MEA en 2005, le concept de services écosystémiques a pris de l'ampleur et a été largement adopté par la communauté scientifique, mais aussi par les décideurs et gestionnaires des territoires. Ce concept émergent est considéré comme le fondement de plusieurs programmes de recherche actuels et un levier d'action fondamental pour garantir la durabilité des écosystèmes naturels. Cependant, le concept de services écosystémiques a fait

l'objet de plusieurs discussions, concernant notamment sa définition, ses cadres conceptuels et ses approches d'évaluation biophysique et de cartographie.

Appliqué au sol, le concept de services écosystémiques vise à mieux faire reconnaître la contribution des sols au bien-être et à l'activité socioéconomique des êtres humains. Il renforce les besoins de connaissance des sols, ce qui semble nécessaire pour les décideurs, qu'ils soient des gestionnaires directs (agriculteurs) ou indirects (marchés, politiques publiques) des espaces agricoles, afin de poursuivre la transition agroécologique et pour gérer les sols en tenant compte des services qu'ils rendent. En effet, les sols fournissent un ensemble de services écosystémiques d'approvisionnement, mais aussi de régulation, culturels et d'autoentretien (MEA, 2005). Les sols cultivés permettent de couvrir la majeure partie des besoins nutritionnels, en fibres et en énergie de la population humaine. Les projections futures, quels que soient les scénarios retenus, montrent que cette prédominance très forte de produits dépendants des sols devrait se maintenir, à la fois pour subvenir aux besoins d'une population en forte croissance et pour répondre aux défis de la transition énergétique. La nécessité de préserver et d'accroître les services d'approvisionnement fournis par les sols est donc un enjeu majeur du XXI^{ème} siècle. Ce besoin impératif de production risque de faire paraître secondaires les autres services assurés par les sols : régulation des flux d'eau ou des populations biologiques par exemple. Or, la préservation de ces services est essentielle à la performance et à la durabilité des écosystèmes productifs, et la mobilisation accrue des processus sous-jacents est un levier d'innovation important pour l'agroécologie. Ainsi, l'évaluation des Services Ecosystémiques rendus par les Sols (SES) est un enjeu important pour raisonner durablement la gestion des sols et des écosystèmes.

De plus, le sol est considéré comme un facteur déterminant du progrès socio-économique des pays ce qui rend son intégration dans les processus décisionnels cruciale (Daily et al., 1997). Plusieurs auteurs comme Robinson et al. (2012) et McBratney et al. (2014) pointent la nécessité d'évaluer les SES et de promouvoir les liens sol-écosystème dans l'élaboration des politiques de gestion des ressources naturelles via un ensemble de concepts liés à la santé, sécurité et qualité des sols. Ces derniers permettent d'éclairer le rôle multifonctionnel du sol dans la génération des flux de services écosystémiques au profit du développement socio-économique de la société. En effet, l'intégration des processus inhérents au sol dans des stratégies de gestion et de développement locales et globales permet d'optimiser les plans d'action en terme de durabilité et de gestion agro-environnementale. Par exemple, gérer les plans de fertilisation selon les aptitudes culturelles du sol permet d'améliorer la valorisation des nutriments par la plante, de réduire l'utilisation

d'intrants chimiques qui représentent une source de pollution potentielle des nappes souterraines et garantir ainsi un environnement plus sain pour la société dans son ensemble.

Le concept de SE a suscité depuis 10 ans de très nombreuses recherches pour définir des méthodes de quantification biophysique cohérentes avec les contextes agro-socio-environnementaux étudiés. Néanmoins, ces approches sont à ce stade très peu normalisées et stabilisées. Ceci est principalement lié à la complexité des nombreux processus sous-jacents mais aussi à leur complémentarité. Ainsi, le sol demeure une composante négligée et très peu prise en compte dans l'évaluation des services écosystémiques et des décisions politiques (Hewitt et *al.*, 2015). La plupart des études d'évaluation des SE sont dépourvues de la composante sol ou cette composante est mal définie ou trop généralisée. Plus encore, plusieurs variables environnementales sont utilisées pour substituer les informations sur les sols (Egoh et *al.*, 2008 ; Deng et *al.*, 2011) comme par exemple les données sur l'utilisation des sols et la couverture des sols (Land Use and Land cover - LULC) qui sont les plus couramment utilisées, en particulier dans les régions où les données pédologiques ne sont suffisamment pas précises et disponibles (Seppelt et *al.*, 2011 ; Schägner et *al.*, 2013 ; Vrebos et *al.*, 2015).

Comme le soulignent Guo et *al.* (2011) et Schägner et *al.* (2013), l'intégration de l'information pédologique dans l'évaluation des services écosystémiques permet d'améliorer la fiabilité et la précision des évaluations biophysiques délivrées. Cependant, la mobilisation de ces données est souvent confrontée à des problèmes de cohérence spatiale, de précision et de disponibilité (Maes et *al.*, 2012 ; Kandziora et *al.*, 2013). La faible disponibilité d'information spatialisée sur les sols aux différentes échelles de décisions agronomiques constitue un verrou majeur, que la cartographie dite numérique des sols peut contribuer à lever (Adhikari et Hartemink, 2016).

La cartographie numérique des sols est une approche efficace de production d'information pédologique qui peut alimenter les travaux d'évaluation des SES. Avec le développement d'outils logiciels issus des avancées de l'intelligence artificielle et de la modélisation numérique et spatiale, la cartographie numérique des sols permet de gérer l'information spatiale existante pour mieux décrire la variabilité spatiale des sols sur de vastes territoires.

Pour caractériser la distribution spatiale des sols à l'échelle nationale, la France dispose d'une base de données nationale mise en place par l'intermédiaire du programme IGCS (Inventaire, gestion et Conservation des sols) et notamment son volet Référentiel Régional Pédologique (RRP) qui vise à cartographier les sols au 1/250 000^{ème}. Ce référentiel est mis en place par département ou par région et répond à un cahier des charges défini au niveau national. Le RRP consiste en une

base de données relationnelle au format DoneSol (INRA, Infosol, 2014) associé à une délimitation spatiale des unités de pédopaysages. La base de données relationnelle met en relation les propriétés pédo-environnementales et la carte des pédopaysages définissant des Unités Cartographiques de Sol (UCS) délimitées dans l'espace. A l'échelle 1/250 000^{ème}, les UCS sont composées d'une ou plusieurs Unités Typologiques de Sol (UTS), qui ne sont pas délimitées géographiquement mais dont les proportions de couverture spatiale au sein des UCS sont explicitement précisées. Le sol étant un objet tridimensionnel, les UTS sont déclinées en strates dans la base de données relationnelle, une strate représentant la variation d'un horizon ou d'un groupe d'horizons diagnostics au sein d'une UTS.

Ces données pédologiques couvrent de larges étendues spatiales d'une façon assez homogène basée sur la caractérisation des processus pédogénétiques des sols. Elles s'appuient principalement sur des observations faites sur le terrain mais elles intègrent aussi de l'expertise pédologique pour assurer une certaine cohérence entre les propriétés du sol attribuées aux différentes strates et les processus pédogénétiques constatés. Toutefois, l'exploration du RRP révèle des problèmes de cohérence spatiale et de précision. En effet,

- La résolution spatiale du RRP est insuffisante pour répondre aux problématiques de gestion agro-environnementale à une échelle locale ;
- La précision des informations délivrées par le RRP est mal connue et difficilement appréhendable ;
- La complexité de manipulation de cette base de données est en soi un verrou méthodologique pour mobiliser les informations sur les sols et alimenter des outils d'aide à la décision.

Par ailleurs, l'évaluation biophysique des services écosystémiques nécessite de disposer d'une information spatialisée sur les sols afin de proposer des estimations intégratives des caractéristiques du milieu. Selon la méthode d'évaluation retenue, qu'elle soit basée sur la modélisation ou autre, l'information pédologique constitue le fondement du processus d'évaluation biophysique. Elle peut par exemple, être injectée dans des modèles simulant l'évolution dynamique des processus des sols dans un contexte de gestion agro-environnementale donné. Les sorties du modèle permettent ensuite de caractériser les indicateurs des services écosystémiques instruits. Cependant, la disponibilité de cette information pédologique est souvent limitée et nécessite un certain effort de mise à disposition aux échelles requises pour les évaluations biophysiques. Il est donc nécessaire de générer des prédictions de propriétés du sol à

une résolution spatiale fine, qui peuvent éventuellement être agrégées et remodelées aux résolutions et formats adéquats. Pour obtenir les prédictions des propriétés du sol à une résolution fine à partir du RRP, 2 étapes sont nécessaires *i)* désagréger les UCS pour la spatialisation des UTS et *ii)* prédire les propriétés quantitatives du sol à partir de la base de données relationnelle pour alimenter les modèles mécanistes.

Dans ce contexte, la désagrégation spatiale du RRP semble être une approche pertinente pour adapter l'information pédologique aux problématiques d'évaluation des services écosystémiques. Cette technique est surtout conseillée pour caractériser la distribution spatiale des sols à grandes échelles où seulement peu de données sont disponibles et bien renseignées.

Comme la définissent Odgers et *al.* (2014), la désagrégation spatiale est une approche de « descente d'échelle » permettant de produire une variabilité spatiale à partir d'une information à plus large étendue spatiale. Elle permet d'affiner le niveau de détails en proposant des estimations spatialisées d'une information donnée à une échelle supérieure à celle d'origine. De ce fait, la désagrégation spatiale permet de produire des données décrivant plus finement les sols pour confronter l'insuffisance des données disponibles. Elle représente un processus d'optimisation de l'information pédologique existante selon des critères assurant à la fois la cohérence de l'organisation horizontale de la couverture pédologique et la cohérence de l'organisation verticale des strates définissant chaque UTS. Par exemple, la désagrégation spatiale produit des estimations pédologiques cohérentes avec l'organisation spatiale des sols le long d'une toposéquence (les sols hydromorphes sont prédits en bas du versant alors que les sols bien drainés sont prédits en haut et milieu du versant). Plus encore, elle permet de respecter l'occurrence des processus pédogénétiques caractéristiques des évolutions particulières des sols comme le lessivage d'argile en profondeur. Dans ce cas, elle permet de prédire des teneurs plus élevées en argiles caractérisant les horizons illuviés des sols qualifiés de lessivés. Un autre atout promoteur de la désagrégation spatiale, c'est qu'elle assure une prédiction conjointe de plusieurs propriétés du sol tout en conservant les interrelations existantes entre elles et qui sont-elles mêmes révélatrices des processus pédogénétiques spécifiques.

Par ailleurs, la désagrégation spatiale des sols permet d'établir une relation statistique entre les variables locales et les prédicteurs de grande échelle. Elle permet donc de décrire la variabilité spatiale des sols en mobilisant un large panel de facteurs environnementaux et de facteurs pédogénétiques par l'implémentation possible des règles spécifiques sur les lois de distribution des sols dans un contexte donné. Pour l'essentiel, la désagrégation spatiale permet à la fois de

délimiter géographiquement les UTS et d'estimer les distributions spatiales des propriétés du sol à grande échelle. Elle délivre 2 types d'informations essentielles : *i)* des cartes de la distribution spatiale des propriétés qualitatives des sols à savoir le développement de profil ou type de sol, le matériau parental et l'hydromorphie des sols *ii)* des cartes de la distribution spatiale des propriétés quantitatives des sols à savoir les 5 fractions granulométriques du sol, la teneur en carbone organique, la charge en éléments grossiers, le pH du sol mesuré à l'eau et la capacité d'échange cationique (CEC).

La performance de cette technique ainsi que la précision des cartes de sols délivrées peuvent être évaluées par l'intermédiaire de plusieurs indicateurs de mesure de la qualité de prédiction. Cette phase de validation consiste à mesurer le degré de concordance entre des observations faites sur le terrain et les données prédites. Ceci permet d'une part d'estimer la précision et la fiabilité de ces données pédologiques spatialisées et d'autre part de quantifier les incertitudes qui peuvent être injectées dans le processus d'évaluation des SES.

Objectifs de la thèse

L'objectif général de la thèse est de développer une méthodologie d'extraction et de « reformulation » de l'information pédologique contenue dans les bases de données existantes, ce qui permettrait ensuite d'injecter des paramètres pédologiques adaptés à la résolution d'intérêt (parcelle agricole, exploitation, commune, bassin versant...) et d'une précision connue dans les modèles et indicateurs d'évaluation des services écosystémiques des sols agricoles.

Cet objectif suppose de répondre aux 4 questions de recherche suivantes :

- 1) Quelle approche de cartographie numérique pertinente peut être mobilisée pour améliorer la résolution spatiale des cartes des sols existantes et mettre en évidence la distribution spatiale des propriétés des sols à l'échelle régionale ?
- 2) Comment valider les cartes des sols obtenues d'une façon à prendre en compte l'incertitude liée aux données pédologiques dans l'évaluation des services écosystémiques des sols ?
- 3) Comment évaluer des services écosystémiques des sols dans le contexte des agrosystèmes par des approches biophysiques ?
- 4) Quel est l'effet de la source de l'information pédologique sur l'évaluation des services écosystémiques des sols sur une large étendue spatiale ?

Structure du mémoire

Ce mémoire s'organise en 8 chapitres (Figure 0.2)

- Le chapitre 1 est une synthèse bibliographique portant sur les services écosystémiques des sols en clarifiant les concepts qui en découlent et les liens avec l'information pédologique. Pour l'essentiel 5 points sont abordés : *i)* la définition du concept du service écosystémique et son origine *ii)* le cadre conceptuel relatif à l'évaluation des SES et les ambiguïtés associées *iii)* les méthodes d'évaluation biophysique des SES *iv)* l'information pédologique mobilisée dans l'évaluation des SES selon les approches retenues et *v)* la cartographie des SES.
- Le chapitre 2 présente l'ensemble des matériels et méthodes utilisées, à savoir le site d'étude, les bases de données mobilisées et les modèles mis en œuvre.
- Le chapitre 3 expose la modélisation de l'organisation spatiale des sols à haute résolution spatiale moyennant un algorithme de désagrégation spatiale des unités cartographique des sols en Bretagne.
- Le chapitre 4 est consacré à la prédiction tridimensionnelle des propriétés des sols nécessaires pour l'évaluation des SES à l'échelle régionale.
- Le chapitre 5 porte sur l'évaluation de la qualité prédictive des cartes des propriétés qualitatives et quantitatives des sols en utilisant plusieurs indicateurs et paramètres statistiques.
- Le chapitre 6 est consacré à l'évaluation biophysique des SES et leurs interactions éventuelles sur une échelle temporelle moyenne en mobilisant une approche de modélisation dynamique des processus biophysiques du sol.
- Le chapitre 7 est dédié à la comparaison de l'effet de la source de l'information pédologique sur l'évaluation des services écosystémiques des sols agricoles.
- Une conclusion générale et des perspectives sont présentées dans le dernier chapitre.

Le corps des chapitres 3,4,5 et 6 est constitué d'articles scientifiques publiés, soumis ou en voie de soumission pour des revues internationales à comité de lecture. Ils sont rédigés en anglais. Toutes les références citées dans cet écrit sont regroupées en fin de document.

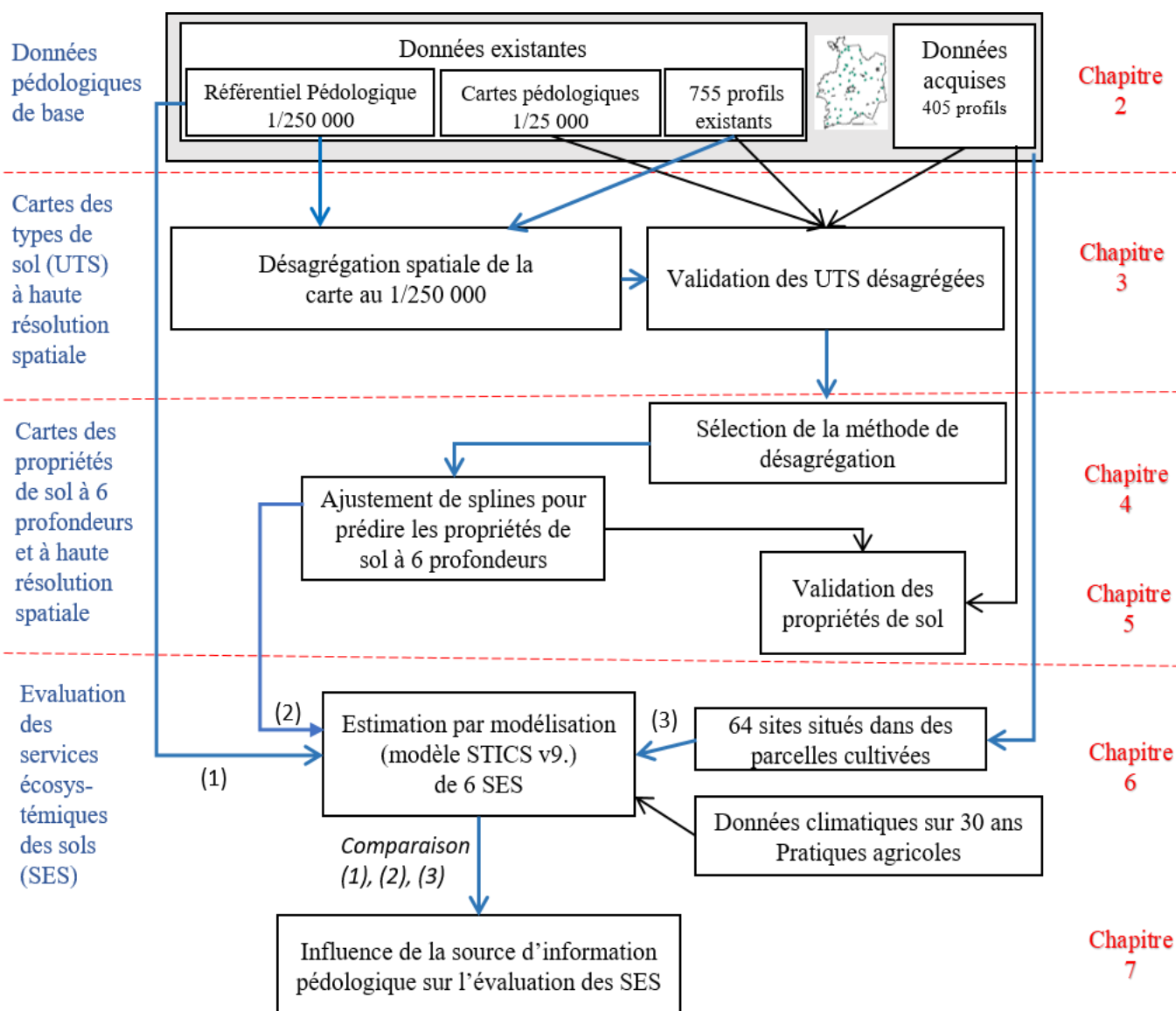


Figure 0.2 : Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 1 :
Synthèse bibliographique sur les services écosystémiques fournis par les sols : concepts et liens avec l'information pédologique

1.1. Le concept de services écosystémiques : enjeux et définition

La nouvelle ère géologique « Anthropocène » a été définie par Crutzen et Stoermer (2000) sur la base de l'influence prépondérante de l'activité humaine et souligne l'importance des impacts environnementaux liés aux activités anthropiques. Depuis la révolution industrielle, la détérioration de notre environnement et des écosystèmes naturels ne fait qu'accentuer le danger sur la persistance humaine et le reste de la biosphère. En parallèle, les conditions humaines en termes de santé, pauvreté, alimentation et niveau de vie ne cessent de s'améliorer dans un contexte d'explosion démographique et d'épuisement des ressources terrestres. Ce phénomène où les humains se portent de mieux en mieux tandis que leur environnement se détériore, est défini par les experts comme « le paradoxe de l'écologiste » (Boudes, 2008).

D'un point de vue rationnel et anthropocentré, la protection de l'environnement est nécessaire pour sauvegarder l'empreinte écologique et garantir la durabilité des écosystèmes naturels. Il existe des limites écologiques liées au caractère fini de la planète, qu'il est indispensable de respecter pour que les générations futures puissent profiter de ses ressources. Conscients de l'importante pression exercée sur l'environnement, les scientifiques de diverses disciplines alertent régulièrement les décideurs politiques et les gestionnaires directs et indirects des écosystèmes naturels sur les catastrophes potentielles qui pourraient être générées par la poursuite des tendances actuelles (Boudes, 2008).

En réponse aux crises environnementales et aux différents enjeux du 21-ème siècle, les Nations Unies ont lancé en 2001 l'évaluation des écosystèmes pour le Millénaire (MEA). Cette évaluation a été conduite sur 4 ans pour appréhender les conséquences de l'évolution des écosystèmes sur le bien-être humain et construire des bases scientifiques solides qui visent essentiellement la conservation de la biodiversité naturelle et l'exploitation durable des ressources planétaires. Plusieurs écosystèmes naturels complexes ont été instruits allant des écosystèmes naturels intacts, sans aucune intervention humaine tels que les forêts naturelles jusqu'aux écosystèmes sous contrôle intensif de l'homme tels que les agrosystèmes (MEA, 2005).

Le MEA a adopté le concept des « services écosystémiques » pour évaluer et caractériser les bénéfiques que les êtres humains tirent du fonctionnement des écosystèmes. Il définit les services écosystémiques par « les bénéfiques fournis par les écosystèmes, dont les services de prélèvement comme la nourriture, l'eau, le bois, les ressources génétiques ; les services de régulation tels que

la régulation du climat, des inondations, des maladies de la qualité de l'eau ou le traitement des déchets ; les services culturels comme les activités récréatives, les bénéfices liés à l'esthétique, au spirituel, et les services de soutien tels que la formation des sols, la pollinisation ou le cycle des nutriments » (MEA, 2005).

Les services écosystémiques traitent donc le lien de dépendance entre les êtres humains et leur milieu (Figure 1.1). Ils sont repartis en quatre grandes catégories : les services d'approvisionnement, les services de régulation, les services culturels et les services de support.

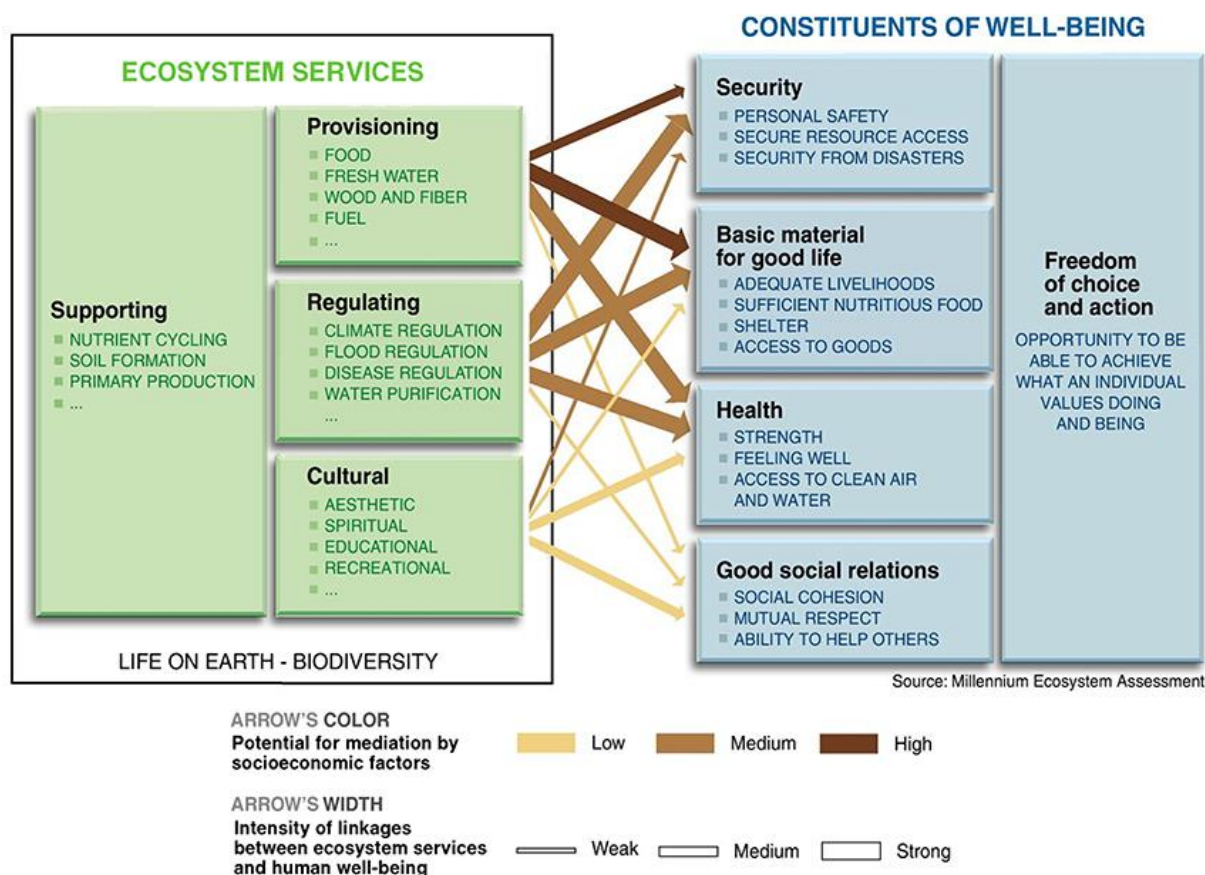


Figure 1.1 : Relations entretenues entre les services écosystémiques et les constituants du bien-être humain (De MEA, 2005)

1.2. Origine et émergence du concept de service écosystémique

Baveye et al. (2016) ont retracé les principales étapes de l'émergence du concept de « service écosystémique ». Ce concept a été proposé dans les années 1970 sous l'impulsion de naturalistes engagés dans la conservation de la Nature (Westman, 1977). Pour ces auteurs, il s'agissait

d'identifier la responsabilité du système économique dans l'exploitation non durable des ressources planétaires et la surexploitation des terres datant du 20^{ème} siècle. Avec l'évolution des approches économiques et écologiques de la nature, le concept de service écosystémique a constamment incité la prise en compte de l'environnement dans les décisions politiques.

Ce concept s'est considérablement développé vers la fin des années 1990 à la suite des travaux économiques de Costanza et *al.* (1997) et Daily (1997). Ces deux publications-clés ont permis de souligner les liens de dépendance entre l'être humain et les écosystèmes naturels pour sensibiliser et faire réagir le grand public. Dans le champ scientifique, l'adoption de ce concept par de nombreux chercheurs de diverses disciplines s'est concrétisée à travers l'évolution du nombre des publications scientifiques portant sur ou considérant les services écosystémiques depuis la parution de ces deux publications fondatrices, avec une accélération indéniable à la suite de la publication du rapport de l'évaluation des écosystèmes pour le Millénaire en 2005 (Figure 1.2).

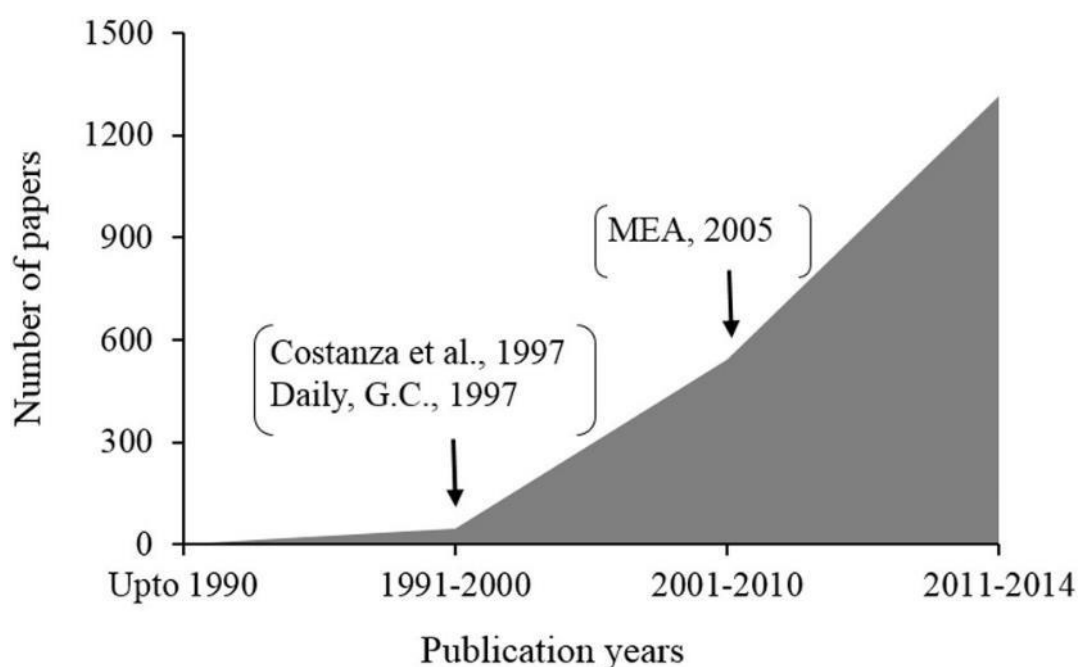


Figure 1.2 : Nombre de publications sur les services écosystémiques entre 1990 et 2014 avec l'expression "ecosystem services" dans les corps des articles (De Adhikari et Hartemink., 2016)

Depuis 2005, le concept de services écosystémiques a véritablement pris de l'ampleur au niveau mondial. Il a fait l'objet de plusieurs rapports institutionnels et décisionnels. Par exemple le MEA est décliné au niveau européen avec le programme MAES (Mapping and Assessment of Ecosystem Services) (Maes et *al.*, 2013) et au niveau national avec le programme EFESE (Evaluation Française des Services Ecosystémiques) (EFESE., 2017). En parallèle, la Convention

sur la Biodiversité Biologique (CBD) a désigné un comité scientifique permanent pour suivre le fonctionnement des écosystèmes appelé The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES).

1.3. Un concept encore sujet à débat

Bien que la définition des services écosystémiques et leur classification en 4 grandes catégories proposées par le MEA en 2005 soient communément admises, ces dernières ont cependant été l'objet de nombreuses discussions (Fisher *et al.*, 2009). Depuis la fin des années 2000, la communauté scientifique cherche à clarifier ce concept afin d'harmoniser son utilisation, d'assurer la comparabilité des travaux et centraliser d'une manière efficace et cohérente l'ensemble des connaissances acquises (Fisher *et al.*, 2009).

Actuellement, la coexistence de plusieurs terminologies assez larges, formées à travers de nombreuses expressions telles que « bénéfiques », « contributions », « processus », « conditions » et « fonctions écologiques » génère une grande ambiguïté dans la définition précise des services écosystémiques, ce qui limite la considération formelle de toute leur diversité et complexité. Le concept de services écosystémiques reste majoritairement confiné à la recherche scientifique, bien qu'il soit en voie d'adoption croissante par de nombreuses organisations et institutions pour consolider et promouvoir la prise en compte des services écosystémiques dans des politiques d'aide à la décision.

Devant la difficulté d'implémenter un consensus sur la définition du concept de services écosystémiques, plusieurs cadres conceptuels sur les services écosystémiques ont vu le jour. Certains se reposent sur le modèle en cascade ou la chaîne de fourniture des biens ou avantages, développé par Haines-Young et Potschin en 2010, afin de caractériser la relation entre le fonctionnement du système et le bien être humain (Figure 1.3). Cette cascade distingue 4 composants principaux reliant les deux extrémités de la chaîne de production en 3 étapes : 1) les structures et les processus écologiques génèrent des fonctions écologiques ; 2) ces fonctions écologiques assurent la fourniture des services écosystémiques et 3) ces services rendent des bénéfices aux êtres humains. Néanmoins, l'adoption de ce concept par plusieurs scientifiques révèle des nuances concernant la définition de chaque composant et la distinction entre le composant et la relation. Ainsi, des réinterprétations variées de ce cadre conceptuel peuvent générer des incohérences et des ambiguïtés.

Pour l'essentiel, 2 approches assez contrastées se sont développées : l'approche biophysique ou naturaliste et l'approche socio-économique. La première approche, principalement adoptée par les écologues, se focalise sur les relations entre les processus écologiques des écosystèmes avec l'ensemble des fonctions et/ou services écosystémiques. Elle admet que le fonctionnement et le maintien des écosystèmes naturels sont assurés par le biais de nombreuses fonctions écologiques associées à des processus biophysiques (Costanza et *al.*, 1997 ; De Groot et *al.*, 2002). Ces fonctions écologiques désignent des phénomènes propres à l'écosystème qui résultent de la combinaison de l'état des écosystèmes, des structures et des processus écologiques et qui se déroulent avec ou sans la présence de l'Homme. Les services écosystémiques sont ainsi la résultante des interactions entre les processus physico-chimiques et écologiques des écosystèmes (Dominati et *al.*, 2010). Par ailleurs, l'aspect naturaliste de l'approche en question reconnaît que les processus et les fonctions existent indépendamment de l'être humain et deviennent des services écosystémiques dès qu'une appropriation individuelle ou sociétale s'impose (Maris, 2014). Les services écosystémiques sont donc des relations entretenues avec le monde naturel (Maris, 2014) dont l'évaluation biophysique prend en compte les processus sous-jacents via des indicateurs qui ne sont forcément pas directement liés au bien-être humain.

Contrairement à l'approche biophysique, l'approche socio-économique met en avant la composante correspondant aux avantages tirés par l'être humain. Cette voie est privilégiée par les économistes qui promeuvent la distinction entre service écosystémique et avantage (Fisher et *al.*, 2009). Elle vise à rendre opérationnelle l'utilisation des services écosystémiques en proposant des outils d'évaluation monétaire des services écosystémiques. Cependant, l'approche socio-économique n'intègre pas la connaissance des processus sous-jacents des services écosystémiques, induisant ainsi des évaluations très différentes selon le scénario ou la méthode mobilisés (Maris, 2014).

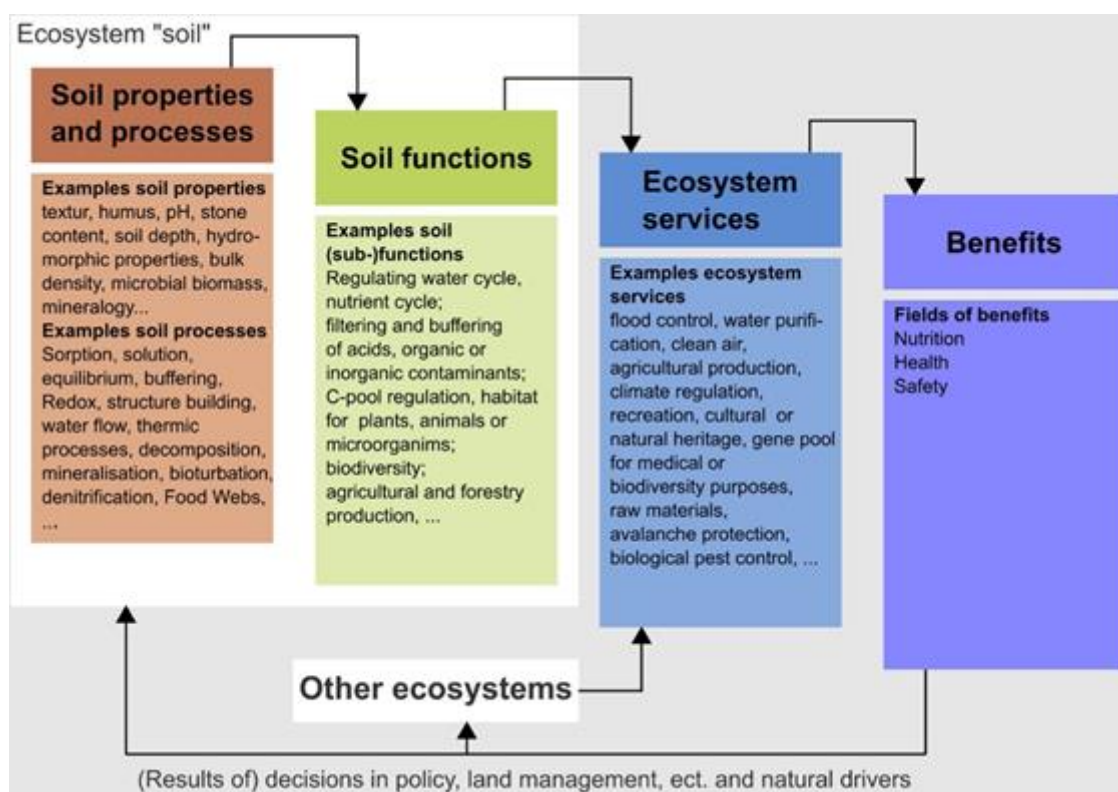


Figure 1.3 : Modèle conceptuel des liens existant entre les processus biophysiques, les fonctions écologiques, les services écosystémiques et le bien-être humain. (De Greiner et al., 2018).

Dans l'optique de développer une évaluation multidimensionnelle et intégrative de tous les éléments constituant le milieu naturel, Cowling et al. (2017) proposent un cadre conceptuel croisé en couplant les aspects biophysiques et socio-économiques. Le concept de service écosystémique peut donc s'entendre comme un concept au carrefour de l'écologie, de l'économie et des sciences sociales en privilégiant des approches systémiques et opérationnelles. Cela a pour ambition de promouvoir la complémentarité existante entre les disciplines dans le développement des stratégies de gestion des écosystèmes naturels.

Clairement, le cadre conceptuel de services écosystémiques évolue selon les perceptions écologiques et socio-économiques. Pour le rendre plus appréhendable par les usagers et les bénéficiaires, Nahlik et al. (2012) ont distingué 2 grands types de définition issues de EFES-EA (2017) : i) « les services écosystémiques sont des composants biophysiques de l'écosystème dont sont dérivés des avantages » et ii) « les services écosystémiques sont directement les avantages ». D'autres définitions intermédiaires entre les 2 définitions de référence compliquent l'harmonisation des notions. De ce fait, l'adoption du concept de services écosystémiques comme

un référentiel incontournable des politiques publiques, agricoles ou environnementales est confrontée à des problèmes de standardisation de cadres conceptuels ou analytiques.

Plusieurs travaux ont tenté d'harmoniser le cadre analytique de services écosystémiques. Parmi eux, on note les travaux de Groot et *al.* (2002), Wallace (2007), Fisher et *al.* (2009) et Maes et *al.* (2012) qui ont proposé de « conceptualiser les SE comme les composants biophysiques des écosystèmes utilisés par l'Homme, activement ou passivement, pour son bien-être » (EFESE-EA, 2017). Ces travaux se situent dans la lignée des travaux internationaux de la classification internationale des services écosystémiques dénommée « Common International Classification of Ecosystem Services » (CICES) (Agence européenne de l'environnement - AEE, 2013) et, dans sa continuité, le programme MAES (Maes et *al.*, 2013).

1.4. Les typologies de classification des services écosystémiques

Dans la littérature scientifique, la définition et la classification de services écosystémiques sont étroitement dépendantes du contexte d'application et de l'écosystème retenu. Pour l'essentiel, il existe 3 grandes typologies couramment utilisées dans la littérature à savoir la classification standard du MEA (2005), la classification TEEB (The Economics and Ecosystem and Biodiversity, 2010) et la classification CICES. Cette dernière retient les catégories de services d'approvisionnement, régulation et culturel. Elle considère que les services de soutien sont implicitement évalués à travers les services de régulation. D'autres classifications à application locale sont aussi développées par les communautés scientifiques pour adapter les typologies de services écosystémiques au contexte local et aux objectifs de l'étude menée comme par exemple la classification Swiss Federal Office for the Environment (FOEN) (Drobnik et *al.*, 2018).

1.5. Les approches d'évaluation des services écosystémiques

L'évaluation biophysique des services écosystémiques se fait essentiellement par le biais d'indicateurs. Comme le définit l'organisation de coopération et de développement économique (OCDE, 2003), un indicateur est « une mesure d'un phénomène environnemental utilisé pour décrire ou évaluer les conditions ou changements environnementaux ». Dans le champ

scientifique, plusieurs types d'indicateurs sont proposés pour appréhender les services écosystémiques à savoir les indicateurs de composition, structures, processus et fonctions des écosystèmes. En effet, les indicateurs de structure reposent souvent sur l'occupation du sol (Eigenbrod et *al.*, 2010) où chaque type d'occupation du sol est associé à un niveau de fourniture d'un service écosystémique donné. Les indicateurs de processus et de fonctions permettent de représenter d'une manière assez simplifiée les processus écologiques à l'origine des services écosystémiques. Par exemple, le service de régulation du climat peut être estimé via le stock et/ou la variation du stock de carbone qui est la résultante de la fonction de stockage de carbone mobilisant un ensemble de processus biophysiques à savoir la minéralisation, la fragmentation et la décomposition de la matière organique. Cependant, l'existence de plusieurs interactions complexes entre les déterminants du milieu, les processus biophysiques et les services écosystémiques limite le développement d'indicateurs pertinents et intégratifs de toutes les interactions possibles avec les composantes du milieu physique.

Dans la quasi-totalité des études scientifiques, trois approches sont couramment utilisées pour dériver des indicateurs d'évaluation biophysique des services écosystémiques :

- L'utilisation des proxies : cette approche consiste à attribuer une valeur de services écosystémiques à chaque type d'occupation du sol (Eigenbrod et *al.*, 2010 ; Fisher et *al.*, 2011). L'étude réalisée par Calzolari et ses co-auteurs en 2016 illustre l'utilisation des proxies dans l'évaluation des services écosystémiques. En effet, ces chercheurs ont appréhendé le service de fourniture d'habitat pour la biodiversité via un indicateur de la qualité biologique des sols spécifique à l'abondance des microarthropodes. Leur approche consiste à associer un niveau de fourniture de service (faible, moyen et fort) à chaque type d'occupation de sol préalablement identifié (prairies permanentes, terres cultivées avec intégration des prairies dans les rotations et autres occupations de sol).
- L'application de modèles : deux grands types de modèles peuvent être appliqués pour estimer les services écosystémiques : i) les modèles mécanistes fondés sur les équations empiriques ou les modèles mathématiques décrivant les processus biophysiques tels que l'Universal Soil Loss Equation (Wischmeier et Smith, 1978), largement utilisé et ii) les modèles économiques basés sur les théories de transferts de bénéfices (Costanza et *al.*, 1997), les coûts de remplacement (Busch et *al.*, 2012) et l'attribution des prix sur le marché (Naidoo et Ricketts, 2006).

- Les mesures directes : bien que de nombreux scientifiques insistent régulièrement sur l'adoption de cette approche comme outil de surveillance à long terme des écosystèmes naturels, cette dernière reste marginalement développée par la communauté des services écosystémiques. Ceci est principalement lié à la lourdeur et au coût des protocoles de mesures directes des processus et des fonctions biophysiques à l'origine des services écosystémiques (Eigenbrod et *al.*, 2010). Par exemple, afin d'évaluer le service de régulation du climat via la fonction de séquestration de carbone dans les sols, Ellili et *al.* (2019) ont réalisé des mesures diachroniques des teneurs en carbone organique et de densités apparentes des sols pour estimer les stocks de carbone à deux dates différentes et appréhender ensuite le service en question grâce à la variation annuelle du stock de carbone dans les sols. La spatialisation de cet indicateur à l'échelle d'un paysage agricole a permis d'identifier les grandes tendances d'évolution en fonction des caractéristiques pédo-topographiques du paysage agricole. Toutefois, ces auteurs ont souligné l'importance de prendre en compte les incertitudes associées aux mesures réalisées sur le terrain dans l'interprétation de cet indicateur biophysique.

1.6. Le sol est un capital naturel à l'origine de plusieurs services écosystémiques

1.6.1. Le sol est le centre d'intérêt de la recherche scientifique actuelle

Afin de comprendre le rôle multifonctionnel des sols au sein des écosystèmes naturels, plusieurs concepts et notions ont émergé dans le champ scientifique (Hartemink, 2015) tels que la qualité et la santé des sols (Doran, 2002 ; Karlen et *al.*, 2003). Grâce à leur diversité et leur rôle déterminant dans le maintien de l'équilibre écologique des écosystèmes, les sols n'ont pas seulement été désignés comme un objet d'étude scientifique ou un champ de recherche, mais également comme le centre d'intérêt d'un certain nombre de mobilisations collectives d'une dimension internationale.

Au niveau européen, le commissaire européen à l'environnement a lancé une initiative pour définir un cadre de protection des sols afin de les intégrer dans la législation des états membres (CE, 2006). Il a souligné que les sols constituent un levier important pour répondre aux grands enjeux de la société par la mobilisation accrue de ses sept fonctions clés à savoir : (i) production

d'aliment et de biomasse, (ii) stockage, filtrage et transformation des composés organiques et minérales (iii) habitats pour les êtres vivants (iv) environnement culturel (v) source de matières premières (vi) réservoir de carbone et (vii) archives du patrimoine géologique et archéologique.

Au niveau mondial, Mc Bratney et al. (2014) ont récemment proposé un cadre intégratif appelé « soil security », visant à maintenir et à optimiser les fonctionnalités des sols afin de valoriser leurs contributions aux avantages environnementaux et sociaux. Ces chercheurs ont défini la sécurité des sols comme « ... le maintien et l'amélioration des ressources mondiales en sols pour produire des aliments, des fibres et de l'eau douce, contribuer à la durabilité énergétique et climatique, ainsi qu'à la préservation de la biodiversité et à la protection globale de l'écosystème ». Le cadre de la sécurité des sols peut donc être considéré comme un élément lié au sol dans l'approche globale des services écosystémiques définie par le MEA (2005). Les rôles des sols dans la fourniture de services écosystémiques ont été identifiés parmi les objectifs de développement durable des Nations Unies pour 2015-2030 qui vise à « ... protéger, restaurer et promouvoir l'utilisation durable des écosystèmes terrestres... » (Nations Unies, 2015). Néanmoins, il reste difficile de passer de ces cadres théoriques généraux à des approches opérationnelles.

Par ailleurs, le concept de la qualité des sols introduit vers la fin du 20^{ème} siècle, a été largement utilisé dans la littérature scientifique pour combler les faiblesses de l'ancien concept de « fertilité des sols », majoritairement réservé à la fonction de production agricole. La qualité des sols est initialement définie comme « la capacité d'un sol à fonctionner en maintenant la productivité biologique, la qualité de l'environnement et la santé des plantes et des animaux » (Doran et Parkin, 1994). D'autres définitions l'ont enrichi au cours des temps comme l'intégration de la notion de durabilité (Doran et Zeiss, 2000) et l'importance de considérer la qualité du sol dans un contexte donné (Doran et Safley, 1997) et une utilisation donnée (Martin et al., 1999). Actuellement, la qualité des sols a été adoptée pour analyser le rôle multifonctionnel des sols dans la fourniture des services écosystémiques (Greiner et al., 2017 ; Bünemann et al., 2018). En outre, lutter contre les phénomènes d'érosion, de compaction, de perte de la matière organique et de baisse de la biodiversité à l'origine de la dégradation des sols nécessite une caractérisation globale des propriétés du sol permettant d'appréhender cette qualité. Toutefois, ce concept émergent n'est actuellement pas stabilisé et standardisé dans un contexte de divergence des définitions, des approches d'évaluation et de cartographie des indicateurs associés (Bünemann et al., 2018). Ceci est tout aussi justifiable par la diversité des objectifs visés et des contextes étudiés mais aussi par l'évolution des approches d'évaluation des indicateurs de qualité des sols, initialement basée sur

l'expertise scientifique, et qui cherche actuellement à impliquer les décideurs et les bénéficiaires. En parallèle, l'aptitude des indicateurs de qualité des sols ne cessent d'évoluer au cours des temps allant de la mobilisation d'un nombre minimum de variables « minimum data sets » vers de nouveaux indicateurs multicritères mobilisant les nouvelles technologies telles que la spectroscopie (Soltani et *al.*, 2018) (Figure 1.4). Néanmoins, un consensus est implémenté pour fixer une liste non exhaustive mais insubstituable des caractéristiques du sol permettant d'appréhender la qualité des sols, à savoir la teneur en matière organique, la texture et la masse volumique (Garrigues et *al.* 2012).

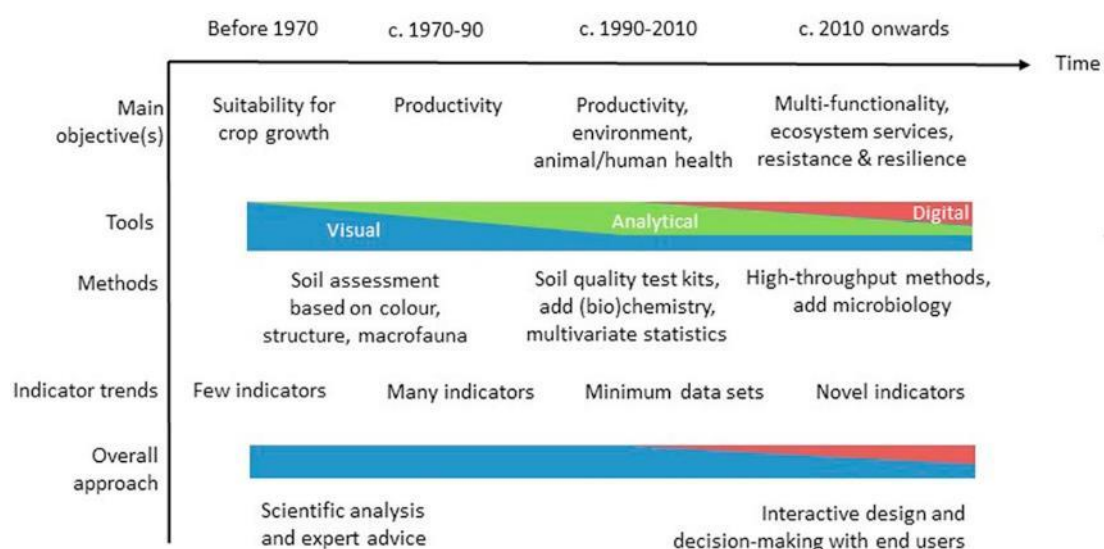


Figure 1.4 : Principaux objectifs, outils et approches de l'évaluation de la qualité des sols à travers l'histoire (De Bünemann et al., 2018).

1.6.2. La controverse entre SES et contribution des sols à la fourniture des services écosystémiques

Le MEA (2005) a clairement démontré dans son rapport, le lien existant entre les écosystèmes et le bien-être humain, en mobilisant le concept de services écosystémiques. Dans le cadre du MEA, la contribution des sols aux services écosystémiques se limite aux services de support avec une attention particulière portée à la fertilité des sols comme un facteur déterminant de nombreux services d'approvisionnement (MEA, 2005).

Comme de nombreuses études ont transposé le cadre conceptuel développé dans le cadre du MEA aux différents cas d'écosystème, la prise en compte du sol dans l'évaluation des services écosystémiques reste assez marginale et simpliste. Elle repose essentiellement sur quelques composants biotiques et néglige l'ensemble des processus pédogénétiques à l'origine du

développement des sols. L'évaluation des services écosystémiques liés aux sols révèle plusieurs difficultés dont certaines sont associées à la controverse entre les concepts de « services écosystémiques des sols » et « la contribution des sols à la fourniture des services écosystémiques » avec l'équivalence des concepts pour certains scientifiques comme De Groot et al. (2018) ou Díaz et al. (2018).

La différence entre les deux concepts précédemment cités dépend notamment de la définition de l'écosystème et des différentes composantes qui le constitue. Chaque concept est associé à un contexte spécifique, à une échelle donnée et à des méthodes d'évaluation biophysique bien déterminées. En effet, le concept de « services écosystémiques des sols » est privilégié à une échelle locale considérant le sol comme un écosystème à part entière caractérisé par de nombreux processus et fonctions écologiques. Les services écosystémiques sont donc le résultat direct des processus biophysiques du capital naturel sans considérer ses interactions avec le milieu naturel et les facteurs endogènes. Cette conceptualisation est à l'origine de plusieurs ambiguïtés notamment entre les processus et les services écosystémiques dérivés.

Au contraire, le concept de « la contribution des sols à la fourniture des services écosystémiques » vise à faire reconnaître le rôle multifonctionnel des sols dans des contextes assez variés et à des échelles spatiales assez larges. Il admet que le sol est le composant central de l'écosystème naturel, qui est en interaction permanente avec le milieu naturel ce qui rend ses fonctions écologiques largement influencées par les conditions biotiques et abiotiques du milieu (Bouma, 2014). Cette conceptualisation est largement soutenue par la communauté scientifique et les gestionnaires des écosystèmes naturels pour analyser d'une façon intégrative l'effet des pratiques de gestion agro-environnementale sur la durabilité des écosystèmes.

1.6.3. Contribution des sols à la fourniture des services écosystémiques

De nombreux chercheurs positionnent le sol comme un capital naturel caractérisé par ses propres stocks de matière, d'information et de biodiversité, mais aussi par les flux qu'ils génèrent, permettant la fourniture d'un ensemble de services écosystémiques (Dominati et al., 2010, Figure 1.5). Dans ce cadre, les services écosystémiques des sols sont définis comme les avantages que les êtres humains retirent des sols et qui sont étroitement liés à la capacité de la biodiversité fonctionnelle à remplir conjointement ses fonctions écologiques (Dominati et al., 2010). Le concept de « capital naturel » est couramment utilisé par la communauté scientifique des services

écosystémiques, pour évaluer et quantifier la contribution des sols à la fourniture des services écosystémiques (Robinson *et al.*, 2009 ; Hewitt *et al.*, 2015).

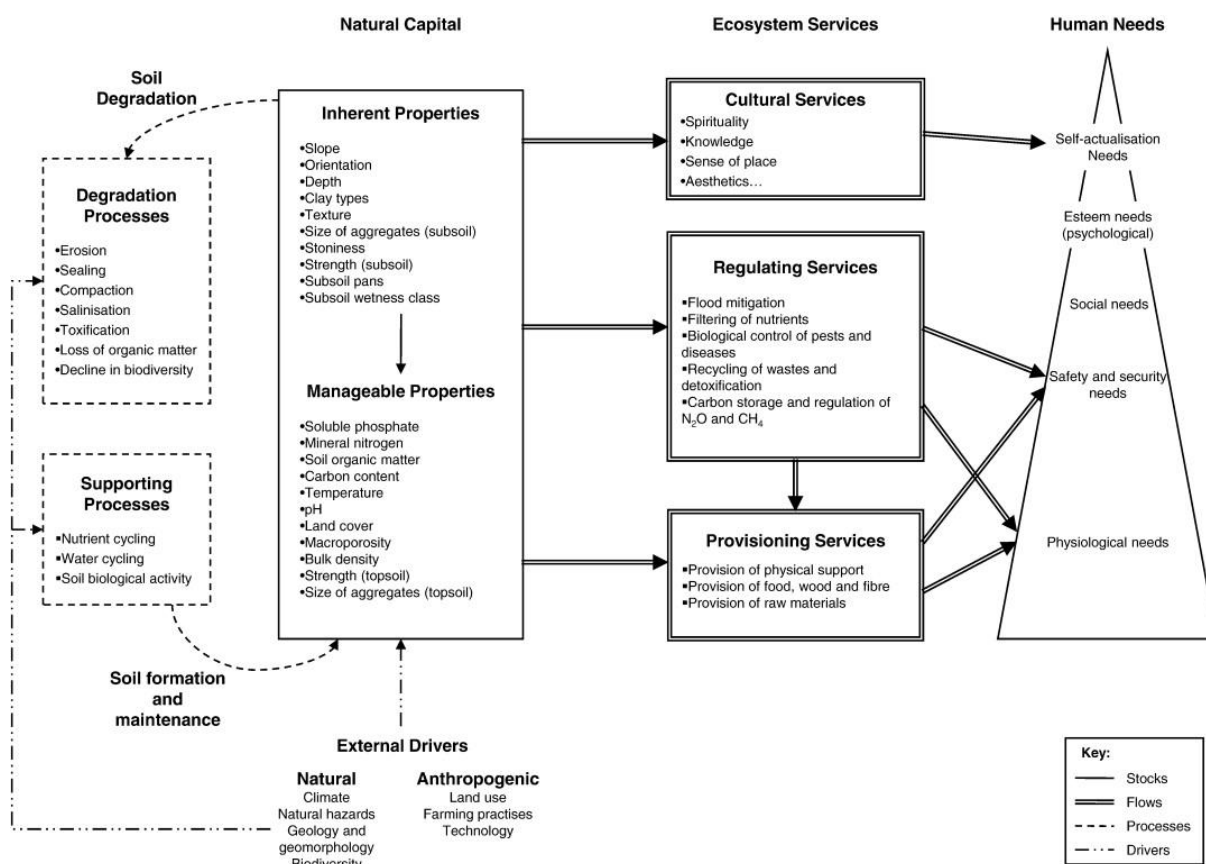


Figure 1.5 : Cadre conceptuel de la fourniture de services écosystémiques spécifique au capital naturel du sol (De Dominati *et al.*, 2010).

Bien que de nombreux cadres conceptuels ont souligné l'importance des sols, ces derniers ont longtemps été peu considérés dans l'évaluation des services écosystémiques avec l'absence d'un consensus permettant d'identifier, classer et valoriser ces services (Robinson *et al.*, 2009). Daily *et al.* (1997) sont les premiers à avoir classer les services écosystémiques des sols en identifiant six services écosystémiques clés : contrôle du cycle hydrologique, support physique des plantes, fourniture d'éléments nutritifs aux plantes, dégradation de la matière organique, maintien de la fertilité du sol et régulation du cycle des éléments majeurs. Cette première tentative de classification a été ultérieurement enrichie par les catégorisations développées par le MEA et la TEEB, où les services ont été classés en fonction du type de service rendu (services de soutien, d'approvisionnement, de régulation et culturel). Des typologies plus intégratives de la composante sol ont été proposées dans des cadres spécifiques aux sols comme par exemple par

Andrews *et al.*, 2004 ; Lavelle *et al.*, 2006 ; Barrios, 2007 ; Palm *et al.*, 2007 ; Haygarth et Ritz, 2009 ; Dominati *et al.*, 2010 ; Faber et van Wensem, 2012. D'autres spécifiques aux agroécosystèmes ont été proposés (Sandhu *et al.*, 2010a ; Sandhu *et al.*, 2010b ; Swinton *et al.*, 2007 ; Zhang *et al.*, 2007). Ainsi, la pluralité des approches est révélatrice de la richesse du concept, apte à décrire une réalité complexe, permettant de prendre en compte l'interdisciplinarité naturelle et sociale du milieu.

Par ailleurs, le cadre conceptuel développé dans le cadre du projet européen RECARE constitue un des prototypes les plus complets de la littérature scientifique, qui est principalement focalisé sur les sols (Schwilch *et al.*, 2016). Ce dernier a été développé afin de rendre opérationnel le concept de services écosystémiques dans l'atténuation des menaces de dégradation des sols en Europe (Figure 1.6). En cohérence avec les catégories du MEA (2005), les sous-catégories de l'approche économique TEEB (2010), le modèle en cascade Haines-Young and Potschin (2010), la définition du capital naturel (Dominati *et al.*, 2014), la typologie CICES (2013) et le projet MAES (Maes *et al.*, 2013), ce cadre conceptuel distingue clairement tous les éléments de base essentiels au développement d'un cadre conceptuel des services écosystémiques allant de l'identification des enjeux associées à la dégradation des sols (érosion, salinisation, compaction, contamination, diminution de la matière organique des sols) jusqu'à la distinction détaillée des bénéficiaires ciblés (les gestionnaires locaux, les économistes, les institutions, les décideurs, et la société dans son ensemble). Cette conceptualisation permet donc de répondre à la fois à deux questions fondamentales : i) comment le capital naturel du sol influence la fourniture de services écosystémiques des sols et ii) comment la pression anthropique et les décisions en matière de gestion agro-environnementale influencent l'équilibre dynamique du capital naturel du sol.

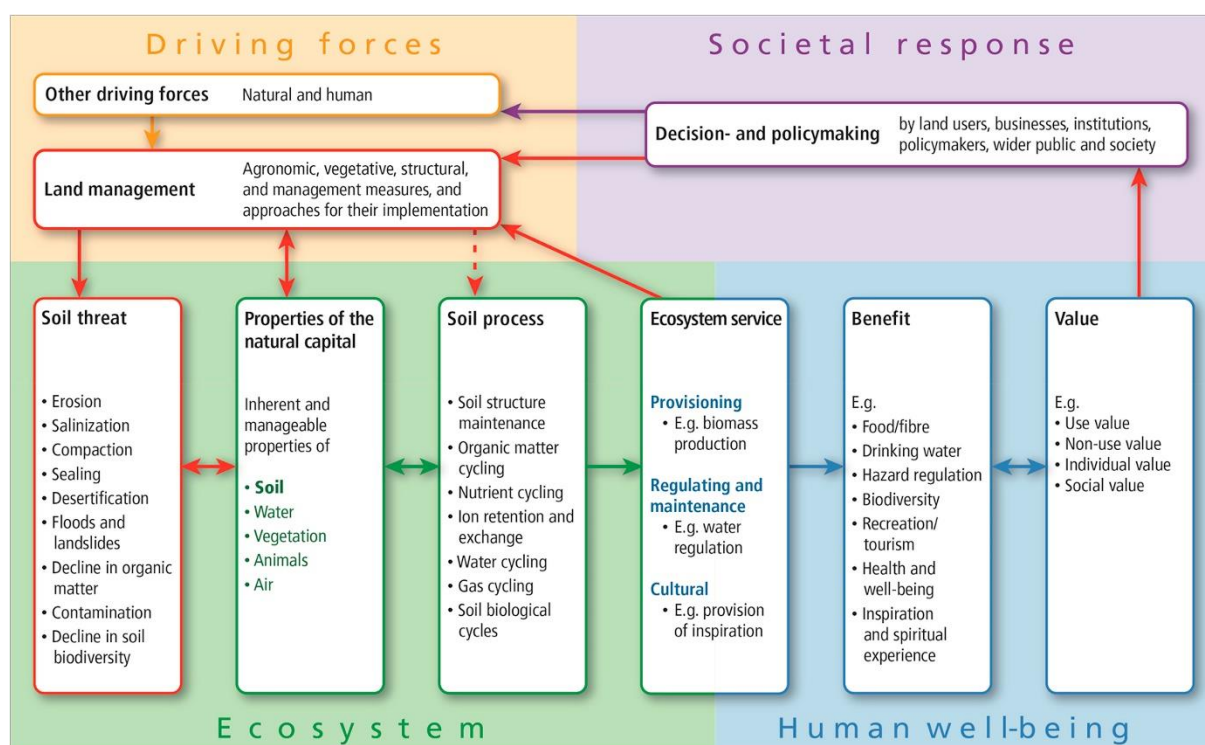


Figure 1.6 : Le modèle conceptuel des services écosystémiques développé dans le cadre du projet européen RECARE qui vise à atténuer les menaces de dégradations des sols européens (De Schilch et al., 2016).

1.6.4. Les services écosystémiques des sols

Les sols représentent un capital naturel complexe et dynamique qui supporte des processus physiques, chimiques et biologiques nécessaires au maintien de la vie planétaire. Ils assurent de nombreuses fonctions diversifiées à l'origine de plusieurs services écosystémiques (Figure 1.7). Ils constituent également un des réservoirs les plus importants de biodiversité (Blum, 2005). Grâce à leur diversité, les sols fournissent de nombreux produits pharmaceutiques et traitements médicaux actuels et futurs (Minton, 2003 ; D'Costa et al., 2006 ; Turbé et al., 2010). Ils agissent sur le cycle de l'eau en absorbant, filtrant, stockant et atténuant les flux d'eau. Ils fournissent des éléments nutritifs, de l'eau mais aussi un environnement physique propice à la production primaire. Cette production constitue le fondement des activités économiques dans divers agroécosystèmes et couvre 90% des besoins nutritionnels de la population humaine (FAOSTAT, 2004). En outre, les sols agissent sur la décomposition de la matière organique endogène et exogène et l'atténuation des contaminants (Andrews et al., 2004 ; Dominati et al., 2010). Ils jouent un rôle primordial dans la régulation du climat, notamment via la séquestration du CO₂ atmosphérique et le stockage de carbone dans leur porosité (Lavelle et al., 2006 ; Haygarth et

Ritz, 2009). En effet, les sols constituent le grand gisement terrestre de CO₂ où ils contiennent environ 2.344 Gt de carbone organique dans leurs trois premiers mètres de profondeur dont 1.500 Gt dans le premier mètre (Jobbágy and Jackson, 2000 ; Guo and Gifford, 2002). Ils sont ainsi le deuxième plus grand réservoir de carbone derrière les océans (38.400 Gt), contenant presque quatre fois plus de carbone que la biosphère (560 Gt) ou l'atmosphère (760 Gt) (Lal, 2004 ; Stockmann et *al.*, 2013). De faibles variations des stocks de carbone organique des sols peuvent donc influencer significativement la concentration en CO₂ de l'atmosphère et, par conséquent, le réchauffement climatique (Kirschbaum, 2000). Par exemple, une libération de 10% du carbone organique des sols serait l'équivalent de 30 ans d'émission de gaz à effet de serre d'origine humaine (Kirschbaum, 2000).

Par ailleurs, les couleurs des sols et ses formations variées sont reconnues depuis longtemps comme la source d'inspiration pour de nombreux artistes et philosophes. Plus encore, les sols ont joué un rôle majeur dans la vision du monde et les croyances religieuses de certaines sociétés. Ils sont aussi le fondement de l'éducation dans diverses disciplines, telles que les sciences agricoles, géologiques, géographiques et biologiques. Ils constituent à la fois une plate-forme d'activités de divertissement (Decaens et *al.*, 2006) et une archive archéologique de l'histoire humaine et de son empreinte écologique. Ainsi, la diversité de services des sols leur confère un rôle crucial dans le bien-être humain ce qui rend la compréhension de ses processus biophysiques et ses fonctions écologiques primordiale pour agir durablement sur l'environnement.

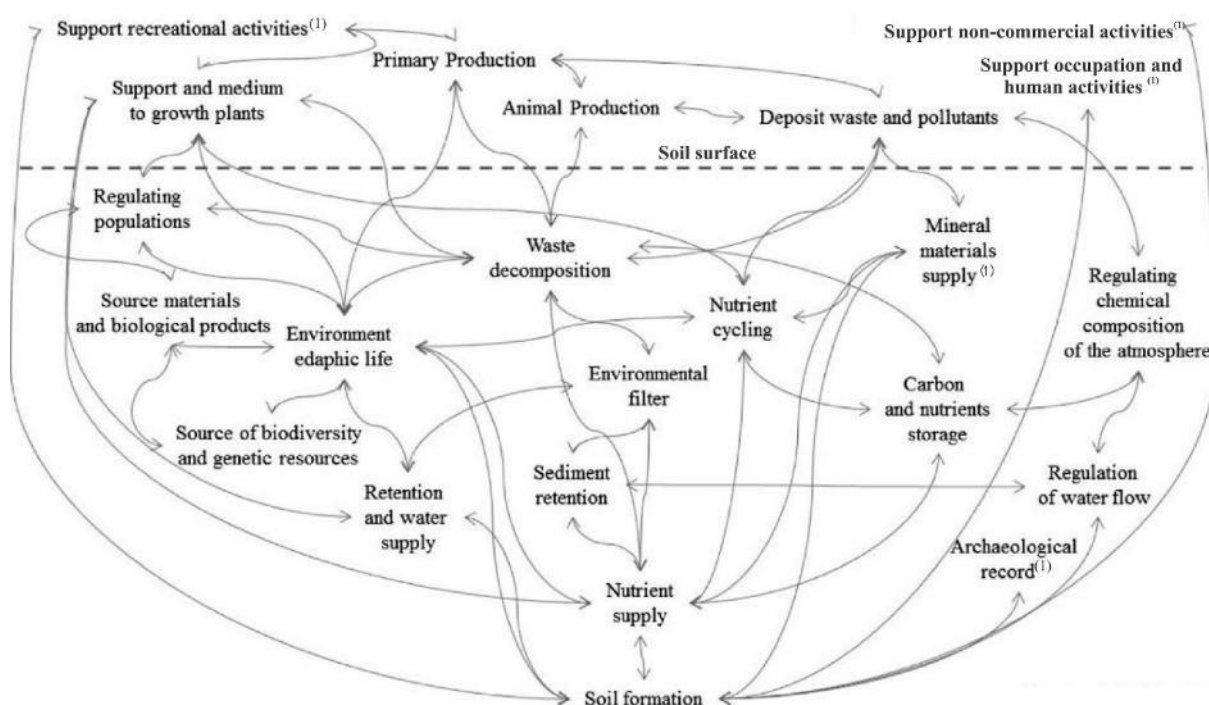


Figure 1.7 : Les fonctions des sols et ses interrelations (De Pardo et al., 2016).

1.7. Le sol est un composant central des écosystèmes anthropisés

1.7.1. Définition d'un écosystème agricole vs agrosystème

Les définitions d'un écosystème agricole sont nombreuses et assez variées. La pluralité de ces définitions s'explique notamment par la diversité de disciplines qui se sont accaparées cette notion, chacune l'abordant selon son propre angle de perception. D'un point de vue écologique et agronomique « un écosystème agricole » désigne l'ensemble « sol-plante-animaux » incluant les êtres vivants (biodiversité, animaux d'élevage), et les aménagements paysagers de la parcelle (haie, bord parcelle). C'est cette définition qui a été privilégiée dans les travaux de l'EFESE-EA pour définir les systèmes agricoles et leurs composants centraux. D'un point de vue géographique, le géographe français Georges Bertrand a proposé le concept d'un « agrosystème » pour rendre compte des interrelations existantes entre l'environnement et l'homme au sein d'un milieu agricole configuré, géré et aménagé par l'homme. Les différents éléments constituant l'agrosystème trouvent leur équilibre grâce aux interventions humaines. « L'agrosystème correspond donc, par définition, à la déstructuration des équilibres naturels et à leur remplacement

par des équilibres secondaires, instables, directement liés au type et au rythme de la mise en valeur » (Beroutchachvili et Bertrand, 1978). L'agriculture remplace, de ce fait, une bonne partie du système moyennant des pratiques agricoles dans le but de produire de la biomasse et d'assurer la souveraineté alimentaire. Cependant, l'agroécologie, qui est une tendance transdisciplinaire, vise à conserver l'équilibre écologique des milieux naturels tout en mobilisant les processus écologiques et réduisant la dépendance vis-à-vis les intrants.

1.7.2. L'intérêt porté au sol dans un contexte agricole

L'engagement du secteur agricole vers des voies permettant de concilier plus efficacement et durablement les enjeux socioéconomiques et environnementaux apparaît aujourd'hui comme une nécessité incontournable. La mobilisation du concept de services écosystémiques aux cas des agrosystèmes permet d'une part d'analyser d'une manière pluridisciplinaire les compromis entre la maximisation des performances économiques et productives des systèmes agricoles et d'autre part l'optimisation des enjeux environnementaux générés. Ainsi, ce concept est de plus en plus largement utilisé pour désigner un nouveau modèle agricole respectueux de l'environnement tout en mobilisant et optimisant les processus inhérents aux sols.

Deux volets essentiels permettent de comprendre les performances agronomiques des écosystèmes agricoles : les composantes associées au sol (propriétés physiques, chimiques et biologiques) et les composantes anthropiques (pratiques agricoles : fertilisation, labour et intrants exogènes). Les sols réagissent différemment, en fonction de la gestion et des intrants exogènes apportés. En effet, les propriétés inhérentes du sol (texture, charge en éléments grossiers et profondeur du sol) sont relativement statiques à l'échelle de temps humaine. A l'inverse, les propriétés dynamiques du sol (pH du sol, teneur en matière organique, teneur en eau et teneur en éléments nutritifs) varient au cours des temps et leur évolution dynamique est en grande partie dépendante des pratiques de gestion agricole.

L'étude multivariée des relations complexes entre les pratiques de gestion agricole et l'évolution des propriétés des sols a fait l'objet de plusieurs études scientifiques (Schulte et *al.*, 2014 ; Schulte et *al.*, 2015) afin d'identifier les voies de maintenance ou d'amélioration potentielles des fonctions de terres agricoles. Les processus du sol tels que la dégradation de la matière organique, les échanges de chaleur et de gaz, la lixiviation des éléments nutritifs et le stockage d'eau sont étroitement liés aux modifications des propriétés du sol et donc à la capacité du sol à remplir ses

fonctions. Ces dernières sont conditionnées par l'évolution des propriétés dynamiques des sols qui réagissent différemment en fonction des pratiques agricoles exercées.

1.7.3. Exemples de cadres conceptuels d'évaluation des services écosystémiques spécifique aux systèmes agricoles

- Cas du programme MAES

Le programme MAES est la déclinaison du MEA à l'échelle européenne, qui résulte d'une obligation législative conduisant tous les états membres à faire un inventaire et une cartographie des services écosystémiques au sein de leur territoire. Il est initié en 2011 pour répondre à la CDB, engager l'enrayement de la perte de biodiversité à l'horizon 2020, en s'appuyant sur une stratégie de conservation de la biodiversité organisée en 6 volets principaux. Plus particulièrement, le 5^{ème} volet vise à améliorer la connaissance des écosystèmes naturels et leurs services écosystémiques. Il appelle donc tous les états membres à cartographier et évaluer les états des écosystèmes et les services qui en découlent sur leur territoire national et à proposer ensuite une évaluation monétaire de ces services pour promouvoir leur intégration dans des politiques nationales et européennes. Un comité d'experts scientifiques supervise et accompagne le développement des cadres d'analyse dans tous les états membres pour assurer la cohérence des évaluations réalisées.

En cohérence avec la typologie CICES, le collectif d'experts du projet MAES a développé une version révisée du modèle en cascade initié par Haines-Young et Potschin (2010) (MAES et al., 2016). Ces scientifiques ont couplé d'une manière harmonisée le modèle en cascade reliant le fonctionnement d'un écosystème au bien-être humain et le modèle DPSIR (Drivers-Pressures-State-Impact-Response) décrivant les interactions entre la société et son environnement et qui est largement utilisé dans la conception et la mise en œuvre de la législation environnementale en Europe (Niemeijer et de Groot, 2008).

Dans un contexte agricole, le cadre conceptuel développé met en relation les systèmes socioéconomiques et l'écosystème agricole par le biais de flux de services écosystémiques et des facteurs exogènes exerçant diverses pressions sur les composantes de l'écosystème (Figure 1.8). Ce dernier est principalement constitué des interactions mutuelles entre les communautés d'êtres vivants et les conditions biotiques et abiotiques du milieu. Ce cadre d'analyse a été aussi transposé aux différents cas de système tels que les écosystèmes terrestres à savoir le milieu urbain, les

prairies, les étangs, mais aussi pour les écosystèmes marins et les écosystèmes d'eau douce tels que les lacs et les rivières.

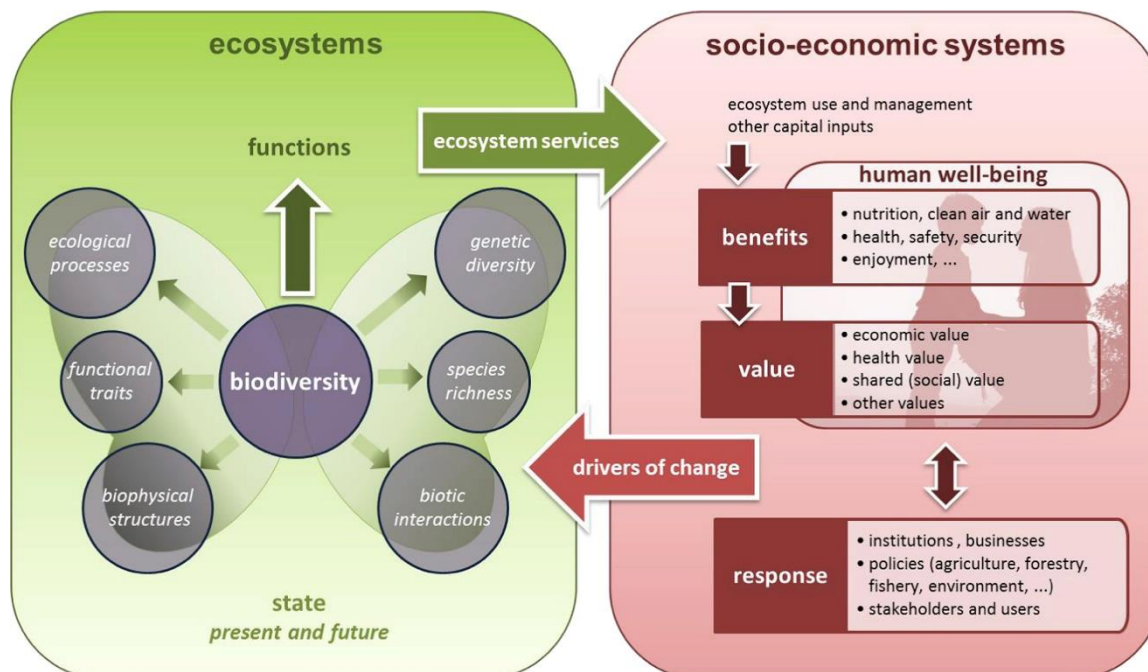


Figure 1.8 : Cadre conceptuel du groupe de travail MAES (2013) qui reconnaît à la biodiversité une place centrale (De Maes et al. (2013).

- Cas du programme EFES-EA

Afin de répondre à l'obligation européenne de dresser l'état des écosystèmes naturels, la France a décliné en 2012 le programme EFES (Evaluation Française des Ecosystèmes et des Services Ecosystémiques) pour proposer une démarche d'évaluation des services écosystémiques fournis par différents types d'écosystèmes sur le territoire national. Cela a eu pour ambition d'éclairer le rôle de la biodiversité dans le maintien de l'équilibre écologique des écosystèmes et d'orienter les projets de planification et de développement territoriale vers une voie respectueuse de la durabilité des écosystèmes. Le volet spécifique aux systèmes agricoles a été attribuée à un comité d'experts de l'INRA afin d'approfondir la connaissance du fonctionnement des écosystèmes agricoles et d'élaborer des politiques prenant mieux en compte la biodiversité et ses services écosystémiques dérivés d'où l'appellation EFES-EA (EA= écosystèmes agricoles). Dans ce contexte et dans la lignée de nombreux travaux européens et internationaux, le collectif d'experts de l'EFES-EA a cherché à évaluer plusieurs services écosystémiques en utilisant la classification CICES comme typologie de référence. Ainsi 14 services écosystémiques diversifiés ont été

instruits et étudiés dans ce cadre, comme par exemple la pollinisation des espèces cultivées, la stabilisation des sols et le contrôle de l'érosion, la structuration des sols, l'atténuation naturelle des pesticides des sols, la régulation du climat global et le potentiel récréatif des agrosystèmes. La considération de la production agricole comme un service d'approvisionnement a été largement discutée et finie par être conceptualisée comme un bien agricole dérivé des interrelations entre les services de régulation et les intrants anthropiques. Le cadre conceptuel développé par le collectif d'experts de l'EFESE-EA repose sur la conceptualisation des services écosystémiques comme composante de l'écosystème dont l'homme dérive des avantages dans le but d'améliorer son bien-être. Cette conceptualisation suppose que l'agrosystème est constitué de deux types de systèmes (écologique et socio-économique) qui sont en interactions permanentes avec les composantes biophysiques du paysage et l'environnement socio-économique (Figure 1.9). En ce qui concerne le système écologique, ce dernier est défini par un ensemble de composantes incluant les entités d'état ou éléments de structure, les déterminantes biophysiques, les processus écologiques et les services écosystémiques. De l'autre côté, le système socio-économique est principalement défini par les gestionnaires de l'écosystème. Les flux entrants sortants à la frontière des deux systèmes écologique et socio-économique désignent les processus de transformation des services écosystémiques en avantages par le biais de facteurs capitaux ainsi que la pression anthropique exercée par les facteurs exogènes sur le système écologique.

Selon la conceptualisation développée dans le cadre de l'EFESE-EA, les services écosystémiques sont donc définis comme « des processus écologiques ou des éléments de la structure de l'écosystème dont l'Homme dérive des avantages, activement en mobilisant du capital matériel (énergie, eau, produits phytosanitaires...) et/ou cognitif (connaissances, par ex. pratiques agricoles) ou passivement (ex. avantage dérivé du SE de régulation du climat) » (EFESE-EA, 2017).

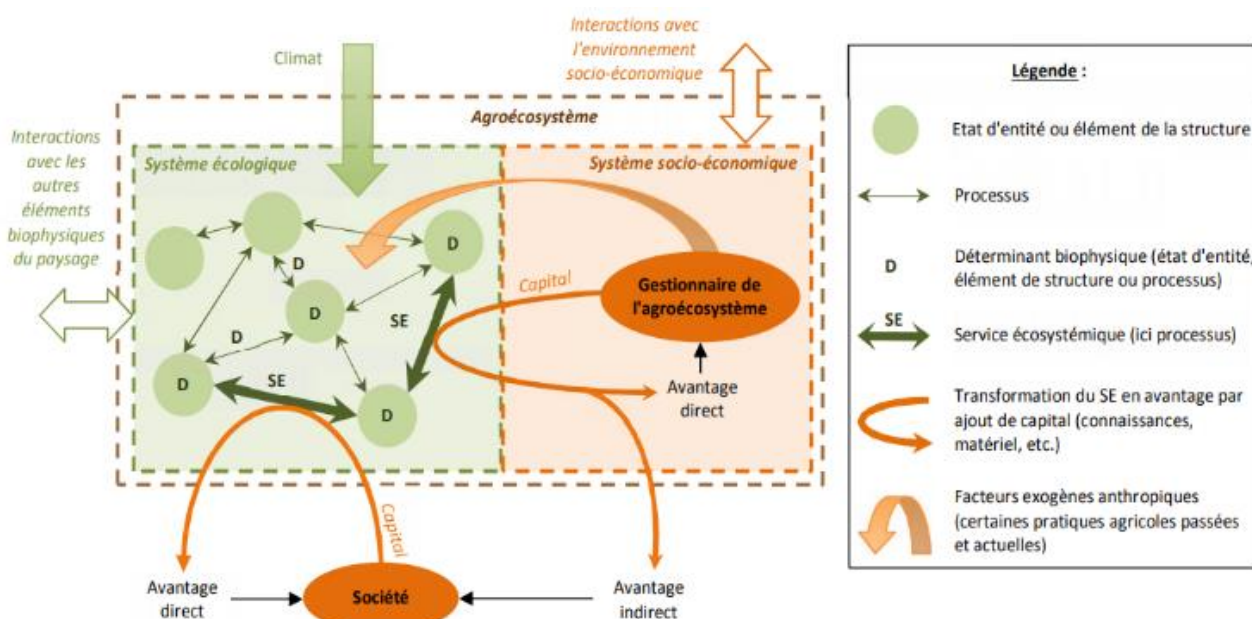


Figure 1.9 : Représentation des concepts clés manipulés dans le cadre de l'EFESE-EA appliqués à la fourniture de deux SE : l'un rendu directement à la société, l'autre rendu directement au gestionnaire de l'agroécosystème et dont la société dérive un avantage indirect. (EFESE-EA, 2017).

Sur le plan fonctionnel, les avantages dérivés spécifiquement des services écosystémiques sont déconnectés de l'écosystème et peuvent être matériels (des biens) ou immatériels (des services socioéconomiques). Ces avantages retirés, qu'ils soient directs ou indirects, sont au service de deux catégories de bénéficiaires à savoir le gestionnaire direct de l'agrosystème (l'agriculteur) et la société dans son ensemble. En effet, l'agriculteur dérive des avantages directs liés à la production agricole. De même, la société dérive des avantages directs associés à la régulation du climat et indirects associés à la conduite d'une agriculture durable exigeant un minimum d'intrants exogènes susceptibles de polluer l'environnement.

En outre, le groupe de travail de l'EFESE-EA a retenu une liste non exhaustive d'indicateurs biophysiques pertinents pour quantifier et spatialiser les services écosystémiques rendus par les systèmes agricoles en considérant leur état actuel à l'échelle de la France entière. Pour l'essentiel, deux types d'approches ont été explorées selon la nature du service écosystémique instruit. Une approche basée sur la modélisation dynamique des interactions entre les trois compartiments sol-plante-climat a été privilégiée pour évaluer les services écosystémiques associés au fonctionnement du système sol-plantes-animaux de la parcelle. Les simulations dynamiques ont été conduites à l'aide de deux modèles développés par l'INRA avec deux domaines d'application différents : le modèle de gestion agro-environnementale STICS (Brisson et al., 1998) pour simuler les systèmes de grandes cultures et le modèle PaSim (Graux et al., 2011) pour simuler les systèmes

prairiaux. La deuxième approche se base essentiellement sur le développement d'indicateurs de configuration et de composition du paysage » pour appréhender les services écosystémiques dépendants des caractéristiques du paysage tels que le service de pollinisation des espèces cultivées.

1.8. Les indicateurs d'évaluation biophysique des services écosystémiques liés aux sols

La définition du « capital naturel sol » est partagée par de nombreux scientifiques qui s'intéressent à l'évaluation des services écosystémiques liés aux sols. Cependant, cette évaluation est confrontée à des problèmes d'harmonisation des indicateurs biophysiques en l'absence d'un référentiel normalisé qui fixe formellement leur caractérisation. Deux approches différentes sont couramment utilisées pour dériver les indicateurs biophysiques. La première consiste à exprimer le service écosystémique par l'état d'entités en s'appuyant sur une mesure directe d'une ou plusieurs propriétés du sol à une date donnée ou en estimant la différence entre une mesure de propriété du sol à une date donnée et une autre mesure, estimée par modélisation, à une date future selon un scénario bien déterminé. Pour illustrer, on s'appuie sur les études de Grimaldi et *al.*, 2014, Lavelle et *al.*, 2014, et Albizua et *al.*, 2015, qui ont évalué plusieurs services écosystémiques liés à la régulation du climat, la régulation du cycle de l'eau, et la fourniture d'un support pour la biodiversité par le biais de nombreuses propriétés du sol telles que la teneur en carbone organique du sol, l'infiltrabilité du sol, le pH et la concentration en azote. D'autres caractéristiques plus intégratives estimées par des règles de pédotransfert, comme par exemple le réservoir utile en eau pour la plante, sont aussi utilisées pour appréhender les services écosystémiques liés au cycle de l'eau et la rétention de l'azote par les interfaces du sol. En outre, Van Wijnjen et *al.* (2012) et Rutgers et *al.* (2012) ont choisi d'évaluer l'atténuation naturelle des polluants par les sols en s'appuyant sur un indicateur qui compare le taux de minéralisation d'azote à deux états différents de l'écosystème (potentiel vs référentiel) par une approche de modélisation.

Par ailleurs, considérer l'état d'entités pour appréhender les services écosystémiques ne semble pas être une approche pertinente pour qualifier le niveau de fourniture des services écosystémiques et tirer des renseignements solides en matière de gestion agro-environnementale.

Ceci est tout aussi justifiable par l'aspect dynamique de certaines propriétés du sol comme le pH, la teneur en carbone organique et la concentration en azote qui sont très variables dans le temps et largement influencées par les conditions abiotiques du milieu. De ce fait, la mobilisation des jeux de mesures à un instant donné dans la caractérisation des services écosystémiques génère des interprétations non reproductibles et surtout non transposables à d'autres écosystèmes.

En revanche, la deuxième approche de caractérisation des indicateurs des services écosystémiques s'appuie sur la capacité de l'écosystème à produire un service donné. Dans la littérature, plusieurs scientifiques ont adapté cette proposition pour fournir des évaluations biophysiques reproductibles et cohérentes avec les pratiques de gestion de l'écosystème. Par exemple, dans les études de Fu et *al.* (2015) et Qin et *al.* (2015), le service du contrôle de l'érosion des sols a été appréhendé par le biais d'un indicateur correspondant à la différence entre l'érosion potentielle et l'érosion actuelle, évalués par le modèle empirique de quantification de l'érosion des sols Revised Universal Soil Loss Equation (RUSLE). Dominati et ses co-auteurs (2014) ont estimé le service associé à la disponibilité de l'azote (et du phosphore) dans le sol par le biais d'un indicateur correspondant à la différence entre la quantité d'azote (ou de phosphore) perdu dans le sol en absence d'un processus de rétention et la quantité d'azote (ou de phosphore) réellement lixivié. Dans sa démarche, L'EFESE-EA a également privilégié cette approche pour évaluer les services associés à la régulation de la qualité de l'eau vis-à-vis de l'azote, de nitrate et du carbone organique dissout.

Par ailleurs, indépendamment de l'approche retenue, une liste incontournable des propriétés du sol est souvent mobilisée pour estimer les services écosystémiques. Adhikari et Hartemink, (2016) rassemblent dans leur revue les propriétés du sol les plus fréquemment citées et recommandées dans la littérature. Les propriétés dynamiques, instables dans le temps, sont souvent priorisées pour évaluer conjointement plusieurs services écosystémiques. Parmi elles, on trouve la teneur en matière organique, le pH, la masse volumique, la teneur en phosphore, la capacité d'échange cationique et la teneur en azote. Les caractéristiques statiques ou inhérentes du sol à savoir la texture, l'épaisseur du sol et la minéralogie de l'argile sont aussi utilisées pour qualifier plusieurs services écosystémiques fournis. Ce constat rejoint celui de Bünemann et *al.* (2018) qui ont rassemblé dans leur revue bibliographique les propriétés du sol les plus citées dans l'évaluation de la qualité des sols.

1.8.1. Les données pédologiques mobilisées dans l'évaluation des services écosystémiques liés aux sols

Les sols contribuent à la fourniture des services écosystémiques mais à des degrés différents selon les territoires et les contextes. Sur le plan fonctionnel, la mobilisation de l'information pédologique dans l'évaluation des services écosystémiques est en grande partie dépendante de la disponibilité et de la qualité des bases de données pédologiques existantes. Par exemple, malgré la position centrale qu'occupe la biodiversité du sol dans le cadre conceptuel proposé par le programme MAES, cette dernière reste très peu représentée par les indicateurs biophysiques retenus. En effet, dans le contexte des systèmes agricoles, le groupe de scientifiques MAES a développé une grille d'indicateurs biophysiques basée sur des données facilement accessibles et disponibles dans les bases de données européennes. L'évaluation du service d'approvisionnement, principalement fondé sur la production agricole, est caractérisée à partir des données collectées dans le cadre du réseau de surveillance de la Politique Agricole Commune (PAC) telles que les statistiques de rendements régionales. Les services de régulation sont appréhendés via des indicateurs associés au contrôle de l'érosion et à la rétention de l'azote sans tenir compte des caractéristiques du sol. Au niveau spatial, la mosaïque paysagère diversifiée permettant de diminuer les risques parasites a permis d'évaluer le service de pollinisation indispensable au maintien de la production agricole. D'un autre côté, les structures paysagères des terres agricoles, caractérisées par des aménagements de gestion agro-environnementale spécifique au contexte européen, ainsi que leur histoire millénaire sont à l'origine des services écosystémiques culturels. Par ailleurs, plusieurs auteurs (Daily et al., 1997 ; Robinson et al., 2009 ; Dominati et al., 2014) pointent l'importance de l'intégration de l'information pédologique dans l'amélioration de la fiabilité des évaluations des services écosystémiques rendus. Les études réalisées ont clairement souligné le rôle déterminant des sols dans la génération des flux de services écosystémiques, tels que le contrôle de l'érosion des sols, la rétention et le stockage de l'eau, l'atténuation des polluants et la régulation du climat. Par conséquent, l'ajout de données pédologiques dans les processus d'évaluation biophysique des services écosystémiques permet d'améliorer significativement la capacité d'appréhension des services rendus tout en intégrant une large variabilité de contextes pédologiques.

Différentes sources de données pédologiques sont mobilisées pour caractériser les services écosystémiques fournis. Ces données diffèrent par leurs méthodes d'échantillonnage et d'analyse physico-chimique mais aussi par leur résolution spatiale et leur précision. Par exemple, Van

Wijnen *et al.* (2012), ont utilisé des cartes des sols nationales au 1/50 000 pour caractériser la distribution spatiale de la teneur en carbone organique et du pH par une simple rasterisation de la carte originale à une résolution spatiale de 100 x 100 m. D'autres propriétés biologiques et chimiques de 283 profils de sol ont été utilisées pour dériver des propriétés plus complexes en utilisant des méthodes de régressions linéaires généralisées. L'ensemble des cartes pédologiques générées a été ensuite utilisé pour alimenter un modèle d'évaluation de l'atténuation naturelle des polluants par les sols et cartographier l'indicateur biophysique sur le territoire national à une résolution spatiale de 100 x 100 m. De même, Calzolari *et al.* (2016) ont mobilisé des données pédologiques répertoriées dans des bases de données anciennes pour dériver les propriétés du sol nécessaires à l'évaluation de plusieurs services écosystémiques. Des analyses physico-chimiques d'environ 3300 profils de sol et une carte des sols régionale au 1/50 000 ont été utilisées pour cartographier les propriétés des sols, moyennant des méthodes géostatistiques, à une résolution spatiale de 1 km x 1 km. Les cartes de la teneur en carbone organique et les 3 fractions granulométriques de la texture ont été validées dans des études antérieures, pour quantifier leur qualité prédictive à l'échelle régionale. Les indicateurs des services écosystémiques instruits sont estimés à l'aide d'équations empiriques à une résolution spatiale de 1 km x 1 km, pour être ensuite agrégées au niveau des unités de pédo-paysages. Au niveau national, l'EFESE-EA a mobilisé des informations pédologiques grossières (cartes des sols au 1/1 000 000^{ème} couvrant la France entière) pour dériver les informations nécessaires au paramétrage du modèle STICS. Les indicateurs des services écosystémiques sont ensuite exprimés en fonction d'un large panel de données de sortie du modèle et évalués à l'échelle des entités pédoclimatiques ayant des caractéristiques climatiques et pédogénétiques similaires. Au regard de la nature des données pédologiques mobilisées dans toutes les études précédemment citées, la résolution spatiale et la qualité de l'information du sol sont identifiées comme un verrou majeur de l'évaluation des services écosystémiques. Pour cette raison, un autre groupe de chercheurs s'est appuyé sur des données ponctuelles en collectant des échantillons de sol, selon un protocole d'échantillonnage prédéfini pour analyser ensuite les propriétés du sol nécessaires à l'évaluation des services écosystémiques. Par exemple, Grimaldi *et al.* (2014) ont échantillonné 267 profils de sol répartis tout au long d'un transect permettant de couvrir la variabilité pédo-paysagère présente et selon 3 profondeurs de prélèvements en contexte de front pionnier amazonien. L'ensemble des caractéristiques des sols (teneur en carbone organique, densité apparente, propriétés physico-chimiques et infiltrabilité) a été utilisé pour appréhender les services écosystèmes instruits. De

même, Hewitt et ses co-auteurs (2015) ont développé une méthode appelée « a stock adequacy method » pour évaluer le capital naturel du sol tout en mobilisant des informations pédologiques collectées dans le cadre d'études pédologiques. Dominati *et al.* (2014) ont mobilisé des données ponctuelles pour paramétrer un modèle mécaniste appelé Soil Plant Atmosphere System Model (SPASMO), simulant l'évolution dynamique des propriétés du sol qui régissent la fourniture des services écosystémiques. Cette modélisation dynamique des interactions sol-plantes permet d'explorer l'influence des pratiques de gestion agro-environnementale sur l'évolution des propriétés du sol et d'aider à la quantification biophysique des services écosystémiques fournis. D'autres scientifiques comme Rutgers *et al.* (2012), Williams et Hedlund (2013), Dominati *et al.* (2014), Lavelle *et al.* (2014), et Albizua *et al.* (2015) ont échantillonné plusieurs profils de sol dans des contextes très différents pour appréhender de nombreux services écosystémiques à partir d'un jeu de mesures des propriétés du sol.

Par ailleurs, les études réalisées par Guo *et al.* (2000) et Nelson *et al.* (2009) ont confirmé l'importance d'injecter les propriétés du sol dans le paramétrage des modèles de contrôle de l'érosion et des flux d'eaux. Plusieurs attributs du sol tels que le matériau parental sont communément retenus comme des variables d'entrée pertinentes pour l'évaluation et la modélisation des services (Reyers *et al.*, 2009 ; Dymond *et al.*, 2012). Comme le sol est un capital naturel tridimensionnel qui varie dans le temps et dans l'espace, tenir compte de cette variabilité pédo-dynamique dans la modélisation et la cartographie des services écosystémiques est nécessaire pour améliorer la fiabilité des évaluations biophysiques et affiner les supports de décision délivrés (Schägnler *et al.*, 2013). Toutefois, les concepteurs et les utilisateurs des modèles doivent être vigilants par rapport à la précision des données pédologiques mobilisées. Cela permet non seulement d'approfondir la compréhension des interactions entre les attributs du sol et le reste de l'écosystème mais aussi d'améliorer la qualité des données de sortie du modèle qui peuvent être mobilisées dans l'évaluation des services écosystémiques (Andrew *et al.*, 2015).

1.8.2. La cartographie des services écosystémiques

La cartographie des services écosystémiques est désormais une discipline émergente pour fournir des supports de communication facilement appréhendables par les décideurs et les gestionnaires des territoires. Elle propose des avancées en matière de planification territoriale permettant d'appréhender la variabilité spatiale des services écosystémiques en fonction des variables environnementales mais aussi leur tendance d'évolution en fonction des scénarios construits. La

cartographie des services écosystémiques est une question largement évoquée dans la littérature à différentes échelles spatiales allant d'une exploitation agricole de quelques hectares (Grimaldi et *al.*, 2014) à des échelles beaucoup plus larges telles que l'échelle nationale (Nelson et *al.*, 2009) et européenne (Larondelle et *al.*, 2013). Cependant, il existe très peu d'études qui ont concrètement produit des cartes caractérisant la répartition spatiale des services écosystémiques avec un certain déséquilibre entre services (Adhikari et Hartemink, 2016). En effet, l'étude bibliographique de Adhikari et Hartemink, (2016) met en évidence la surexposition de certains services écosystémiques, notamment les services de régulation, aux dépens d'autres. Le service de régulation du climat via le stockage du carbone est le plus souvent spatialisé. Sa prédominance est notamment expliquée par l'importance de ce service vis-à-vis des enjeux actuels mais aussi par la maîtrise d'un certain nombre de processus qui le conditionnent. Cependant, malgré sa simplicité apparente, la cartographie du service de régulation du climat révèle des ambiguïtés associées au choix d'indicateur à spatialiser qu'il soit le stock de carbone en un état statique valable à une date donnée ou le processus de stockage dynamique via la variation du stock de carbone par rapport à une référence.

Concrètement, il existe deux grands types d'approches de spatialisation des services écosystémiques. La première repose sur la modélisation statistique moyennant des méthodes d'apprentissage. Elle consiste à mettre en place une méthode de régression en exprimant la variable d'intérêt (indicateur de services) en fonction d'un ensemble de variables explicatives (données climatiques, données sur les sols et données sur la trajectoire historique de l'occupation des sols et les pratiques agricoles). Cette modélisation statistique présente deux atouts principaux. D'une part, elle permet d'expliquer les variations de chaque indicateur de services écosystémiques à partir d'une combinaison de données disponibles d'une façon exhaustive sur l'ensemble des territoires. Construire une régression consiste en effet à sélectionner selon un critère spécifique un certain nombre de variables, permettant ainsi d'identifier les facteurs déterminants des variations des indicateurs de services. D'autre part, en estimant des valeurs sur l'ensemble de la localité, la régression permet d'obtenir une cartographie pour l'ensemble d'un site d'étude à partir de quelques points d'échantillonnage. Pour modéliser une variable quantitative à partir de variables qualitatives et quantitatives, les méthodes classiques les plus répandues sont les arbres de régression (Algorithme Classification And Regression Trees CART (Breiman et *al.*, 1984)) et la régression linéaire multiple.

D'un autre côté, la deuxième approche de spatialisation des services écosystémiques consiste à mobiliser des outils cartographiques spécialisés avec des modes de fonctionnement assez différents tels que InVEST (extension SIG, Sharp et *al.*, 2014), Solves (application web, Sherrouse et Semmens, 2010) ou ARIES (application web, Villa et *al.*, 2009). Ces outils reposant essentiellement sur des modèles mécanistes se développent et se diffusent afin de créer des interfaces faciles d'utilisation pour que les différents acteurs du territoire puissent posséder une représentation de la distribution spatiale des services écosystémiques. Par exemple, l'outil InVEST permet de modéliser une série de services écosystémiques pour analyser les enjeux généraux et le changement associés aux évolutions d'occupation du sol sur la fourniture des services écosystémiques. Il a été utilisé, aux États-Unis, pour évaluer et cartographier, à une résolution spatiale de 30 x 30 m, 3 services écosystémiques (la régulation de l'érosion, la régulation du climat via la séquestration du carbone et la régulation de la qualité de l'eau) selon 3 scénarios de politiques publiques et de réchauffement climatique (Nelson et *al.*, 2009).

1.8.3. Les incertitudes associées à l'évaluation et la cartographie des services écosystémiques des sols

Une autre question liée aux incertitudes des services écosystémiques est rarement abordée dans la littérature scientifique. Dans un contexte d'opérationnalisation des cartes de la distribution spatiale des services écosystémiques, la qualification de ces incertitudes est intéressante de fait qu'elle permet de limiter les conséquences non désirées des décisions politiques prises à partir des cartes fournies. La cartographie des services écosystémiques est une schématisation générale de tous les processus sous-jacents aux services écosystémiques. Elle accumule donc les incertitudes associées aux données d'entrée mais aussi celles associées aux processus d'évaluation biophysique telles que la modélisation et les fonctions de pédotransfert permettant d'alimenter les modèles mécanistes et écologiques. Toutefois, il n'est pas possible d'obtenir une estimation de cette incertitude pour toutes les méthodes de modélisation et de cartographie des services écosystémiques. Ceci s'explique notamment par la complexité des processus mais aussi par le fait que l'incertitude n'est pas quantifiée systématiquement pour toutes les données mobilisées.

Conclusion du chapitre 1

Ce chapitre a présenté une synthèse bibliographique portant sur l'évaluation des services écosystémiques des sols et les concepts qui en découlent, en mettant l'accent sur l'information pédologique nécessaire à cette évaluation. Le concept de service écosystémique est un concept émergent dans le champ scientifique et est en voie d'adoption croissante par de nombreuses institutions dans des projets de développement et de planification territoriale. Cependant, ce concept n'est à ce jour pas stabilisé, car il persiste dans la littérature des divergences de cadre conceptuel, de définitions et des méthodes d'évaluation biophysique associées. De nombreux travaux, à l'instar du programme européen MAES et de son volet EFESE en France, ont néanmoins cherché à évaluer et à cartographier les services écosystémiques rendus par différents écosystèmes, pour en faire l'inventaire et renforcer leur prise en compte dans la décision publique. Malgré sa position centrale au sein de l'environnement, le sol reste une composante globalement peu prise en compte dans les méthodes d'évaluation biophysique des services écosystémiques. Ainsi, les informations pédologiques mobilisées sont souvent basiques et généralement approchées par des données d'occupation et d'usage des sols.

Conscient de l'importance des sols, le groupe de travail de l'EFESE-EA (2017) a privilégié l'intégration des propriétés intrinsèques et extrinsèques des sols dans l'évaluation des services écosystémiques des terres agricoles. Pour cela, le collectif d'experts a développé un cadre conceptuel positionnant les sols et les services écosystémiques dérivés comme des composantes essentielles de l'écosystème agricole. C'est ce cadre conceptuel qui a été privilégié dans les travaux de la thèse pour évaluer les services écosystémiques des sols agricoles dans un contexte breton. En outre, une phase de mise à disposition de l'information sur les sols a été adoptée pour extraire l'information nécessaire à l'évaluation des services écosystémiques. Cette évaluation biophysique repose essentiellement sur la modélisation dynamique des processus des sols, pour laquelle les données pédologiques en entrée sont obtenues par des méthodes de désagrégation spatiale des bases de données pédologiques existantes. Cette mise à disposition de l'information pédologique, qui alimente le modèle d'évaluation des services écosystémiques, ainsi que l'évaluation de la qualité prédictive de cette information, fait l'objet des chapitres 3, 4 et 5.

L'aspect lié aux interactions entre les services écosystémiques des sols n'a pas été abordé dans l'état de l'art général de la thèse. Néanmoins, les types d'interactions ainsi que les méthodes de caractérisation associées sont explicitées dans l'introduction du chapitre 5.

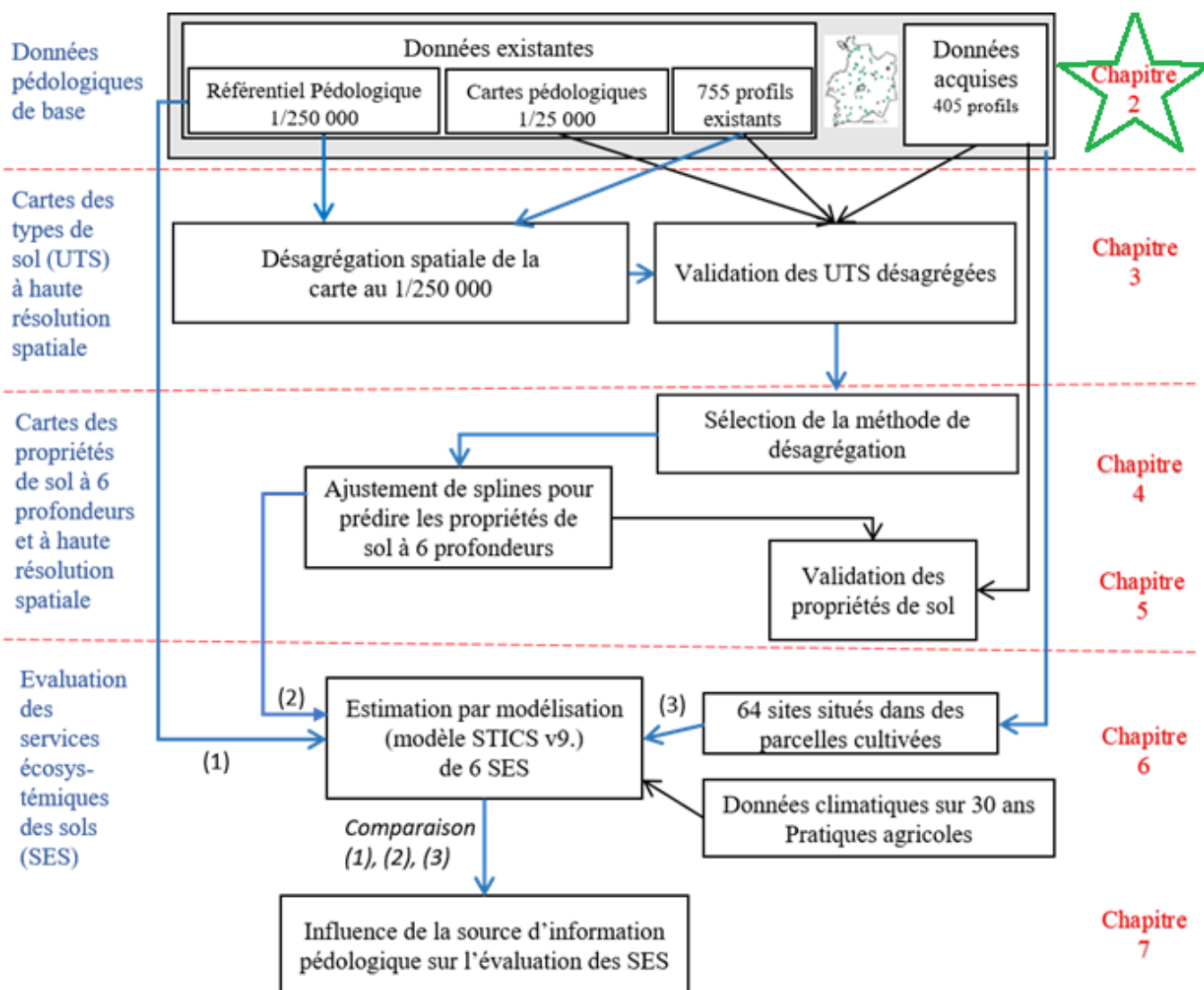


Figure 2.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 2 : Matériels et Méthodes

2.1. Présentation du site d'étude

La zone d'étude correspond à la région Bretagne, qui est composée de quatre départements administratifs (les Côtes-d'Armor, le Finistère, l'Ille-et-Vilaine et le Morbihan) et s'étend sur une superficie totale de 27 040 km², ce qui représente 5 % du territoire national. Dans le cadre de la thèse, une attention particulière a été portée sur le département d'Ille-et-Vilaine, qui représente 25% de l'ensemble du territoire breton. Les quatre départements de la région Bretagne couvrent une variété importante de substrats géologiques et une variabilité pédopaysagère marquée par un réseau bocager plus ou moins dense. La complexité géologique génère un relief assez contrasté avec une grande diversité d'aménagement du paysage. Le point culminant de la Bretagne s'élève à 387 m d'altitude, avec plusieurs autres sommets des monts d'Arrée qui atteignent des altitudes similaires variant entre 380 et 384 m d'altitude. Le reste de la péninsule bretonne présente des altitudes globalement peu élevées qui ne dépassent pas les 300 m d'altitudes. La topographie, elle aussi est étroitement liée aux formations géologiques marquée par des pentes contrastées avec localement des pentes fortes et des cours d'eau encaissés. Quant au contexte hydrologique, la Bretagne présente peu de grands cours d'eau avec de nombreuses têtes de bassin versants. Plus spécifiquement, le nord de la Bretagne compte de nombreux fleuves côtiers souvent courts. Toute cette diversité géologique, topographique et hydrologique sont à l'origine d'une grande diversité des sols, qui se distinguent par leur organisation spatiale le long des versants, leurs épaisseurs et leurs stocks de carbone organique.

2.1.1. Contexte climatique

La Bretagne présente un climat océanique tempéré, caractérisé par de faibles amplitudes thermiques, avec des étés plutôt frais et des hivers souvent pluvieux, ventés et doux. Le climat breton est caractérisé par des humidités importantes et des précipitations assez fréquentes mais peu abondantes. Il n'est cependant pas homogène et présente de fortes variabilités, notamment entre les zones côtières et l'intérieur des terres. Sur l'année, la température moyenne en Ille-et-Vilaine est de 11°C, avec des précipitations annuelles moyennes de 669 mm (Figure 2.1) (Base de données CLIMATIK de l'INRA, 2019).

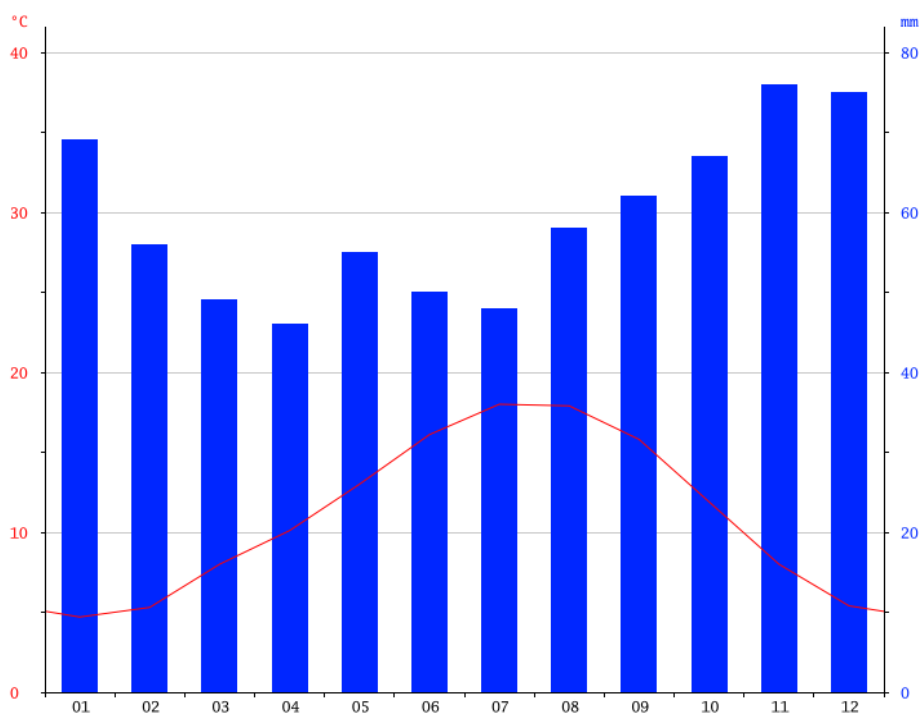


Figure 2.1 : Diagramme ombrothermique (Base de données CLIMATIK de l'INRA, station de Saint Jacques de la Lande située à 48 m d'altitude)

2.1.2. Contexte géologique

De point de vue géologique, la Bretagne s'étend sur un massif ancien, le Massif Armoricaïn (Figure 2.2). La structure du Massif armoricaïn est d'origine polyphasée et marquée par de nombreux évènements thermo-tectoniques caractérisant de longues périodes sédimentaires, entrecoupées d'épisodes tectoniques, métamorphiques, plutoniques et volcaniques. Le Massif armoricaïn a été soumis à deux orogénèses principales qui marquent la géologie bretonne actuelle et permettent de distinguer trois grands domaines géologiques (domaines nord, centre et sud-armoricaïns) séparés par deux zones de cisaillement majeurs (le Cisaillement Nord Armoricaïn et le cisaillement Sud Armoricaïn) (Figure 2.2, BRGM, 2009). L'orogénèse cadomienne datant de la fin du Néoprotérozoïque jusqu'au début du Cambrien, a donné lieu à la chaîne cadomienne qui couvre la quasi-totalité de la partie nord de la Bretagne. Cette chaîne constitue un segment de la chaîne panafricaine européenne et affleure en Bretagne sous forme de granitoïdes, de diorites et gabbros et des migmatites du dôme. La deuxième orogénèse est l'orogénèse varisque ou hercynienne qui représente un cycle orogénique paléozoïque. Elle a débuté au Dévonien et s'est achevée avec le Permien pour donner lieu à la chaîne hercynienne qui affleure notamment dans la

partie sud de la Bretagne. Les roches y sont fortement déformées et métamorphisées (granites d'anatexie, gneiss, micaschistes, schistes bleus), A la fin de cette période d'orogénèse, les grands cisaillements sud et nord-armoricains se forment (BRGM, 2009).

En Bretagne centrale, les roches d'origine sédimentaires, dont principalement les roches schisteuses, dominent. On y trouve également deux grands types de roches du socle : les granites, sous forme de massifs intrusifs, et les grès. De plus, les phénomènes d'émergence de la péninsule bretonne mais aussi sa forte altération due à un climat chaud et humide, sont à l'origine d'une altération massive des roches. Le massif armoricain a également subi de nombreuses transgressions marines qui donnent lieu à des dépôts sédimentaires essentiellement calcaires (BRGM, 2009).

Par ailleurs, le territoire du département d'Ille-et-Vilaine se situe dans la partie orientale du Massif armoricain et plus particulièrement dans les domaines Nord Armoricain et Centre Armoricain. Il est constitué de plusieurs entités géologiques de diverses origines morphologiques et tectoniques (BRGM, 2009). Les roches sédimentaires briovériennes dont principalement les roches silto-gréseuses couvrent la majeure partie de département. La partie nord présente des formations de granite fréquentes et localement regroupées (BRGM, 2009). La partie centrale et méridionale est couverte de dépôts briovériens mais aussi des dépôts sédimentaires faiblement métamorphisés composés de grès, quartzites et schistes. Dans l'ensemble, le département d'Ille-et-Vilaine est principalement constitué de roches granitiques, de schistes, de grès et de quartzites. Ces formations de socle sont localement recouvertes par des formations superficielles constituées de limons d'origine éolienne et d'alluvions fluviales anciennes, notamment dans la vallée de la Vilaine au sud du département, ou récentes (Quaternaire). La cartographie prédictive du matériau parental (Lacoste et al, 2011, Figure 2.3) montre clairement l'abondance des alluvions fluviales récentes et la prédominance des dépôts de limons éoliens dans centre-ouest et le nord-ouest du département. Ces limons, jaunes à brun-jaunes, sont non carbonatés, lités (limons à doublets) et contiennent majoritairement du quartz, des argiles (illites et kaolinites), des vermiculites et des chlorites (Bogdanoff et al., 1996)

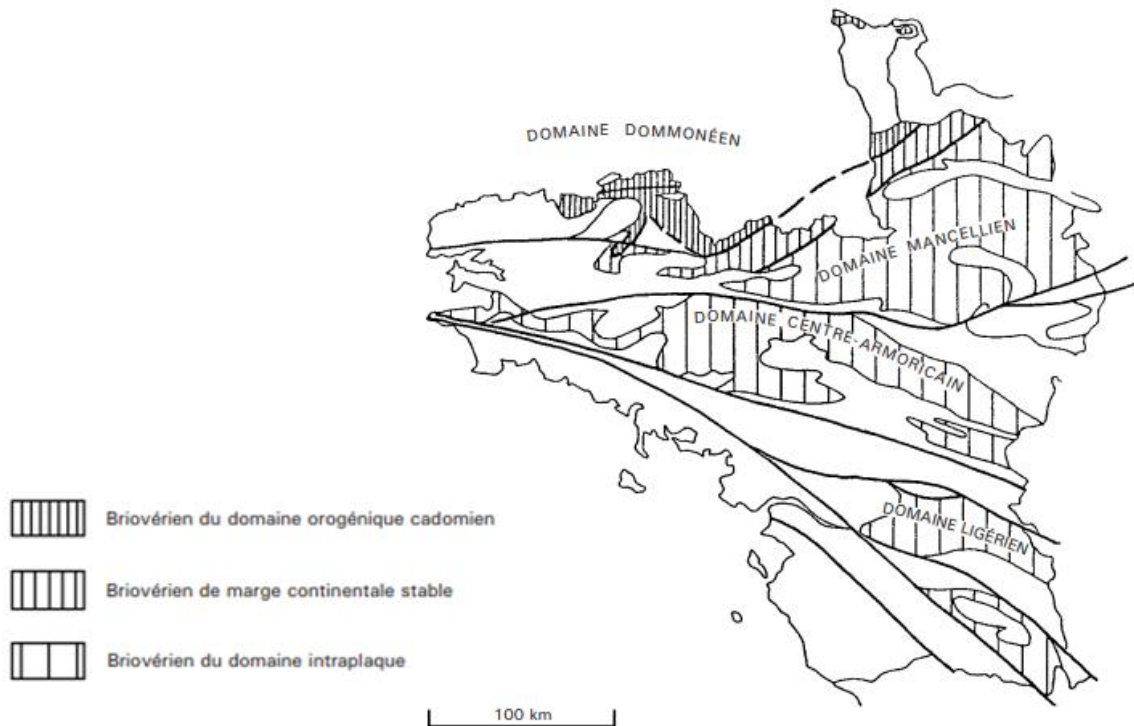


Figure 2.2 : Répartition cartographique des terrains briovériens dans l'ouest de la France (seuil du Poitou, Bretagne et Normandie) (BRGM, 2009).

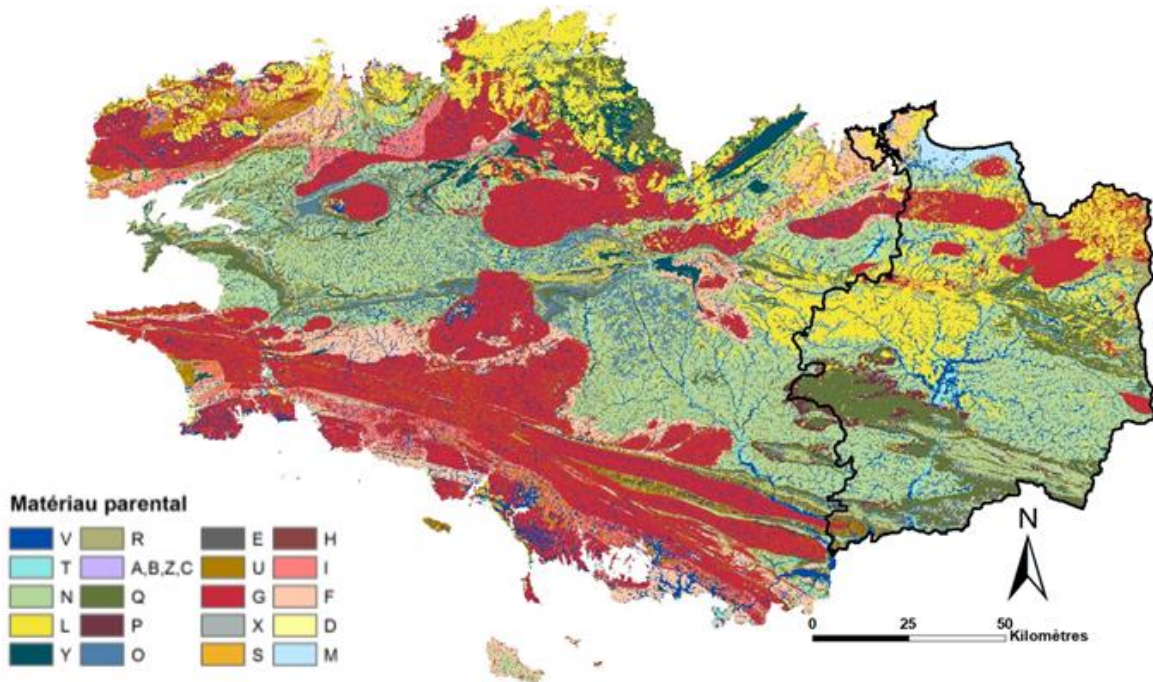


Figure 2.3 : Cartographie prédictive du matériau parental en Bretagne (A=Argile, Altérite, B=Cuirasse ferrugineuse, C=Calcaire, D=Dune sableuse d'origine marine, E=Eboulis de pente, F=Micaschistes, G=Granite I=Gneiss, L=Limons, M=Marais, N=Schiste tendre, O=Schiste moyen, P= Schiste dur, R= Schiste gréseux, S= Sable, T= Terrasse caillouteuse, U= Matériau d'apport colluvial, V= Matériau d'apport alluvial, X= Quartz et Poudingue, Z= Matériau remanié. (Lacoste et al., 2011)

2.1.3. Contexte pédologique

La diversité des matériaux géologiques en Bretagne : roches plutoniques, matériaux métamorphisés, matériaux encaissant briovériens, limons éoliens du Quaternaire (Lacoste et al, 2011, Figure 2.3) génère une grande diversité de sols avec des caractéristiques très variables. Les sols bretons regroupent environ 320 types de sols distincts (Berthier et al., 2013a ; Berthier et al., 2013b ; Le Bris et al, 2013a ; le Bris et al, 2013b) dont majoritairement des : Brunisols, Luvisols, Rankosols, Colluviosols et Fluviosols (Référentiel Pédologique, 2008.). Ces sols se différencient notamment par le matériau parental, les processus pédologiques, l'épaisseur et l'intensité de leur hydromorphie avec la prédominance des sols brunifiés (Figure 2.4). Plus encore, certains sols bretons se distinguent par leur rareté régionale, soit parce que les conditions de leur pédogenèse sont peu fréquentes dans la région, soit parce qu'ils sont issus d'un matériau d'origine relativement rare. Parmi ces sols, on note les sols podzoliques, les sols de tourbes, les sols sableux marins, les sols argileux de marais et les sols des cuirasses ferrugineuses. Globalement, les sols bretons sont riches en carbone organique, Néanmoins, la teneur en carbone organique des sols n'est pas répartie de manière homogène à l'échelle de la Bretagne où elle varie de 2 à 8 % avec une tendance annuelle à la baisse d'environ 0,06% (Bretagne environnement, 2009). Les sols de Bretagne sont généralement acides avec un pH régional moyen de 6,3 (Bretagne environnement, 2009).

A l'échelle du département d'Ille-et-Vilaine, environ 171 types de sol sont distingués (Le Bris et al., 2013) avec la prédominance des sols brunifiés marqués par des processus de lessivage d'argile. Ces sols présentent des épaisseurs variées, allant de quelques centimètres à plus de 3 mètres. Environ 54 % de ces sols présente une épaisseur supérieure à 100 cm. Les horizons organo-minéraux présentent également une large gamme d'épaisseur, comprises entre 5 et 60 cm. En Ille-et-Vilaine, les sols les plus épais sont développés à partir des limons éoliens ou des alluvions et les sols les moins épais sont développés sur des grès et des cornéennes. Ces sols présentent des degrés d'hydromorphie assez variés et qui sont étroitement liés à leurs positions topographiques dans les versants. Les sols hydromorphes occupent généralement les zones de faibles pentes et les positions de bas-fond à proximité des cours d'eau. En revanche, les sols bien drainés occupent les plateaux. Dans l'ensemble, l'organisation spatiale des sols du point de vue de leur épaisseur, hydromorphie, texture et surtout de l'épaisseur des horizons organo-minéraux, est fortement liée à la variabilité du matériau parental des sols mais aussi aux structures paysagères marquées par un réseau bocager dense et un réseau hydrographique très développé.

Plus encore, les sols de département d’Ille-et-Vilaine sont majoritairement de texture limoneuse avec une granulométrie intermédiaire, entre sable et argile. Cette texture favorise la rétention d’eau dans la porosité du sol mais leur confère aussi une certaine fragilité liée à leur sensibilité à l’érosion et à la compaction. Globalement, des gradients de granulométrie distinguent les sols en Ille-et-Vilaine avec des sols sableux issus notamment des granites, et des sols argileux situés dans les zones de marais littoraux et parfois intérieurs, et dans les zones d’altération de certains grès.

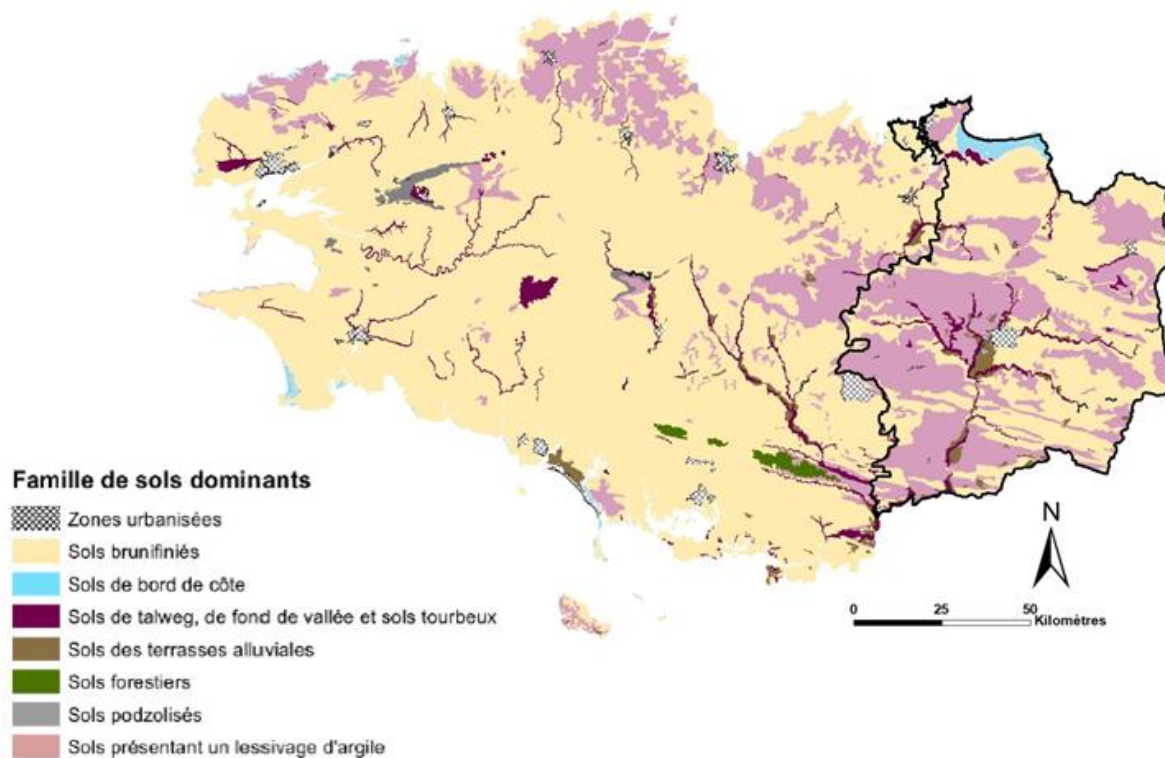


Figure 2.4 : Cartes de types des sols les plus fréquents en Bretagne.

2.1.4. Structure du paysage et occupation du sol

La caractérisation des paysages bretons a été réalisée dans le cadre de programme Sols de Bretagne par le laboratoire COSTEL (Climat et Occupation du Sol par TELédétection) (Le Dû-Blayo et al., 2008). La carte délivrée est une carte de grandes unités paysagères prenant en compte le relief, la végétation et l’occupation des sols. La cartographie des unités paysagères est issue d’un traitement d’images MODIS grâce à des algorithmes de classification pour optimiser la discrimination du paysage breton. Au total, 19 unités paysagères sont distinguées en Bretagne. La Figure 2.5 représente la répartition spatiale de ces grands ensembles sur le territoire breton. Elle révèle des structures paysagères assez contrastées et témoigne encore une fois d’une forte

hétérogénéité : les plaines avec bocage à ragoisse destructuré, qui représentent 22 % de la superficie bretonne, se situent surtout à l'est. Le bocage dense et les prairies sur collines se trouvent principalement sur les hauteurs à l'ouest. Les cultures de légumes, sur les plateaux et sur les plaines, bien qu'elles représentent moins de 2 % de la surface régionale, constituent des surfaces non négligeables sur la côte nord, et marquent profondément le paysage. Les températures douces et peu contrastées liées à la proximité de la mer et la présence de sols limoneux très épais sont propices aux cultures légumières de plein champ.

En Ille-et-Vilaine, le paysage est majoritairement bocager et l'agriculture est orientée vers la polyculture et l'élevage laitier. Le paysage est caractérisé par un bocage dense et des prairies sur collines qui occupent environ 50% du territoire du département. Les plateaux avec des paysages ouverts et des bocages résiduels représentent environ 8% de la surface départementale. Finalement, les cultures de légumes sur plateaux et collines occupent environ 3% de la superficie départementale. Globalement, le département d'Ille-et-Vilaine est représentatif de la diversité des occupations des sols à l'échelle régionale avec la prédominance des plaines avec bocages à ragoisse destructurés.

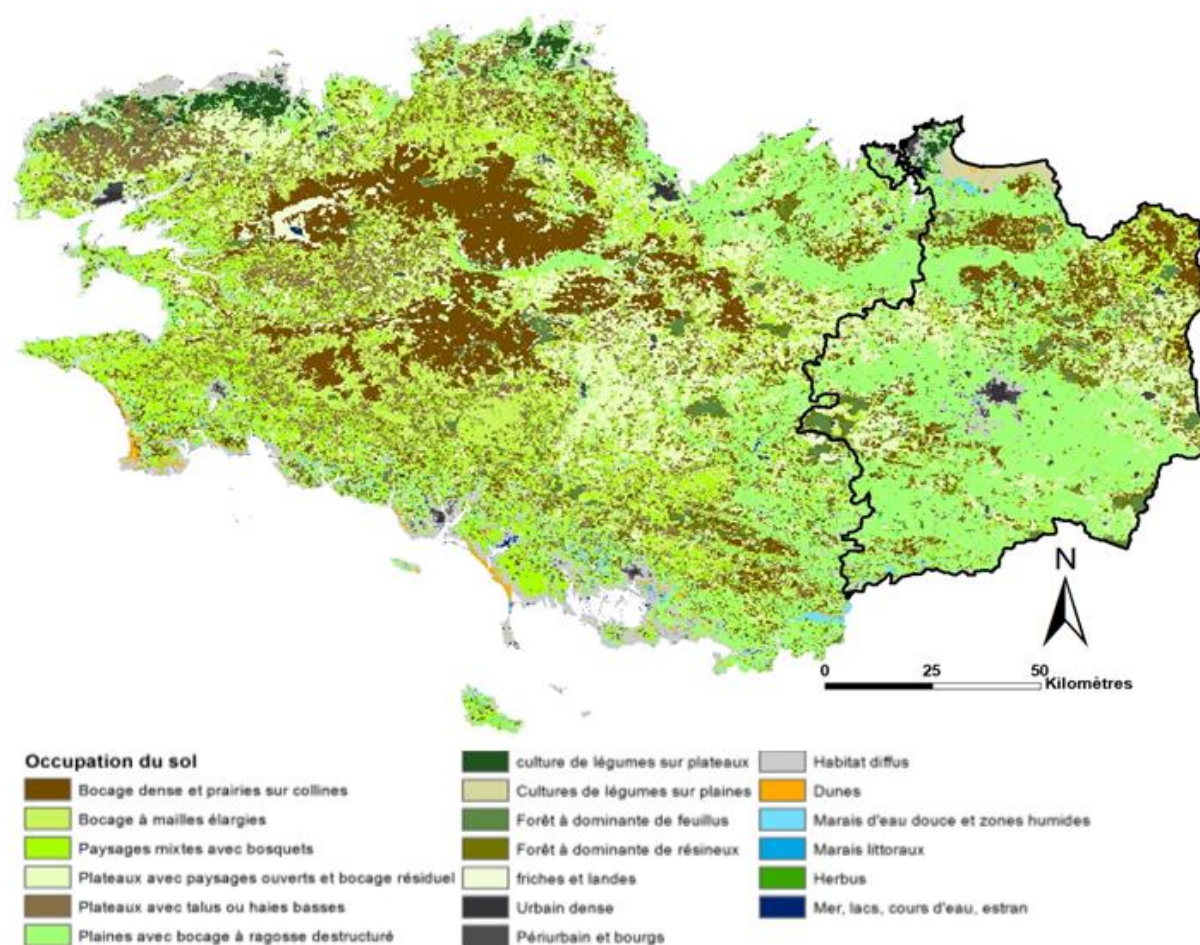


Figure 2.5 : Les unités paysagères de la Bretagne déterminées par télédétection (Le Du-Blayo et al., 2008).

2.2. Les données pédologiques

Dans le cadre de cette thèse, deux types de données pédologiques sont mobilisés i) les données issues d'étude pédologiques préexistantes répertoriées dans la base de données sémantique Donesol 3 du Référentiel Régional Pédologique de Bretagne (RRP) ; et (ii) les données récentes collectées au cours de la thèse sur le territoire du département d'Ille-et-Vilaine à des fins de validation des cartes pédologiques produites.

2.2.1. Le RRP de Bretagne

2.2.2. Structuration de la base de données

Le programme IGCS (Inventaire Gestion et Conservation des Sols) a pour ambition de produire, structurer et diffuser les informations sur la couverture pédologique de la France. Bien qu'il s'agisse d'un programme multi-échelle, l'accent est mis sur la réalisation du volet RRP (Référentiel Régional Pédologique) qui vise à cartographier les sols sur chaque région française à une échelle spatiale de 1/ 250 000^{ème} tout en respectant un cahier des charges normalisé au niveau national. Les données pédologiques sont consignées selon les spécifications requises par le modèle de base de données relationnel Donesol, format de stockage nationale des données pédologiques. Le programme IGCS est coordonné au niveau national par l'Unité Infosol (INRA Orléans) et décliné régionalement grâce à un réseau de partenaires. En Bretagne, AGROCAMPUS OUEST et l'UMR SAS assurent la réalisation, la diffusion et la valorisation du RRP.

La base de données sémantique de Donesol est composée de plusieurs objets ponctuels et surfaciques caractérisant l'organisation spatiale des sols (Figure 2.6). A l'échelle de restitution cartographique 1/ 250 000^{ème}, l'information pédologique est représentée par des Unités Cartographiques de Sols (UCS), délimitées dans l'espace et cartographiées sous forme de polygones, composés d'une ou plusieurs Unité Typologiques des Sols (UTS) non délimitées. Selon le dictionnaire de données Donesol 3, une UCS est formellement définie comme une plage cartographique dont les facteurs pédogénétiques révèlent une certaine homogénéité dans l'organisation spatiale (Infosol, 2014). On distingue deux types d'UCS : une UCS est dite pure lorsqu'elle est composée d'une seule UTS, ou complexe lorsqu'elle est composée de plusieurs UTS avec des proportions surfaciques précises. En outre, selon la même source, une UTS est définie comme une portion de la couverture pédologique présentant un ensemble de caractéristiques pédogénétiques homogènes (Infosol, 2014). Chaque UTS est déclinée en un certain nombre de strates définissant l'organisation spatiale des horizons pédologiques. Les strates disposent d'une description pédologique détaillée des propriétés qualitatives et quantitatives de sol. Cette description est issue des observations et des analyses ponctuelles réalisées sur le terrain en prospectant plusieurs sondages et fosses pédologiques répartis sur le territoire d'étude. Dans l'ensemble, les données contenues dans les strates permettent de caractériser la variabilité pédologique d'une succession de strates typique de chaque UTS. Dans le cas des données quantitatives, les analyses effectuées et/ou l'expertise pédologique sont

mobilisées pour définir une distribution probable des propriétés de sol pour une strate donnée. Les valeurs minimales et maximales ainsi que la valeur modale de chaque propriété sont renseignées. Quant aux données qualitatives, en plus de la modalité dominante, deux modalités supplémentaires sont fournies pour appréhender la variabilité intrinsèque d'une strate donnée.

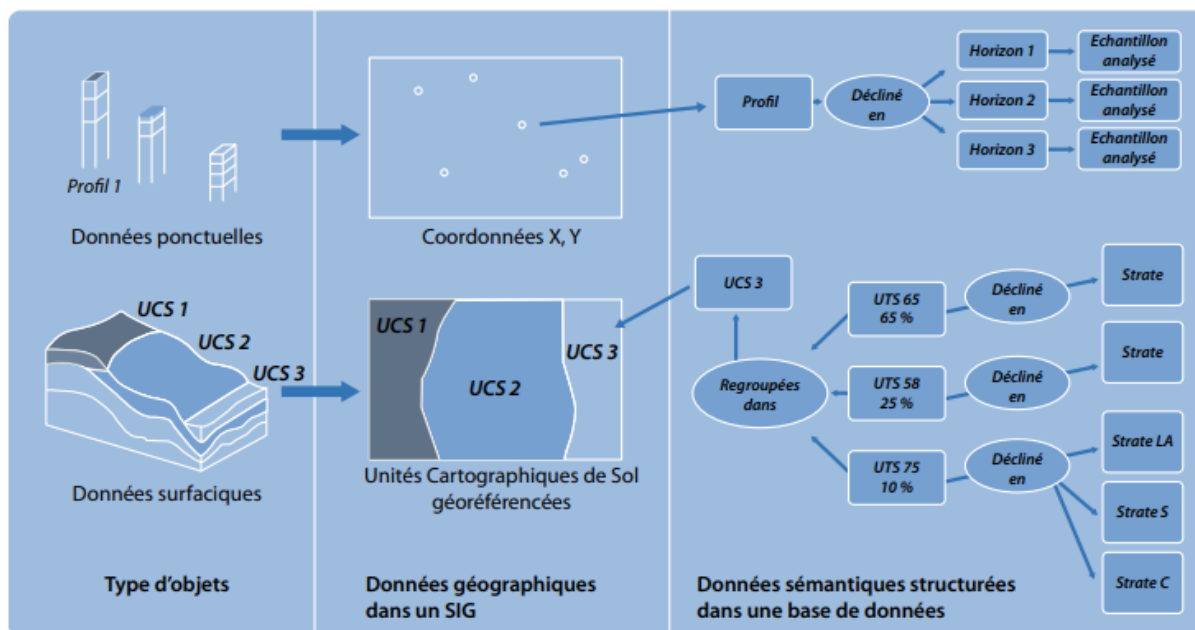


Figure 2.6 : Organisation des principaux objets pédologiques (UCS, UTS, strates, profils et horizons) dans le système d'information pédologique (SIG+DoneSol) (<https://sols-et-territoires.org>).

Le RRP de Bretagne a été réalisé dans le cadre du projet « Sols de Bretagne » (www.sols-de-bretagne.fr) et achevé en 2012 pour délivrer une cartographie exhaustive des sols bretons et une base de données relationnelle permettant d'établir les liens entre les UCS, les UTS et les propriétés pédo-environnementales du territoire régional. Le RRP de Bretagne se compose de 4 bases de données départementales constituées de façon cohérente au niveau régional. Concrètement, deux UCS issues de deux départements mais portant le même numéro d'identification correspondent à une unité cartographique similaire. Seules les proportions des UTS qui la composent peuvent varier à la marge d'une UCS à l'autre. Une liste unique d'UTS a été constituée sur l'ensemble de la région.

Le RRP de Bretagne comprend :

- Une carte des pédopaysages à une échelle du 1/ 250 000^{ème} contenant 341 UCS, dont la majorité sont des UCS composées avec un nombre total de 320 UTS. Les UCS sont

représentées par 1 ou plusieurs polygones et la carte contient 1984 polygones au total (Figure 2.7).

- Une carte des entités physiographiques de Bretagne, au nombre de 29 (9 en Ille-et-Vilaine), qui constituent la clé d'entrée de la légende des cartes des pédopaysages (<http://sols-de-bretagne.fr/donnees-et-outils/CARTES/DONNEES-DE-CADRAGE/Entit%C3%A9s%20physiographiques-de-Bretagne/>)
- 3761 profils pédologiques, échantillonnés entre 2005 et 2012.
- 692 profils pédologiques observés sur fosse.
- Un document de synthèse par département, présentant l'organisation des sols (UTS) au sein des UCS (Berthier et al., 2013a ; Berthier et al., 2013b ; Le Bris et al, 2013a ; le Bris et al, 2013b).

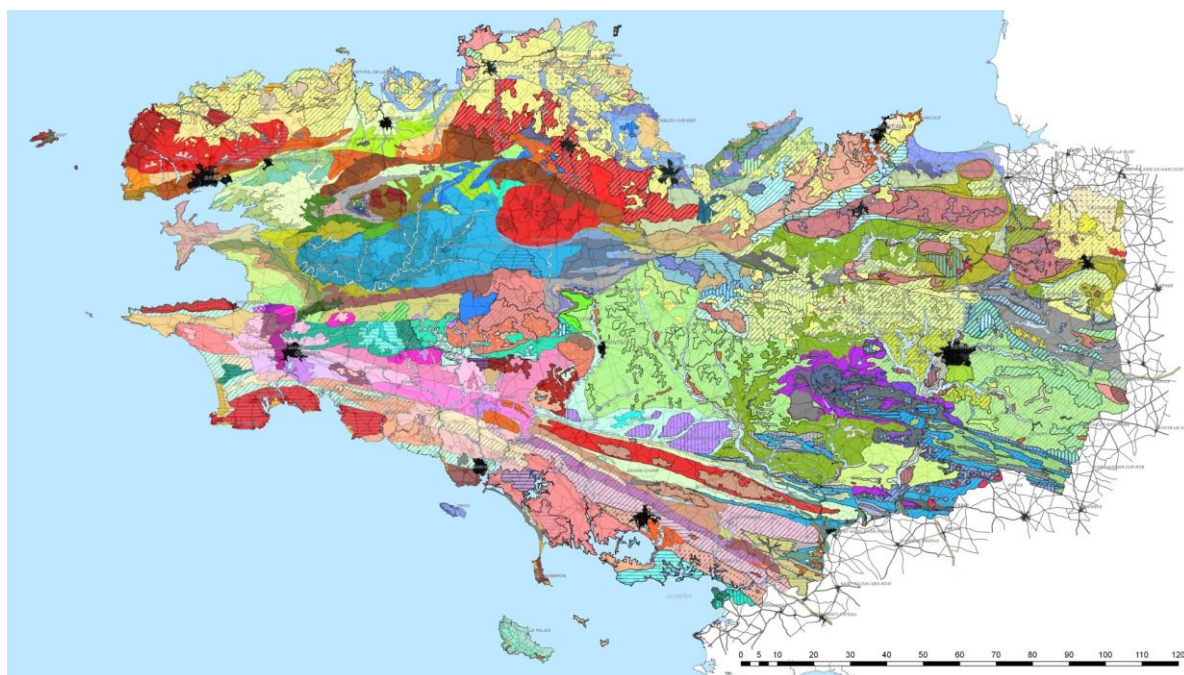


Figure 2.7 : Les unités cartographiques des sols (UCS) en Bretagne (<http://sols-de-bretagne.fr/donnees-et-outils/CARTES/PEDOPAYSAGES/>).

2.2.3. Extraction des données pédologiques du RRP de Bretagne

L'extraction des données pédologiques de la base de données sémantique Donesol 3 constitue la première étape du travail de valorisation de ces données mené dans le cadre de la thèse. Elle consiste à interroger la base de données à l'aide d'une requête en langage SQL permettant d'extraire les données relatives aux objets surfaciques (Figure 2.8). Pour l'essentiel, nous avons

extrait les proportions surfaciques des UTS constituant les UCS et les données pédologiques des strates constituant les UTS à savoir les limites supérieures et inférieures d'apparition de chaque strate, et les valeurs modales de la teneur en carbone organique, de la CEC, des cinq fractions granulométriques, du pH et de la charge en éléments grossiers. Les variables extraites de la base de données Donesol et les noms des tables sont présentés dans le tableau suivant (Tableau 2.1).

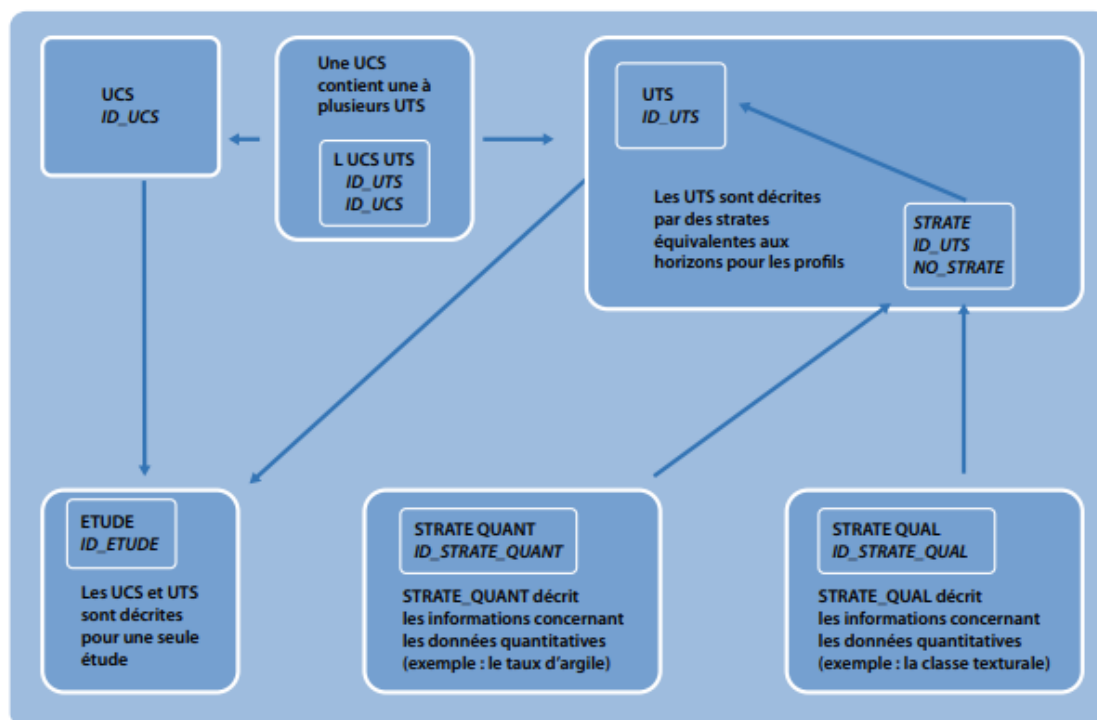


Figure 2.8 : Liens entre les différentes tables de Donesol qui caractérisent les données surfaciques (<https://sols-et-territoires.org/produits-du-reseau/publications-marquantes/>).

Tableau 2.1 : Synthèse des variables extraites de la base de données Donesol 3 du RRP de Bretagne.

Propriétés	Table				
	UCS	UCS-UTS	UTS	Strate	Strate_quant
Argile	no_ucs ;	pourcent	no_uts	no_strate ; prof_appar_moy ; epais_moy	nom_var ; val_mod ; no_strat; val_min; val_max
Limon	nb_uts ;				
Sable	no_int_ucs				
Carbone					
pH					
CEC					
Eléments grossiers					

2.2.4. Analyse du taux de complétude de la base de données

La mise en place d'une méthodologie de cartographie numérique des sols nécessite une caractérisation fine des informations disponibles. Une analyse approfondie des données pédologiques contenues dans le RRP de Bretagne est donc incontournable pour caractériser les potentialités de ces informations et optimiser la caractérisation spatiale des sols. Cela consiste à déterminer le taux de complétude de la base de données Donesol tout en précisant la propriété du sol renseignée et son taux de renseignement pour toutes les strates des UTS.

En accord avec le consortium *GlobalSoilMap* (Arrouays et al., 2014), qui vise une cartographie précise des sols mondiaux, les propriétés de sol seront exprimées en six profondeurs standards 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm et 100-200 cm. Une analyse du taux de renseignement des strates vis-à-vis les propriétés des sols aux profondeurs requises par les spécifications du programme *GlobalSoilMap* constitue donc une étape nécessaire pour harmoniser les prédictions. Pour l'essentiel, ce travail exploratoire de la base de données Donesol 3 permet de qualifier l'aptitude des données du RRP à décrire la variabilité des sols pour être utilisées comme données d'entrée des modèles de cartographie numérique des sols.

Pour caractériser le taux de renseignement du RRP de Bretagne vis-à-vis des propriétés des sols à prédire selon les spécifications requises par le consortium *GlobalSoilMap*, un ensemble d'indicateurs d'analyse des données pédologiques a été calculé. La caractérisation des données pédologiques fournies dans le RRP de Bretagne est effectuée en deux temps : i) une caractérisation globale du taux de renseignement des strates vis-à-vis les valeurs modales, maximales et minimales des propriétés des sols et ii) une caractérisation fine pour analyser le taux de renseignement des strates aux intervalles de profondeurs spécifiés par le consortium *GlobalSoilMap*.

Le premier indicateur est le taux de renseignement global des strates pour l'ensemble des propriétés des sols vis-à-vis de leurs modalités de renseignements à savoir les valeurs minimales, modales et maximales, Cet indicateur est défini par l'équation suivante (Vaysse, 2015, Eq 2.1) :

$$Taux\ info_{propriété} = \left(\frac{Nstrates_{propriété}}{Nstrates_{total}} \right) \times 100 \quad (Eq\ 2.1)$$

avec $Nstrates_{propriété}$ correspondant au nombre de strates renseignées pour une propriété de sol donnée et $Nstrates_{total}$ correspondant au nombre total de strates présentes dans le RRP.

Pour réaliser une analyse plus fine, un indicateur de complétude pour chaque propriété de sol aux différents intervalles de profondeur GlobalSoilMap a été calculé selon l'équation suivante (Vaysse, 2015, Eq 2.2) :

$$Taux\ UTS_{propriété \times prof} = \left(\frac{Nuts_{propriété \times prof}}{Nuts_{total\ prof}} \right) \times 100 \quad (Eq\ 2.2)$$

avec $Nuts_{total\ prof}$ correspond au nombre total d'UTS définies dans le RRP à l'intervalle de profondeur considéré et $Nuts_{propriété \times prof}$ correspond au nombre d'UTS disposant d'une information sur la propriété des sols pour l'intervalle de profondeur considéré.

Un intervalle de profondeur est considéré comme totalement renseigné si au moins 5 cm de cet intervalle est inclus entre les bornes de profondeur d'apparition et de disparition des strates des UTS. Ce taux de complétude est calculé pour toutes les modalités de renseignements des données pédologiques à savoir la valeur modale et les valeurs minimales et maximales définissant les gammes de variation de ces propriétés.

2.2.5. Résultats de caractérisation de la base de données Donesol 3 du RRP de Bretagne

Le tableau 2.2 présente les résultats associés aux taux de renseignement global des strates constituant les UTS vis-à-vis des trois modalités de renseignement des propriétés. Globalement la base de données est très bien renseignée surtout pour les valeurs modales où le pourcentage des strates disposant d'une valeur modale est de 100% pour toutes les propriétés de sol.

Au regard des bornes minimales et maximales, les variables texturales sont les plus renseignées avec un taux de renseignement de l'ordre de 71 % pour les cinq fractions granulométriques et atteignant 98% pour la charge en éléments grossiers. Dans le cas de la CEC, la teneur en carbone organique et le pH, le taux de renseignement est plutôt élevé et oscille entre 64% et 71%.

Ainsi, l'aspect informatif des données surfaciques du RRP de Bretagne, vis-à-vis des valeurs modales des propriétés des sols, justifie leur mobilisation ultérieure comme données d'entrée de modèles de cartographie numérique. Cependant, les bornes minimales et maximales ne pourront pas être utilisées pour délivrer des prédictions des propriétés assorties de leurs incertitudes vu leur incomplétude.

En complément des indices de complétude globaux des objets surfaciques, des indices prenant en compte les intervalles de profondeurs spécifiés par le consortium sont également calculés pour

toutes les modalités de renseignements des propriétés des sols et les résultats sont montrés dans le tableau 2.3. Pour ce qui concerne le renseignement de la valeur modale, toutes les propriétés de sol sont renseignées pour les intervalles de profondeurs du programme *GlobalSoilMap*. Le nombre d'UTS diminue fortement avec la profondeur allant de 320 UTS informées en surface jusqu'à 15 UTS informées pour l'intervalle de profondeur 100-200 cm. Dans le cas des bornes minimales et maximales, le taux de renseignement est moyen pour toutes les propriétés avec des taux aussi élevés en surface qu'en profondeur.

Tableau 2.2 : Taux de renseignement des strates du RRP de Bretagne pour neuf propriétés de sol vis-à-vis leurs modalités de renseignement (valeurs modales et les bornes minimales et maximales) (SOC=teneur en carbone organique, CEC=capacité d'échange cationique, A=teneur en argiles, S=teneur en sables, L=teneur en limons, LG=teneur en limons grossiers, LF=teneur en limons fins et EG=charge en éléments grossiers).

Type d'info	Propriété	SOC	CEC	pH	A	S	L	LG	LF	EG
Valeur modale	Taux d'info (%)	100	100	100	100	100	100	100	100	100
Valeur maximale	Taux d'info (%)	71	64	68	71	71	71	71	71	98
Valeur minimale	Taux d'info (%)	64	68	71	71	71	71	71	71	70

Tableau 2.3 : Taux de complétude des UTS pour les différentes propriétés des sols en fonction des intervalles de profondeurs du programme *GlobalSoilMap* (SOC=teneur en carbone organique, CEC=capacité d'échange cationique, A=teneur en argiles, S=teneur en sables, L=teneur en limons, LG=teneur en limns grossiers, LF=teneur en limons fins et EG=charge en éléments grossiers).

Type info	Classe prof	N	SOC	CEC	pH	LF	A	L	S	EG	LG
Modale	[0-5]	320	100	100	100	100	100	100	100	100	100
	[5-15]	317	99	100	100	100	100	100	100	100	100
	[15-30]	314	99	100	100	100	100	100	100	100	100
	[30-60]	278	99	100	100	100	100	100	100	100	100
	[60-100]	163	99	100	100	100	100	100	100	100	100
	[100-200]	15	100	100	100	100	100	100	100	100	100
Type info	Classe prof	N	SOC	CEC	pH	LF	A	L	S	EG	LG
Maximale	[0-5]	320	73	68	72	72	72	72	72	72	72
	[5-15]	317	73	69	72	73	73	73	73	74	73
	[15-30]	314	61	55	65	69	69	69	69	69	64
	[30-60]	278	66	62	67	71	71	71	71	65	72
	[60-100]	163	65	62	67	71	71	71	71	62	71
	[100-200]	15	60	83	40	83	83	83	83	47	83
Type info	Classe prof	N	SOC	CEC	pH	LF	A	L	S	EG	LG
Minimale	[0-5]	320	72	68	72	73	73	73	73	95	73
	[5-15]	317	73	69	72	74	74	74	74	95	74
	[15-30]	314	62	55	65	67	67	67	67	90	67
	[30-60]	278	66	58	69	72	72	72	72	85	72
	[60-100]	163	65	62	67	71	71	71	71	86	71
	[100-200]	15	60	53	40	53	53	53	53	54	53

2.2.6. Jeu de validation

Selon les recommandations de Brus et al., (2011), la validation des cartes pédologiques devrait se faire à l'aide d'un jeu de données indépendant du jeu de calibration, sélectionné par échantillonnage aléatoire stratifié. Cette approche est plus efficace qu'une validation croisée ou

encore une division du jeu de données initial en un jeu de calibration et en un jeu de validation car elle permet d'estimer la qualité de la carte produite indépendamment de tout biais.

A cette fin, nous avons mis en œuvre une méthode d'échantillonnage aléatoire stratifié selon le matériau parental dominant des UCS pour assurer une meilleure répartition géographique des sondages de validation. Dans le cadre de la thèse, pour des raisons budgétaires et logistiques, la collecte du jeu de validation s'est limitée au département d'Ille-et-Vilaine. Le territoire de validation retenu couvre la majorité de la variabilité régionale constatée et représente de ce fait une zone de développement méthodologique pour valider les cartes produites. Le département d'Ille-et-Vilaine est tout d'abord découpé en 11 strates géographiques selon le matériau parental dominant de superficies distinctes. Par la suite 45 points ont été sélectionnés par un tirage aléatoire, proportionnellement à la superficie des strates découpées avec un minimum de 2 points par strate.

Les 45 points sélectionnés ont permis de définir 45 transects topographiques, couvrant la variabilité pédopaysagère du département. Chaque transect est tracé manuellement le long des versants en suivant la direction de la plus forte pente. Puis, le dénivelé total qui correspond à la différence d'altitudes entre le haut et le bas de versant, a été calculé puis divisé par trois pour définir trois sections de dénivelés égaux. Le point initial permet de caractériser la section dans laquelle il se trouve. Un point a été tiré de façon aléatoire au sein de chacune des deux autres sections, le long du transect.

Pour caractériser la variabilité à courte distance, deux points supplémentaires ont été choisis dans des directions aléatoires, sur le terrain, à 20 m du point caractérisant chaque section. Au total 135 sections (45 transects x 3) ont été définies et le plan d'échantillonnage adopté comprend 405 points ou sites répartis sur l'ensemble du territoire d'étude (Figure 2.9).

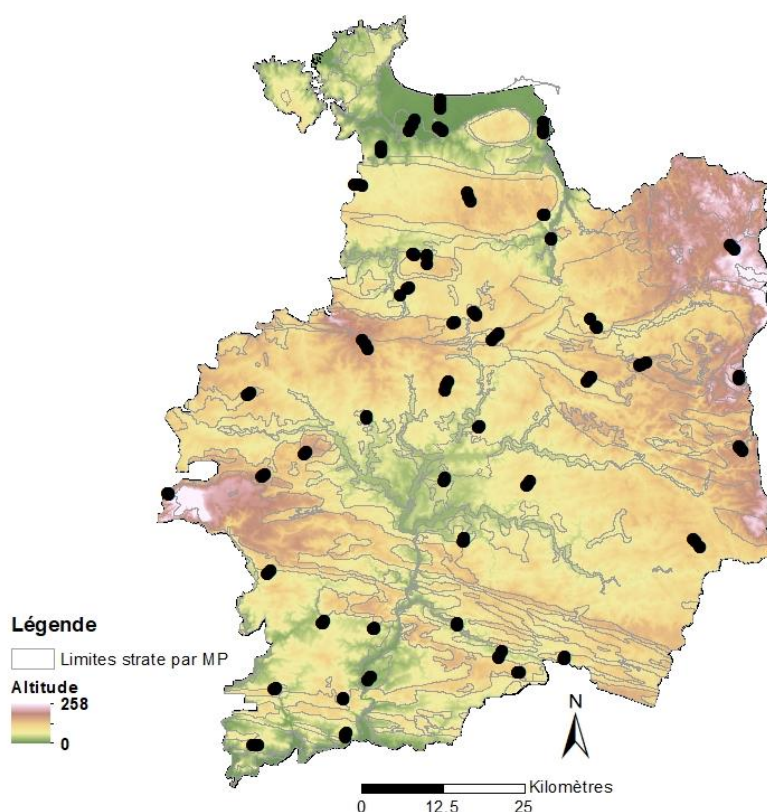


Figure 2.9 : Topographie du département d'Ille-et-Vilaine et localisation des sites de validation sélectionnés par un échantillonnage aléatoire stratifié selon le matériau parental dominant.

2.2.7. Description des sols et analyses

Une campagne de terrain a été menée entre mars et juin 2017. En chacun des 405 points, un sondage à la tarière hélicoïdale classique (longueur de 1,20 m), équipée si nécessaire de rallonges, a permis de décrire et d'échantillonner le sol suivant le plan d'échantillonnage présenté précédemment. Les profils de sol situés sur le transect ont été prospectés jusqu'à 2 m de profondeur alors que les sondages à courtes distances ont été prospectés jusqu'à 1,20 m de profondeur. La position géographique de chaque point a été levée à l'aide d'un GPS de 2 m de précision environ et des photos du paysage et du sol ont été réalisées (Figure 2.10).

La description pédologique des profils de sol prospectés est faite en décrivant leurs successions d'horizons génétiques avec une description détaillée de chaque horizon. Cela suppose de déterminer, l'épaisseur et la couleur des différents horizons (charte Munsell), l'hydromorphie (présence de signes d'engorgement temporaire ou permanent), la texture (pourcentage d'argile, de limon et de sable), le pourcentage de la matière organique et la charge en éléments grossiers.

De plus, une description globale de chaque profil de sol a été faite via l'attribution d'un code synthétique appelé code quatre critères (Rivière et al., 1992) caractérisant le matériau parental, le degré d'hydromorphie, le développement de profil selon le Référentiel Pédologique français (AFES, 2008), et la classe de profondeur du sol.

Les profils de sol prospectés ont été échantillonnés aux six intervalles de profondeurs spécifiés par le consortium *GlobalSoilMap* selon une stratégie d'échantillonnage composite. En effet, pour chaque section et pour chaque intervalle de profondeur, les échantillons du sol prélevés au sein des trois profils (profil situé sur le transect + les 2 profils supplémentaires à courtes distances) ont été mélangés pour former un échantillon composite. Ensuite, un échantillon d'environ 400 g de sol a été prélevé, dans chaque intervalle de profondeur, et mis dans des sacs en plastique pour effectuer les analyses physico-chimiques. Dans le cadre de la thèse, que les échantillons de terre prélevés aux intervalles de profondeurs 5-15 cm, 30-60 cm et 60-100 cm ont été analysés. Les échantillons prélevés aux intervalles de profondeur 0-5 cm, 15-30 cm et 100-200 cm ont été archivés.

Pour effectuer les analyses physico-chimiques nécessaires, les échantillons de sol retenus (367 échantillons) ont été séchés à l'air puis tamisés à 2 mm. Pour déterminer la charge en élément grossiers, chaque échantillon a été pesé avant et après son passage sur le tamis et en fonction du refus sur le tamis, la masse des éléments grossiers a été déduite. Une fraction de chaque échantillon a été broyée à 10 μm pour la détermination de la teneur en carbone organique.

Tous les échantillons tamisés et les sous échantillons broyés ont été envoyés au Laboratoire d'Analyse des Sols INRA d'Arras pour effectuer les analyses granulométriques et mesurer la teneur en carbone organique, le pH et la capacité d'échange cationique.



Figure 2.10 : Exemples de profils de validation sur des schistes briovériens. Pour chacune de ces 4 photos, le profil de sol au milieu représente le profil situé sur le transect et les deux autres représentent les profils à courtes distances (à 20 m).

2.2.8. La caractérisation pédologique du jeu de validation

Les 405 profils de sol prospectés couvrent une grande hétérogénéité pédologique du département d'Ille-et-Vilaine. La caractérisation pédologique de ces profils met en évidence la présence de treize matériaux parentaux différents (schiste briovérien, granite, gneiss, schiste dur, schiste moyen, grès, schiste gréseux, terrasse, limon, matériau d'apport alluvial et colluvial) avec la prédominance du schiste briovérien (23%), suivi par le limon éolien (15%). En ce qui concerne l'épaisseur de sol, les sols profonds dont la profondeur dépasse 80 cm, représentent 46% de l'ensemble des profils prospectés. Quant aux sols moyennement profonds, ayant une profondeur comprise entre 40 et 80 cm, ils représentent environ 36% des profils et le reste, soit 18%, représente des sols peu profonds avec moins de 40 cm de profondeur. Concernant l'hydromorphie, les sols observés sont majoritairement sains (37%), mais on note également la présence considérable de sols hydromorphes qui représentent environ 30% des profils. Finalement, les types de sol observés sont relativement divers. On observe notamment des néoluvisols, luvisols, brunisols, rankosols, colluviosols, thalassosols et des fluvisols. Ces sols se distinguent principalement par leurs processus pédogénétiques.

De plus, les propriétés analytiques du jeu de validation couvrent une large gamme de valeurs des propriétés des sols. Par exemple, le pH varie de 3,95 à 8,58 (moyenne = 6,07) pour l'intervalle de profondeur 5-15 cm. Il varie de 4,41 à 8,9 (moyenne = 6,25) pour l'intervalle de profondeur 30-60 cm et de 4,55 à 9,01 (moyenne = 6,30) pour l'horizon le plus profond (60-100 cm) (Tableau 2.4). La variabilité exprimée par l'écart-type est relativement constante pour les variables texturales, pour les trois intervalles de profondeurs échantillonnés. Une diminution nette de la teneur en carbone organique avec l'augmentation de la profondeur caractérise l'ensemble des échantillons collectés.

Pour permettre une comparaison visuelle, les résultats de projection de la composition granulométrique des données fournies par le RRP de Bretagne et les analyses réalisées sont représentées côte-à-côte (Figure 2.11). La Figure 2.11A situe, dans le triangle des textures GEPPA (1963), la composition granulométrique de l'ensemble des strates du RRP de Bretagne. Les caractéristiques majeures du nuage de points obtenu montrent que plus de 80% des échantillons appartiennent à quatre classes de texture voisines à savoir les classes LAS (Limon argilo-sableux), La (Limons argileux), LSa (Limon sablo-argileux) et L (limon). Cela correspond à la partie très dense du nuage de points ayant des teneurs en argile comprises entre 12 et 25% et des teneurs en limon

comprises entre 40 et 70%. De plus, les classes de textures sableuses (SS, S, Sa) et argileuses (AA, A, As) sont très peu représentées.

La Figure 2.11B représente, dans le même triangle, la composition granulométrique des 367 échantillons prélevés sur le département d'Ille-et-Vilaine. La plupart des analyses granulométriques effectuées se retrouvent dans les classes LAS (Limon argilo-sableux), La (Limons argileux), LSa (Limon sablo-argileux), L (limon). Par conséquent, le plan d'échantillonnage aléatoire stratifié adopté couvre bien la variabilité texturale observée à l'échelle régionale.

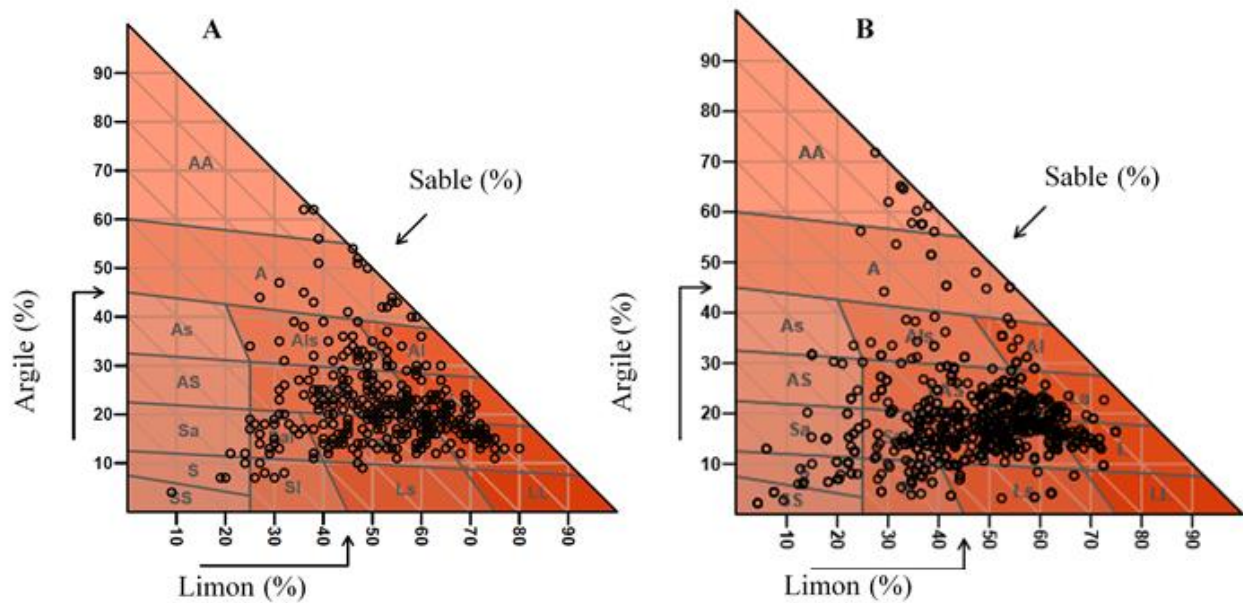


Figure 2.11 : Situation dans le triangle du GEPPA de la composition granulométrique des 1020 strates du RRP de Bretagne (A) et des 367 analyses de terre réalisées en Ille-et-Vilaine (B).

Tableau 2.4 : Paramètres statistiques estimés de différentes propriétés analytiques de trois intervalles de profondeur (5-15 cm (n= 135) ; 30-60 cm (n=125) et 60-100 cm (n=103)).

profondeur	Propriété	minimum	médiane	moyenne	maximum	Ecart type
5-15 (cm)	CEC (cmol ⁺ kg ⁻¹)	2,33	7,03	9,52	112	11,58
	pH	3,95	5,99	6,07	8,58	0,99
	Carbone (‰)	10,40	23,30	30,34	275	2,97
	Eléments grossiers (%)	0,00	12,63	13,40	49,81	9,93
	Sable grossier (%)	0,20	9,30	12,38	48,20	10,90
	Sable fin (%)	1,20	11,90	13,00	40,60	6,24
	Limon grossier (%)	0,70	25,90	26,90	52,30	11,27
	Limon fin (%)	0,25	25,40	24,78	40,50	7,30
	Argile (%)	6,80	19,20	20,90	62,20	8,05
30-60 (cm)	CEC (cmol ⁺ kg ⁻¹)	1,49	4,88	5,90	34,4	4,20
	pH	4,41	6,18	6,25	8,90	1,00
	Carbone (‰)	2,55	7,53	11,84	56,90	12,19
	Eléments grossiers (%)	0,00	14,39	16,11	52,37	12,62
	Sable grossier (%)	0,00	9,90	13,47	5,78	12,46
	Sable fin (%)	0,30	11,70	12,87	54,00	6,65
	Limon grossier (%)	0,40	25,00	26,30	49,80	11,44
	Limon fin (%)	0,21	24,80	24,00	42,00	7,92
	Argile (%)	6,00	19,10	20,50	51,40	8,20
60 -100 (cm)	CEC (cmol ⁺ kg ⁻¹)	2.25	6.06	6.94	28	2.72
	pH	4,55	6,05	6,30	9,01	0,89
	Carbone (‰)	1.07	2.65	3.20	14.90	1.28
	Eléments grossiers (%)	0,21	6,78	9,20	61,03	6,73
	Sable grossier (%)	0	10,40	14,06	80,60	10,12
	Sable fin (%)	0,24	10,80	12,21	62,50	54,86
	Limon grossier (%)	0,40	24,20	26,79	70,88	10,93
	Limon fin (%)	0,15	24,60	24,47	55,30	7,64
	Argile (%)	3,90	21,00	22,45	53,75	7,15

Conclusion partielle

En synthèse, la région Bretagne est marquée par une grande diversité géologique, pédologique et topographique. Les efforts nationaux en matière de restitution cartographique de l'information pédologique à une échelle spatiale de 1/250 000^{ème} mais aussi les efforts régionaux tels que la cartographie prédictive du matériau parental (Lacoste et al, 2011), visent à mieux caractériser l'organisation spatiale des sols ainsi que la variabilité de leurs caractéristiques.

L'analyse exploratoire de la base de données existante associée au RRP de Bretagne, au format Donesol 3, montre clairement un taux de complétude très élevé des objets surfaciques. Cela justifie l'utilisation ultérieure de ces données pédologiques pour une cartographie précise des sols en Bretagne tout en respectant les spécifications du consortium *GlobalSoilMap*. Dans l'ensemble, les données surfaciques (UTS) sont très bien informées pour toutes les modalités de renseignement des propriétés et surtout pour les valeurs modales qui révèlent un taux de complétude maximale (100%). Les indices de complétude calculés assurent une analyse purement quantitative de la base de données Donesol 3 mais n'augurent pas de la fiabilité des valeurs renseignées. Ces données sont issues non seulement des analyses des propriétés des sols, réalisées à différentes périodes depuis une trentaine d'années mais aussi de l'expertise pédologique. Finalement, le jeu de validation collecté à l'échelle du département d'Ille-et-Vilaine montre clairement sa pertinence vis-à-vis de sa représentativité de la variabilité régionale. Cette stratégie d'échantillonnage probabiliste permet de calculer des estimateurs sans biais des paramètres statistiques (moyenne, variance, etc...). Plus encore, le fait que les sites sélectionnés n'interviennent pas dans la création des données du RRP de Bretagne, leur confère un caractère indépendant et ce jeu de données permet une validation objective et fiable, d'un point de vue statistique, des prédictions des propriétés des sols.

2.3. Méthodologie d'évaluation des services écosystémiques des sols

Cette partie est consacrée à la présentation des indicateurs de services écosystémiques instruits ainsi que du modèle utilisé. Elle est structurée en deux volets : i) une présentation générale du modèle de culture utilisé et son organisation modulaire qui permet une simulation dynamique des processus des sols et ii) une présentation générale des indicateurs retenus pour appréhender les services écosystémiques.

2.3.1. Présentation du modèle de culture STICS

Le modèle de culture STICS (Simulateur multIdisciplinaire pour les Cultures Standard) est développé à l'INRA depuis une trentaine d'années, en collaboration avec de nombreuses structures de recherche, des organismes professionnels agricoles et des établissements d'enseignement agronomique (Brisson et al., 2008). Il offre un cadre conceptuel permettant de caractériser les interactions dynamiques entre les différentes composantes de l'écosystème agricole à savoir le sol, la plante, le climat et les pratiques agricoles. Le modèle de culture STICS décrit l'ensemble de processus écophysologiques et agronomiques, associés au fonctionnement de nombreuses espèces cultivées, en grande partie des cultures annuelles. Il représente un modèle mécaniste qui simule à pas de temps journalier le comportement du système sol-culture-atmosphère au cours d'une ou plusieurs années de façon ponctuelle dans l'espace. Il permet donc de retranscrire les relations environnementales complexes via de nombreuses variables de sortie décrivant les processus dynamiques du système sol-plante-climat. Ainsi, le modèle de culture STICS est un modèle générique en termes de cultures modélisées mais aussi en termes d'itinéraires techniques configurés et simulés. Cette généricité est facilitée par l'organisation du modèle en neuf modules correspondant à des processus spécifiques (Figure 2.12).

- Trois modules d'écophysologie des parties aériennes des plantes : modules de développement phénologique, croissance aérienne, élaboration du rendement ;
- Quatre modules sur le fonctionnement du sol en interaction avec les parties souterraines des plantes : modules croissance racinaire, bilan hydrique, bilan azoté, transferts ;

- Un module sur la gestion des interactions entre les itinéraires techniques et le système sol-plante.
- Un module concerne le microclimat qui traite les effets du climat et du bilan hydrique sur l'évolution de la température et de l'humidité de l'air au sein du feuillage.

Chaque module présente des paramètres optionnels permettant à l'utilisateur d'activer ou de désactiver certains processus associés à l'écophysiologie, au fonctionnement du sol ou encore à la gestion de culture.

Dans la suite, nous présentons les modules de base du modèle de culture STICS.

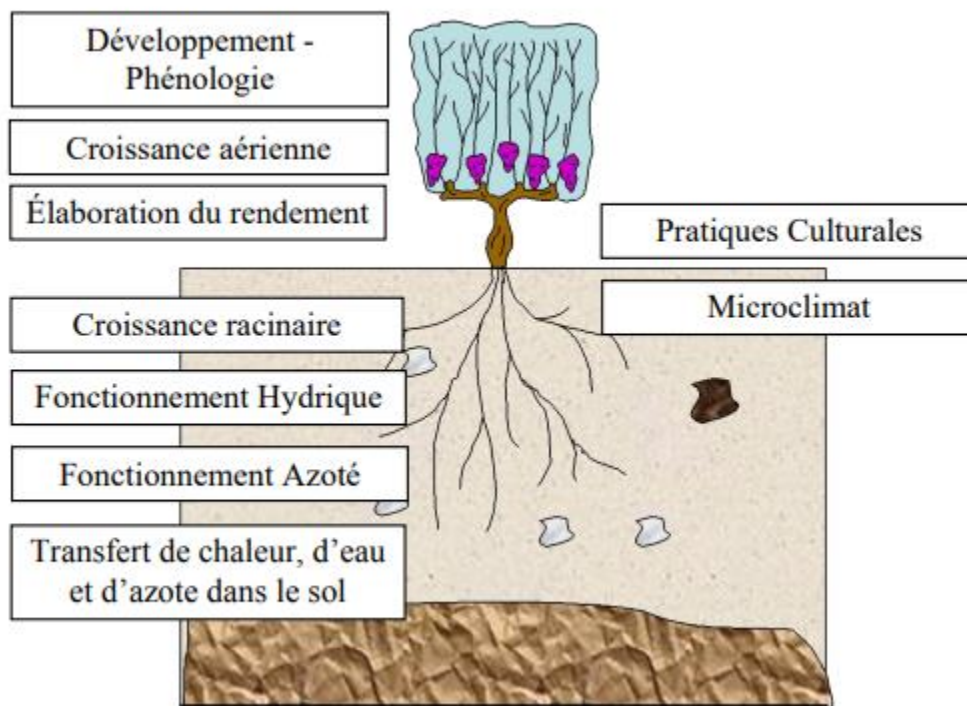


Figure 2.12 : Schéma de l'organisation modulaire du modèle de culture STICS (adapté de GARCIA DE CORTAZAR ATAURI,2006).

2.3.2. Le développement phénologique

Le module associé au développement phénologique simule l'évolution des phases phénologiques des plantes en distinguant deux stades de développement : le stade végétatif et le stade reproductif. Ces deux phases sont simulées de manière indépendante et varient d'une espèce à l'autre. Le passage du

stade végétatif au stade reproductif est conditionné par les durées de deux stades (végétatif et reproductif) qui varient en fonction de l'espèce et la variété. Leur calcul se repose principalement sur la température de la culture et le développement peut être ralenti en présence de conditions défavorables à savoir les stress hydrique et azoté et la photopériode.

2.3.3. La croissance aérienne

Pour les cultures annuelles, la croissance aérienne est basée sur le calcul de l'évolution de l'indice de surface foliaire (LAI) et la biomasse aérienne. L'indice de surface foliaire est dépendant du stade de développement phénologique de la plante, de la température et des stress hydrique et azoté. La croissance en biomasse aérienne est appréhendée à partir de l'efficacité de conversion du rayonnement intercepté. Cette dernière est aussi liée aux phases de développement de la plante et aux conditions de stress endogène et exogène.

2.3.4. L'élaboration du rendement

Le rendement s'intéresse à la quantité et la qualité des organes récoltés en fonction de certains critères tels que : la teneur en eau, la teneur en sucre et la maturité physiologique. Pour les cultures annuelles, l'élaboration du rendement est estimée selon l'indice de récolte dynamique où le nombre et le remplissage des organes récoltés n'a pas d'effet sur le développement végétatif de la plante.

2.3.5. La croissance racinaire

Les racines jouent un rôle essentiel dans l'absorption de l'eau et de l'azote minéral. La croissance racinaire est estimée par la croissance du front racinaire ainsi que la densité racinaire à partir de certaines caractéristiques liées à la température du sol, l'humidité du sol et sa densité apparente. La croissance racinaire caractérise le développement de la partie souterraine de la plante. Cette croissance s'arrête à une phase de développement spécifique de la plante, ou lorsque les racines rencontrent un obstacle de nature physique (roche mère, horizon très compact), ou de nature chimique (conditions d'anoxie).

La distribution racinaire se fait par couche élémentaire de 1 cm en fonction de la proportion des racines présentes dans chaque horizon et des contraintes liées au sol.

2.3.6. Le fonctionnement hydrique

Le fonctionnement hydrique des sols repose sur le calcul du bilan d'eau en fonction des besoins en eau des plantes, à différents stades phénologiques, et la quantité d'eau réellement disponible dans la matrice du sol. Un déséquilibre dans le bilan hydrique se manifeste par une insatisfaction des besoins hydriques des plantes. Cela déclenche un stress hydrique chez la plante et limite ainsi sa croissance et le développement de sa partie aérienne. La quantité d'eau disponible pour la plante provient des apports d'eau par les précipitations, l'irrigation et aussi les remontées capillaires. Les pertes en eau correspondent principalement aux pertes en eau par évaporation du sol, ruissellement et drainage. Les besoins en eau des plantes et l'évaporation directe sont estimées dans le modèle STICS par une approche dite "résistive", bien détaillée par Brisson et al. (1998). Quant à la transpiration, elle est spécifique à chaque plante et dépend en grande partie de la teneur en eau du sol et de ses phases de développement physiologique.

2.3.7. La dynamique de l'azote

Le module de fonctionnement azoté repose sur le bilan d'azote du système sol-plante. Ce bilan fait intervenir d'une part les apports d'azote dans le système via les apports d'engrais et la minéralisation de la matière organique et d'autre part les processus de volatilisation, de dénitrification et de lixiviation. Les besoins des plantes en azote sont définis par la courbe maximale de dilution d'azote. Ce module permet de calculer les différents indices de stress azoté et pour cette raison il mobilise le concept de courbe critique de dilution (Lemaire et al., 1997).

2.3.8. Transferts de chaleur, d'eau et de nitrates

Le module de transfert de chaleur, d'eau et de nitrates permet de simuler la température du sol à partir de la température à la surface calculée dans le module "Microclimat" et d'une conductivité thermique liée au milieu, et les quantités d'eau et d'azote dans le sol.

Dans le cas d'un transfert d'eau, le sol est représenté par une succession de réservoirs qui gèrent le stock d'eau dans le sol. Ce dernier est caractérisé par deux réservoirs distincts (la microporosité et la macroporosité). Le transfert de l'eau dans la microporosité du sol est calculé par couche élémentaire de 1 cm. Pour chaque couche de sol, l'eau s'infiltré pour la remplir, jusqu'à atteindre sa teneur en eau à la capacité au champ. L'eau excédentaire rejoint la couche sous-jacente jusqu'à atteindre la dernière couche, ou l'eau en excès est évacuée à la base du profil de sol par drainage (Figure 2.13). Comme le transfert de l'eau, le transfert de nitrate est également modélisé selon l'analogie de type réservoir. Le taux de nitrate présent dans une couche élémentaire, se mélange avec l'eau qui la rejoint. L'eau excédentaire, avec la nouvelle concentration en nitrate quitte la couche élémentaire pour rejoindre la couche sous-jacente, et ainsi de suite (Figure 2.13).

En synthèse, le module de circulation d'eau et de nitrate tient compte, pour cinq couches de sol au maximum, de caractéristiques hydrodynamiques : la teneur en eau à la capacité au champ, la teneur en eau au point de flétrissement permanent mais aussi de la densité apparente de la terre fine et la charge en éléments grossiers (Figure 2.13).

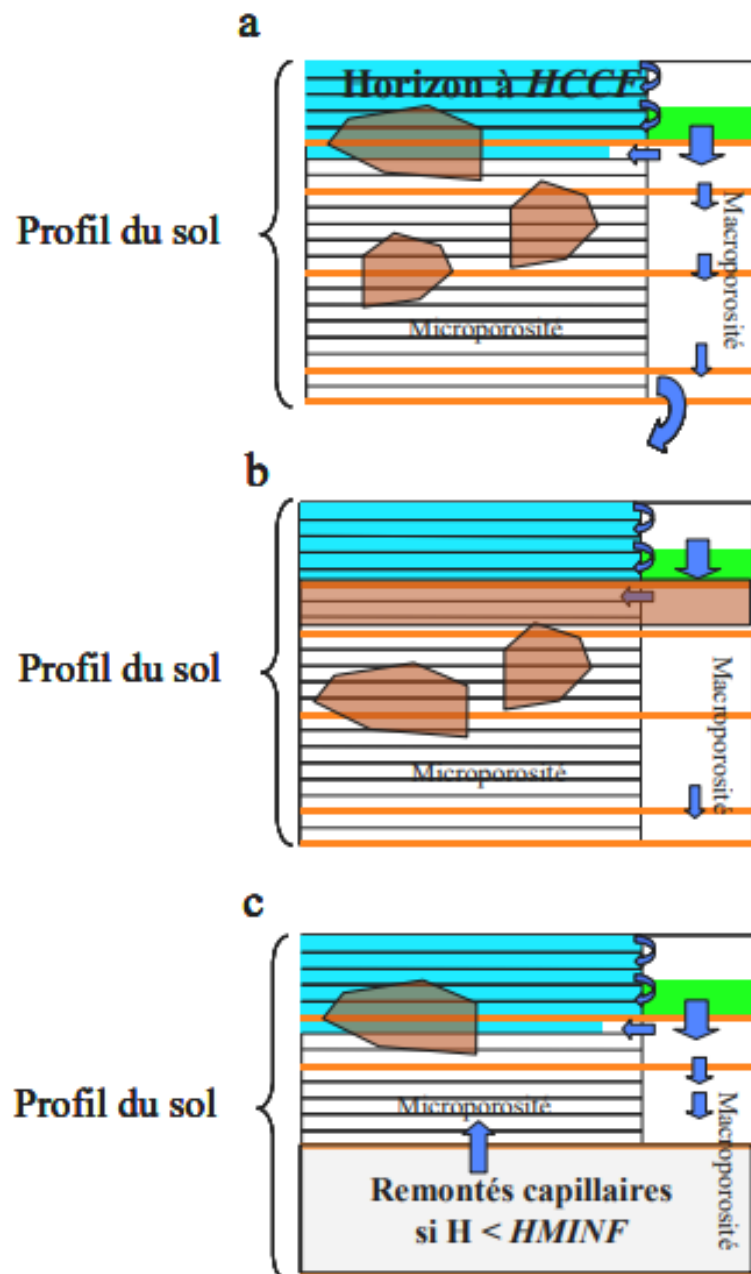


Figure 2.13 : Conception des différents compartiments du sol selon le module de transfert de STICS. a) Description de la circulation de l'eau en sol profond et filtrant ; b) description de la circulation de l'eau en sol profond et tassé ; c) description de l'eau en sol superficiel sur substrat poreux. HCCF est l'humidité à la capacité au champ de chaque couche; HMINF l'humidité au point de flétrissement permanent de chaque couche ; et H est l'humidité pondérale de l'horizon. (adapté de GARCIA DE CORTAZAR ATAURI,2006).

2.3.9. Gestion des pratiques culturales

Les techniques culturales représentent une composante essentielle dans la modélisation STICS. Elles sont associées au pilotage d'irrigation, à la gestion du plan de fertilisation minérale et organique et à la restitution des résidus organiques dans le sol. Plus encore, les techniques culturales caractérisent les conditions physiques de surface du sol en termes de ruissellement, de croûte de battance, de présence d'un mulch synthétique ou végétal. Ce module tient compte des techniques qui ont un effet considérable sur la culture telles que la date de semis, la densité de semis et la date de récolte définie en fonction de la phase de maturité physiologique, de la teneur en sucre et de la teneur en eau de la culture.

2.3.10. Microclimat

Le module Microclimat permet de contrôler plusieurs processus liés au couvert végétal et au fonctionnement du sol. Ce module conditionne plusieurs processus à savoir le développement phénologique, l'évapotranspiration, la photosynthèse, la dynamique de la matière organique dans le sol, la germination et la levée des plantes. D'un autre côté, grâce au bilan énergétique journalier, STICS via son module du microclimat permet de calculer l'humidité de l'air mais aussi la température des cultures. Cette dernière mobilise plusieurs variables d'état dont principalement le rayonnement net journalier et l'évaporation des plantes.

2.3.11. Les services écosystémiques instruits et le choix des indicateurs

Les indicateurs de services écosystémiques ont été évalués à partir de variables issues de la modélisation STICS du système sol-plante-climat. Ces indicateurs sont liés à plusieurs services écosystémiques qui sont apparus comme intéressants à étudier dans le contexte agricole breton.

Ces six indicateurs ont été choisis pour leur pertinence spécifique au contexte agricole breton orienté vers la polyculture et l'élevage intensif et pour leur complémentarité. Certains de ces services écosystémiques sont emblématiques, tels que celui de la régulation du climat, ou encore le service de régulation de la qualité de l'eau dont les enjeux sont à la fois connus de la communauté scientifique et du grand public. Les autres services écosystémiques, bien que moins emblématiques, présentent

de forts enjeux dans le contexte agro-environnemental local. Finalement, le choix des indicateurs de services écosystémiques a également été déterminé vis-à-vis de la littérature scientifique et des données disponibles.

- La régulation du climat est définie par la capacité des sols à séquestrer du carbone organique. Pour quantifier le service en question, la variation annuelle du stock de carbone dans le sol en considérant une profondeur de référence de 30 cm, a été utilisée comme indicateur.
- La régulation de la qualité de l'eau. Ce service est défini par la capacité du sol à retenir l'azote dans sa porosité matricielle. L'indicateur choisi correspond au ratio entre la quantité totale d'azote dans le système (minéralisation + apport d'azote minéral) et la quantité d'azote lixivié par drainage vers la nappe souterraine à l'échelle une année calendaire. Cet indicateur caractérise donc le niveau de fourniture actuel du service en tenant compte des apports d'intrants réalisés.
- La fourniture d'azote à la plante. Ce service est déterminé en fonction de la quantité d'azote disponible pour la plante dans le sol et la quantité d'azote réellement utilisée par la plante. L'indicateur choisi correspond à un ratio qui caractérise la capacité du sol à couvrir les besoins azotés des plantes pendant une année calendaire sans aucune fertilisation minérale. Il ne tient compte que des apports d'azote inhérents au système via le processus de minéralisation de la matière organique endogène au système.
- Stockage et restitution de l'eau aux plantes cultivées. Ce service est appréhendé via un indicateur exprimant la quantité d'eau effective utilisée par la plante durant une année calendaire. Il caractérise la quantité d'eau restituée à la plante sous forme d'eau verte et correspond à la quantité d'eau annuelle transpirée par la culture cultivée.
- Production de la biomasse végétale. Ce service est estimé via le rendement annuel de cultures de rente. La masse sèche de cultures à 0% d'humidité issue de la modélisation STICS, est convertie en énergie équivalente pour harmoniser les rendements de différentes cultures et faciliter ainsi la comparaison annuelle de ces rendements.
- Stockage et restitution de l'eau bleue. L'indicateur retenu correspond à la quantité d'eau annuelle percolée à la base du profil de sol. Cet indicateur caractérise le rendement en eau

annuel, estimé en utilisant les différentes composantes du bilan hydrique qui décrit les échanges sol-eau.

Conclusion partielle

La deuxième partie du chapitre matériels et méthodes a porté essentiellement sur le modèle STICS, utilisé pour caractériser l'évolution dynamique des processus des sols et les indicateurs biophysiques considérés pour appréhender les six services écosystémiques instruits. Une présentation détaillée des indicateurs retenus en fonction des variables de sortie du modèle de culture STICS, de l'itinéraire technique et des données climatiques, est faite dans le chapitre 6 de la thèse.

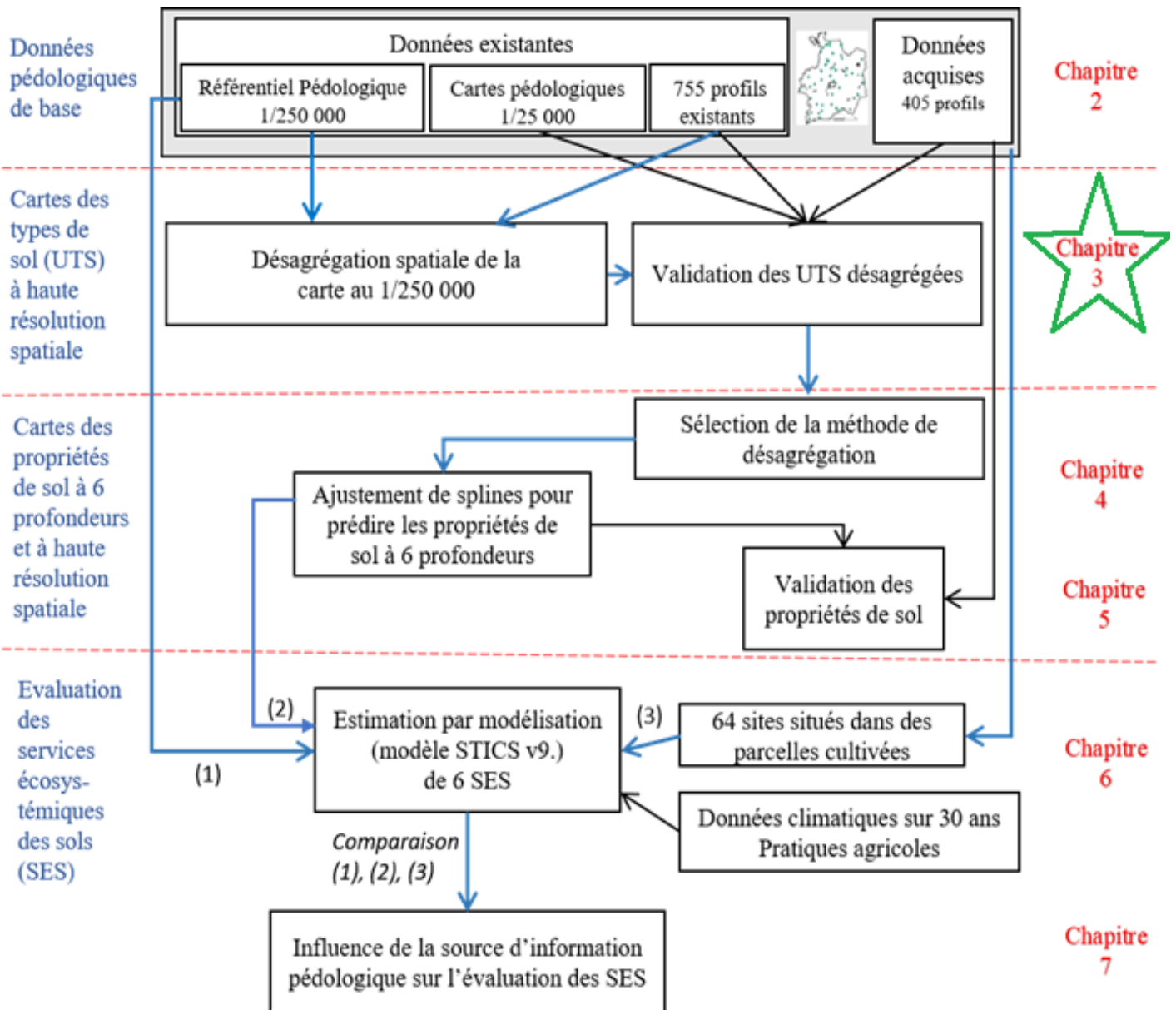


Figure 3.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 3 :

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

Introduction du chapitre 3

Situés à l'interface de plusieurs compartiments de l'environnement, les sols ont une importance cruciale pour répondre aux enjeux environnementaux actuels, notamment en termes de sécurité alimentaire et de lutte contre le changement climatique. Ainsi, la connaissance de la distribution spatiale des sols sur de larges étendues spatiales est nécessaire pour gérer durablement les écosystèmes naturels ou anthropisés et proposer des plans d'aménagement adaptés aux caractéristiques pédo-environnementales du milieu. Par exemple, une bonne gestion des plans de fertilisation, en amont des cours d'eau, selon les aptitudes culturales des sols, permet de réduire l'occurrence du phénomène d'eutrophisation (Smith et Schindler, 2009). Cependant, la faible disponibilité d'une information spatialisée et précise sur les sols constitue un verrou de la considération formelle des sols dans des stratégies de gestion agro-environnementale. En effet, aux larges étendues spatiales, l'information pédologique est généralement disponible d'une façon exhaustive, mais à une faible résolution spatiale.

Dans la région Bretagne, la caractérisation de la distribution spatiale des sols a été réalisée par l'intermédiaire du programme national IGCS (Inventaire, Gestion et Conservation des Sols) et notamment son volet Référentiel Régional Pédologique (RRP). Les sols bretons sont donc cartographiés à une échelle spatiale de 1/ 250 000^{ème} et identifiés en unités de pédopaysages dites Unités cartographiques de sol (UCS), spatialement délimitées. Chaque UCS est composée d'unités typologiques des sols (UTS) non cartographiées dans l'espace, mais dont les proportions sont explicitement données. Une voie prometteuse pour mieux améliorer la représentation de la distribution spatiale des sols consiste à cartographier les UTS en précisant leurs limites spatiales. Dans cet objectif, la désagrégation spatiale semble être une technique efficace pour améliorer la distribution spatiale des sols (Odgers et al., 2014). Cette approche de « descente d'échelle » a été largement utilisée dans la bibliographie internationale pour répondre à la nécessité d'une caractérisation fine des sols aux larges étendues spatiales (Odgers et al., 2014 ; Holmes et al., 2015 ; Chaney et al., 2016 ; Vincent et al., 2018 ; Chen et al., 2018 ; Møller et al., 2019).

L'objectif de ce chapitre vise à vérifier l'hypothèse que l'implémentation de règles sur les lois de distribution des sols dans la procédure de désagrégation améliore la précision des cartes pédologiques

désagrégées produites. Cela suppose d'évaluer l'apport des relations sol-paysages implémentées dans l'algorithme DSMART par Vincent et al. (2018). Pour répondre à cet objectif, nous avons comparé des cartes de sol désagrégées dérivées de l'algorithme DSMART classique, de l'algorithme DSMART avec des observations de sol supplémentaires et de l'algorithme DSMART alimenté par les relations sol-paysages dans le département d'Ille-et-Vilaine, qui représente sur une superficie de 6 775 km² à l'est de la Bretagne, en France, en mobilisant les données du RRP de Bretagne.

Ce chapitre est le fruit d'une collaboration avec Budiman Minasny et Brendan Malone menée à l'occasion d'un séjour de mobilité de la doctorante à l'université de Sydney dans le cadre du parcours d'excellence d'Agreenium entre Mars et Juin 2018. Les résultats de ce chapitre ont fait l'objet d'un article scientifique soumis à la revue SOIL de l'EGU.

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

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 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

3.1. 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; 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; Lacoste et al., 2011; Lemercier et al., 2012; Nauman and Thompson, 2014; Holmes et al., 2015; Vaysse and Lagacherie, 2015; Ellili 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.

3.2. Materials and methods

3.2.1. 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) (Figure 3.1). 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 except in the western part of the department where it talls 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).

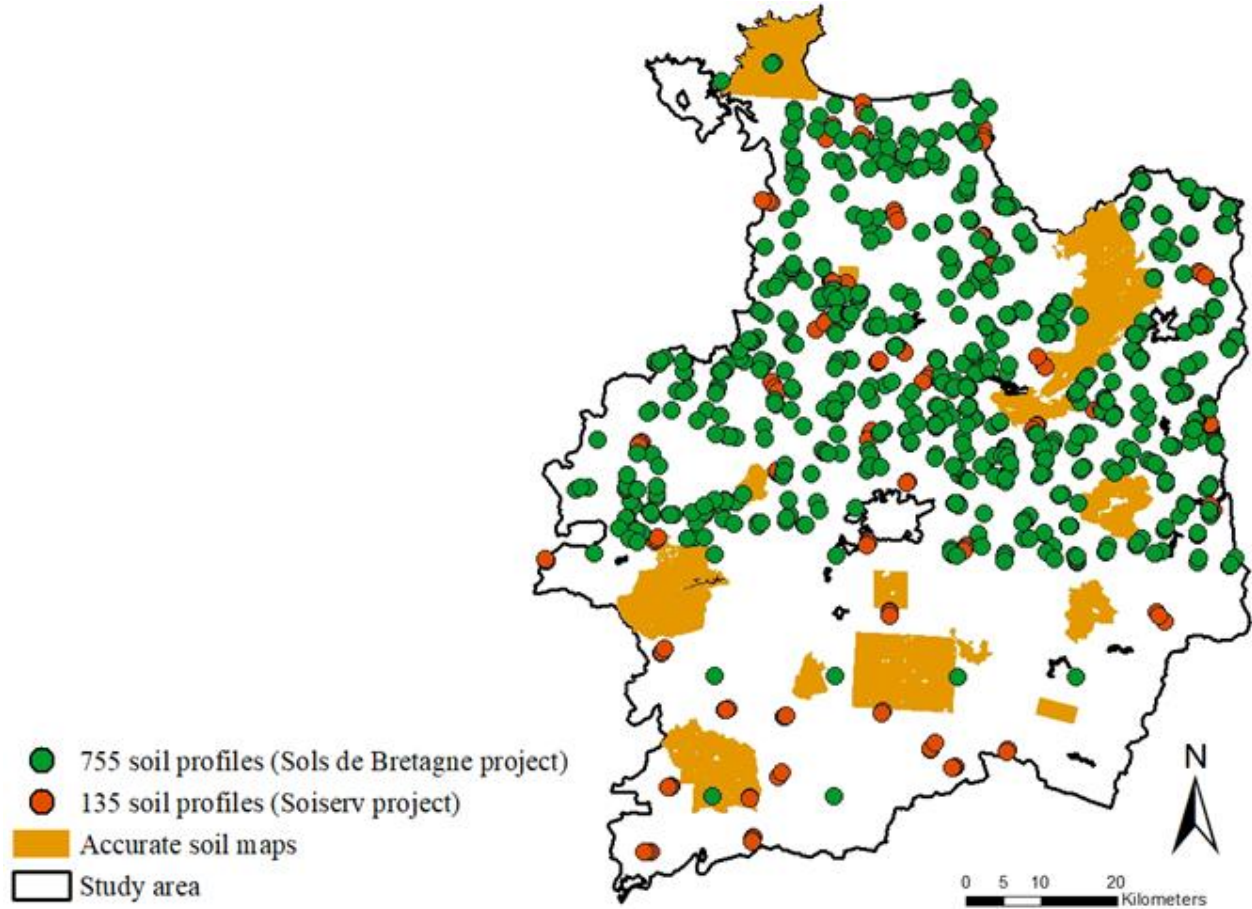


Figure 3.1 : Location of the study area and the validation datasets.

3.2.2. Soil data

3.2.2.1. 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).

3.2.2.2. Soil validation data

To assess the quality of disaggregated soil maps, three validation datasets were used (Figure. 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 et al., 2019, submitted).
- 755 legacy soil profiles collected between 2005 and 2008 during the “Sols de Bretagne” programme (INRA Infosol, 2014). These profiles were sampled to characterize hydromorphic soil conditions and soil landscape heterogeneity.

Existing detailed soil maps (1:25,000) covering 87,150 ha, 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 50 m spatial resolution.

3.2.3. 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 3.1) were considered as covariates in the disaggregation process at a 50 m spatial resolution. Terrain attributes included elevation, slope, Compound Topographic Index (CTI) (Beven and Kirkby, 1979, Merot et al., 1995) and Topographic Position Index (TPI) (Vincent et al., 2018) that together were derived from a 50 m resolution Digital Elevation Model (IGN, 2008). These attributes were computed using ArcGIS 10.1 (ESRI, 2002) and MNT surf software (Squidant, 1994).

Environmental attributes describing soil parent material (Lacoste et al., 2011) and hydromorphic soil conditions via waterlogging index (Lemercier et al., 2013) 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.

Table 3.1 : 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	

3.2.4. Disaggregation procedure: DSMART algorithm

3.2.4.1. 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 (Figure 3.2):

- 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.

3.2.4.2. 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. The model was then extrapolated over the study area.

3.2.4.3. 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, hydromorphic 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.

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 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 unique calibration dataset to build decision trees.

3.2.4.4. 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 (1st STU, 2^{sd} STU, and 3rd 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.3.1:

$$CI = 1 - (P_{1^{st} \text{ STU}} - P_{2^{sd} \text{ STU}}) \quad (\text{Eq 3.1})$$

Where $P_{1^{st} \text{ STU}}$ and $P_{2^{sd} \text{ STU}}$ denote respectively the highest probability of occurrence for 1st 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 1st 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.

3.2.5. 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.

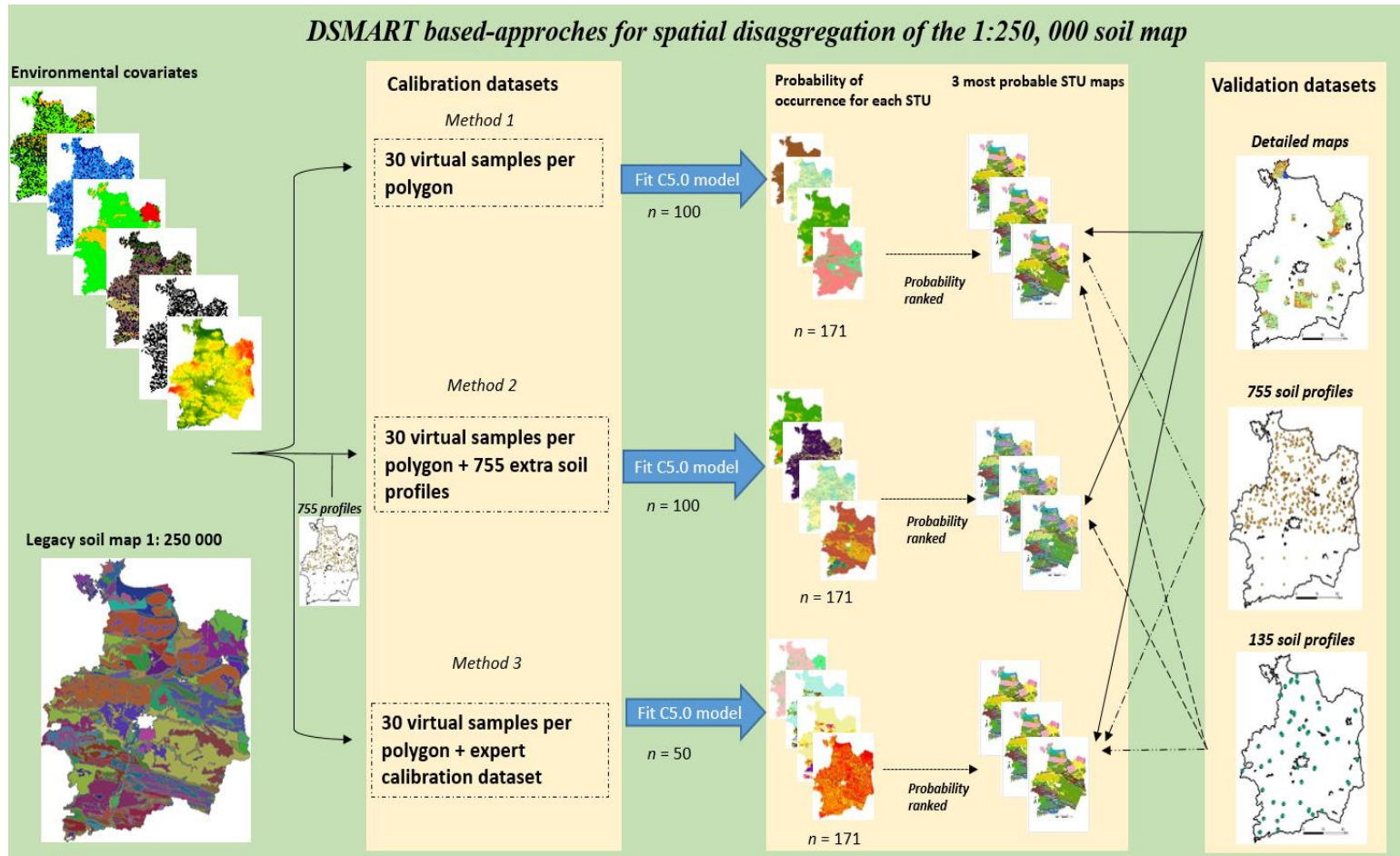


Figure 3.2 : Schematic of the DSMART based approaches algorithm. The steps in DSMART are: 1) construct the calibration dataset; 2) train C5.0 model; 3) estimate STU maps and their associated probabilities of occurrence.

3.2.6. 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 $A_j = \sum_{i=1}^n a_{ij}$ 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 3.2). *Completeness* measures how well zones from the second map fit inside regions from the first map (Eq 3.5). The final value of *V measure* is calculated as the weighted harmonic mean of homogeneity and completeness (Eq 3.8). 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 Stepinski, 2018). These indicators qualify local associations by highlighting the region's inhomogeneities (Eq 3.3, Eq 3.4), or zone's inhomogeneities (Eq 3.6, Eq 3.7). 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 (Eq 3.3) 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

3.3) 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 2.3.

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 3.4 and Eq 3.7), and a global indicator C (Eq 3.5) measures a homogeneity of a given region in terms of zones.

$$h = 1 - \sum_{j=1}^m \left(\frac{A_j}{A}\right) \left(\frac{\text{Variance of regions in zone } j=S_j^R}{\text{Variance of regions in the domain}=S^R}\right) \quad (\text{Eq 3.2})$$

$$S_j^R = - \sum_{i=1}^n \left(\frac{a_{ij}}{A_j}\right) \log \left(\frac{a_{ij}}{A_j}\right) \quad (\text{Eq 3.3})$$

$$S^R = - \sum_{i=1}^n \left(\frac{A_i}{A}\right) \log \left(\frac{A_i}{A}\right) \quad (\text{Eq 3.4})$$

$$c = 1 - \sum_{i=1}^n \left(\frac{A_i}{A}\right) \left(\frac{\text{Variance of zones in region } i=S_i^Z}{\text{Variance of zones in the domain}=S^Z}\right) \quad (\text{Eq 3.5})$$

$$S_i^Z = - \sum_{j=1}^m \left(\frac{a_{ij}}{A_i}\right) \log \left(\frac{a_{ij}}{A_i}\right) \quad (\text{Eq 3.6})$$

$$S^Z = - \sum_{j=1}^m \left(\frac{A_j}{A}\right) \log \left(\frac{A_j}{A}\right) \quad (\text{Eq 3.7})$$

$$V_\beta = \frac{(1+\beta)hc}{(\beta h)+c} \quad (\text{Eq 3.8})$$

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.

3.3. Results

3.3.1. 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 3.3 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.

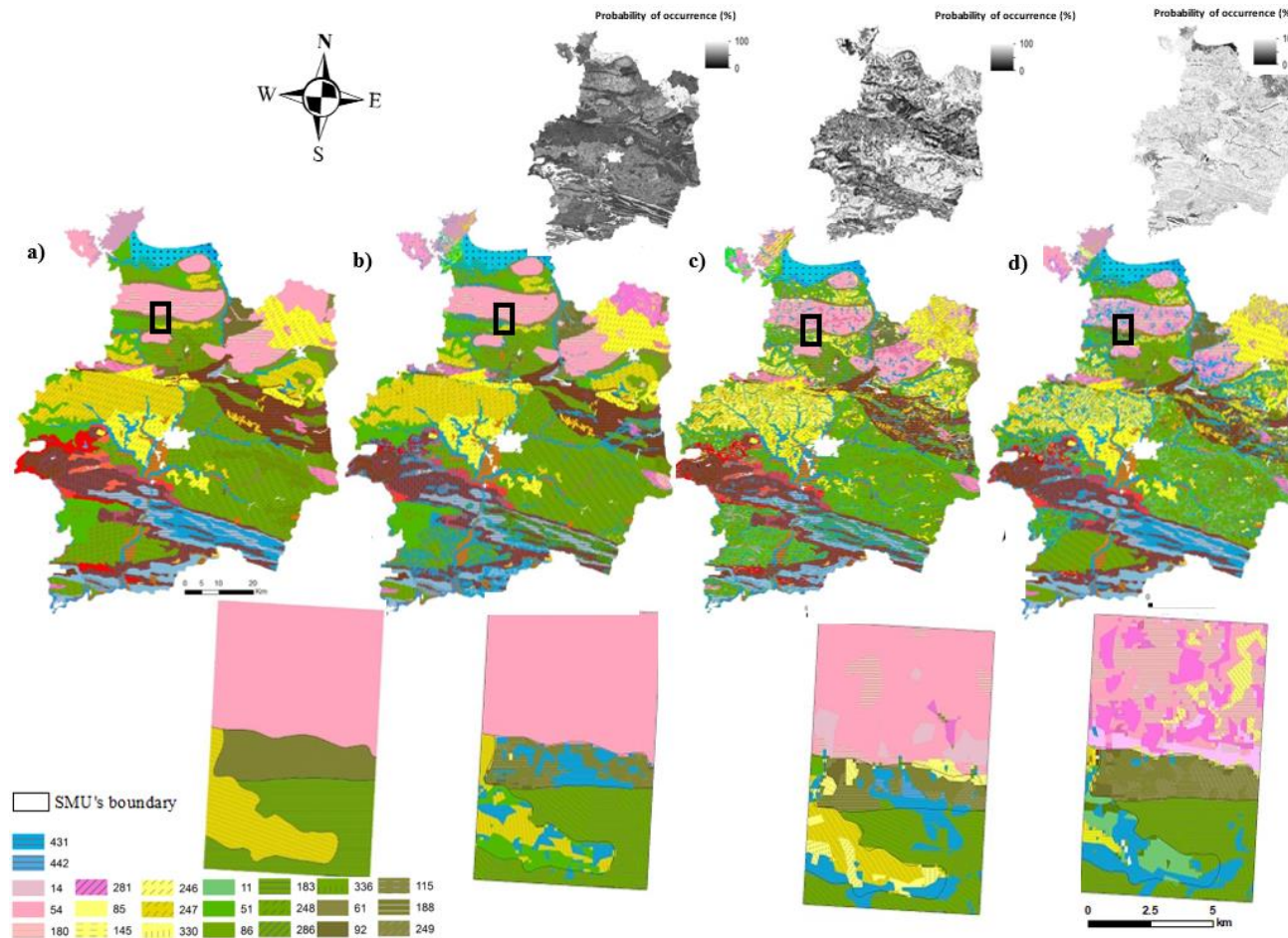


Figure 3.3 : Digital soil map of the most probable STU and their associated probability of occurrence for the whole study area and for a focus zone, a) Legacy soil map: most probable STU for each SMU, b) original DSMART approach; c) DSMART with expert rules; d) DSM.

Figure 3.4 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 (Figure 3.5). 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, 0.68 for DMSART with expert rules and 0.28 for DSMART with extra soil observations maps. Meanwhile, the average confusion index reached 0.8 for the original DSMART approach while DSMART with extra soil observations and DSMART with expert rules 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.

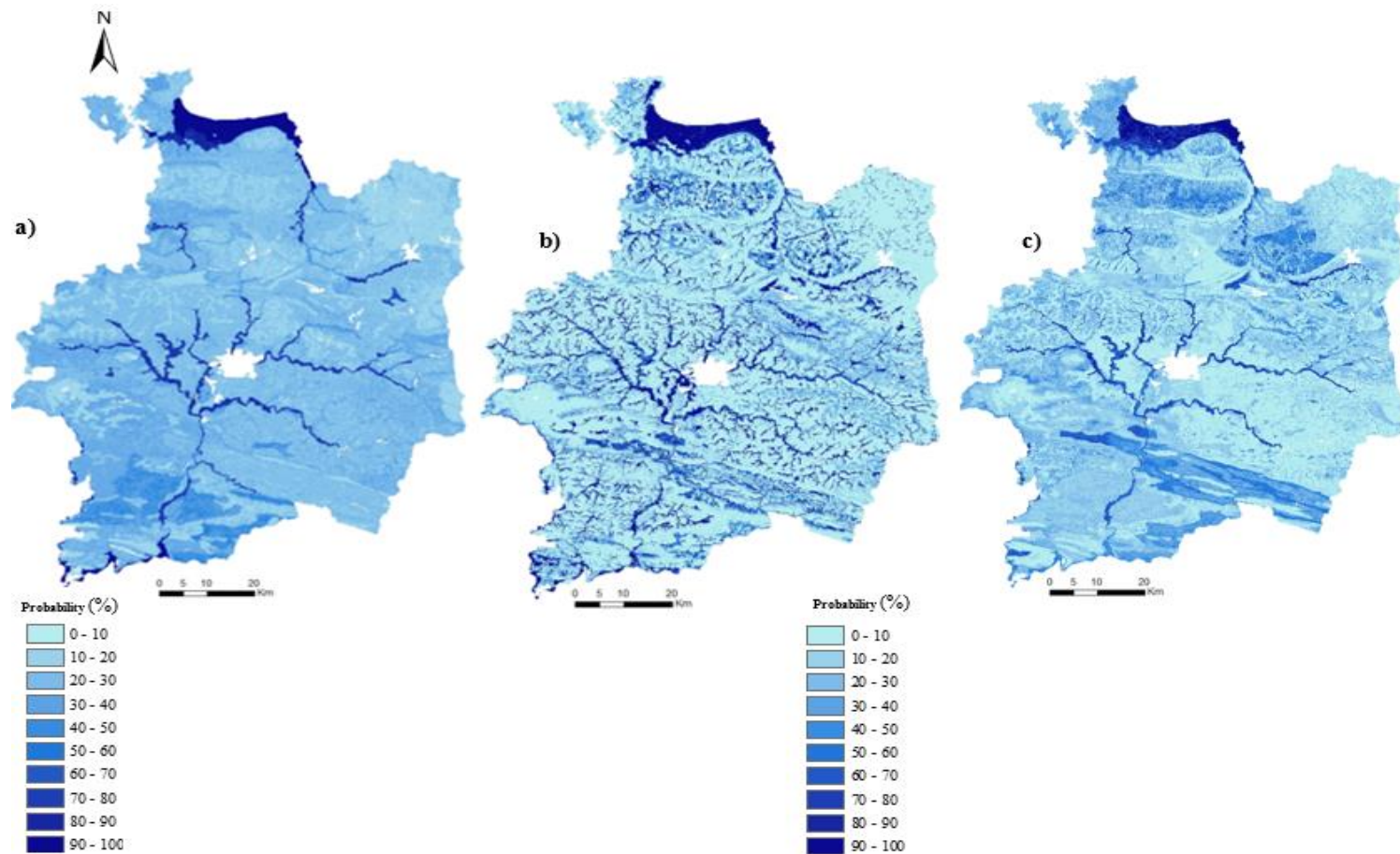


Figure 3.4 : Global probability of hydromorphic soils over the study area derived from a) original DSMART, b) DSMART with soil landscape relationships and c) DSMART with extra soil observations. The probabilities of the three STU with highest prediction occurrence are summed if they are hydromorphic.

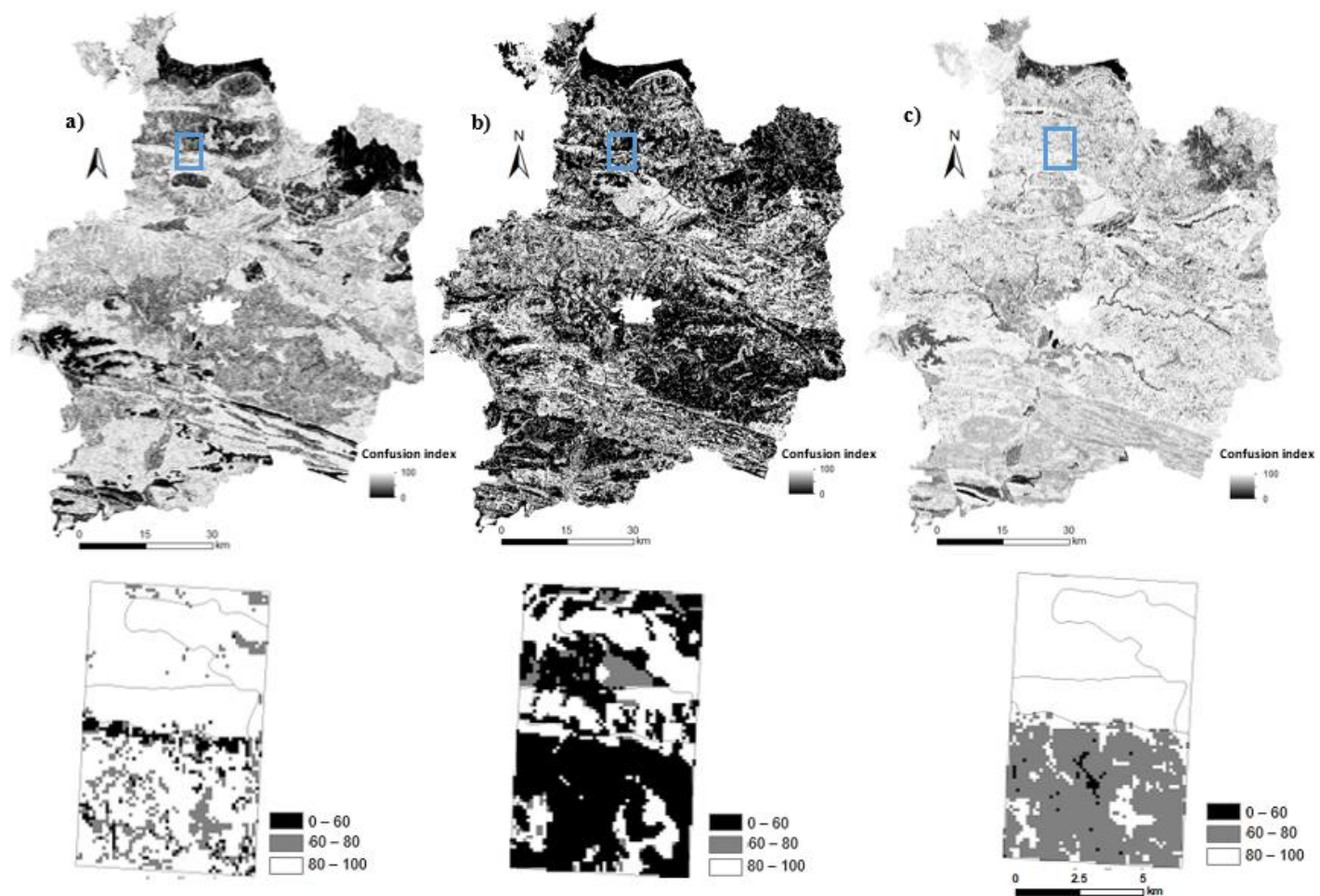


Figure 3.5 : Confusion index maps for a) Classic DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations.

Figure 3.6 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.

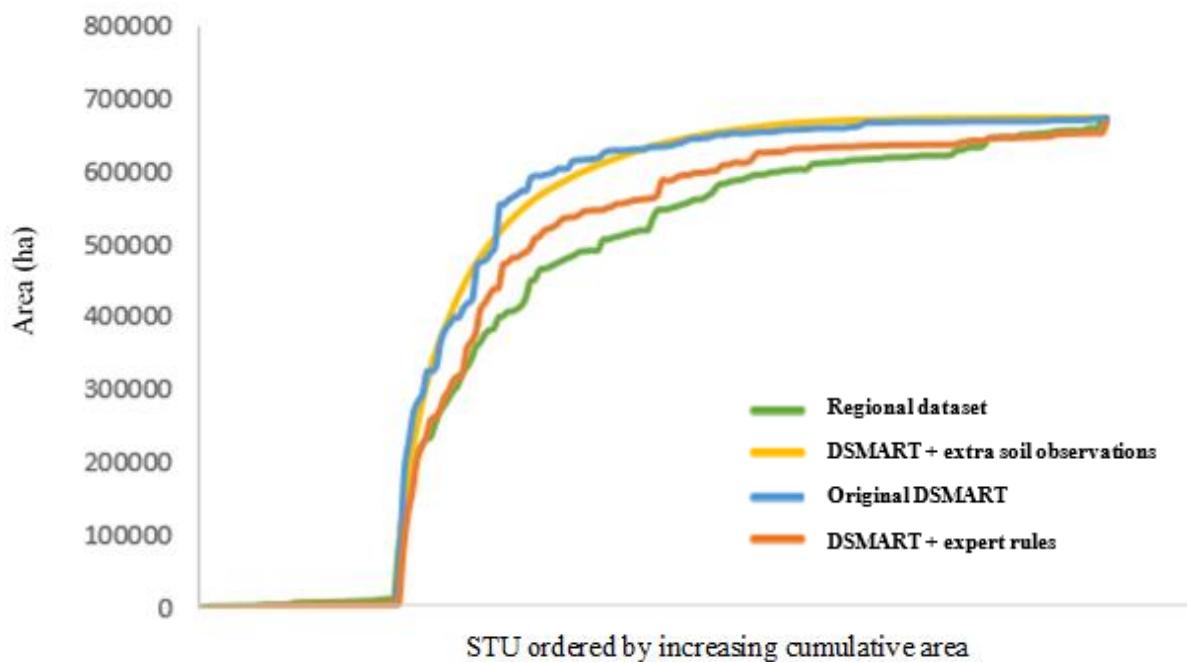


Figure 3.6 : Cumulative area of the 171 STUs estimated from the regional soil database and predicted by different DSMART based approaches.

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 3.2).

Table 3.2 : 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 (km ²)	Rank	Predicted area (km ²)	Rank	Predicted area (km ²)	Rank	Predicted area (km ²)
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

3.3.2. Covariates importance in the decision trees

Figure 3.7 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 (Figure 3.7a), 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 (Figure 3.7b) 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 (Figure 3.7c) 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.

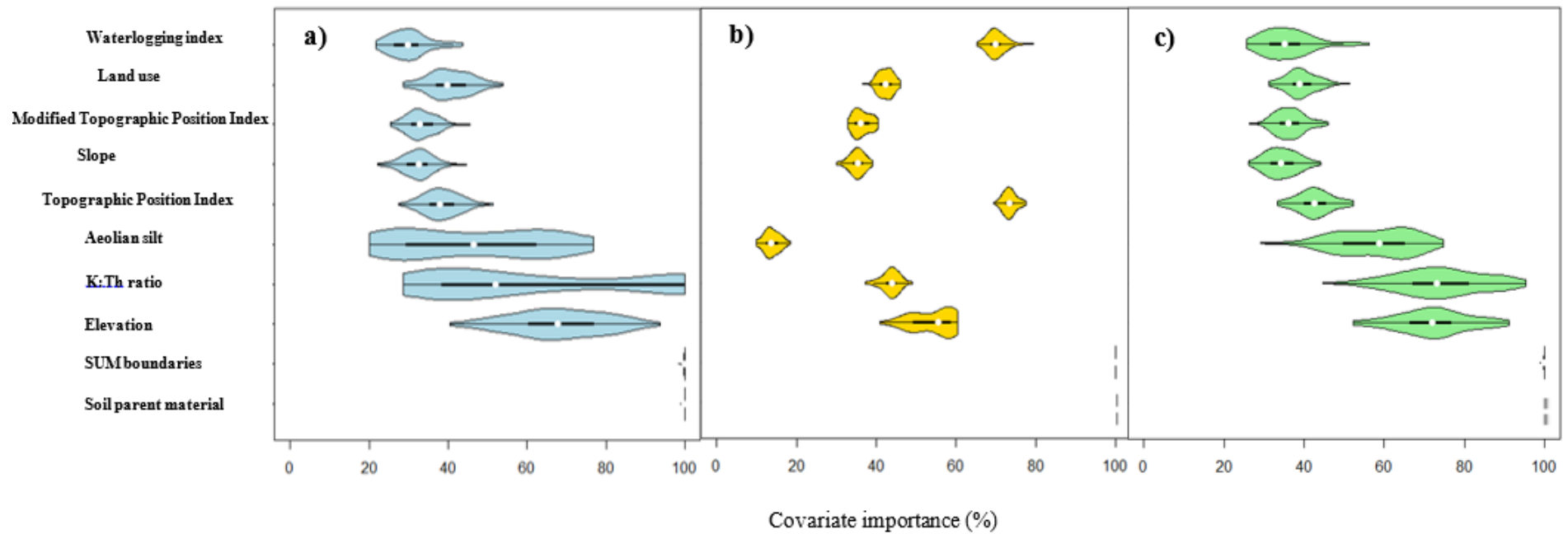


Figure 3.7 : Violin plots of the relative importance of each environmental covariate used in a) Original DSMART approach; b) DSMART with expert rules; c) DSMART with extra soil observations.

3.3.3. 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 3.3). 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 3.3). Using a 3 x 3 window of pixels markedly improves the global accuracies, which increased for the two validation datasets (Table 3.3). 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 3.3). 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 3.3).

Table 3.3 : 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				

3.3.4. Comparing disaggregated maps

Figure 3.8 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 $h_1 = 0.49$, $c_l = 0.58$ and $V_l = 0.53$ 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 Figure 3.8b map seemed to be more homogeneous than the map Figure 3.8a 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_l score. The two inhomogeneity maps (Figures 3.8a and 3.8b) 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 .

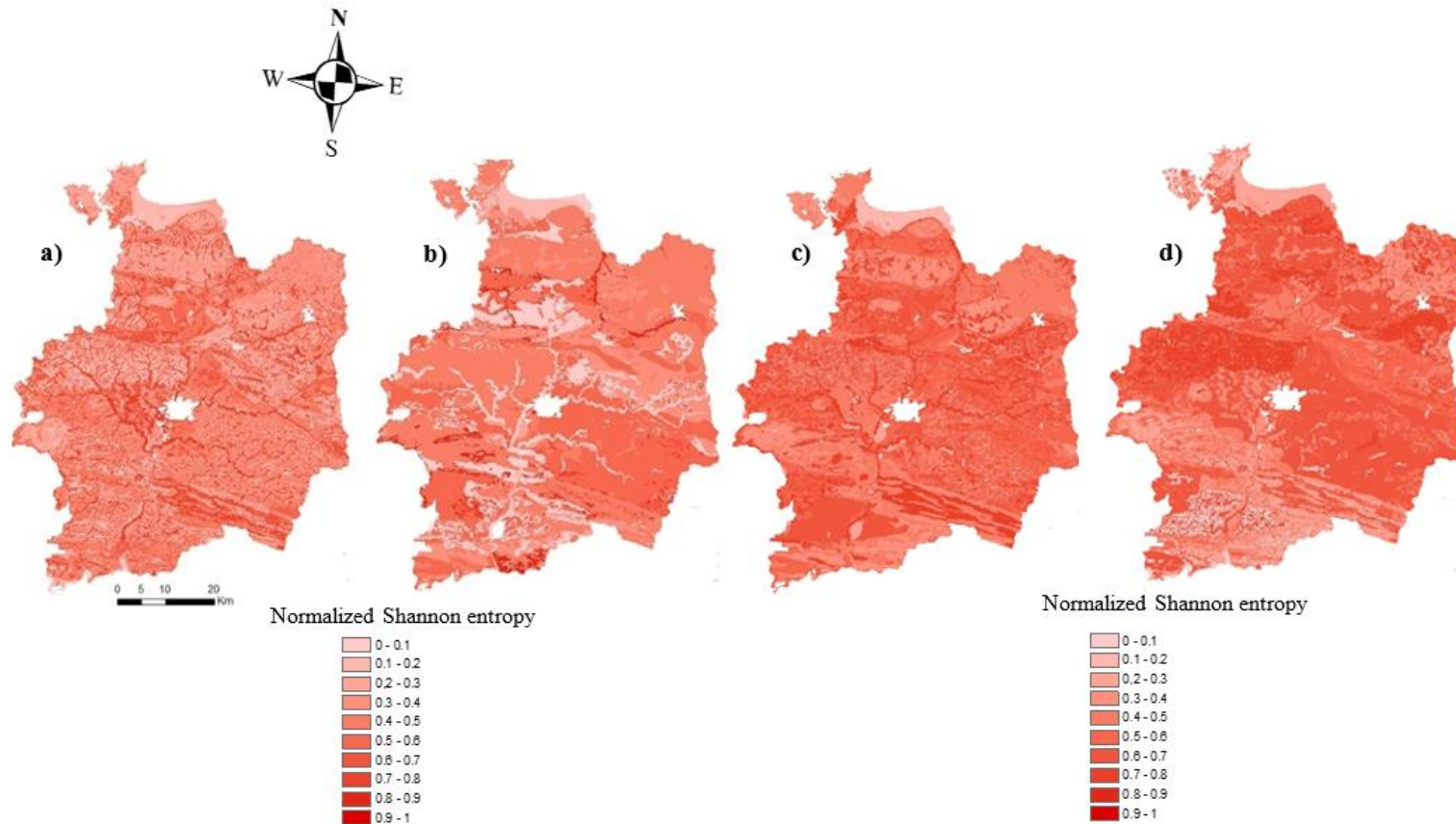


Figure 3.8 : Spatial association between disaggregated maps of Ile-et-Vilaine department. a) map of inhomogeneity of DSMART with soil landscape relationships map in terms of original DSMART map b) map of inhomogeneity of original DSMART map in terms of DSMART with soil landscape relationships map c) map of inhomogeneity of DSMART with extra soil observations 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.

3.4. Discussion

3.4.1. Performance of the disaggregation procedures

Produced disaggregated soil maps closely resemble the abundant soils in the original soil map (Holmes et al., 2015; Figure 3.3). 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 3.4). 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.

Table 3.4 : 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 (km ²)	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

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.

3.4.2. 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).

3.4.3. 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.

3.4.4. 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.

3.4.5. 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.

3.5. 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).

Conclusion du chapitre 3

Le chapitre 3 pose le cadre méthodologique et opérationnel adopté pour améliorer la caractérisation spatiale des sols bretons en désagrégeant les unités pédo-paysagères du département d'Ille-et-Vilaine à une résolution spatiale de 50 m. Pour cela, nous avons mobilisé trois versions de l'algorithme DSMART pour affiner la caractérisation spatiale des sols. La première correspond à la version originale de l'algorithme DSMART développé par Odgers et al. (2014). La deuxième est une version modifiée en ajoutant des profils de sol supplémentaires au jeu de données de calibration du modèle. La dernière est une version modifiée de l'algorithme DSMART par Vincent et al. (2018) intégrant des règles *a priori* décrivant des lois de distribution des sols en Bretagne.

Bien que les cartes des sols issues des trois méthodes, montrent des différences visuelles nettes de la distribution spatiale des types de sol, les résultats de validation, qu'ils soient obtenus par comparaison à des données ponctuelles où à des cartes précises, montrent des performances assez similaires. Par exemple, en utilisant 135 profils de sol collectés selon un échantillonnage probabiliste et une approche de validation pixel à pixel, les performances de validation sont de l'ordre de 18,1% pour DSMART classique, 19,8% pour DSMART avec des règles sol-paysages décrivant les lois de distribution des sols en Bretagne et 16,9% pour DSMART avec 755 profils de sol supplémentaires. Cela est notamment expliqué par le nombre important d'UTS présentes sur le site d'étude (171 UTS). Certaines UTS présentent des processus pédogénétiques similaires, ce qui rend difficile leur distinction par de simples règles d'apprentissage. Plus encore, le résultat de validation à l'échelle de chaque pixel, s'est considérablement amélioré en utilisant l'approche de validation par voisinage (3x3) et encore plus en utilisant l'approche de validation sémantique (regroupement expert des 171 UTS en 78 groupes).

Dans l'ensemble, les cartes des types de sols, générées par les trois approches de désagrégation spatiale retranscrivent en grande partie l'organisation spatiale des sols en Ille-et-Vilaine. Elles permettent d'adresser la distribution spatiale des sols à l'échelle du département tout en délimitant spatialement les UTS. L'inspection visuelle des cartes des sols, issues des trois méthodes, montre clairement que la version originale de l'algorithme DSMART favorise la prédiction de l'UTS dominante de chaque unité pédopaysagère, ce qui conduit à lisser la distribution spatiale des sols. Au contraire, l'implémentation des règles expertes dans la procédure de désagrégation spatiale du RRP

de Bretagne permet de prédire des types de sols cohérents avec leurs positions le long des toposéquences et les processus pédologiques tels que les processus d'accumulation des alluvions le long des réseaux hydrographiques. Un constat similaire mais de moindre intensité est fait suite à l'ajout des profils supplémentaires des sols à la procédure de désagrégation spatiale.

La désagrégation spatiale du RRP mobilisant des règles *a priori* de distribution des sols spécifiques au contexte local conduit à des résultats cohérents avec les caractéristiques topographiques et géomorphologiques du secteur d'étude. Cette approche est donc retenue comme la plus pertinente pour améliorer la connaissance de la distribution spatiale des sols à large étendue spatiale où peu de données pédologiques sont existantes d'une manière exhaustive. Par ailleurs, nos résultats garantissant une certaine cohérence verticale et horizontale de couverture pédologique, méritent d'être mobilisés dans la prédiction des propriétés de sol. C'est pourquoi nous utilisons la carte des types de sols (UTS), obtenue selon la méthode de Vincent et al. (2018) de désagrégation spatiale, pour appréhender les distributions spatiales des propriétés de sol à l'échelle de département d'Ille-et-Vilaine dans le chapitre suivant (chapitre 4).

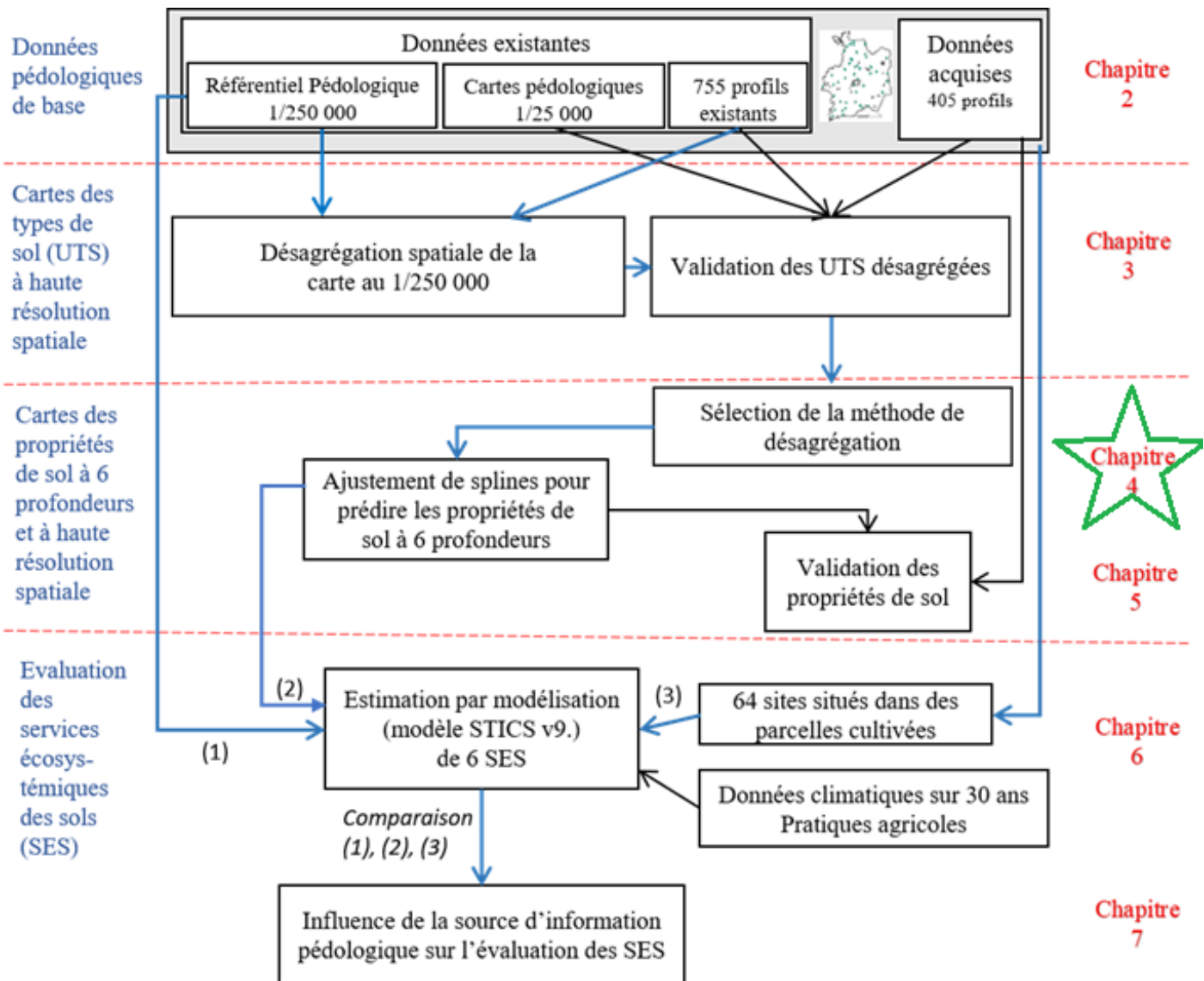


Figure 4.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 4 :
**Mapping soil properties at multiple
depths from disaggregated legacy soil
maps**

Introduction du chapitre 4

Après avoir comparé les méthodes de désagrégation spatiale, la version modifiée de l'algorithme DSMART par Vincent et al. (2018), intégrant des règles *a priori* de distribution des sols dans le paysage, a été retenue pour désagréger le RRP de Bretagne. Cette approche mobilise un large panel de variables pédo-environnementales et de données pédologiques anciennes (RRP de Bretagne). Pour affiner la distribution spatiale des sols sur une large étendue spatiale : Vincent et al. (2018) ont ainsi délimité spatialement les UTS tout en respectant l'occurrence et l'organisation spatiale des principaux types de sol spécifiques au contexte breton. Cependant, l'information spatialisée des types des sols ne permet pas d'accéder directement aux propriétés de base qui informent les caractéristiques des sols à différentes profondeurs : ces dernières peuvent être incorporées dans des outils d'aide la décision et des modèles de gestion agro-environnementale. D'autre part, pour répondre aux objectifs du projet mondial de cartographie des sols *GlobalSoilMap* (Arrouays et al., 2014), qui vise une cartographie précise des sols selon des spécifications bien déterminées, les propriétés des sols doivent être prédites à six profondeurs standard jusqu'à la profondeur d'apparition de la roche mère.

Ce chapitre est consacré au développement et la mise en place d'une méthodologie permettant de prédire les propriétés des sols aux intervalles requis par les spécifications *GlobalSoilMap*. Cette approche se base sur les cartes de types des sols désagrégées, obtenues dans le chapitre 3, et sur l'utilisation des régressions splines à conservation de masse pour harmoniser les prédictions pédologiques. Les fonctions splines sont mobilisées pour interpoler des propriétés de sol, au départ renseignées par horizons pédologiques d'épaisseurs variables, vers des couches de sol d'épaisseur standard et identiques pour tout point de l'espace.

La méthodologie implémentée permet de prédire simultanément plusieurs propriétés des sols tout en conservant les interrelations existantes entre propriétés des sols sur l'ensemble d'un profil de sol. Ce chapitre fait l'objet d'un article en voie de soumission dans la revue *Soil Research*.

Mapping soil properties at multiple depths from disaggregated legacy soil maps

Abstract

Predicting several soil properties simultaneously while preserving the relationships among them remains a great challenge in the digital soil-mapping (DSM) framework, particularly over large areas, where only legacy soil maps at coarse spatial resolution are available. Spatial disaggregation seems to be an effective technique to extract pedological information by downscaling the original information to produce soil maps at finer resolutions. In a previous study, legacy soil maps of Brittany (France) were disaggregated at a 50 m spatial resolution using the DSMART (Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees) algorithm and pedological knowledge. The present study had two main objectives: (i) assess the preservation of relationships among soil properties by applying the equal-area spline method, which estimates soil properties at standard depths, and (ii) combine disaggregated soil maps and spline-function results to estimate spatial patterns of nine soil properties for six regular soil-depth intervals down to 200 cm across Brittany, an area of 27 040 km². To this end, soil properties were first generated for standard soil-depth intervals using spline functions. The weighted mean of each soil attribute was then calculated for each grid cell based on reference soil-property values of the three most probable predicted soil types, and their associated probabilities of occurrence were used as weights. To assess the ability of spline functions to preserve soil-property relationships, multiple statistical analyses were performed using original and splined soil datasets. Bivariate and multivariate analysis highlighted that spline functions preserved soil-property relationships. Derived digital soil maps showed strong spatial patterns: SOC and silt contents generally decreased with depth, while sand content and coarse fragment percentage consistently increased with depth. In addition, experimental semivariogram analysis of SOC content showed high spatial variability over short distances for all soil-depth intervals except the deepest (100-200 cm), while silt content showed high semivariance for the deepest soil layers. This study can be considered an example of harmonisation to common output specifications, which generates a geo-database of quantitative soil properties that describe

lateral and vertical soil variation for regular depth intervals. These predictions can be incorporated into environmental models to help decision makers manage landscapes.

4.1. Introduction

Addressing environmental issues requires accurate information about spatial patterns of soil types and properties. Consequently, providing quantitative soil information of known accuracy is a great challenge to satisfy the needs of end-users, especially landscape managers and stakeholders (Chaney et al., 2016, Ellili et al., 2019). In most environmental and agricultural research, accurate and continuous soil data are increasingly incorporated in soil-landscape modelling to monitor natural resources and ecosystems (Odgers et al., 2012). However, soil data are usually not available at the adequate spatial resolution, particularly over large areas, where only legacy soil maps at coarse spatial resolution are available.

From a pedological viewpoint, soil attributes vary either continuously or sharply down a soil profile (Ponce-Hernandez et al., 1986) as well as across the landscape. Soil-depth functions, which may result from different methods, predict soil properties for specific depth intervals. They are usually used to estimate the variation in soil properties with depth. However, in classic pedology, soils are generally described according to their morphological horizons by extracting a bulk soil sample from each horizon and analysing its main properties. In legacy soil databases, in which horizon-based information is used to describe the variation in soil properties with depth, upper and lower limits of the horizons may differ among sites. When continuous representation of the variation in soil properties with depth is needed (e.g. simulation modelling of soil processes), the discontinuous nature of depth functions is a serious problem when attempting to predict soil properties at specific depths within a soil profile. Moreover, the ability of current methods to predict continuous soil properties is usually unknown and rarely investigated.

Bishop et al. (1999) highlighted that the discontinuity of depth functions derived from bulk horizon data lends to inaccuracies when attempting to predict a soil attribute at a specific soil depth within a soil profile. Many attempts have been made to derive continuous depth functions of soil properties. Jenny (1941) made the earliest known attempt, drawing freehand curves between data points that represented the mid-point depth of a given horizon. More sophisticated approaches have been used, such as fitting exponential decay functions (Russell and Moore, 1968). However, the main disadvantage of these methods is that each local variation in the soil profile affects the shape of the fitted function at all depths. Consequently, the low flexibility of these functions results in variable

quality of fit over soil depth (Malone et al., 2009; Odgers et al., 2012). More recently, Minasny et al. (2006) fitted exponential decay functions to soil organic carbon (SOC) data in the soil profile to map carbon storage in the Lower Namoi Valley, Australia, and achieved an acceptable quality of fit.

To compensate for the lack of flexibility in depth functions, the use of spline functions seems to be a good alternative. In fact, certain spline functions, such as smoothing splines (Erh, 1972) and equal-area splines (Ponce-Hernandez et al., 1986), allowed a series of independent local functions to be fit over small intervals of a soil profile. Bishop et al. (1999) improved the approach of Ponce-Hernandez et al. (1986) by fitting quadratic polynomial splines to soil horizons. Their modified equal-area quadratic splines effectively predicted depth functions for soil pH, electrical conductivity, clay content, SOC content, and gravimetric water content at -33 kPa. However, their method required input data from contiguous soil horizons. One decade later, Malone et al. (2009) solved this issue by generalising Bishop et al. (1999)'s approach to be able to use soil input data from non-contiguous soil horizons.

Using soil-depth functions allows soil attributes to be predicted at specific soil depths. Spline functions are applied to individual soil observations and predict attributes only for a single geographic point. To characterise three-dimensional (3D) variation in soil properties, equal-area quadratic splines require intensive sampling and analysis of soil profiles. Thus, vertical variation in soil properties can be predicted, but lateral variation remains unaddressed. Combining digital soil mapping (DSM) techniques and soil-depth functions appears to be a good option to capture both lateral and vertical variations in soil properties. DSM predicts soil properties based on their relationships with environmental variables (Minasny et al., 2008) to address soil variability, even in areas with limited soil data (McBratney, 2003). Malone et al. (2009) combined equal-area smoothing splines and neural network models to map SOC storage and available water capacity based on limited soil data in the lower valley of the Namoi River, Australia. Lacoste et al. (2014) also combined DSM techniques and equal-area splines to derive 3D maps of SOC stock at high spatial resolution across an agricultural landscape in Brittany, France. Other researchers (Bishop et al., 2015; Vaysse et al., 2015) adopted the same strategy to map soil properties at specific soil depths across a defined study area by using existing soil databases. More importantly, *GlobalSoilMap* (Arrouays et al., 2014) specifications recommended using equal-area splines to produce soil-property maps for six regular depth intervals (0-5, 5-15, 5-30, 30-60, 60-100, and 100-200 cm).

The aim of this study was to develop a method to assess the ability of spline functions to preserve relationships among soil properties. We did this by i) fitting spline functions to soil typological unit (STU) soil information to express soil attributes for six regular depth intervals and ii) performing bivariate and multivariate statistical analysis of soil attributes before and after fitting splines to assess the goodness of fit and degree of preservation of soil-property relationships. Furthermore, maps of nine soil properties of major interest for agronomic and environmental purposes were generated at the regional scale by combining DSM techniques with equal-area splines. STU soil attributes were first standardised for regular depth intervals down to 200 cm and then mapped using disaggregated STU maps obtained from a previous study at a spatial resolution of 50 m (Vincent et al., 2018, Ellili et al., 2019). The overall workflow was as follows:

- i) Extract soil-property information for all strata of each STU and identify the nature of the bottom STU strata (i.e. lithic or paralithic contact)
- ii) Fit equal-area spline functions to soil properties using STU features to obtain estimates for standard depth intervals
- iii) Assess preservation of soil-property relationships after fitting equal-area spline functions
- iv) Assess the ability of splined data to correctly predict a complex variable that depends on additional soil properties (Effective Cation-Exchange Capacity=ECEC)
- v) Map soil properties over the study area by combining disaggregated STU maps and fitted equal-area soil-property values

4.2. Materials and methods

4.2.1. The study area

The study area was the Brittany region, in north-western France. It covers approximately 27 040 km², with a variety of physical and geographic features. Its climate is oceanic, with mean annual precipitation ranging from 650 mm in the east to 1300 mm in the west, and mean annual temperature ranging from 11-13°C (CLIMATIK database: INRA, 2019). Agriculture, especially annual crops interspersed with temporary and permanent grasslands, represents the major land use. Brittany has a dense hydrographic network and natural wetlands.

The relief is relatively gentle and strongly correlated with geological formations, with elevations of 0-382 m (Figure 4.1). Brittany is part of the Armorican Massif, which has complex geology: sedimentary rocks, rocks metamorphosed to differing degrees (sandstone, schist), metamorphic rocks (gneiss), igneous rocks (granite), and loess deposits. This high geological diversity generates a wide range of soils. According to the World Reference Base of Soil Resources (IUSS Working Group WRB, 2014), the main soils include Cambisols, Stagnic Fluvisols, Histosols Podzols, Luvisols, and Leptosols. Soils are organised along toposequences: relatively well-drained deep soils in upslope and plateau locations, shallower well-drained soils on mid-slopes, and soils with marked redoximorphic features in valleys.

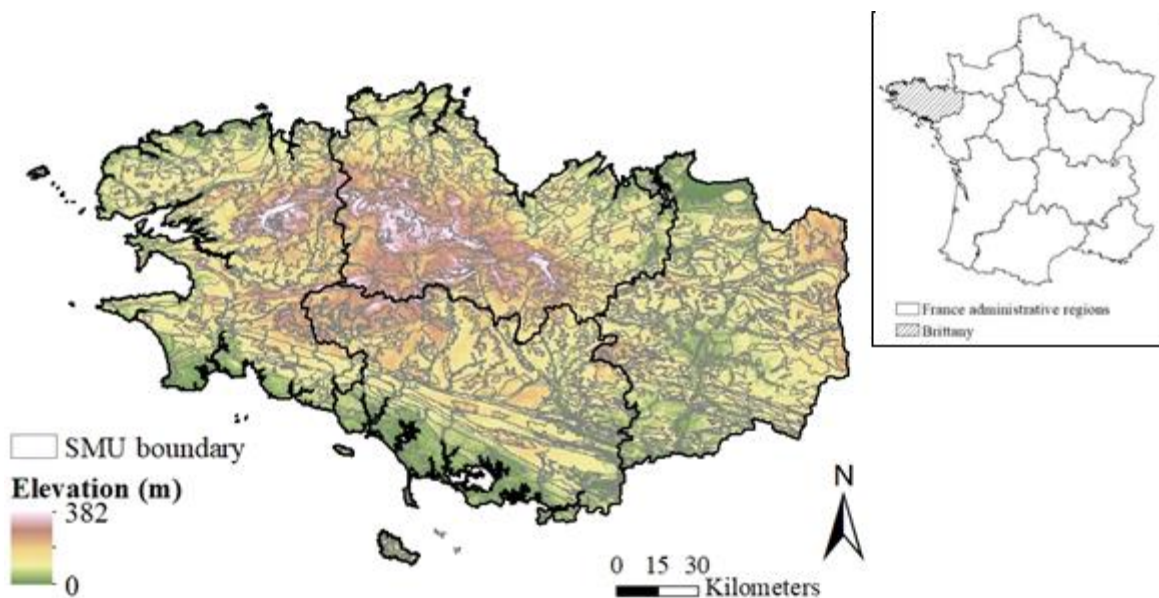


Figure 4.1 : Location of the Brittany study area, showing boundaries of soil map units (SMU) and elevation.

4.2.2. Regional soil database at 1:250 000 scale

Soil mapping in Brittany is represented in a regional geographic database called the “Référentiel Régional Pédologique” (RRP) available at 1:250 000 scale (Lemerrier, 2010). It contains a set of polygons with crisp boundaries, commonly called soil map units (SMUs), which are defined as areas with homogeneous soil-forming factors, such as morphology, geology, and climate. Each SMU contains known proportions of several STUs, each of which is described in the RRP by a set of strata that describe the vertical variation in the soil. The strata are spatial horizons describing the vertical

organisation of STUs. Pedological features of SMU, STU, and strata, including depth, thickness, SOC content, ECEC, pH, and five particle-size fractions, are contained in a relational database called DoneSol (INRA Infosol, 2014). Furthermore, each STU has a representative soil profile with a full pedological description. These profiles, which cover a wide pedological diversity in Brittany, are contained in a separate database. In Brittany, 341 SMUs, 320 STUs, and 1020 strata are currently defined in 1984 polygons.

4.2.3. Assessing and mapping soil properties at the regional scale

4.2.3.1. Disaggregation of complex mapping units

The first step of the workflow (Figure 4.2) was performed in previous studies. Using the DSMART algorithm (Disaggregation and harmonisation of Soil Map Units Through Resampled Classification Trees) (Odgers et al., 2014) and soil-landscape relationships, Vincent et al. (2018) disaggregated the existing legacy 1:250 000 soil map of Brittany at a resolution of 50 m. Disaggregation yielded estimates of the probability of occurrence of the 320 STUs in each pixel. Finally, Vincent et al. (2018) retained only the three STUs with the highest probabilities of occurrence for each pixel and produced soil-type grids for the region indicating the first, second, and third most probable STU.

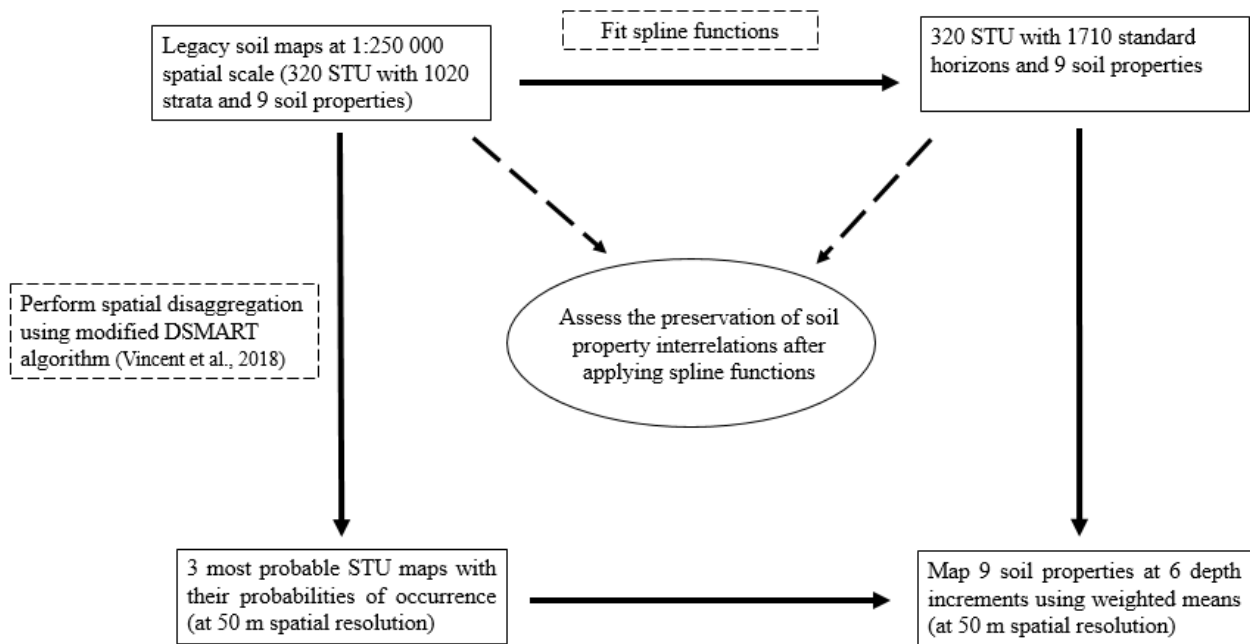


Figure 4.2 : The workflow used to derive disaggregated soil-property maps at the regional scale. STU = soil typology unit.

4.2.3.2. Modelling depth variation using spline functions

Because STU strata were described according to morphological features, they have variable upper and lower depths. Strata were vertically interpolated to express vertical variation in soil properties of each STU for *GlobalSoilMap* standard depth intervals by applying equal-area spline functions (Bishop et al., 1999; Malone et al., 2009) using the GSIF package (Hengl et al., 2013) of R software (R Core Team, 2019). The equal-area spline function respects the mean of the target soil property and ensures continuous variation with soil depth (Malone et al., 2009). Its result is a set of interpolated values that reflect the target property's mean for regular depth intervals down to 200 cm. To avoid spline-function loopholes, which consist of overestimating error at the extremes of a given STU, we followed previous recommendations (Odgers et al., 2012; Vaysse et al., 2015) and added two pseudo-strata to each STU. The first was created by subdividing the first stratum into two elementary strata, which created a thin topsoil layer from 0-3 cm. The second was added from the bottom stratum down to 200 cm, but only for STUs that had no stratum down to 200 cm but whose depth could reach 200 cm. When present, the bottom pseudo-stratum had the same characteristics as

the stratum just above it. More importantly, STU depth can be defined as the distance from the soil surface to the contact with bedrock (lithic) or weathered rock (paralithic), which contains fissures that make it easier for roots to penetrate the bedrock (Soil Survey Staff, 1993). In the present study, STU depth was determined using a previously developed classification tree (Vaysse et al., 2015; Styc and Lagacherie, 2019) applied to the bottom horizon of the representative soil profile of each STU. In the classification tree, each horizon is classified as “censored”, “may be censored”, or “not censored” (Figure 4.3), the first two of which meant that the bottom horizon could be duplicated down to 200 cm. In general, the bottom horizon was classified as “not censored” when it had lithic or paralithic contact (R, M, F, C/R, C/M, C/D horizons) and “censored” when it was identified as a pedogenetic horizon (A, B, E horizons). Since ambiguities in the C horizon mean that it cannot be directly classified as “may be censored” or “not censored”, the tree applies pedological rules that assess additional soil properties such as weathering, internal structure, classification, and compactness. The tree thus uses a binary approach to assess the C horizon and classifies it as “may be censored” when at least one of three rules is satisfied.

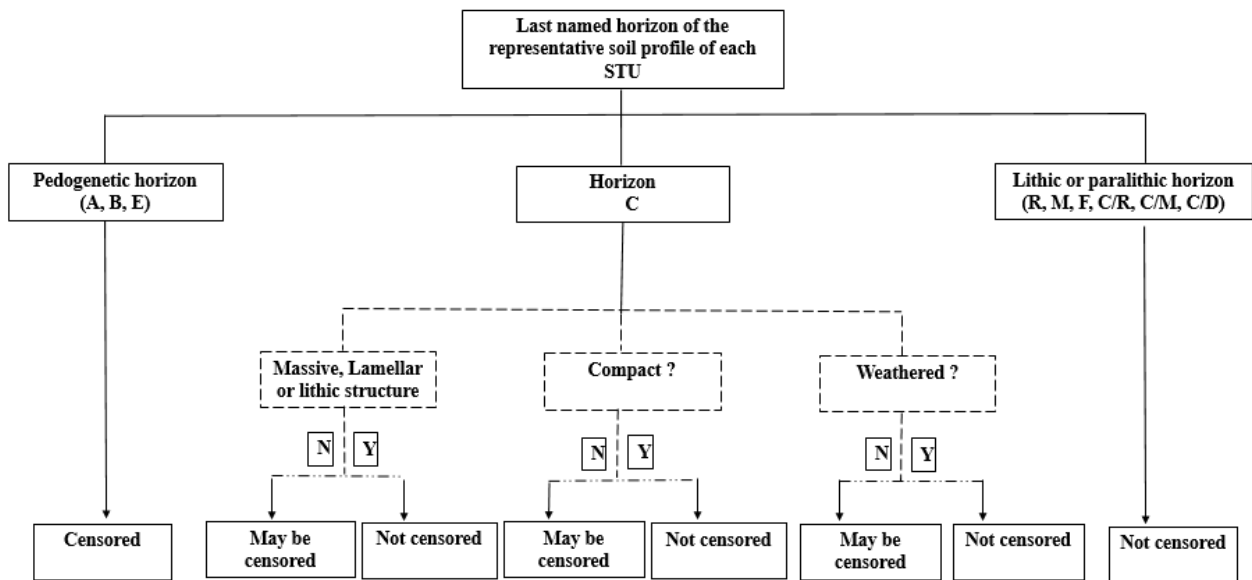


Figure 4.3 : The classification tree used to characterise lithic and paralithic horizons. STU = soil typological unit (adapted from Vaysse, 2015).

4.2.3.3. Correlation analysis among soil properties

Statistical analyses were performed to assess the degree to which fitting spline functions preserved soil-property relationships. As far as we aware, this is the first time that preservation of soil attribute relationships using spline functions has been investigated for legacy soil maps. First, a Pearson correlation matrix was calculated to demonstrate the degree of interaction between soil properties before and after fitting spline functions. The matrix was calculated for two soil datasets: (i) all STU strata identified in the DoneSol database (1020 strata) and (ii) soil properties estimated for the six regular depth intervals down to 200 cm (1710 horizons). A two-tailed parametric t-test was also performed to assess the significance of correlation coefficients. Next, a normalised principal component analysis (PCA) was performed on both original and splined soil-property values to visualise organisation of soil data and correlations among target soil properties simultaneously. The main advantage of this statistical method was that the relative relationships among soil attributes was preserved and all variables were projected onto a factorial plan.

4.2.3.4. Prediction of a complex soil variable (ECEC)

To avoid the problem of multicollinearity among soil attributes of linear models, a stepwise multiple linear regression (Hocking, 1976) was used to determine the best combination of soil attributes in the original and splined datasets to predict ECEC. Models were selected using the Akaike information criterion. This approach assessed the ability of the splined soil data to accurately predict a complex soil attribute that depends strongly on explanatory soil variables. Overall, the method achieved three objectives simultaneously by comparing (i) the performance of models that best predict a target variable (ECEC) as a function of additional soil properties for both soil datasets, (ii) selected explanatory soil properties before and after fitting spline functions, and (iii) the partial coefficient of each soil property for both models based on their associated confidence intervals.

4.2.3.5. Digital soil mapping of soil properties

The first step of the workflow in this study standardised the depth of the STU characteristics to be expressed for six regular depth intervals down to 200 cm. By using the three most probable STU maps, with their associated probabilities of occurrence at a 50 m spatial resolution (Vincent et al.,

2018; Ellili et al., 2019), and STU soil-property estimates for six regular depth intervals, several soil-property maps were generated to estimate the spatial pattern of each soil property in Brittany. To this end, the weighted mean of each soil attribute was calculated for each grid cell based on reference soil-property values of the three most probable STU, and their associated probabilities of occurrence were used as weights, as follows (Eq 4.1):

$$\hat{y}(x_{ij}) = \frac{\sum_{k=1}^3 (P_{STU_{kj}, x_{ij}}) * y(STU_{kj}, x_{ij})}{\sum_{k=1}^3 (P_{STU_{kj}, x_{ij}})} \quad (\text{Eq 4.1})$$

where $\hat{y}(x_{ij})$ is the predicted soil property for grid cell (x_{ij}) and soil depth $j=1, 2, 3, 4, 5, 6$; $y(STU_k, x_{ij})$ is the reference soil property estimated at soil depth j associated with STU $k=1, 2, 3$ predicted at all (x_{ij}) ; and $(P_{STU_{kj}, x_{ij}})$ is the probability of occurrence of STU $_k$ in the given grid cell (x_{ij}) .

The workflow produced maps of nine soil properties – SOC; particle-size distribution of fine sand, coarse sand, fine silt, coarse silt, and clay; coarse fragments; pH; and ECEC – at a fine spatial resolution of 50 m over the entire area from legacy soil maps.

4.2.3.6. Spatial correlation of soil properties

To analyse the spatial correlation of disaggregated soil properties, we calculated experimental semivariograms of each predicted variable for the six standard depth intervals. To this end, we first generated a grid with a resolution of 50 m covering the entire area and randomly selected 30 000 cells from it using the Spatially Balanced sampling tool of ArcGis 10.6 GIS software (ESRI, 2012). Next, soil-property estimates were extracted for each cell of the selected dataset using the multiple extraction tool of ArcGIS. Before semivariogram analysis, SOC content was log-transformed, and ECEC was square-root-transformed. By applying the variogram function in the R package “gstat” (Pebesma, 2004), several experimental semivariograms were generated for all soil properties and all standard depth intervals to characterise the spatial pattern of the soil-property maps produced.

4.3. Results

4.3.1. Spline-function outputs

As an example of spline-function outputs, the equal-area splines fit to strata silt content for four STUs depended strongly on the initial silt content of each stratum (Figure 4.4). Abrupt changes in content between adjacent strata influenced the shape of the spline curve considerably. However, the spline curves were constrained at the top and bottom strata because adding the two thin pseudo-strata minimised overfitting.

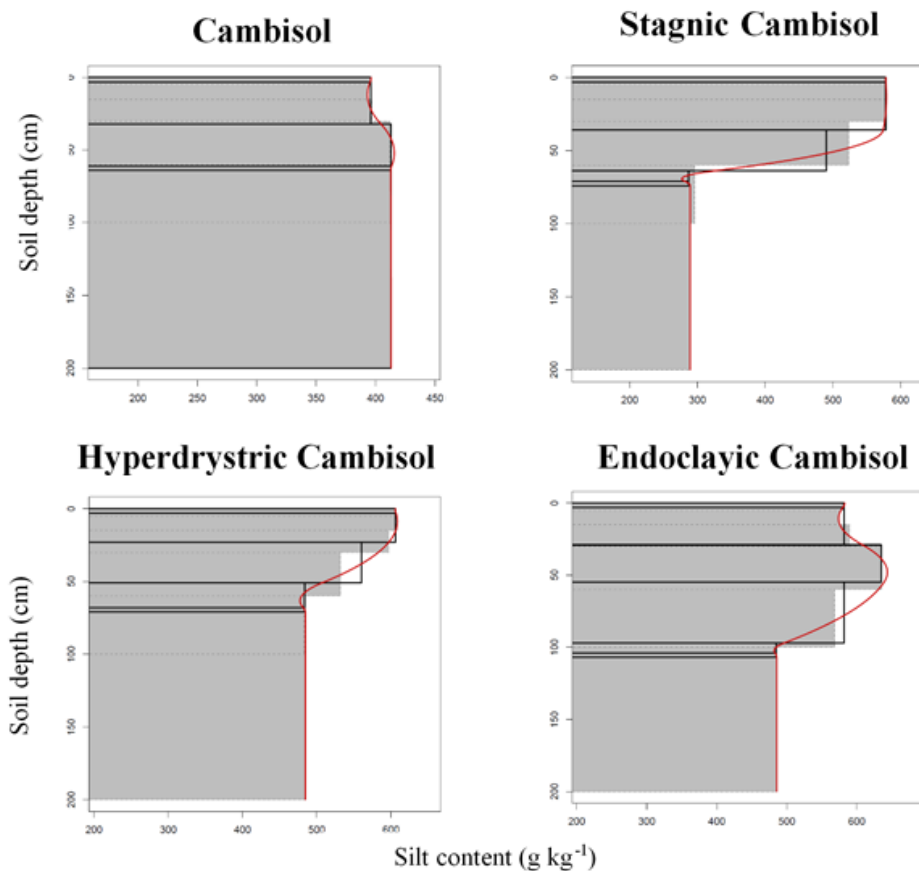


Figure 4.4 : Strata silt content (solid grey rectangles), equal-area spline curves (red), and estimated mean silt content for *GlobalSoilMap* regular soil-depth intervals (open black rectangles) for four soil typological units.

4.3.2. Correlation between soil properties before and after fitting smoothing spline functions

For both Pearson correlation matrices, well-known correlations between soil properties were significant (Table 4.1). For instance, SOC content correlated positively with clay content and ECEC but negatively with pH and sand content ($P < 0.001$). The correlation between ECEC and clay content was also positive ($P < 0.001$). In contrast, pH was weakly but not significantly correlated with ECEC and clay content. The correlation matrices showed similar correlation coefficients between soil properties in both the original and splined datasets, indicating that applying spline depth functions to pedological strata preserved soil attribute correlations.

4.3.3. Multicollinearity among soil properties

The first two dimensions of the PCA performed on 1020 strata features explained 57% of the total variance (Figure 4.5a). The main contributing variables were sand content (30%) and total silt content (25%) on the first dimension and clay content (26%), CEC (22%), and SOC content (15%) on the second. Similarly, the first two dimensions of the PCA performed on 1710 standard horizon features explained 58% of the total variance (Figure 4.5b). The main contributing variables were sand content (30%) and total silt content (23%) on the first dimension and SOC content (15%), total silt content (17%), and ECEC (17%) on the second. Thus, the first PCA dimension for the two datasets (34.3% and 34.2%, respectively) depended on soil-texture fractions, contrasting coarse texture and relatively fine texture. The second PCA dimension for the two datasets (22.7% and 23.3%, respectively) depended on physico-chemical properties and fine texture, particularly clay content.

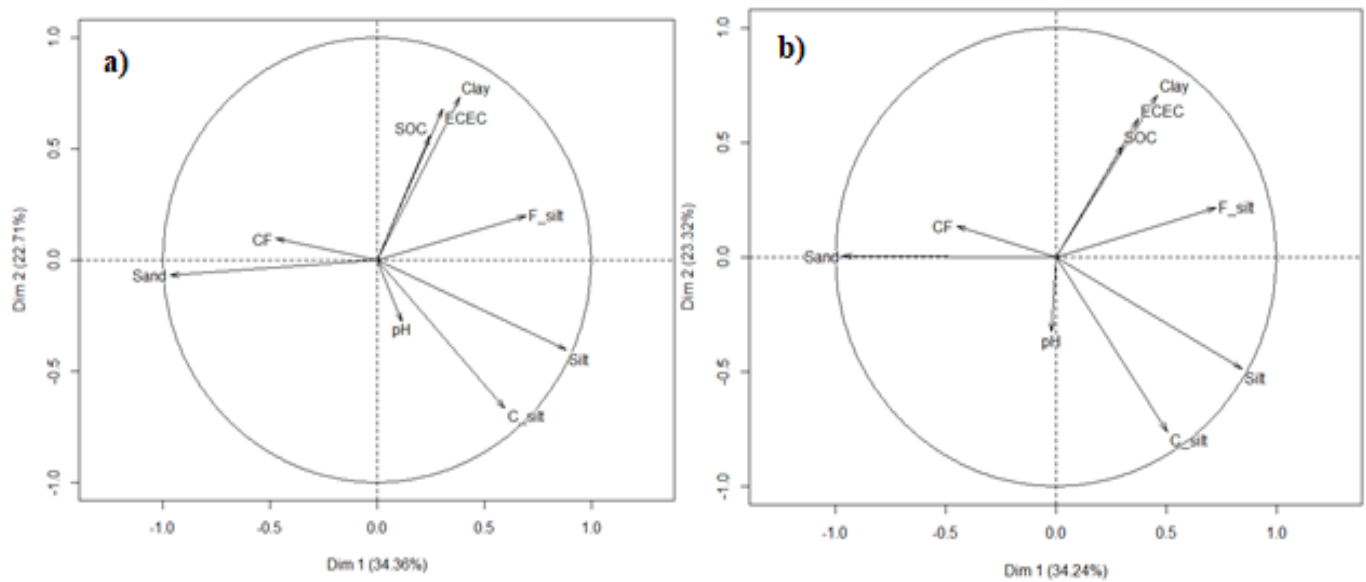


Figure 4.5 : Principal component analysis based on physico-chemical and texture characteristics of soil typological unit strata (a) before and (b) after fitting spline functions. SOC: soil organic carbon content, ECEC: effective cation exchange capacity, sand: sand content, CF: coarse fragments, F-silt: fine silt, C: silt: coarse silt, and silt: silt content.

Table 4.1 : Pearson’s correlation matrices between soil properties before (1020 strata) and after fitting spline functions (1710 standard horizons). SOC = soil organic carbon, CEC = cation exchange capacity. * P < 0.05; ** P < 0.01; *** P < 0.001.

Dataset	Property	SOC	Sand	Clay	Silt	Coarse fragments	Fine silt	Coarse silt	pH
Before	Soil organic carbon	1							
	Sand	-0.17***	1						
	Clay	0.26***	-0.53***	1					
	Silt	0.03	-0.83***	-0.04	1				
	Coarse fragments	-0.11**	0.33***	-0.12**	-0.30***	1			
	Fine silt	0.10*	-0.71***	0.33***	0.62***	-0.08	1		
	Coarse silt	-0.04	-0.51***	-0.30***	0.80***	-0.32***	0.02	1	
	pH	-0.14**	-0.07	0.01	0.07	-0.25***	-0.21***	0.25***	1
	CEC	0.48***	-0.26***	0.48***	-0.01	-0.17***	0.12**	-0.11**	0.04
After	Soil organic carbon	1							
	Sand	-0.19***	1						
	Clay	0.30***	-0.55***	1					
	Silt	0.03	-0.82***	-0.02	1				
	Coarse fragments	-0.13***	0.31***	-0.10***	-0.30***	1			
	Fine silt	0.13***	-0.71***	0.39***	0.59***	-0.11***	1		
	Coarse silt	-0.05	-0.49***	-0.31***	0.80***	-0.29***	-0.02	1	
	pH	-0.15***	0.01	-0.05	0.02	-0.21***	-0.26***	0.22***	1
	CEC	0.40***	-0.28***	0.50***	0.00	-0.21***	0.18***	-0.14***	0.01

4.3.4. Stepwise multiple linear regression

The best regression model for the 1020 strata features (before fitting spline functions) contained SOC content (P < 0.001), clay content (P < 0.001), and pH (P < 0.01) and explained 37% (multiple R²) of

variation in ECEC (Adj. $R^2 = 0.36$, $F = 193$, $P < 0.005$). The best regression model for the 1710 standard horizons (after fitting spline functions) also contained SOC content, clay content, and pH (all ($P < 0.001$)) and explained 33% of variation in ECEC (Adj. $R^2 = 0.32$, $F = 213$, $P < 0.001$). The 90% confidence intervals of both models appeared similar, which confirms that fitting spline functions to pedological strata information preserved the relationships among soil attributes (Table 4.2). For instance, for SOC content, the 90% confidence interval before and after fitting spline functions was 0.42-0.52 and 0.21-0.34 g kg^{-1} , respectively.

Table 4.2 : Regression coefficients of effective cation exchange capacity (ECEC) regression models before and after fitting spline functions. SOC = soil organic carbon. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

Model	Variable	Estimate	Standard error	t-value	Significance	5% limit	95% limit
Before splines	Constant	-3.33	1.34	-2.48	0.01*	-4.55	-1.11
	SOC (g kg^{-1})	0.47	0.03	14.80	<0.05***	0.42	0.52
	Clay (%)	0.33	0.02	14.38	<0.05***	0.30	0.37
	pH	0.71	0.22	3.15	0.001**	0.34	1.08
After splines	Constant	-2.3	0.67	-1.02	0.30	-2.61	0.62
	SOC (g kg^{-1})	0.27	0.02	13.74	<0.05***	0.21	0.34
	Clay (%)	0.33	0.01	17.56	<0.05***	0.30	0.37
	pH	0.55	0.15	3.47	<0.005***	0.29	0.82

4.3.5. Soil-property mapping at multiple depths

Depth functions successfully preserved relationships among physico-chemical soil properties. This implies the reliability of soil-property estimates and their spatial patterns over the study area; the estimates remained highly informative and reflected information in the legacy soil maps.

In soil-property maps, SOC content tended to decrease down the soil profile across most of Brittany, being higher in surface layers (0-5 and 5-15 cm) than in subsurface layers (60-100 and 100-200 cm) (Figure 4.6). Furthermore, the spatial pattern of SOC content showed a strong increase from east to west and north to south. Overall, soils not used for annual crops – located in permanent grasslands, wooded areas, and along watercourses – tended to have higher SOC content than those used for

annual crops. In surface layers (0-5 and 5-15 cm), cropping areas (northern, western, and central Brittany) had the lowest SOC content (usually 10-40 g kg⁻¹), while grasslands and wooded areas had SOC content in excess of 150 g kg⁻¹.

Likewise, silt content always decreased as depth increased, except for soils developed from Aeolian loam deposits in northern Brittany (Figure 4.7). This trend was also related to the depth of these soils, which did not exceed 100 cm. For surface layers (0-30 cm), silt content usually ranged from 30-60% but exceeded 70% for soils developed from Aeolian loam, which coincide with areas of annual crops. For subsurface layers, particularly 30-60 cm and 60-100 cm, the lowest silt contents (< 30%) were located in areas of natural vegetation in south-western and north-central Brittany.

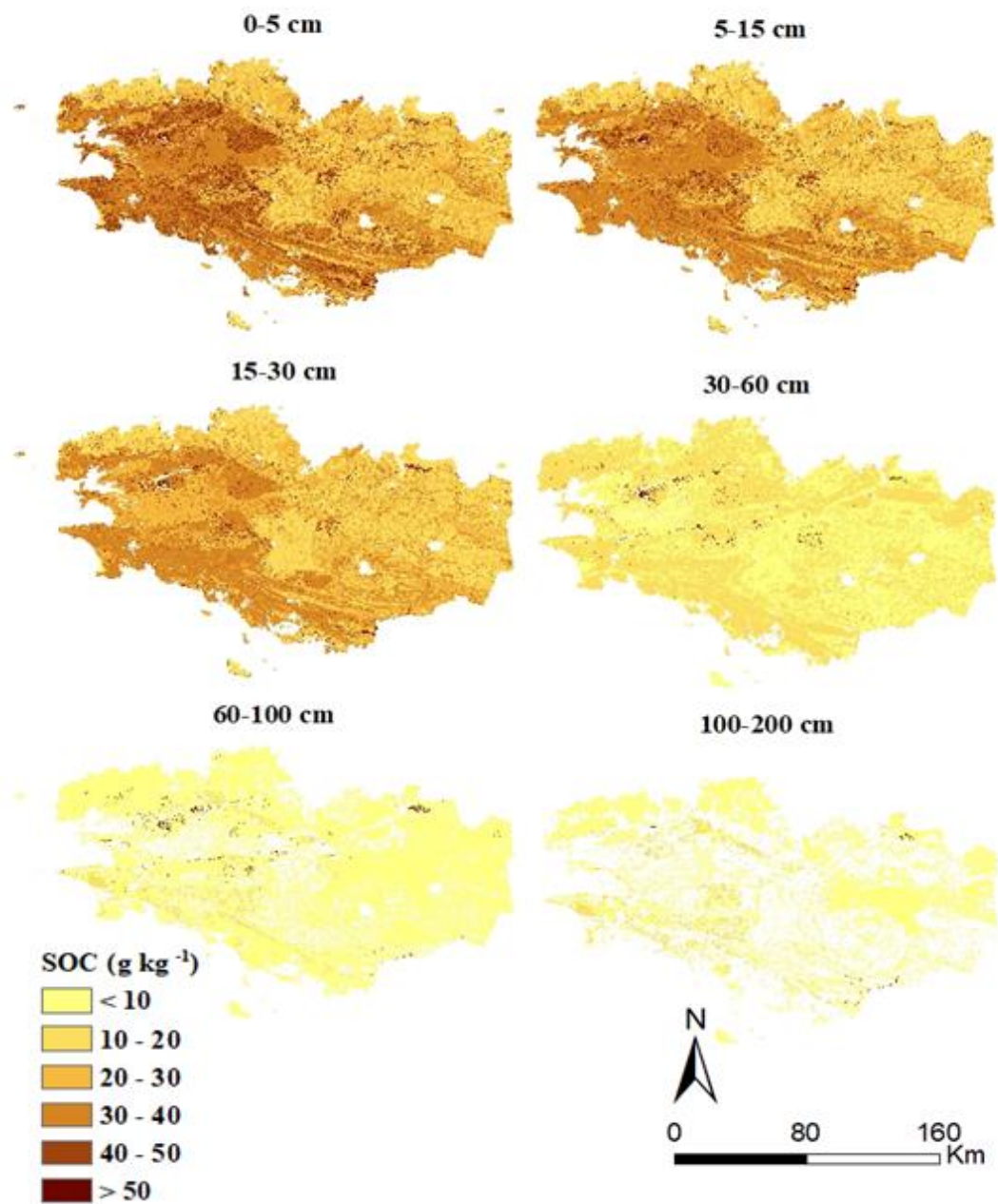


Figure 4.6 : Disaggregated maps of soil organic carbon (SOC) content in Brittany at the six soil-depth intervals.

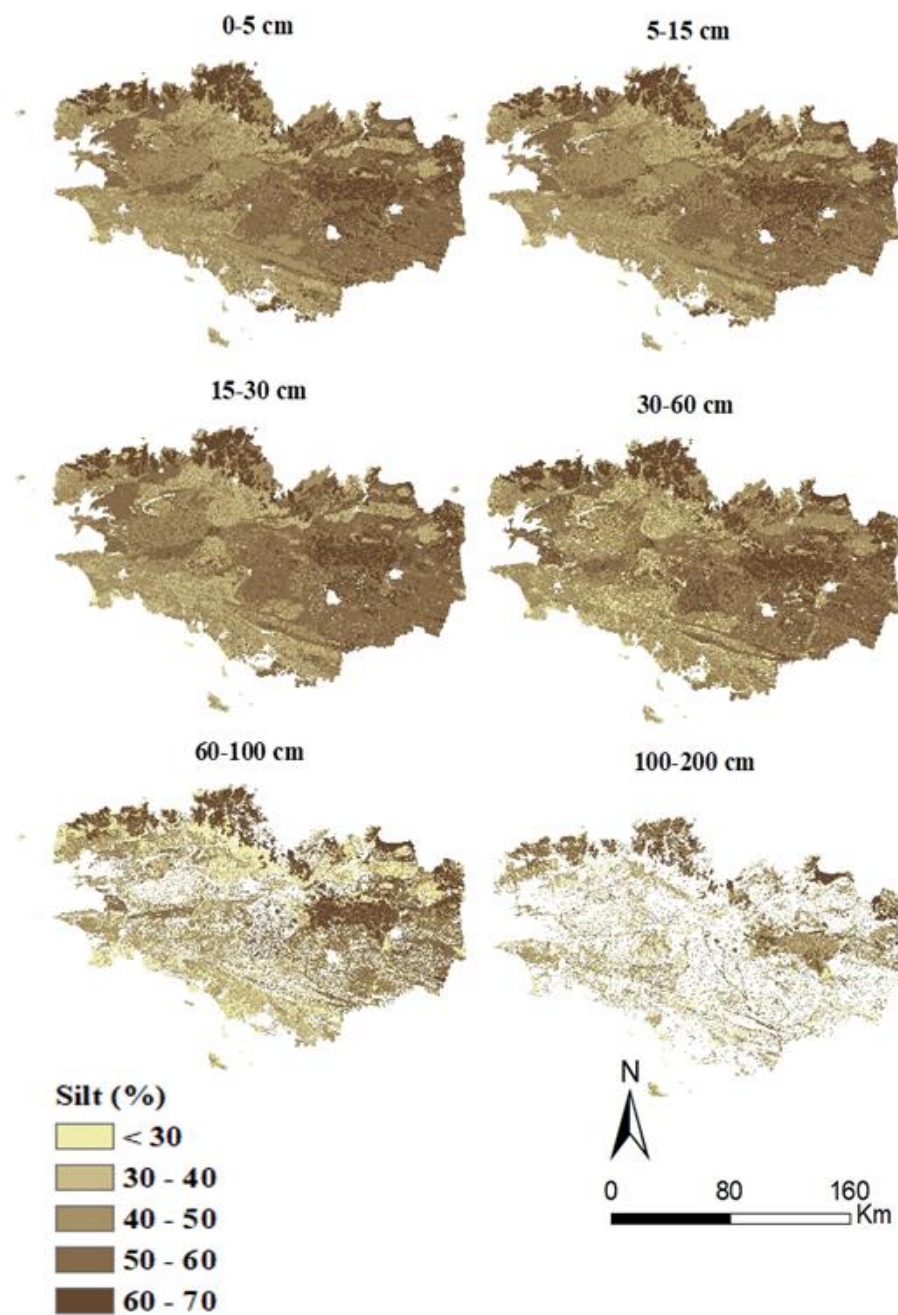


Figure 4.7: Disaggregated maps of silt content in Brittany for the six soil-depth increments.

Overall, the disaggregated data covered a wide range of soil property values (Table 4.3). For instance, median SOC content decreased from 30 g kg⁻¹ at 0-5 cm to 4 g kg⁻¹ at 100-200 cm. Most soil properties followed the same trend, except for sand content, CEC, and coarse fragments, which increased down soil profiles.

Table 4.3 : Descriptive statistics of the disaggregated soil maps of the soil properties studied.

Soil property	Depth interval (cm)	Min	1 st quantile	Median	Mean	3 rd quantile	Max
Soil organic carbon (g kg ⁻¹)	0-5	3	20	30	32	37	761
	5-15	4	19	28	30	34	663
	15-30	2	19	25	28	32	789
	30-60	0	7	10	13	14	807
	60-100	0	3	4	6	5	583
	100-200	0	2	4	5	7	686
Clay (%)	0-5	2	15	17	18	21	65
	5-15	1	16	17	18	21	66
	15-30	1	15	17	18	20	68
	30-60	1	14	16	16	19	69
	60-100	2	15	17	18	21	65
	100-200	1	12	16	16	19	59
Sand (%)	0-5	1	23	28	32	43	88
	5-15	0	23	29	32	41	88
	15-30	0	23	29	32	42	90
	30-60	2	23	31	33	42	99
	60-100	0	23	29	32	41	88
	100-200	2	22	34	33	42	94
Silt (%)	0-5	5	41	50	50	57	73
	5-15	4	42	50	50	57	74
	15-30	4	42	51	50	57	74
	30-60	2	43	50	50	58	77
	60-100	1	33	44	44	55	79
	100-200	1	34	43	42	49	75
Fine silt (%)	0-5	1	21	25	26	30	45
	5-15	1	21	25	25	29	45
	15-30	1	21	25	26	30	44
	30-60	1	21	24	25	30	40
	60-100	1	17	21	22	25	41
	100-200	1	17	21	20	23	42
Coarse silt (%)	0-5	3	20	22	25	26	64
	5-15	3	20	22	24	26	65
	15-30	2	20	22	25	26	65
	30-60	0	20	23	25	26	63
	60-100	1	16	20	22	25	63
	100-200	1	16	21	22	24	58

Coarse fragments (%)	0-5	0	3	6	7	9	55
	5-15	0	3	5	6	8	55
	15-30	0	3	6	7	10	58
	30-60	0	4	8	11	15	83
	60-100	0	6	10	13	15	92
	100-200	0	5	9	11	14	98
Cation exchange capacity (cmol ⁺ kg ⁻¹)	0-5	2	10	12	12	14	69
	5-15	2	10	13	13	15	39
	15-30	2	10	11	11	13	40
	30-60	1	4	5	6	7	32
	60-100	1	3	4	4	5	95
	100-200	1	3	5	5	5	58
pH	0-5	4	6	6	6	6	9
	5-15	4	6	6	6	6	9
	15-30	4	6	6	6	6	9
	30-60	4	5	6	6	6	9
	60-100	4	5	5	5	6	9
	100-200	4	5	5	5	6	9

Overall, the experimental semivariograms highlighted significant differences among soil properties and soil-depth intervals. For all soil properties, subsurface layers, particularly 60-100 and 100-200 cm, varied more than surface layers (0-5, 5-15 and 15-30 cm), which had nearly the same semivariogram shape (Figure 4.8). Furthermore, subsurface layers had a “nugget effect” (i.e. value of the semivariogram at zero lag size) twice as large as those of surface layers. For instance, the nugget effect for sand content was nearly 50 (g kg⁻¹)² at 0-5, 5-15, and 30-60 cm and reached 100 (g kg⁻¹)² in the bottom layer (100-200 cm). In contrast, the range of the nugget effect varied by soil property and depth interval and was a mean of 30 000 m. The sill:nugget ratio also varied greatly by soil property and depth interval (e.g. for SOC content, from 2.0 for surface layers (0-60 cm) to 2.5 for the bottom layer).

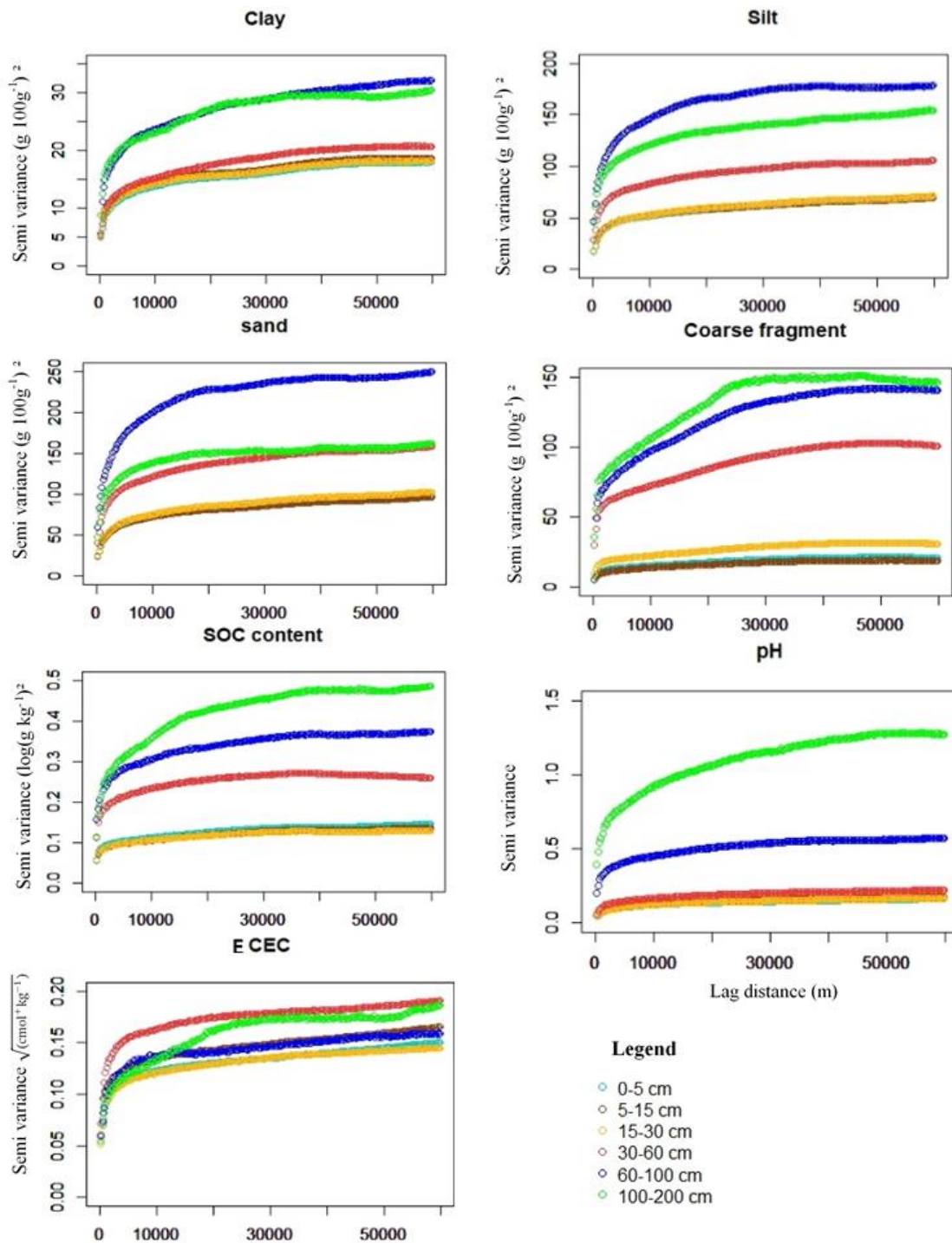


Figure 4.8 : Experimental semivariograms derived from disaggregated soil maps for seven soil properties for six regular soil-depth intervals.

4.4. Discussion

4.4.1. Equal-area spline method: advantages and limitations

Reconstructing STU information as a function of depth by fitting spline functions to soil-property values was a convenient way to estimate soil attributes for regular depth intervals. Several studies (Malone et al., 2009; Odgers et al., 2012; Lacoste et al., 2014; Vaysse et al., 2014; Stycs and Lagacherie 2019) have used this method to standardise many soil properties by depth. In the present study, it effectively preserved existing relationships between all soil properties, as well as the pedological information. Due to their flexibility, the quadratic polynomial spline functions captured local variation in each soil interval (Ponce-Hernandez et al., 1986; Bishop et al., 1999). More importantly, equal-area splines predict several soil properties simultaneously for a set of common depth intervals. These predictions are similar for a given soil depth and reflect specific pedological process. Soil-property estimates can be used in environmental models and decision-making tools to derive consistent proxies of ecosystem functions and help guide stakeholders' choices.

The equal-area spline method estimated variation in a given soil property with depth satisfactorily for many kinds of soil profiles, but it also had some limitations. For instance, soil properties (e.g. SOC content, clay content) of some soil profiles changed abruptly with depth, which sometimes generated negative values at points close to the abrupt boundary. The simplest solution was to insert two thin pseudo-strata at the top and bottom strata to restrict the abrupt transition in the spline function. Despite its potential effectiveness, most studies do not follow this approach because it requires additional time and effort to harmonise the input data.

4.4.2. Spatial structure of disaggregated soil-property maps

The disaggregated digital soil maps depicting soil properties for all depth intervals were generally consistent with landscape features, including land use. For instance, SOC content peaked in wooded areas and permanent grasslands but was lowest in cropping areas. More importantly, Histosols, especially in north-western Brittany, showed high SOC content in deeper soil layers. Thus, SOC content was influenced mainly by structural factors, such as parent material, topography,

hydromorphic soil conditions, and some management practices (e.g., soil tillage, crop cover, and land use (Ellili et al., 2019)).

As expected, the spatial pattern of silt content highlighted well the geological and pedological knowledge about soils in Brittany. Silt content increased with depth for soils developed in deep Aeolian loam deposits, especially in north-central Brittany, and was lowest in soils in the north and south, where soils are derived from less-weathered granite. In addition, silt content decreased with depth over the entire study area. Predictions of sand content and coarse fragments by soil-depth interval also appeared to be consistent with existing knowledge about the study area, particularly the map of predicted soil parent material created in a previous study (Lacoste et al., 2011).

We found high variability in the spatial pattern of each soil attribute studied, all of which had a high nugget effect. According to the geostatistical literature (Journel and Huijbregts, 1978), the nugget effect is related to measurement errors or spatial variation at distances smaller than the sampling interval, or both. However, the experimental semivariograms showed higher variability for deeper layers than surface layers, which can be explained mainly by the degree of weathering of soil parent material, which varies greatly among the contrasting parent materials in Brittany.

4.4.3. Legacy soil data and DSMART limitations

DSM of soil properties from legacy soil-type maps using the DSMART algorithm has the advantage of explicitly representing spatial patterns of several soil properties simultaneously. Odgers et al. (2015) generated maps of clay content and pH for *GlobalSoilMap* depth intervals by applying the PROPR (Digital Soil Property Mapping Using Soil Class Probability Rasters) algorithm. It uses all derived soil-class-prediction rasters and their associated probability-of-occurrence rasters to estimate soil properties for each grid cell.

In the present study, only the three STUs with highest probability of occurrence were used to predict soil properties. Although the DSMART algorithm attempts to predict the most probable STU for a specific soil-landscape context, its accuracy varied greatly and depended strongly on the pedological complexity of the study area. In addition, the estimated probabilities of occurrence of STUs in each grid cell depended greatly on SMU proportions, which were derived from pedological knowledge. Overall, these issues involve limitations in the legacy soil data used in the study. For one, STUs were

not sampled equally across the landscape, which leads to unequal representation of soil samples among STUs, especially when some STUs are more abundant than others. In addition, differences in sampling designs followed to sample legacy soil profiles may have influenced the determination of SMU proportions. For instance, some projects sampled soil profiles to study redoximorphic soil conditions, while other projects did so to study agricultural soils. Finally, the decrease in data availability with depth could also have biased our predictions, especially those for deeper soil intervals.

Conclusion and perspective

In the present study, we developed a method to produce soil-property maps from legacy soil data for the regular intervals defined by *GlobalSoilMap* specifications. By combining spline functions and the DSMART algorithm, it achieved continuity in both lateral and vertical variation in soil properties. Moreover, our predictions for each depth interval were consistent with the geological and pedological heterogeneity across the landscape. However, some challenges remain beyond creation of the soil-property maps themselves, particularly the assessment of uncertainties in soil-property estimates. More sophisticated approaches are required to estimate uncertainties that propagate from legacy soil maps, disaggregated STU maps, spline functions, and soil-property maps.

Conclusion du chapitre 4

Le chapitre 4 a permis de répondre à deux objectifs principaux : (i) d'évaluer la pertinence de l'ajustement de régressions splines à conservation de masse en vue de préserver les corrélations entre les propriétés des sols ; (ii) de prédire un ensemble de propriétés de sol à des profondeurs standards en cohérence avec les spécifications du consortium *GlobalSoilMap* à l'échelle de la Bretagne entière. Nos résultats montrent que les régressions splines préservent les corrélations existantes entre les propriétés pédologiques. En effet, les corrélations calculées pour chaque paire de propriétés de sol avant et après l'ajustement des fonctions splines ne sont pas significativement différentes. De plus, l'analyse multivariée effectuée sur les deux jeux de données de propriétés de sol, avant et après ajustement des fonctions splines, montre des projections similaires des variables pédologiques sur le premier plan factoriel de l'ACP, avec des inerties totales équivalentes.

Par ailleurs, la mobilisation conjointe des propriétés de sol (la teneur en carbone organique, les cinq fractions granulométriques, la charge en éléments grossiers, le pH, et la CEC) exprimées aux profondeurs requises par les spécifications *GlobalSoilMap* et la caractérisation spatiale des UTS, permet de proposer une cartographie des propriétés des sols, aux six intervalles de sol standards, sur l'ensemble de territoire breton, à une résolution spatiale de 50 m.

Les cartes des propriétés de sol obtenues, sont plausibles avec les caractéristiques inhérentes du milieu et la connaissance *a priori* de l'ensemble du territoire. Plus encore, les cartes dérivées satisfont une double cohérence horizontale et verticale de la couverture pédologique. Cependant, la qualité prédictive de ces cartes mérite d'être évaluée, afin d'estimer les niveaux d'incertitude qui peut être injectée dans les processus de gestion agro environnementale ou encore d'évaluation des services écosystémiques des sols. C'est pourquoi, la qualité prédictive de l'information pédologique aux différentes profondeurs requises par les spécifications *GlobalSoilMap* a été investiguée dans le chapitre suivant de la thèse.

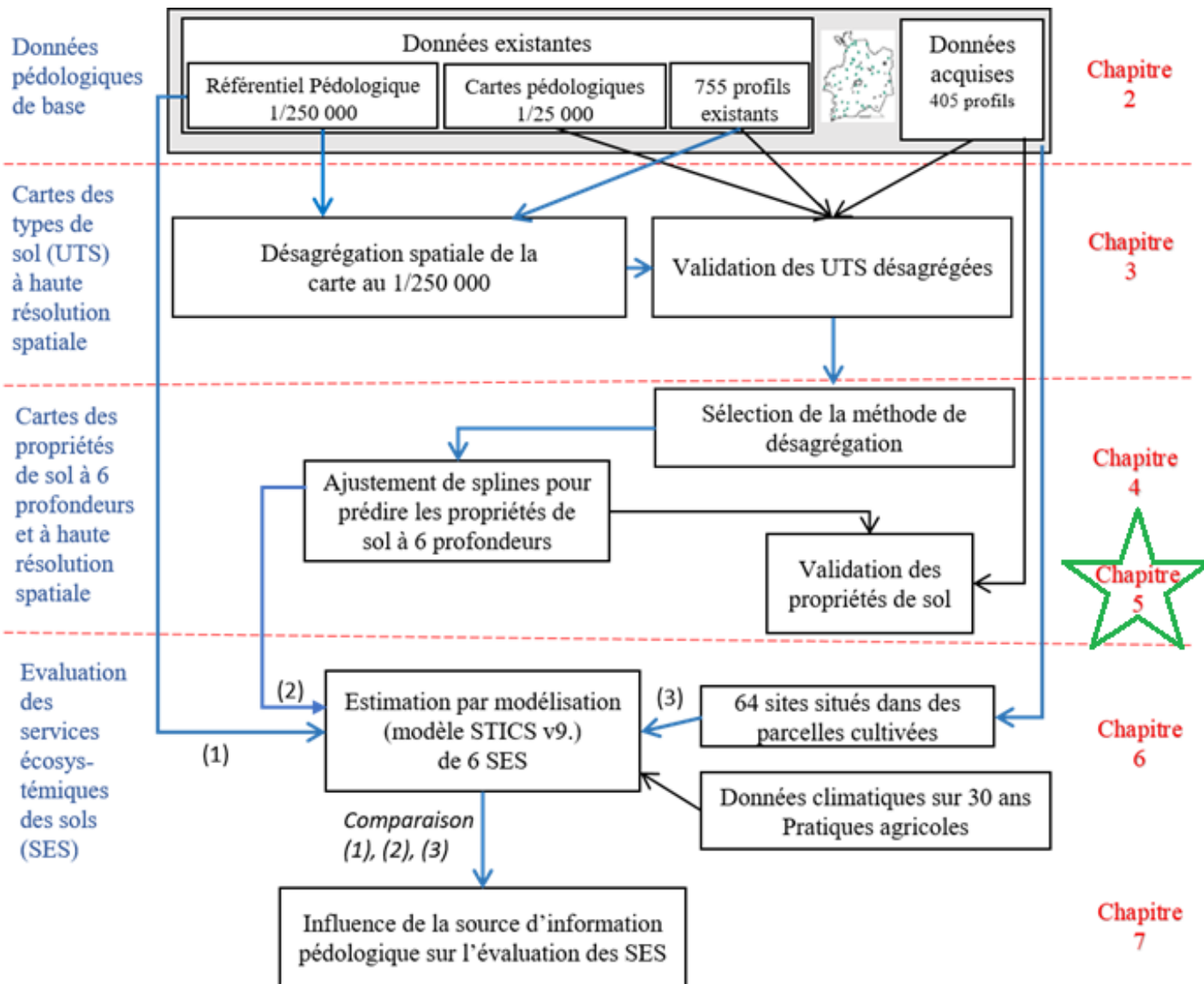


Figure 5.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 5 :
**Validation of digital maps derived from
spatial disaggregation of legacy soil
maps**

Introduction du chapitre 5

Dans les chapitres précédents, le RRP de Bretagne au 1/250 000^{ème} a été désagrégé à une résolution spatiale de 50 m par application de l'algorithme DSMART et en mobilisant des règles *a priori* de distribution des sols dans les paysages. Les UTS spatialisées sont ensuite utilisées pour appréhender la distribution spatiale de plusieurs propriétés de sol aux profondeurs requises par les spécifications *GlobalSoilMap* à l'échelle de la Bretagne entière.

Pour juger concrètement des cartes pédologiques obtenues, il nous faut développer une méthodologie de validation permettant d'évaluer quantitativement leur qualité prédictive. Sur le plan opérationnel, la validation consiste à mobiliser un jeu de données indépendant pour comparer les valeurs prédites et les valeurs observées, à la fois des types des sols et des propriétés des sols aux différentes profondeurs. L'incertitude de prédiction ainsi quantifiée pourra être incorporée dans des outils d'aide à la décision et des modèles de gestion agro-environnementale.

Dans ce contexte, le présent chapitre vise à évaluer la qualité prédictive des cartes de sols désagrégées à l'échelle du département d'Ille-et-Vilaine. Pour l'essentiel, deux approches ont été développées pour répondre à cet objectif. La première consiste à évaluer la qualité prédictive des variables catégorielles des sols, associées aux critères cartographiques considérés à savoir le matériau parental, la classe d'hydromorphie, le type de sol, la classe de profondeur de sol et finalement l'UTS prédite. La deuxième approche, consiste à évaluer la qualité prédictive des propriétés quantitatives de sol à savoir les cinq fractions granulométriques, la teneur en carbone organique, la capacité d'échange cationique et le pH. Ce chapitre a fait l'objet d'une publication scientifique publiée en 2019 dans la revue *Geoderma*¹.

¹Ellili Bargaoui Y., Walter C., Michot D., Saby N, Vincent S., Lemerrier B. 2019 - Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma*, vol. 356, 113907

Validation of digital maps derived from spatial disaggregation of legacy soil maps

Abstract

Spatial disaggregation of soil map units involves downscaling existing information to produce new information at a finer scale than that of the original source. Currently, it is becoming a powerful tool to address the spatial distribution of soil information over large areas, where legacy soil polygon maps are the only source of soil information. Because of the high expense of additional resampling, few studies have sought to validate disaggregated soil maps using independent sampling. This study implemented spatial disaggregation approach to measure the quality of soil property predictions derived from disaggregated soil maps, using stratified simple random sampling of a study area of 6 775 km² (11 strata and 135 soil profiles). In a previous study, the existing legacy soil polygon map of Brittany (France) at 1:250,000 scale was spatially disaggregated at 50 m resolution using an algorithm called Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART), which uses soil-landscape expert rules of soil distribution in space. By fitting equal-area spline functions, soil properties were then estimated at six depth intervals according to *GlobalSoilMap* specifications. To validate disaggregated soil maps, two approaches were developed according to the soil attribute nature (continuous or categorical). For categorical soil properties (soil parent material, soil drainage class, soil type and soil depth class), the overall strict purity (the degree to which all classification criterion are respected) by the most probable STU (Soil Typological Unit) map was estimated at 34%, while the overall average purity reached 70%. The overall partial soil-type purity reached 60%, the overall partial parent material purity reached 78% and the overall partial soil drainage class as well as soil depth class purities reached 65% and 78%, respectively. Continuous soil properties (clay content, fine silt content, coarse silt content, total silt content, fine sand content, coarse sand content, coarse fragments, Cation Exchange Capacity (CEC) and pH) were validated at two soil depth intervals (5-15 and 30-60 cm) using 260 soil samples. In general, soil property predictions were unbiased except for coarse fragments and CEC in the 5-15 cm layer. Validation statistics (R^2 , RMSE, RRMSE and ME) were better for the 30-60 cm layer except for soil particle-size distribution. Thus, differences in prediction accuracies among strata (the

validation support) denote areas where more soil data or better soil prediction models are needed to improve the disaggregation process.

Keywords: Digital soil mapping, spatial disaggregation, soil sampling, validation, France

5.1. Introduction

Soil information derived from legacy maps is often at a spatial resolution too coarse to be useful for soil management decisions and for solving a variety of agricultural and environmental issues (Bui and Moran, 2001). Spatial disaggregation of legacy soil polygon maps has been introduced recently to produce enhanced soil information at a finer scale than that of the original source (Yang et al., 2011; Kerry et al., 2012; Odgers et al., 2014; Subburayalu et al., 2014; Feng et al., 2016; Møller et al., 2019; Ellili et al., 2019; Zeraatpisheh et al., 2019). Such disaggregation techniques are based on machine-learning algorithms modelling the relation between a soil class and a suite of environmental covariates, which are assumed to represent soil forming factors. This approach has been shown to be a powerful tool to deliver soil information over large areas where complex soil polygons, which contain several STU (Soil Typological Unit), are the only source of soil information (Bui and Moran, 2001). Nevertheless, disaggregated digital soil maps, like available conventional soil maps, are inherently uncertain, and their quality has been insufficiently investigated.

To measure the quality of soil maps, it is recommended to compare predictions to independent data not used in the modelling (Chatfield, 1995, Brus et al., 2011) (i.e. “external” or “test” accuracy). To this end, one needs to define the sampling scheme for validation, which comprises two important aspects: sampling design and sampling support.

For the validation strategy, three common approaches are usually found in the literature. The first consists of applying internal validation by splitting the data. This approach randomly selects validation subsamples from the calibration dataset to be used to estimate the accuracy of the fitted model. Generally, these “holdback” data represent a small part (20-30%) of the full dataset. This validation approach was used, for example, by Ramirez et al. (2014) to assess the quality of digital soil maps across a study area covering a total area of 5 km² in Sao Paulo state in Brazil. It was also used by Veronesi et al. (2012) to validate 3D soil compaction maps of the Czech University of life Sciences farm (21 ha) located in the Czech Republic. The second approach consists of using a k-fold cross-validation. It differs from the previous method in that the splitting procedure is repeated several times, which makes it more efficient (Hastie et al., 2009; Stoorvogel et al., 2009; Biswas and Zhang, 2018). Leave-one-out cross-validation is the most common form of cross-validation, ensuring that each sample point can be used as a validation dataset: one sample point is left, while the rest is used

to calibrate the model. This approach was explored by Malone et al. (2009) to validate digital soil maps depicting soil organic carbon (SOC) content and available water capacity across a 1500 km² area in Sydney, Australia. Lacoste et al. (2014) also used cross-validation on their training datasets to assess the quality of high-resolution 3D maps of SOC content over a heterogeneous agriculture landscape in Brittany, France. In contrast to the two previous methods, the third method involves collecting new samples using probability sampling to assess model accuracy.

The first two approaches are often used because quality measures and their standard errors can be obtained easily without additional sampling. However, they may provide unbiased and valid estimates of map accuracy only if the locations of sampling sites are selected by probability sampling, even if the subset is randomly selected from the dataset (Brus et al., 2011). Indeed, spatial auto-correlation will generally occur within the prediction errors, and the calibration dataset itself can be biased because the splitting process is not statistically optimised. Therefore, Brus et al. (2011) recommended collecting additional independent data according to a probability sampling strategy. Nevertheless, due to the high expense of additional resampling, few studies have validated the accuracy of spatial model using independent sampling as done by Walter (1990), Kempen et al. (2009), Brus et al. (2011) and Keskin and Grunwald (2018). In 2009, Kempen et al. explored the use of multinomial logistic regression for digital soil mapping to update an existing soil map and used additional fieldwork to assess the quality of the updated soil map. Brus et al. (2011) validated a soil-class map of the Drenthe province of the Netherlands (268,000 ha) using 150 validation locations selected by stratified simple random sampling. Stoorvogel et al. (2009) also used random transect sampling to assess the accuracy of a quantitative soil map of the Nioro study area (81,600 ha), Senegal, depicting SOC content of the topsoil.

Another important aspect for validating a digital soil map is the type of sampling units and sample support. In most digital soil mapping studies, using a point support is the standard practice to assess the quality of the soil maps produced. Therefore, even at large areal extents, digital soil maps are validated over a small area using bulk soil samples and soil cores rather than management-related spatial supports (e.g. entire fields or management units). In the digital soil mapping literature, few studies have considered a non-point support in the validation procedure. For example, Bishop et al. (2015) assessed the quality of digital soil maps depicting clay content using an independent validation dataset collected following stratified random sampling. They investigated three spatial supports: i)

point, ii) 48 m blocks and iii) soil-land use complexes. They found that point supports yielded the lowest measures for assessing digital soil map quality, while soil-land use complexes achieved the highest. In another study, Stoorvogel et al. (2009) validated a map of topsoil SOC content in Senegal using validation samples collected at a block support of 30 x 30 m, which corresponds to the mean size of fields in their study area. Each composite soil sample had a support of 900 m² and was derived by thoroughly mixing five subsamples taken within a 12 m radius to capture soil short-range variability. Saby et al. (2008) clearly showed a strong relation between within-site variability and site area for soil monitoring sites in Europe. To integrate short-range spatial variation, soil inventory programs usually follow a composite sampling approach. For example, in the framework of the Land Use and Cover Area frame Statistical survey (LUCAS) a harmonized soil dataset was collected over the extent of the European Union following a composite sampling approach (Ballabio et al., 2016). At the national scale, the French soil monitoring network (RMQS) adopted a composite sampling strategy as well (Saby et al., 2014).

The quality of digital soil maps can be assessed by several measures, which depend mainly on whether the soil information is categorical or continuous. To assess the quality of soil maps in the Netherlands, Marsman and De Gruijter (1986) calculated multiple quality indicators for categorical information such as the partial purity (the degree of concordance between observed and predicted of a classification criterion) of several classification criterion (subgroup, sand classes, loam classes and groundwater classes), the mean purities of soil classification criterion and the strict purity. The accuracy of continuous soil property predictions by soil maps is generally depicted by common statistics of the cumulative spatial error such as the mean error, the variance error, absolute error and the mean squared error.

The objective of this study was to develop a method to measure the accuracy of a 50 m resolution disaggregated soil maps and their derived soil properties maps over a large area (6 775 km²), using an independent validation dataset whose sampling locations were selected using a probability sampling strategy. Like Odgers et al. (2015) continuous soil properties were derived from disaggregated soil maps to address the spatial distribution of continuous soil properties (pH, SOC content, soil distribution size, Cation Exchange Capacity (CEC)) over regular depths up to 200 cm. Moreover, the accuracy of disaggregated maps depicting predicted STU (Odgers et al., 2015; Chaney et al., 2016, Møller et al., 2019) as well as categorical soil attributes (soil type, soil drainage class,

parent material and soil depth class) were assessed to determine how the accuracy of soil maps varies with classification criteria selected. In our context, some classification criterion like soil depth and soil drainage behaviour are relevant to characterize the agronomic potential of soils. For example, insufficient soil drainage increases the risk of soil compaction and reduces nutrient availability. In addition, predictions of soil properties are the main inputs of decision-making tools to sustainably manage and solve environmental issues. Thus, maps of functional soil properties with known accuracy are needed to provide a simple guide for non-soil specialist agriculture and stockholders. Overall, two approaches were developed to validate both categorical maps depicting predicted soil typological units (STUs) with their associated classification criterion and continuous maps depicting soil properties at two soil depths.

5.2. Materials and methods

5.2.1. Study area

The study area is the Ille-et-Vilaine department, in eastern Brittany, France (NW France, 47° 40' to 48° 40' N, 1° to 2° 20' W) (Figure 5.1). It has a total area of 6 775 km², which is drained mainly by its major rivers (Ille and Vilaine rivers) and their tributaries. The central and coastal parts of the study area have low elevation, usually less than 50 m above sea level in the coastal zone and the valleys and less than 100 m elsewhere. The western part of the department has higher elevations, peaking at 256 m. The mean annual rainfall is about 650 mm and the annual temperatures average 11.4 °C. The study area is part of the Armorican Massif (BRGM, 2009), whose geology is complex: intrusive rocks (granite, gneiss and micaschist) in northern and north-western zones, sedimentary rocks (sandstone, Brioverian schist) in central and southern zones, and superficial deposits (aeolian loam, alluvial and colluvium deposits) overlaying bedrock formations with decreasing thickness from north to south. This high geological heterogeneity generates high soil variability over short distances. The soils in the study area include Cambisols, Stagnic Fluvisols, Histosols, Podzols, Luvisols and Leptosols according to the World Reference Base of Soil Resources (IUSS Working Group WRB, 2014). The main land uses are annual crops (e.g. maize, wheat, barley) and temporary or permanent

grasslands, but the study area also includes woods and natural areas. In this study, the built environment was not considered.

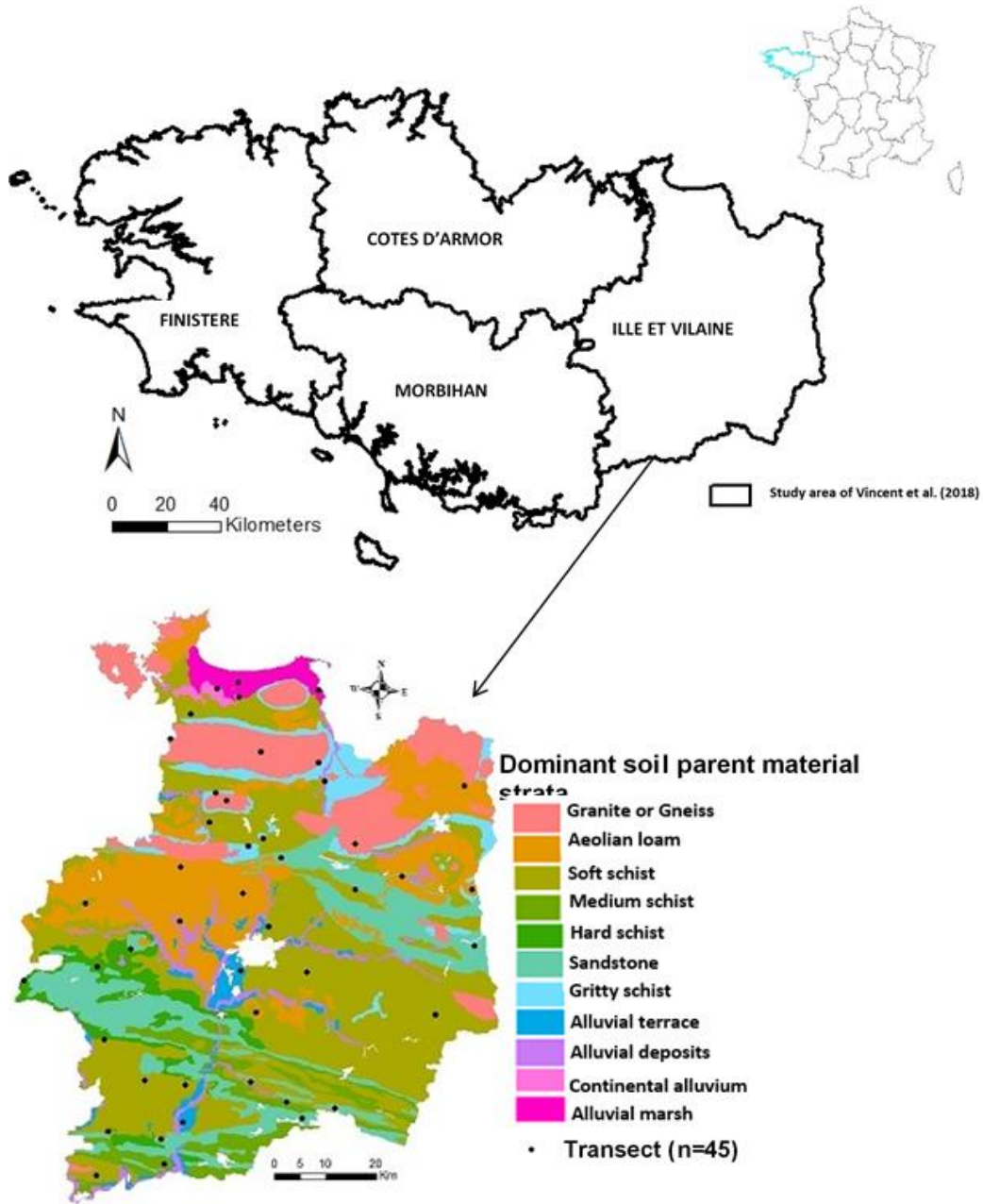


Figure 5.1 : Locations of the 45 transects by stratified simple random sampling based on dominant soil parent material strata.

5.2.2. Disaggregated soil map

To address soil spatial distribution in Brittany, a regional database at 1:250,000 scale called the “Référentiel Régional Pédologique” was developed in 2012 in the “Sols de Bretagne” project. This regional database defines a set of polygons with crisp boundaries commonly called Soil Map Units (SMUs). SMUs are complex, namely each SMU contains several soil types called Soil Type Units (STUs) in known proportions. SMUs are defined as areas with homogeneous soil-forming factors, such as morphology, geology, and climate. Each STU is vertically organized into strata. The strata are spatial horizons describing the vertical structuration of STUs. Pedological features of SMU, STU and strata including depth and thickness, soil organic carbon content, CEC, pH, and 5-particle size fractions are collated in a relational database called Donesol (INRA Infosol, 2014). The Brittany coarse soil map contains 341 soil map units and 320 STUs

In an earlier study (Vincent et al., 2018), the existing 1:250,000 legacy soil map of Brittany was disaggregated at 50 m resolution using the algorithm DSMART and soil-landscape relations. This study yielded a set of rasters depicting the three most probable STUs and their probability of occurrence within each pixel of 0.25 ha. These three most probable STUs allowed capturing the most variability in predictions. The most probable STU in a grid cell is the one that was most frequently predicted based on 50 iterations (Figure 5.2). Similarly, the 2nd STU and the 3rd STU were respectively, the 2nd and the 3rd most frequently predicted STU in a grid cell. Therefore, the three most probable STUs are those predicted with the three highest probabilities of occurrence. Restricted to the study area, 98 SMUs were disaggregated to spatially delineate 158 STUs occurring within the SMUs. In our study, independent soil data was used to validate the three most probable soil maps with their derived soil properties maps, covering the study area.

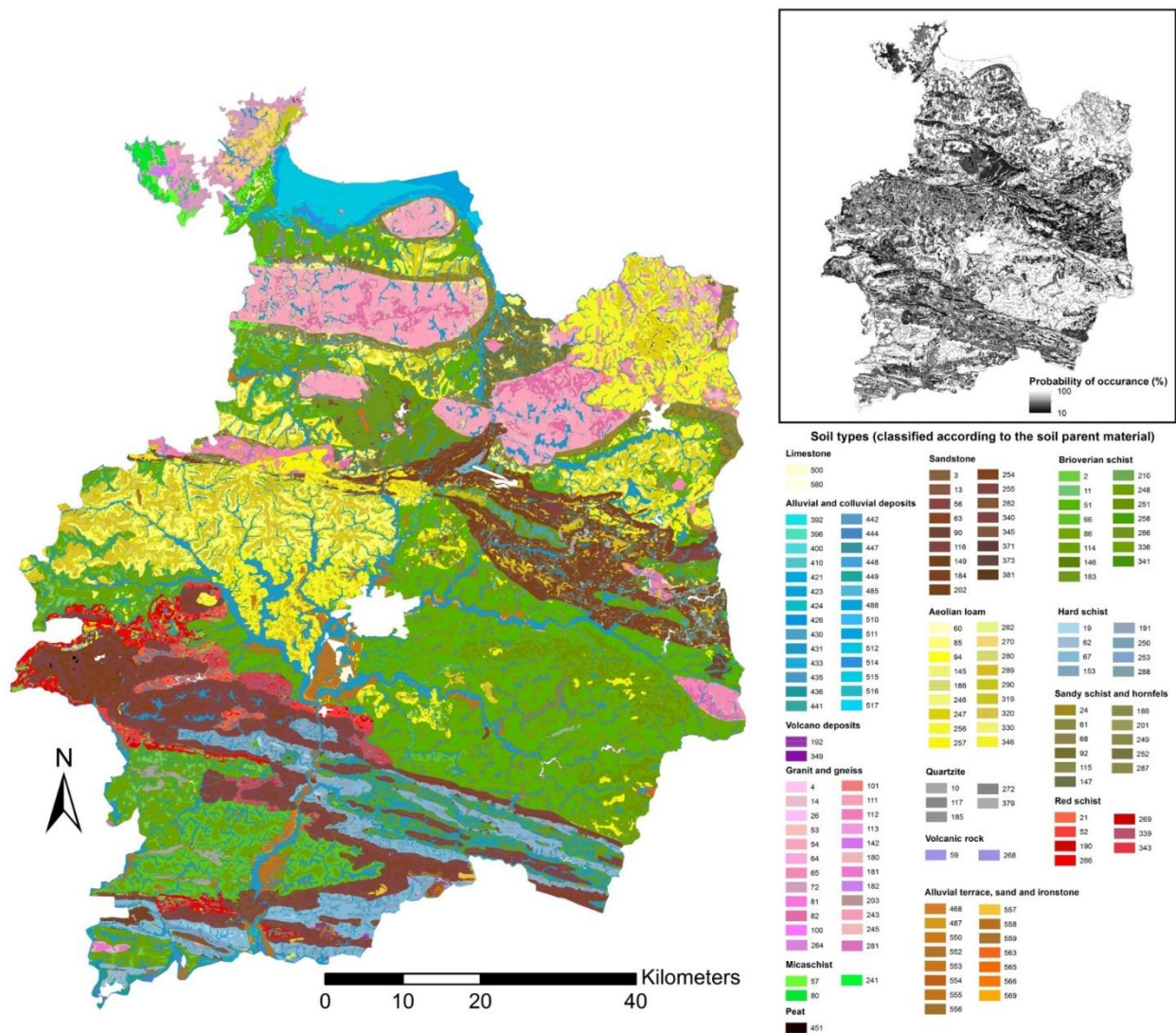


Figure 5.2 : The 50 m resolution disaggregated soil map showing the most probable soil typological unit and associated probability of occurrence.

5.2.3. Soil property estimation

Continuous soil properties were mapped at the regional scale from the 1:250,000 disaggregated map (Vincent et al., 2018) according to *GlobalSoilMap* depth intervals (0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm and 100-200 cm). The first step of our study standardized the depth of STU horizons by fitting equal-area spline functions (Bishop et al., 1999) using the R package GSIF (Hengl,

2013). The equal-area spline function respects mean values of soil properties and ensures continuous variation in soil properties with depth (Malone et al., 2009). The result is, for each STU, a set of interpolated values of soil properties for the required depth intervals up to 200 cm that equals the mean of the intervals.

Soil properties were estimated as the weighted mean of modal values of the reference soil property, whose weights were the probabilities of occurrence of the relevant STU (Eq 5.1). In our study, we estimated soil properties in space using the property values and probabilities of occurrence of the three most probable STUs, as follows:

$$\hat{y}(x_i) = \frac{\sum_{k=1}^3 (P_{STU_k, x_i}) * y(STU_k, x_i)}{\sum_{k=1}^3 (P_{STU_k, x_i})} \quad (\text{Eq 5.1})$$

where $\hat{y}(x_i)$ is the predicted soil property for grid cell (xi), $y(STU_k, x_i)$ is the reference soil property value associated with STU k=1, 2, 3 predicted at all (xi) and (P_{STU_k, x_i}) is the probability of occurrence of STUk at the given grid cell (xi).

5.2.4. Development of soil dataset for map validation

Sampling locations were selected using a stratified simple random sampling strategy, in which the mapped area was subdivided into subareas called strata, and from each stratum a simple random sample was selected. Eleven strata were obtained by grouping the 96 SMUs of the original soil map by similar dominant soil parent material: granite or gneiss, Aeolian loam, soft schist, sandstone, gritty schist, alluvial terrace, alluvial deposits, continental alluvium, and alluvial marsh (Figure 5.3). A total of 45 sampling locations were selected with a per-stratum sample size proportional to the area of each stratum. Each stratum had a minimum of two sampling locations (Table 5.1). Locations for which sampling permission was denied or which proved otherwise impossible to sample were replaced with randomly selected locations within the same stratum.

Table 5.1 : Characteristics of the 11 strata used for stratified simple random sampling to produce the validation dataset, dominant parent material, percentage of the study area and number of transects sampled.

Stratum	Dominant parent material	Percentage of study area (%)	Transects (n=45)
1	Granite or gneiss	12	5
2	Aeolian loam	17	7
3	Soft schist	35	13
4	Medium schist	5	2
5	Hard schist	4	2
6	Sandstone	14	6
7	Gritty schist	5	2
8	Alluvial terrace	2	2
9	Alluvial deposits	3.6	2
10	Continental alluvium	0.4	2
11	Alluvial marsh	2	2

At each selected site, a transect along the hillslope was defined and divided into three sections: upslope, midslope and downslope (Figure 5.3). A principal sampling location was then randomly selected within each section and two additional points were selected in a random direction at 20 m distance from the main (principal) location. This approach was followed to capture the local short-range spatial variability. The set of three locations, characterising each section, defined a plot in our study (Figure 5.3). The validation dataset comprises a total of 135 plots.

Validation sites were located using a global positioning system (GPS) with a mean absolute error of 2 m in each direction. Within each section, the main site was described from auger borings up to 200 cm in depth, and the two other soil profiles were described up to 120 m in depth. Soil morphology was described at each sampling site to determine each horizon: upper and lower limits, organic matter content, soil moisture, compacity, matrix and mottle colours according to Munsell soil colour chart, coarse fragment percentage, and soil texture class according to the GEPPA texture triangle (Baize, 2000). Finally, soil depth, soil type, soil drainage class and soil parent material were defined according to Rivière et al. (1992) and the French soil classification system (Baize and Girard, 2008). The soil type refers to the identification of diagnostic horizons depicting pedogenetic processes. The soil type can be for example Cambisol, Fuvisol, Albeluvisol. Meanwhile, the STU nomenclature

reflects different information at the same time as the weathering degree of soil parent material, the redoximorphic conditions, and the soil depth. In our study, three drainage classes were distinguished: well drained, moderately drained and poorly drained. Similarly, only two soil depth classes were distinguished: deep (≥ 60 cm) and shallow (< 60 cm). After field description, each of the 405 soil profiles was individually allocated to a suitable STU.

Soil samples for physico-chemical analyses were collected at six depth intervals according to *GlobalSoilMap* specifications: 0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm. For each plot and each soil depth interval, a composite soil sample was collected from the three points sampled within the 20 m radius.

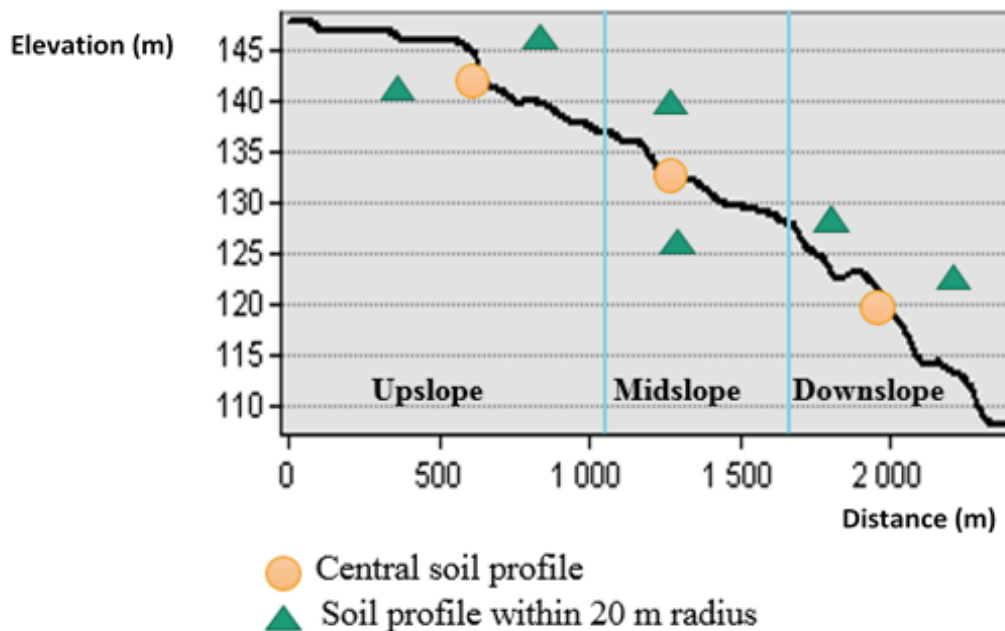


Figure 5.3 : Diagram of the transect sampling strategy along the hillslope.

5.2.5. Laboratory analysis

In our study, for budgetary reasons, only two soil layers were considered (5-15 cm and 30-60 cm). The 5-15 cm soil layer allowed characterizing soil properties of topsoil genetic horizons while the second soil layer 30-60 cm allowed characterizing soil properties of deeper soil horizons. Overall, considering the maximum soil depth, a total of 260 soil samples were air dried then sieved to 2 mm; the resulting fine earth and gravel were weighed to determine the percentage of coarse material

present in soils. All samples were analysed according to standard methods to determine their particle-size distribution in five classes (NF X 31-107), SOC content by dry combustion (Thermoscientific Finnigan EA 1112 Flash elemental analyser) (NF ISO 10694), CEC (NF X 31-130) and pH 1:5 H₂O (soil pH in water with 1:5 soil-to-water ratio, NF ISO 10390).

5.2.6. Measures of map quality

Figure 5.4 shows the general diagram followed to compute different maps quality measures according to the type of soil property map produced. The following sections detailed each approach as well as the quality indicators retained.

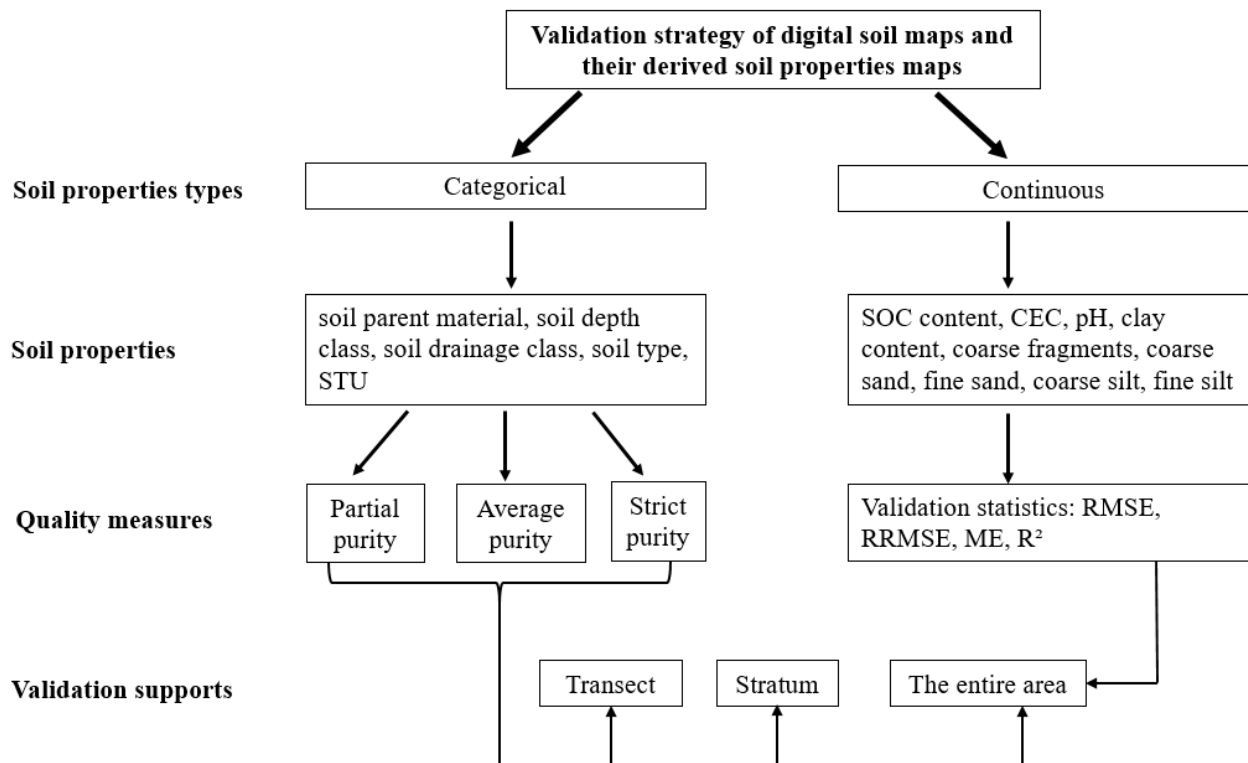


Figure 5.4 : Diagram of the validation strategy to compute quality measures of digital soil maps and their derived soil properties maps.

5.2.6.1. Quality measures for continuous soil properties

The prediction performance of continuous soil properties was assessed by several quality measures relating the observed value of a soil property derived from in situ sampling to its corresponding

prediction for each depth interval. These statistics comprise the root-mean-squared error (RMSE) (Eq 5.2), relative-root-mean-squared error (RRMSE) (Eq 5.3), mean error (ME) (Eq 5.4), and coefficient of determination (R^2) (Eq 5.5).

For stratified simple random sampling (De Gruijter et al., 2006), the quality indicator is estimated as the weighted mean per stratum:

$$\text{RMSE} = \sum_{h=1}^H w_h \widehat{rmse}_h = \sum_{h=1}^H w_h \sqrt{\frac{1}{n_h} \sum_{i=1}^{nh} (\hat{y}(x_i) - y(x_i))^2} \quad (\text{Eq 5.2})$$

$$\text{RRMSE} = \sum_{h=1}^H w_h \widehat{rrmse}_h = \sqrt{\frac{1}{n_h} \sum_{i=1}^{nh} \frac{1}{y(x_i)^2} (\hat{y}(x_i) - y(x_i))^2} \quad (\text{Eq 5.3})$$

$$\text{ME} = \sum_{h=1}^H w_h \widehat{me}_h = \frac{1}{nh} \sum_{i=1}^{nh} (\hat{y}(x_i) - y(x_i)) \quad (\text{Eq 5.4})$$

For R^2 , there is no estimator for stratified simple random sampling.

$$R^2 = \frac{\sqrt{\sum_{i=1}^n (y(x_i) - \bar{y}(x_i))^2}}{\sqrt{\sum_{i=1}^n (y(x_i) - \bar{y}(x_i))^2}} \quad (\text{Eq 5.5})$$

where W_h denotes the relative area of stratum h , $h=1 \dots H=11$, $y(x_i)$ denotes the observed value of the soil property at validation site (x_i), $\hat{y}(x_i)$ denotes the predicted value of the soil property for the 50 m cell containing x_i and nh is the number of validation soil samples collected at each depth interval within each stratum.

5.2.6.2. Quality measures for categorical soil properties

5.2.6.2.1. Validation supports

In our study, quality measures of categorical soil attributes were estimated at three different scales:

- Section, to detect whether the quality of soil maps varied by hillslope position
- Mapping stratum, the basis of the stratified simple random sampling
- Global considering the entire study area.

Thus, we calculated measures both globally over the entire study area and on average within each stratum. This procedure provided an overall assessment of predictive performance according to the type of support.

5.2.6.2.2. Quality measures

Categorical soil attributes were validated using three disaggregated soil maps depicting the 1st, 2nd or 3rd most probable STUs, respectively. We used several methods to calculate purity values, considering four classification criteria: parent material, soil type, soil drainage class (i.e. redoximorphic conditions of soils) and soil depth class. All quality measures were detailed in the following section:

- Partial purity: the percentage of field observations of a given criterion that equal the value on the disaggregated map. Partial purity was calculated for each soil profile and equalled 1 if the predicted classification criterion equalled the observed classification criterion or 0 if not.
- Strict purity: the percentage of field observations for which all four-classification criteria equal those of the disaggregated map. This measure was calculated for each soil profile.
- Mean purity: the arithmetic mean of the partial purities of the four criteria.

The partial purity of each classification criterion equalled 1 if correctly predicted for at least one of the three profiles prospected within a 20 m radius. Hence, each section had a single purity value for each criterion, corresponding to a binary value of 0 or 1.

At the transect level, the partial, strict and mean purities were calculated by averaging the purity values of all sections included.

At the stratum level, partial purities were calculated for each of the four criteria by averaging the corresponding purity values of all n transects included in a given stratum. As with partial purity, the strict purity of the 1st, 2nd and 3rd STU maps equalled the mean of all strict purities for all n transects within the stratum. Mean purity was derived from the partial purities and equals the arithmetic mean of the four partial purities. The standard deviation (Se) of each partial purity within strata was also computed using Eq 5.6:

$$\text{Se} (P_{hk}) = \sqrt{\sum_{i=1}^{n_h} \frac{1}{(n_h-1)} (P_{ihk} - \bar{P}_{hk})^2} \quad (\text{Eq 5.6})$$

where P_{hk} denotes the partial purity of stratum h for a given map k (1st, 2nd and 3rd predicted STU), P_{ihk} denotes the purity value of all transects within stratum h of map k , n_h denotes the number of sections in stratum h and \bar{P}_{hk} denotes the mean purity value in the given stratum.

For the stratified simple random sampling, the overall partial purity of each classification criterion (parent material, soil type, soil drainage class, and soil depth class), the overall strict purity and the

overall mean purity were estimated as the mean of the strata purities weighted by their respective area using Eq 5.7 (De Gruijter et al., 2006):

$$\bar{P}_k = \sum_{h=1}^H w_h P_{hk} \quad (\text{Eq 5.7})$$

where \bar{P}_k denotes the overall purity, $w_h = Ah/A$ where A denotes the study site area and Ah denotes the stratum area, P_{hk} denotes the mean of purity values (partial, average and strict) in stratum h and k denotes the STU map. The associated standard error (Se) was estimated using Eq 5.8:

$$\text{Se}(\bar{P}_k) = \sqrt{\sum_{h=1}^H \frac{1}{(n_h)} (w_h^2 s_h^2)} \quad (\text{Eq 5.8})$$

where s_h^2 denotes the variance of purity values in stratum h calculated using Eq 6.

5.3. Results

5.3.1. Soil spatial variability over short distances

Overall, 92% of the plots showed only one parent material, thus short-range variability appeared less pronounced for this criterion; similar percentages were observed for soil drainage class and soil depth class (Figure 5.5). In contrast, STUs generally appeared highly variable, as only 40% of the validation plots had a single STU, while 55% and 5% of the validation plots had two or three STUs, respectively. Overall, short-range spatial variation in STU allocated was higher than for the other classification criteria.

5.3.2. Purity measures

In general, the 1st most probable STU map had the highest overall purities (Table 5.2). For instance, overall purity of soil parent material was estimated at 78%, 44% and 31% for the 1st, 2nd and 3rd STU maps, respectively. Likewise, overall purity of soil depth class was higher for the 1st STU maps (78%) than for the 2nd (56%) or 3rd STU maps (48%). For overall purity of soil type, the 1st STU map reached 60% and those of the 2nd and 3rd STU maps reached 31% and 26%, respectively. For drainage class purity, the 1st STU map remained the most accurate (65%), followed by the 2nd (35%) and 3rd STU maps (29%). STU purity appeared to be low for the three STU maps, being 23%, 5% and 3% for the 1st, 2nd and 3rd STU maps, respectively.

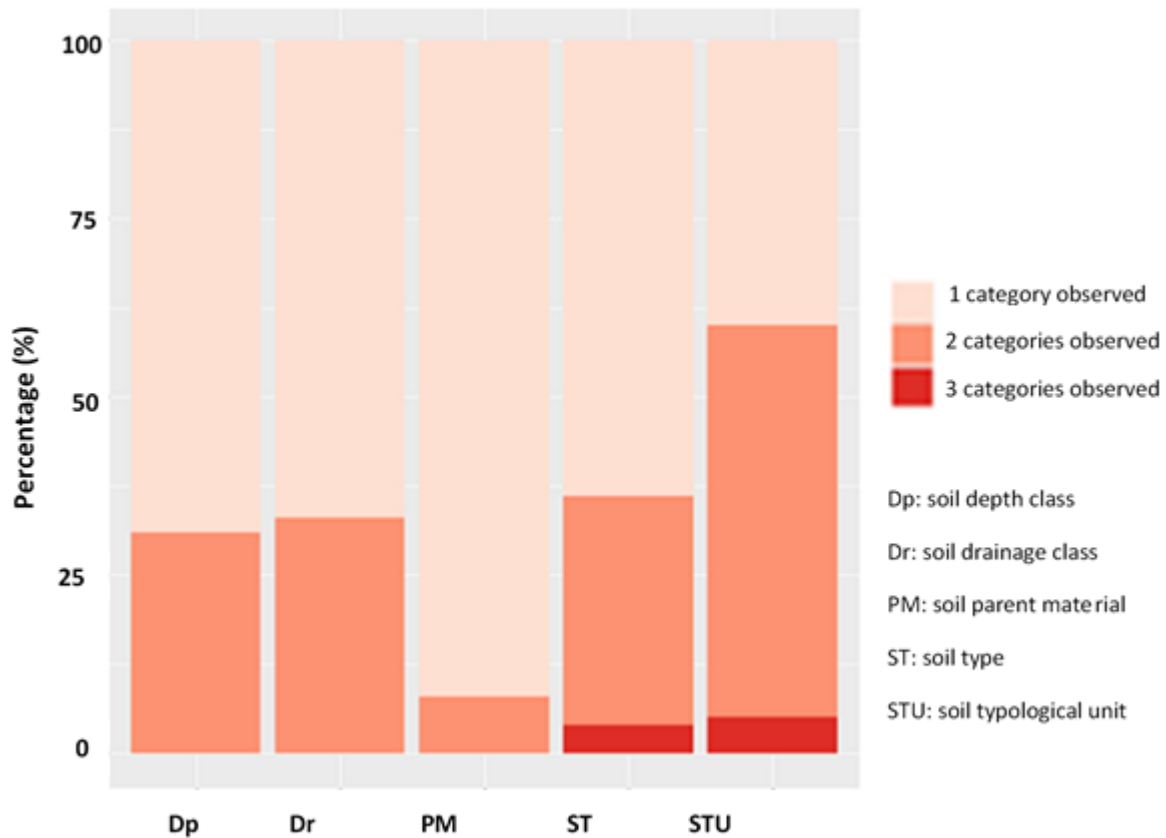


Figure 5.5: Percentage of validation plots in which 1, 2 or 3 categories of a given soil classification criterion (soil depth, drainage class, parent material, soil type, soil typological unit) were observed within a 20 m radius.

Table 5.2 : Estimated partial, strict and mean purities (%) for the three soil maps over the entire study area.

Soil map	Partial purities					Strict purity	Mean purity
	Parent material purity	Drainage class purity	Soil type purity	Soil depth class purity	STU purity		
STU1	78 (6.6)	65 (4.5)	60 (4.5)	78 (4)	23 (4.4)	34	70
STU2	44 (4.7)	35 (3.3)	31 (3.6)	56 (4.9)	5 (2)	12	42
STU3	31 (4.1)	29 (2.9)	26 (3.5)	48 (4.8)	3 (1.8)	5	34

Numbers in brackets are design-based estimates of standard deviations for the 45 transect purity values.

Over the entire study area, strict and mean purities had a positive linear relation with the mean probability of STU occurrence (Figure 5.6). For the three soil maps, strict purity appeared to be relatively low (Table 2, Figure. 6). The highest strict purity (34%) was reached by the most probable STU map followed by the 2nd STU map (12%). The 3rd most probable STU map had the lowest strict purity (5%). Meanwhile, mean purity was almost three times that of strict purity for all soil maps except for the 1st STU map, for which it was only two times that of strict purity. Mean purity equalled 70%, 42% and 34% for the 1st, 2nd and 3rd STU map, respectively.

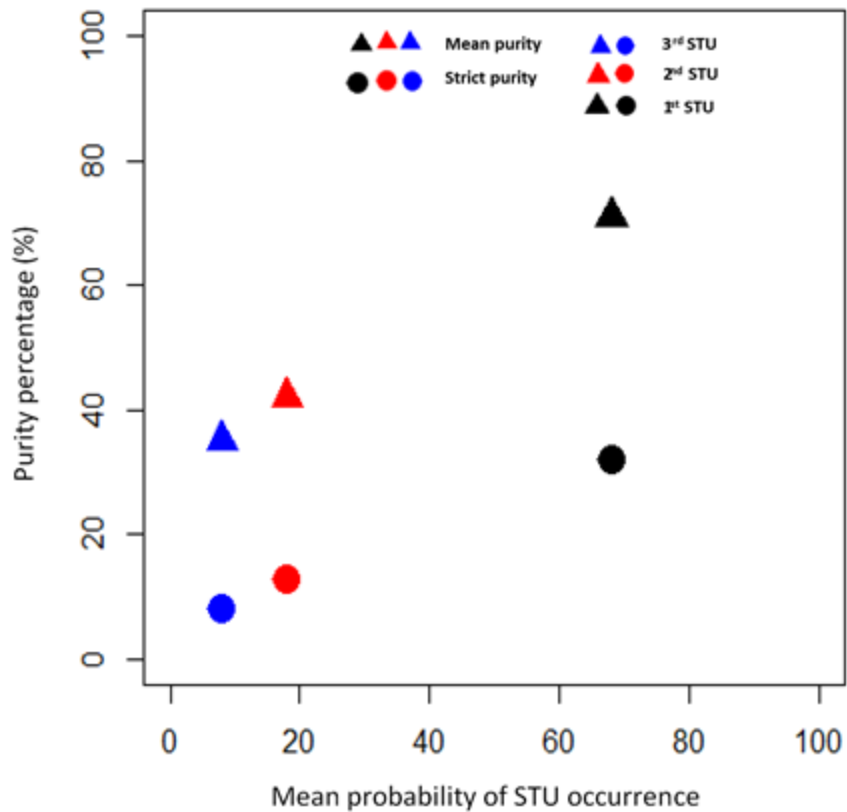


Figure 5.6 : Relation over the entire study area between strict and mean purities (%) and probability of occurrence of the three most probable soil typological units (STUs).

Aeolian loam, hard schist and alluvial terrace strata had the lowest strict purities (< 20%) (Table 5.3). In contrast, medium schist, continental alluvium and alluvial marsh strata had the highest strict purities (> 67%). Mean purity exceeded 65% for all strata except hard schist and alluvial deposits, for which it was 54% and 58%, respectively. Continental alluvium (92%) and alluvial marsh strata

(96%) had the highest mean purities. For drainage class, the partial purity for granite or gneiss, soft schist, medium schist, sandstone, alluvial terrace, continental alluvium and alluvial marsh strata were the highest (> 65%). Aeolian loam, gritty schist and alluvial deposit strata had partial purities $\geq 50\%$, while that of hard schist was estimated at 33%.

Table 5.3 : Design-based estimates of purities of soil classification criteria, strict purity and mean purity for selected strata for the 1st most probable STU map.

Stratum	Parent material purity	Drainage class purity	Soil type purity	Soil depth class purity	Strict purity	Mean purity
Granite or gneiss	72 (46)	72 (25)	67 (21)	78 (17)	39	72
Aeolian loam	81 (40)	52 (37)	52 (32)	81(37)	20	66
Soft schist	77 (42)	67 (19)	59 (27)	80 (25)	33	71
Medium schist	100 (0)	83 (23)	83 (23)	83 (23)	67	87
Hard schist	83 (40)	33 (0)	50 (23)	50 (23)	17	54
Sandstone	73 (45)	73 (43)	53 (38)	67 (0)	40	67
Gritty schist	67 (51)	50 (23)	83 (23)	83 (23)	33	71
Alluvial terrace	100 (0)	83 (23)	17 (23)	83 (23)	17	71
Alluvial deposits	83 (40)	50 (23)	33 (0)	67 (0)	33	58
Continental alluvium	100 (0)	67 (47)	100 (0)	100 (0)	67	92
Alluvial marsh	100 (0)	83 (23)	100 (0)	100 (0)	83	96

Numbers in brackets are design-based estimates of standard errors for the 45 transect purity values.

For soil type purity, soil type was generally predicted well, except for alluvial terrace strata, considering both the 1st and 2nd probable STU (Figure 5.7). Alluvial terrace strata had the lowest purity (17%), which markedly affected the strict purity (17%), whereas the mean purity remained high (71%). For alluvial marsh and continental alluvium strata, the distribution functions were on the

maximum, indicating that all soils developed from these parent materials were correctly predicted by the 1st most probable STU.

When comparing the distribution functions of soil type purity within strata, medium schist and gritty schist strata had similar distributions for the 1st STU. Meanwhile, medium schist and alluvial terrace strata had similar distributions for the 2nd STU. Furthermore, hard schist, sandstone and alluvial deposit strata had a wide range of partial purities. This variation was more pronounced for granite or gneiss, hard schist and gritty schist strata for the 2nd STU (Figure 5.7).

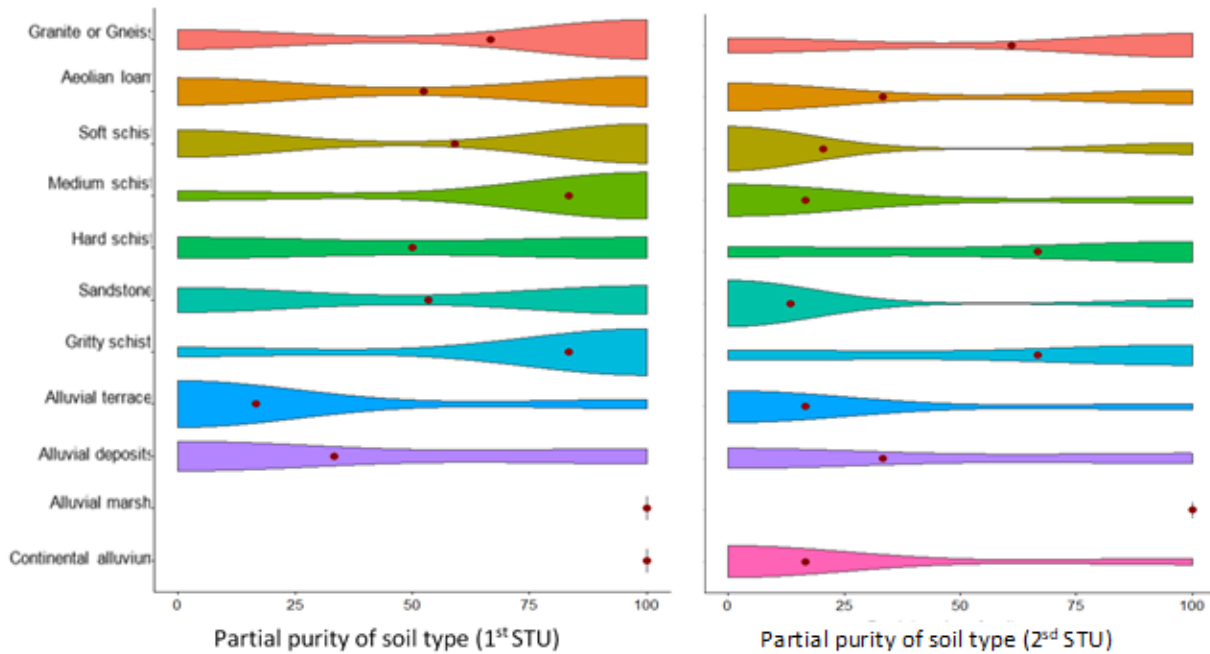


Figure 5.7 : Distribution functions of soil type purity and their mean values (red dots) by stratum for the (left) 1st and (right) 2nd soil typological unit (STU) predicted.

At the local scale, the partial purity of each classification criterion did not vary significantly among hillslope positions (upslope, midslope and downslope) for any of the three disaggregated soil maps (Table 5.4). The χ^2 test comparing the proportions of good prediction of each classification criterion was not significant ($\alpha=0.05$). This suggested that the effect of hillslope position in the disaggregation process is not important for all classification criteria (Table 5.4). In addition, confidence intervals of good soil type predictions overlapped for all sections and soil maps considered (Figure 5.8); therefore, there was no significant effect of hillslope position ($\alpha=0.05$).

Overall, the disaggregation algorithm predicted soil spatial distribution along hillslope positions with the same performance, and the 1st most probable STU has the best validation measures over the study area.

Table 5.4 : Results of the χ^2 test performed to compare proportions of good prediction of classification criteria along hillslope positions for the three soil typological unit (STU) maps ($\alpha=5\%$).

	1 st STU map	2 ^{sd} STU map	3 rd STU map
Classification criterion	Probability (p value)		
Parent material	0.40	0.41	0.65
Soil type	0.80	0.30	0.14
Drainage class	0.96	0.68	0.25
Soil depth class	0.30	0.48	0.51

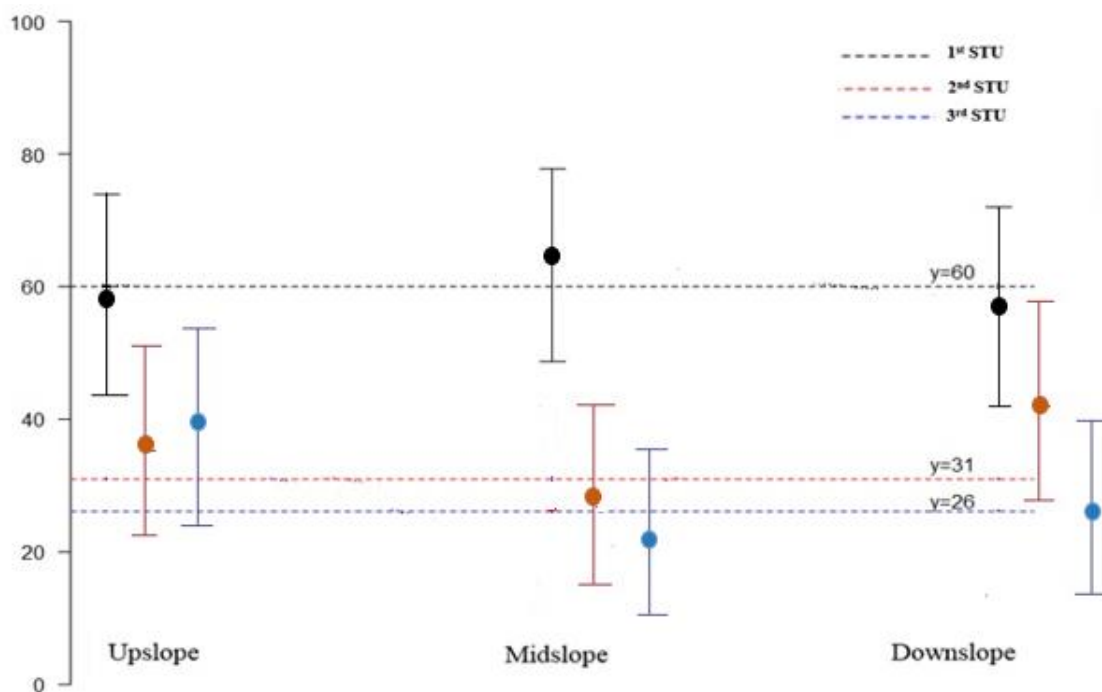


Figure 5.8 : Mean percentage of good soil type prediction and their 95% confidence intervals for the three most probable soil typological units (STUs) according to hillslope position. Dashed lines denote the overall soil type purity for the entire study area.

5.3.3. Descriptive statistics of continuous soil properties for the validation dataset

In general, the validation dataset covered a wide range of soil property values. For instance, pH ranged from 3.95 to 8.58 (mean = 6.07) at 5-15 cm and from 4.41 to 8.9 (mean = 6.25) at 30-60 cm (Table 5.5). The variability expressed by the standard deviation was relatively constant for all texture fractions and for both depth intervals. SOC content varied more at 5-15 cm, particularly for alluvial deposits, medium schist, and granite or gneiss (Figure 5.9). SOC content decreased sharply and varied less with increasing depth, except for alluvial marsh. The mean of SOC contents, except those of soil sampled from alluvial marsh, did not differ significantly at 5-15 cm even though they clearly differed among parent materials.

Considering both soil depth intervals, hard schist parent material had the lowest pH, whereas marsh parent material had the highest pH. In addition, for medium schist parent material, the distribution of associated pH values was wider at 5-15 cm than at 30-60 cm.

Clay content had a wider distribution for marsh parent material at 5-15 cm, and for medium schist and alluvial deposit parent materials at 30-60 cm. Overall, clay content increased slightly with increasing depth for all parent materials except aeolian loam, hard schist and alluvial marsh.

Table 5.5 : Descriptive statistics of soil properties for the validation dataset at 5-15 and 30-60 cm (n = 135 and 125, respectively).

Depth interval (cm)	Soil property	minimum	median	mean	maximum	standard deviation
5-15	CEC (cmol+ kg-1)	2.33	7.03	9.52	112	11.58
	pH	3.95	5.99	6.07	8.58	0.99
	SOC content (‰)	10.40	23.30	30.34	275	2.97
	Coarse fragments (%)	0.00	12.63	13.40	49.81	9.93
	Coarse sand (%)	0.20	9.30	12.38	48.20	10.90
	Fine sand (%)	1.20	11.90	13.00	40.60	6.24
	Coarse silt (%)	0.70	25.90	26.90	52.30	11.27
	Fine silt (%)	0.25	25.40	24.78	40.50	7.30
	Clay (%)	6.80	19.20	20.90	62.20	8.05
30-60	CEC (cmol+ kg-1)	1.49	4.88	5.90	34.4	4.20
	pH	4.41	6.18	6.25	8.90	1.00
	SOC content (‰)	2.55	7.53	11.84	56.90	12.19
	Coarse fragments (%)	0.00	14.39	16.11	52.37	12.62
	Coarse sand (%)	0.00	9.90	13.47	5.78	12.46
	Fine sand (%)	0.30	11.70	12.87	54.00	6.65
	Coarse silt (%)	0.40	25.00	26.30	49.80	11.44
	Fine silt (%)	0.21	24.80	24.00	42.00	7.92
	Clay (%)	6.00	19.10	20.50	51.40	8.20

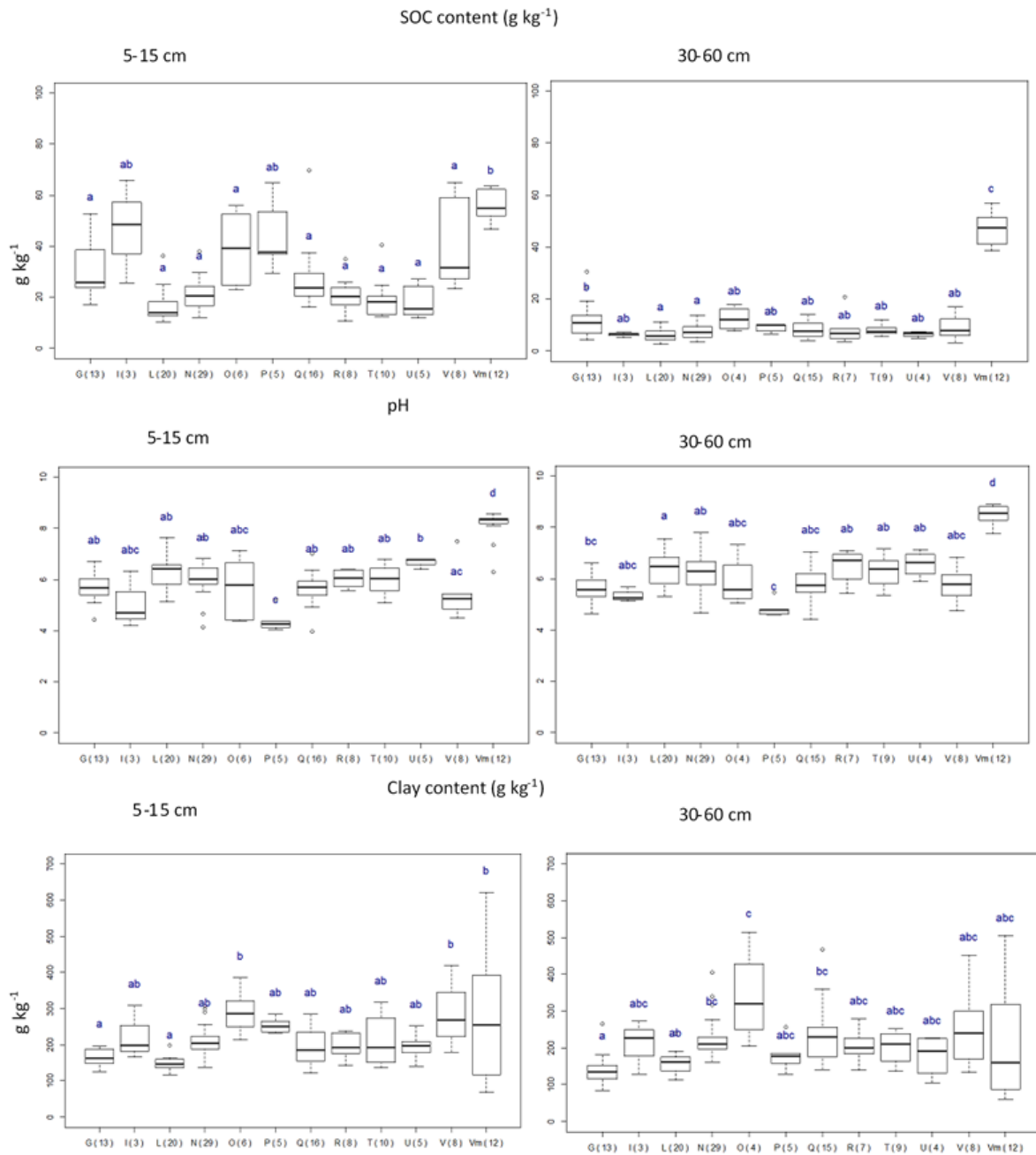


Figure 5.9 : Distribution of SOC content, pH and clay content according to observed parent material for the validation dataset. G: granite, I: gneiss, L: aeolian loam, N: soft schist, O: medium schist, P: hard schist, Q: sandstone, R: gritty schist, T: alluvial terrace, U: colluvium deposits, V: alluvial deposits, Vm: marsh. Whiskers represent 1.5 times the interquartile range. The number in brackets is the number of soil samples collected from each parent material. Means of groups sharing a letter in the group label do not differ significantly at $\alpha=5\%$.

5.3.4. Validation of continuous soil property predictions

At 5-15 cm, MEs of the models were close to zero, suggesting unbiased predictions for pH (0.29), sand content (-0.20%) and fine silt content (-0.20%) (Table 5.6). The MEs of soil texture fractions increased slightly with increasing depth, but those of CEC and SOC content strongly decreased.

R^2 was larger for pH and clay content at both 5-15 and 30-60 cm. For instance, R^2 for clay content was moderate (0.65 and 0.37 at 5-15 and 30-60 cm, respectively). SOC content and coarse fragment percentage had the smallest R^2 . For SOC content, the R^2 was estimated at 0.07 at 5-15 cm and 0.05 at 30-60 cm soil layer. Meanwhile, the R^2 for coarse fragment percentage was estimated at 0.13 and 0.05 at 5-15 cm and 30-60 cm soil depths, respectively.

RRMSE was less than 1 for all soil properties except coarse fragments at 30-60 cm. RRMSE and R^2 were generally of opposite magnitude. For instance, R^2 was 0.43 for pH at 5-15 cm, with an RRMSE of about 0.13. An opposite trend characterised SOC content, which had the lowest R^2 (0.07) and an RRMSE of 0.65.

Overall, sand, fine silt and total silt contents were underestimated regardless of depth interval, whereas coarse fragments were consistently overestimated at both depth intervals. Therefore, the quality of soil property predictions depended on the property considered and was generally better at 30-60 cm than at 5-15 cm except for particle-size distribution.

Table 5.6 : Accuracy indicators of soil property prediction at 5-15 and 30-60 cm (n = 135 and 125, respectively, root-mean-squared error (RMSE), relative-root-mean-squared error (RRMSE), mean error (ME) and coefficient of determination (R²)).

Depth interval (cm)	Soil property	ME	RMSE	RRMSE	R ²
5-15	CEC (cmol+ kg-1)	-3.95	5.12	0.90	0.28
	pH	0.29	0.80	0.13	0.43
	SOC content (g kg-1)	2.15	14.00	0.65	0.07
	Coarse fragments (%)	9.53	12.84	0.86	0.13
	Sand (%)	-0.20	9.94	0.50	0.41
	Coarse silt (%)	0.74	9.20	0.41	0.25
	Fine silt (%)	-0.20	6.70	0.39	0.05
	Total silt (%)	-1.20	9.80	0.23	0.28
	Clay (%)	1.47	5.50	0.26	0.65
30-60	CEC (cmol+ kg-1)	0.06	2.15	0.43	0.34
	pH	0.45	0.78	0.12	0.55
	SOC content (g kg-1)	0.32	5.70	0.69	0.05
	Coarse fragments (%)	8.15	13.98	1.54	0.06
	Sand (%)	-2.06	10.32	0.59	0.39
	Coarse silt (%)	0.19	10.18	0.53	0.17
	Fine silt (%)	-1.88	6.00	0.39	0.25
	Total silt (%)	-1.69	10.95	0.28	0.13
	Clay (%)	3.75	7.53	0.27	0.37

5.4. Discussion

5.4.1. Soil map quality measures

In our study, the quality of disaggregated soil property maps was tested using the common accuracy measures of partial, strict and mean purity (Wilding et al., 1965; Beckett and Webster, 1971; Walter, 1990; Kempen et al., 2009). Partial purity reveals the percentage of the disaggregated map for which

each classification criterion equals the field observation. This measure is a strict assessment in which each error in the validation set is given equal weight. Moreover, it does not consider existing pedological similarities between categories of each classification criterion. For example, the partial purity of soil type is the same regardless of whether there is confusion in the prediction of two taxonomically distant or taxonomically similar soils. Even the limits of these quality measures, they remained relevant to assess the accuracy of disaggregated soil maps, which are mainly used in soil management decisions as well as in the decision-making tools.

The main shortcoming of strict purity is that incorrect prediction of a single classification criterion has the same influence as that of all classification criteria. The lowest strict purity reported in our study resulted from incorrect prediction of one or more classification criteria of soil profiles. For instance, strict purity for the alluvial terrace stratum was 17%, while its associated partial purities exceeded 83%, except for soil type purity (17%). The hard schist strata followed the same trend: its strict purity was 17%, while its partial purities were $\geq 50\%$ except for drainage class purity (33%). Therefore, low strict purity was driven by the low partial purity of a single classification criterion although the partial purities for the three other criterion were high.

5.4.2. Performance of categorical soil attributes prediction by spatial disaggregation

As collecting new soil samples to validate disaggregated soil maps is expensive and time consuming, particularly for large areas, few studies have validated the digital soil maps produced (Minasny et al., 2013). Most of the studies that did were based on relatively small datasets (e.g. $n=150$ (Kempen et al., 2009); $n=123$ (Collard et al., 2014), $n=53$ new + 10 legacy soil profiles (Nauman et al., 2014)), which may have limited their ability to cover all soil distributions. An alternative approach is to use legacy soil profiles, as done by Odgers et al. (2014) and Nauman and Thompson (2014). Although commonly used in digital soil mapping, this approach may have significant hidden sampling bias because rare soils are poorly represented, and legacy datasets were usually not designed probabilistically.

In our study, the overall purity of soil type was estimated at 60% by the 1st most probable STU. Similar findings were reached by Kempen et al. (2009, 58%) and Collard et al. (2014, 65.9%) using multinomial logistic regression and probabilistic sampling designs. The overall strict purity, assessed

using independent soil data, was estimated at 34% (1st STU) and reached 51% considering the three most probable STUs. Odgers et al. (2014) reported similar results using 285 validation legacy profiles and the DSMART algorithm; 22.5% of profiles were predicted correctly with the most probable STU and 50% were predicted as being one of the three most probable STUs. Other studies, such as Holmes et al. (2015), reported low strict purity (20%, 1st STU), while Subburayalu and Salter (2013) reported strict purity that reached 23% using decision trees and 26% using random forest model. However, comparing our partial purity results to those reported in previous studies remains difficult because of the complexity of soils in our study area and differences in the taxonomic classification system considered.

The accuracy of traditional soil maps is also assessed using statistics such as purity and the kappa index (Kempen et al., 2009). Previously, using stratified random sampling, Marsman and De Gruijter (1986) reported strict purity values of 8.0-10.7% depending on the mapping method. Much higher strict purities (39-95% depending on the level of classification) were reported by Wilding et al. (1965). However, the comparison between digital soil maps and conventional maps remains unsatisfactory because the taxonomic classification systems differ, and the validation sampling design was not always defined in each study.

The accuracy of the disaggregated soil map appeared to be lower than that of local existing 1:25,000 soil maps (Walter, 1990). These maps depict the southern part of the Ille-et-Vilaine department and cover almost 2% of our study area. Accurate maps were obtained from field description of soil profiles selected following a dedicated sampling design and produced using the local Armorican Massif auger method (Rivière et al., 1992). To assess the quality of existing 1:25,000 soil maps, Walter (1990) validated the three most abundant SMUs by randomly selecting 120 validation sites per SMU. Soils were described using the Armorican Massif auger method, and the quality of soil maps was assessed according to the partial purity of each classification criterion, and the strict and mean purities. The study of Walter (1990) and ours differed greatly in the partial purity of soil type. The partial purities of the existing map exceeded 84% for the three SMUs vs. 60% in our study. Despite differences in soil drainage and soil depth class categories, strict purities and mean purities were of the same order of magnitude as those in our study. The mean purities for the three SMUs ranged from 61-77% vs. 72% in our study, while the strict purities for three SMUs ranged from 12-37% vs. 32% in our study. Therefore, the disaggregated map cannot replace high-resolution soil maps

based on intensive sampling but is an efficient way to address the spatial variability of dominant soil types across the study area. An important advantage of disaggregated soil maps is their ability to cover large areas even where soil data are scarce but environmental covariates are available at high spatial resolution. Thus, the disaggregated map is an effective way to produce maps of soil properties for large areas, with a measure of prediction uncertainty from the probability layers.

5.4.3. Accuracy of quantitative soil property predictions by spatial disaggregation

For soil maps depicting quantitative soil properties, the quality of prediction was assessed using the common statistical parameters R^2 , RMSE, ME and RRMSE. In general, the variability of predicted soil properties within the study area seemed plausible when compared to those of previous studies, such as that of Lacoste et al. (2014). For SOC content, like Lacoste et al. (2014), results were better (lower absolute ME and RMSE) at 30-60 cm than at 5-15 cm. Using an independent validation dataset, these authors validated high-resolution 3D SOC distribution maps across a heterogeneous agricultural landscape in Brittany. Overall, they found an increase in SOC prediction accuracy with soil increasing depth. For example, the RMSE was 12.6 g kg^{-1} at 0-7.5 cm but markedly lower (2.2 g kg^{-1}) at 60-75 cm. However, our results do not agree with those of Minasny et al. (2013) or Malone et al. (2009), which highlighted a decreasing accuracy of prediction with increasing depth. The influence of agriculture practices (e.g. tillage, fertilisation), differences in soils and the quality of legacy soil data might explain this difference. Even though human influences were proxied using land cover in the DSMART algorithm, this covariate appears to be too poor for large land use classes to represent human influences in our study. Spline functions fitted to STUs to estimate soil properties at *GlobalSoilMap* depth intervals might also explain some of the inaccuracy of predicted soil properties.

Odgers et al. (2015a) performed a similar study in which they applied the Digital Soil Property Mapping Using Soil Class Probability Rasters algorithm (PROPR) to address the spatial pattern of clay content over their study area (South Australia) using soil class probability rasters. Interestingly, the R^2 of Odgers et al. (2015a) is 67% smaller (0.22) than that in our study (0.65) at 5-15 cm and 52% smaller (0.18) than that in our study (0.37) at 30-60 cm. The RMSE and ME are higher in our study than those of Odgers et al. (2015a) by a factor of ~ 2 . In another study, Odgers et al. (2015b)

mapped pH following the same method as the clay content over the Shire of Dalrymple in Australia. Their study yielded a small R^2 between observed and predicted pH at both 5-15 and 30-60 cm. R^2 reached 0.06 vs. 0.43 in our study at 5-15 cm and 0.10 vs. 0.54 in our study at 30-60 cm. However, the RMSE was relatively similar in the two studies even if the sampling design and strategy were different.

The main shortcoming revealed by this research is that soil classes with low probabilities of occurrence have no influence on predictions of soil properties. Only the three most probable STUs were considered for validation and to derive quantitative soil property maps. Although DSMART attempts to predict the most probable STUs for a specific soil-landscape context, prediction accuracy varies greatly and depends strongly on the soil heterogeneity of the study area. In addition, the resulting rasters depicting the probability of occurrence of STU maps themselves depend greatly on polygon component proportions, which are derived from expert soil knowledge and brings more uncertainty to the predicted STUs. Thus, close examination of the soil regional database remains crucial to increase prediction performance for both disaggregated soil maps and derived soil maps depicting quantitative soil properties.

5.4.4. Improvement and future study

In this study, statistical validation of disaggregated soil maps provided good overall purities. Each classification criterion was judged individually, and an overall measure (strict purity) assessed the overall accuracy of digital soil maps. However, expert soil knowledge was not included when calculating the validation statistics. Soil taxonomic distance, especially for soils occurring in similar environmental conditions or having similar morphological criteria, was not considered but is worth investigating in future mapping studies.

5.5. Conclusion

To validate a soil map, using an independent dataset is a safe option that leads to unbiased estimates of quality measures without using models to assess the uncertainty (Brus et al., 2011). This was applied at the regional scale in north-western France, where the three most probable disaggregated

STU maps at 50 m resolution were available. A total of 260 soil samples, collected at 5-15 and 30-60 cm depths, were analysed to determine their particle-size distributions, CEC, pH, SOC content and coarse fragment percentages. For soil maps depicting quantitative soil properties derived from spatial disaggregation, soil texture fractions remained the least biased at both depth intervals. In general, validation statistics of soil properties depended on the soil property considered as well as on the soil depth interval. For categorical soil maps, the partial purity of soil class, and the strict and mean purities were informative. Overall, the predictive quality of disaggregated soil maps (1st, 2nd and 3rd STUs) varied among strata, but continental alluvium and marsh strata were predicted well for the three soil maps. Furthermore, the 1st most probable STU, with high probability of occurrence, had the best validation measures regardless of soil parent material. Differences in prediction accuracies among strata denote areas where more soil data or better soil prediction models have to be obtained or applied first to improve the disaggregation process.

Acknowledgements

The authors gratefully acknowledge all farmers from the Ille-et-Vilaine department involved in our research. We thank 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 funded by the Soilserv program funded by ANR (Agence Nationale de la Recherche) (ANR-16-CE32-0005-01).

Conclusion du chapitre 5

Le chapitre 5 a présenté les résultats de validation des cartes des sols, désagrégées à une résolution spatiale de 50 m, à l'échelle de département d'Ille-et-Vilaine. Nous avons quantifié la qualité prédictive de l'ensemble des cartes obtenues en mobilisant un jeu de données indépendant, acquis en 2017 dans le cadre de la thèse, couvrant la variabilité des matériaux parentaux des sols au sein de la zone d'étude. Ensuite, des indices de pureté partielle associés aux critères cartographiques retenus (matériau parental, classe d'hydromorphie, développement de profil et classe de profondeur), des indices de pureté moyenne et des indices de pureté stricte ont été calculés à différentes échelles spatiales : la strate de matériau parental dominant à la base de l'échantillonnage du jeu de données indépendant, le transect de trois points échantillonnés le long du versant et le département. En parallèle, des paramètres statistiques, à savoir les coefficients de détermination, les erreurs moyennes, les erreurs quadratiques moyennes, ont été calculés pour évaluer la qualité prédictive de huit cartes de propriétés quantitatives des sols à savoir les cinq fractions de la texture, la teneur en carbone organique, le pH et la CEC. Les résultats obtenus montrent que les indices de pureté varient selon le critère cartographique considéré, mais aussi selon le support de validation retenu. Par exemple, les indices de pureté partielle du matériau parental apparaissent élevés puisqu'ils oscillent entre 67% et 100% selon la strate de matériau parental considérée et atteignent une valeur moyenne globale de 78% à l'échelle de la zone d'étude entière. Au contraire, les indices de pureté de développement de profil sont plus faibles et couvrent une large gamme de variation allant de 17% à 100%. Les indices de pureté moyenne varient entre 58% et 96% où la majorité de la variation des impuretés provient de variations de la pureté partielle de développement de profil. Globalement, les résultats obtenus sont intéressants dans un contexte d'amélioration de l'information spatialisée sur les sols, mais aussi pour promouvoir l'intégration d'une information précise des sols dans des évaluations agro-environnementales des écosystèmes naturels.

Pour conclure, la désagrégation spatiale de cartes existantes permet des avancées dans la caractérisation fine des sols aux larges étendues spatiales en mobilisant des données pédologiques à faible résolution spatiale et des covariables environnementales couvrant d'une manière exhaustive le territoire d'étude

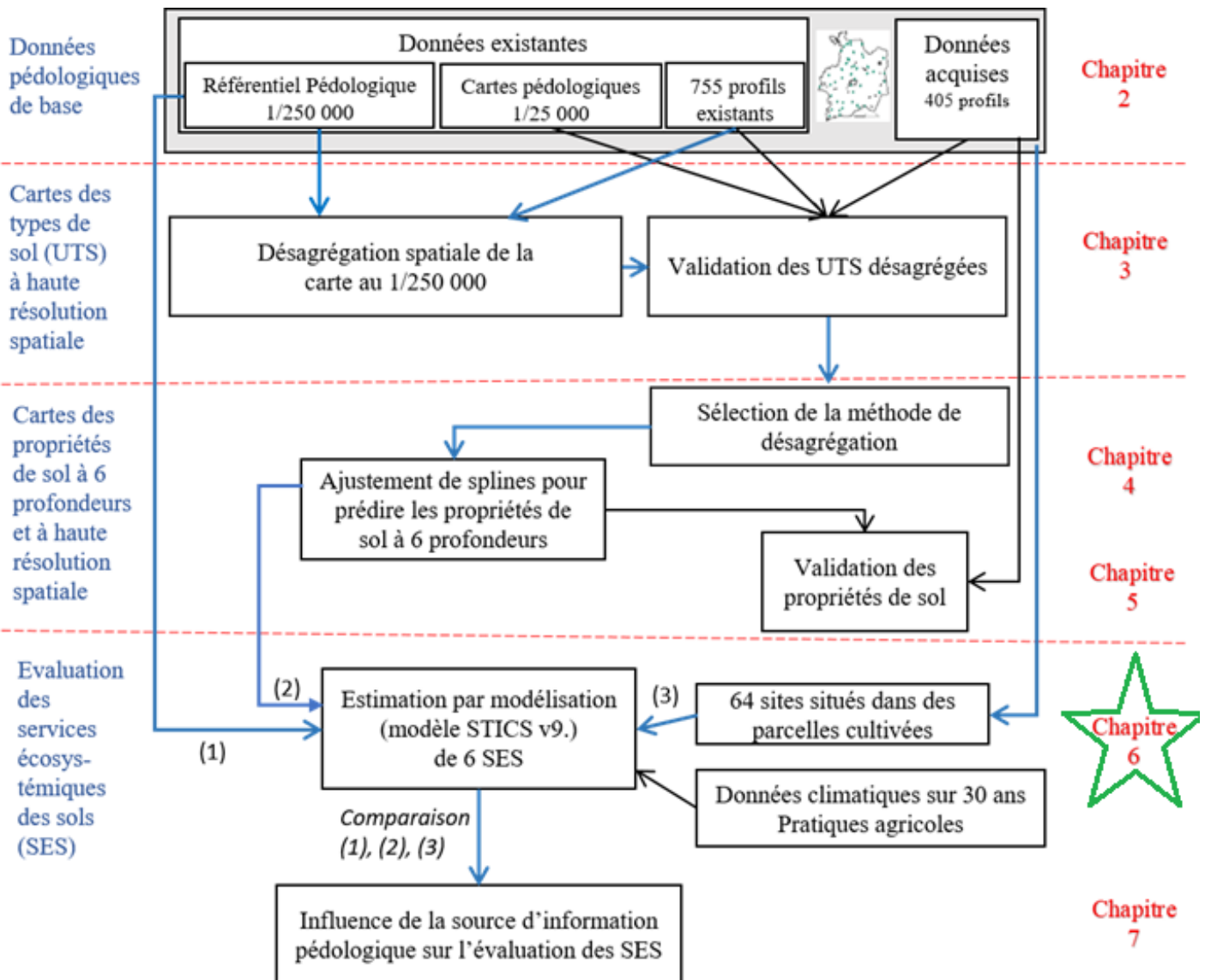


Figure 6.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 6 :
**Assessment of six soil ecosystem
services by coupling simulation
modelling and field measurement of soil
properties**

Introduction du chapitre 6

Le Millenium Ecosystem Assesement (2005) a introduit auprès de la communauté scientifique de science du sol le concept de services écosystémiques pour caractériser l'état des écosystèmes naturels, et analyser le rôle promoteur de la biodiversité dans le maintien et /ou l'amélioration de la durabilité des écosystèmes. Ce concept a considérablement pris de l'ampleur, notamment dans le champ scientifique, où il a été adopté par de nombreux scientifiques pour caractériser les fonctions des sols ainsi que leurs interrelations.

Les méthodes actuelles d'évaluation des services écosystémiques des sols reposent essentiellement sur des indicateurs biophysiques. Ces derniers mobilisent généralement peu de données sur les sols ou l'information intégrée apparaît souvent simpliste. Cela est notamment expliqué par la faible disponibilité de l'information pédologique, mais aussi par des lacunes méthodologiques pour mobiliser l'information pédologique dans le processus d'évaluation des services écosystémiques dérivés.

Dans ce contexte, ce chapitre vise à développer une méthodologie d'évaluation des services écosystémiques des sols et leurs interactions dans un contexte agricole tout en s'appuyant sur des propriétés intrinsèques et extrinsèques des sols. Pour ce faire, nous avons adopté le cadre conceptuel développé dans le cadre de l'EFESE-EA (2017), qui positionne le sol et les services écosystémiques dérivés comme des composants centraux de l'écosystème.

Concernant les données sur les sols mobilisées pour évaluer les services écosystémiques des sols cultivés, nous nous sommes appuyés sur le jeu de données collecté dans le cadre de la validation des cartes désagrégées, en ne retenant que les points situés dans des sols cultivés. Pour l'essentiel, six services écosystémiques sont évalués dans la présente étude par le biais de six indicateurs biophysiques. Ces derniers sont déterminés par une approche de modélisation dynamique des processus des sols sur une échelle temporelle moyenne (30 ans). Plus encore, les interrelations entre les services écosystémiques instruits sont aussi investiguées moyennant différentes approches.

Assessment of six soil ecosystem services by coupling simulation modelling and field measurement of soil properties

Abstract

Soil is a reservoir of natural capital that provides several ecosystem services, ensuring human well-being and sustainable socioeconomic development. Many researchers nevertheless argue that there is no consensus on practical indicators to assess soil ecosystem services (SES). As many policy decisions rely on metrics and indicators to communicate concise and relevant information, an assessment of ecosystem service indicators can help identify gaps hindering policymakers from more fully adopting ecosystem service approaches. The aim of this study was to develop a method to quantitatively evaluate six SES using a set of indicators derived from dynamic soil and crop modelling using the STICS model developed by INRA. In a 6 775 km² study area in north-western France (Brittany), 64 soil sampling points, located in agricultural areas, were selected following a stratified random sampling design based on soil parent material stratification. STICS inputs required climate data, soil property data and soil and landscape management practices. Over a baseline period from 1988-2018, similar crop management practices were simulated that reflected the dominant one performed by conventional farmers across the study site: applying organic and mineral fertilisers to a maize-wheat-catch crop rotation. Also, STICS outputs were used to derive six biophysical indicators characterising six SES. Both mean and annual SES indicators were calculated and analysed. Interrelations among SES indicators and soil properties were investigated using a Pearson correlation matrix, multivariate variance analysis and regression-based methods. The main results revealed that pedogenetic and inter-annual variability were the main drivers of SES provision, particularly for ground water recharge, plant biomass provision, plant water provision and carbon sequestration. All SES were strongly correlated, except carbon sequestration and plant nitrogen provision, which showed weak correlations with the rest of SES. Moreover, analysing interrelations between SES indicators and soil properties such as soil depth, soil texture and its related variables such as available water capacity played a predominant role in ensuring high levels of water-related

SES indicators. Meanwhile, physicochemical soil properties were strongly correlated with carbon sequestration but less so with plant nitrogen provision. Overall, these findings characterise the effect of soil variability on SES provision using modelling and field measures of several soil properties, which can be useful for policy makers.

6.1. Introduction

Soils provide a wide range of goods and services important for human well-being and sustainable socio-economic development, collectively called “ecosystem services” (Costanza et al., 1997). The ecosystem services concept emerged in economics in 1970 (Cardona, 2012) and was considered the central pillar of the Millennium Ecosystem Assessment (MEA) in 2005.

The conceptual framework for classifying, quantifying and modelling ecosystem services is often context-specific. In the literature, three main international classifications are widely used: the MEA (2005), The Economics of Ecosystems and Biodiversity (TEEB, 2010) and the Common International Classification of Ecosystem Services (CICES, 2011). For the first two frameworks, four primary categories of services are distinguished: provisioning (e.g. fresh water, wood, food, and fibre), regulating (e.g. climate, erosion, and floods), cultural (e.g. aesthetic or spiritual values) and supporting (e.g. physical support to plants, animals and human infrastructure). However, the CICES framework distinguishes only provisioning, regulating and cultural categories and considers therefore that cultural services are included in the regulation category.

Soil is defined as the central interface between the lithosphere, biosphere, hydrosphere and atmosphere (Szabolcs, 1994; Dominati et al., 2013; Greiner et al., 2017). Understanding the multifunctional role of soil in ecosystem functioning is crucial (Bouma, 2010; Dominati et al., 2014), and more importance was recently accorded in soil science research to quantifying soil’s contribution to human welfare. Due to its inherent and manageable properties and processes, soil ensures specific functions that set a biological balance among ecosystem components. The conceptual framework of Dominati et al. (2010) considers the soil as natural capital with stocks of matter and biodiversity, but also with generated flows, allowing the provision of several services fulfilling human needs. Robinson et al. (2012) and McBratney et al. (2014) emphasised the need to assess soil ecosystem services (SES) and promote soil-ecosystem connections when developing land-resource policy and management. In the scientific literature, however, assessment of SES, based on fundamental soil features, is rarely investigated or defined (Adhikari and Hartemink, 2016).

Currently, one of the greatest challenges is how to quantify SES (Costanza et al., 1997; Adhikari and Hartemink, 2016; de Groot et al., 2010; Hauck et al., 2015; Baveye et al., 2016). In general, biophysical assessment is performed using indicators (also called “metrics”) (Hauck et al., 2015).

Direct measurement of SES consists of measuring soil properties and processes involved in each SES. For instance, the soil's contribution to climate regulation is assessed mainly by its ability to sequester organic carbon in topsoil horizons (Baveye et al., 2016; Ellili et al., 2019); even if some studies suggest that deeper soil horizons can also contribute significantly to carbon sequestration (Calzolari et al., 2016). To characterise dynamics of soil organic carbon (SOC) over time, resampling soils and correctly interpreting observed changes in SOC are mandatory. Monitoring these changes is hampered, however, by several sources of uncertainty, which are related mainly to land-use change and agricultural practices that greatly influence soil bulk density and SOC content.

In contrast, modelling consists of mathematically describing soil processes that underlie the provision of SES (Nelson et al., 2009). Soil properties that are key to providing a particular SES are identified and related to context-specific environmental variables. Thus, SES assessment can benefit from soil research into dynamic spatio-temporal modelling of soil properties and processes that can support provision of SES (Adhikari and Hartemink, 2016; Calzolari et al., 2016). For instance, several studies have quantified water-quality regulation service by modelling water percolation and the associated filtration of pollutants (Baveye et al., 2016). Other researchers have estimated soil erosion regulation by combining the Universal Soil Loss Equation and the Soil and Water Assessment Tool (Lü et al., 2012). Moreover, at the French national scale, the EFESE-EA "Evaluation Française des Ecosystèmes et des Services Écosystémiques-Ecosystèmes Agricoles" collective expertise has attempted to assess several SES using the STICS ("Simulateur multIdisciplinaire pour les Cultures Standard") dynamic model developed by INRA (Brisson et al., 2008). Essentially, a wide range of provisioning, regulation and cultural SES were assessed over a wide range of pedoclimatic variability and under different management scenarios.

As modelling ecosystem services requires many input variables, direct measurements are recommended (Baveye et al., 2016); however, they are rarely used in the literature because of a lack of soil data (Adhikari and Hartemink, 2016). To compensate for the lack of soil data, several studies have used proxies, which play an important role in spatial planning of SES. The proxy approach consists of using certain environmental variables and soil properties to quantify SES indirectly. The most commonly used proxies are land-use and land-cover data (Adhikari and Hartemink, 2016), which have been useful in regions where soil data are scarce (Vrebos et al., 2015). These data are used as indicators of ecosystem services and properties, and to estimate the spatial provision of

ecosystem services. In several landscapes from Colombia and Brazil, Grimaldi et al. (2014) quantified water cycle and soil erosion regulation services using the rate of water infiltration into the soil. Similarly, Calzolari et al. (2016) assessed eight ecosystem services at the regional scale in Italy by creating multiple indices based on both manageable and inherent soil properties, which are available from legacy soil data. Furthermore, the benefits-transfer approach, based on the spatial distribution of proxies, is widely used to generate maps of ecosystem services (Nelson et al., 2009). This method uses “broad-scale assessments” (Nelson et al., 2009) to extrapolate from point estimates of multiple services to the entire biosphere (Costanza et al., 1997; Nelson et al., 2009). This approach dismisses the spatial heterogeneity of ecosystem services that depend on agricultural practices and local environmental context (Costanza et al., 1997; Nelson et al., 2009).

Despite the large number of studies on biophysical assessment of SES, there is currently no standard or widely recognised method that correctly defines and quantifies each service (Maes et al., 2012). Adhikari and Hartemink, 2016; Greiner et al. 2017; Drobnik et al., 2018). Researchers continue to develop indicator frameworks, which remain a great challenge because of the lack of normative guidelines for selecting ecosystem service indicators (Hauck et al., 2015). For instance, to represent the climate regulation ecosystem service, Grimaldi et al. (2014) used the SOC stock of the 0-30 cm layer on a given date as an indicator while, for the same purpose, Nelson et al. (2009) selected the change in SOC stock over time.

Several studies have attempted to describe relationships among ecosystem services because identifying and understanding the former are crucial for decision makers and stakeholders (Nelson et al., 2009; Wu et al. 2013; Hauck et al., 2015). Understanding SES interrelations offer a relevant support for landscape managers to guide their choices when knowledge about landscape functions is lacking (Fossey et al., submitted). Moreover, knowledge about SES interactions contribute to defining and evaluating policies. For instance, it may help to identify awareness about environmental issues and foster dialogue among stakeholders. According to Bennett et al. (2009), interactions among ecosystem services reveals two types of interactions: synergy and trade-off. Synergy involves mutual improvement of ecosystem services. Thus, benefits of one ecosystem service directly increase those of another service. In contrast, trade-off refers to decreasing one ecosystem service while increasing another. Following a general use, we distinguished three possible interactions: trade-off, synergy and no effect (Bennett et al., 2009; Jopke et al., 2015; Lee and Lantebach, 2016).

In the literature, various methods exist to quantify these interactions using empirical methods, econometric tools, scenario simulations and participatory methods (Nelson et al., 2009; Deng et al., 2016; Lee and Lautenbach, 2016). In 2014, Mouchet et al. explicitly identified a set of quantitative methods to assess interrelations among ecosystem services, describing most methods with their potential framework application. Of these methods, pairwise correlation coefficients are the most popular quantitative method to assess interactions among ecosystem services. Regression-based methods address more mechanistic relationships among SES. Other multivariate analyses were identified as being more flexible, regardless the nature of the ecosystem service indicator, such as Principal Component Analysis (PCA), Factorial Analysis for Mixed Data and Redundancy Analysis. Another promising tool, Structural Equation Modelling (Grace, 2006), was also identified as a relevant way to investigate causal relationships among ecosystem services, ecosystem properties and a set of explanatory variables. Ecosystem service interactions can be also visualised using radial diagrams (Egoh et al., 2008; Calzolari et al., 2016), and bagplots were used (Jokpt et al., 2015) to map SES interactions at the European scale.

The purpose of this study was to assess SES and interrelations in the dominant agricultural system of western France using the STICS model over a period of 30 years. Six SES were selected: climate regulation (CS) via soil carbon sequestration, water quality regulation (WQ), groundwater recharge (GW), water to plant provision (WP), nitrogen (N) to plant provision (NP) and plant biomass provision (YE). These six SES were assessed at 64 observation sites covering the primary soil variability observed in Brittany (France). We explored the influence of inter-annual climatic variability and pedological variability on SES, as well as interrelations between SES to highlight synergy and trade-off situations. Furthermore, interrelations between SES and some soil properties were also investigated to determine the soil properties that most influence the provision of each SES.

6.2. Materials and methods

6.2.1. The study area

The study area is the Ille-et-Vilaine department, in eastern Brittany (NW France, 47° 40' to 48° 40' N, 1° to 2° 20' W). It covers an area of 6 775 km² with gentle topographic features, except in the

western zone, where the elevation peaks at 256 m. The study area is part of the Armorican Massif (BRGM, 2009), whose geology is complex: intrusive rocks (granite, gneiss and micaschist) in northern and north-western zones, sedimentary rocks (sandstone, Brioverian schist) in central and southern zones, and superficial deposits (aeolian loam, alluvial and colluvium deposits) overlaying bedrock_ formations with decreasing thickness from north to south (INRA Infosol, 2014). This high geological heterogeneity generates high soil variability over short distances. Most soils in the study area are Leptosols, Cambisols, Luvisols and Stagnic Fluvisols according to the World Reference Base of Soil Resources (IUSS Working Group WRB, 2014). In specific locations, Podzols, Histosols and Tidalic Fluvisols are observed. Agriculture in the study area is devoted mainly to milk and animal production, and the major crops (wheat, maize and barley) are used for animal feed.

6.2.2. The STICS soil-crop model

In this study, we used the STICS crop model (Brisson et al., 2008), which was designed to simulate simultaneously crop growth and several soil processes connecting water, N and carbon dynamics in the soil-crop-atmosphere continuum. STICS is one-dimensional model with a daily time step for multi-year simulations. It distinguishes multiple compartments of the crop canopy and defines a soil profile with at most five layers. STICS has been calibrated for several crops such as maize and wheat (Brisson et al., 2002). It has been used in several countries, such as Portugal (Fraga et al., 2018) and Brazil (Silva et al., 2019). The French collective expertise EFESE-EA (EFESE-EA, 2017) selected STICS to assess several SES at the French national scale under different management scenarios, soil water conditions and irrigation regimes.

6.2.3. Datasets

6.2.3.1. Climate dataset

Climatic conditions were assumed to be spatially uniform over the study area, and the daily input data required for the model were obtained from the meteorological station of Saint-Jacques-de-la-Lande, near Rennes, located 100 m above sea level. Daily weather variables were retrieved for the baseline period 1988-2018 from the INRA CLIMATIK weather database (INRA, 2019). Data

included daily precipitation (mm), solar radiation (MJ m^{-2}), mean daily wind speed (m s^{-1}), maximum and minimum air temperature ($^{\circ}\text{C}$) and relative air humidity (%). Potential evapotranspiration (mm day^{-1}) was estimated by the Penman-Monteith method. Figure 6.1 shows the mean annual rainfall, global solar radiation and the mean annual temperature over the baseline period.

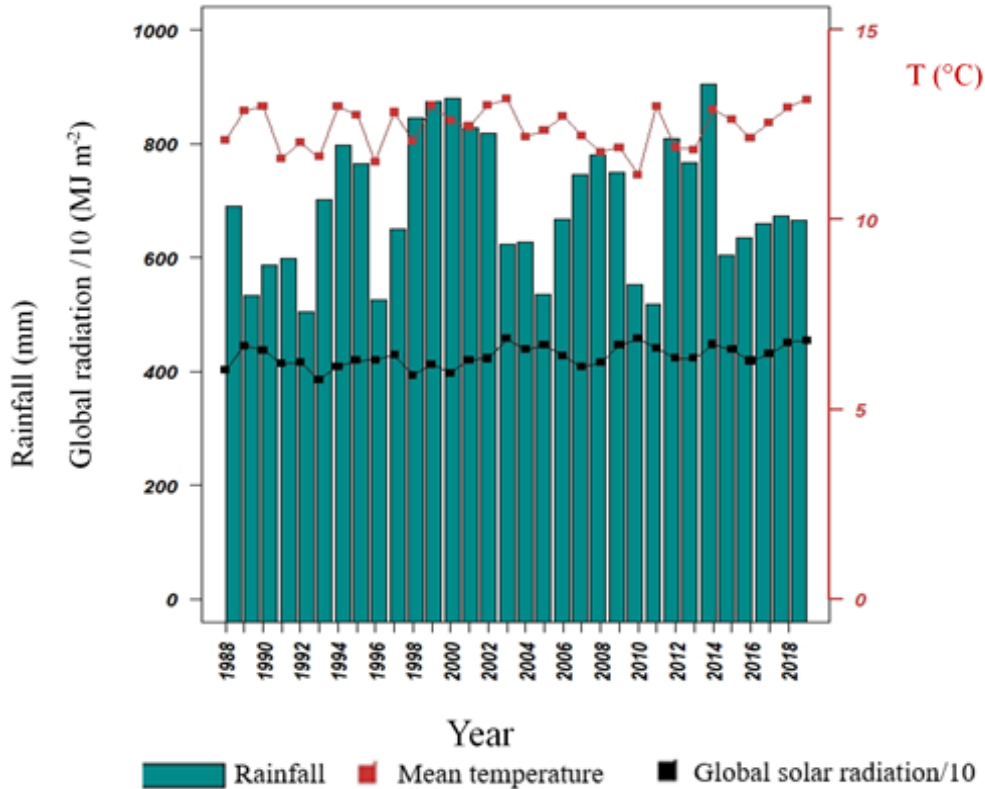


Figure 6.1 : Mean annual rainfall (mm), global solar radiation (MJ m^{-2}) and mean annual temperature ($^{\circ}\text{C}$) of the study area.

6.2.3.2. Soil dataset

In a previous study (Ellili et al., 2019), 135 soil profiles were sampled over the 6 775 km^2 study area following a stratified random sampling design based on an existing 1:250,000 soil parent material map (Figure 6.2). The soils were augered down to 2 m or to paralithic contact, and their morphology was described in detail to relate the soil type to the French Référentiel Pédologique (Baize and Girard, 2008) and to the WRB international classification (IUSS, 2014). The profiles were then sampled at six regular depths according to *GlobalSoilMap* specifications (0-5, 5-15, 15-30, 30-60, 60-100, and

100-200 cm). Of the 135 soil profiles, data of 64 sites located on agricultural land were used for this study. To decrease costs, only three soil layers (5-15, 30-60 and 100-200 cm) were analysed. All soil samples were analysed according to standard methods to determine their SOC content by dry combustion (Thermoscientific Finnigan EA 1112 Flash elemental analyser) (NF ISO 10694)), soil mineral N content, cation exchange capacity (CEC) (NF X 31-130), pH 1:5 H₂O (soil pH in water with 1:5 soil-to-water ratio, NF ISO 10390) and their particle size distribution in three classes (clay, silt, and sand, NF X 31-107). The stony material percentage was also determined after soil samples were air-dried and sieved to 2 mm (Table 6.1).

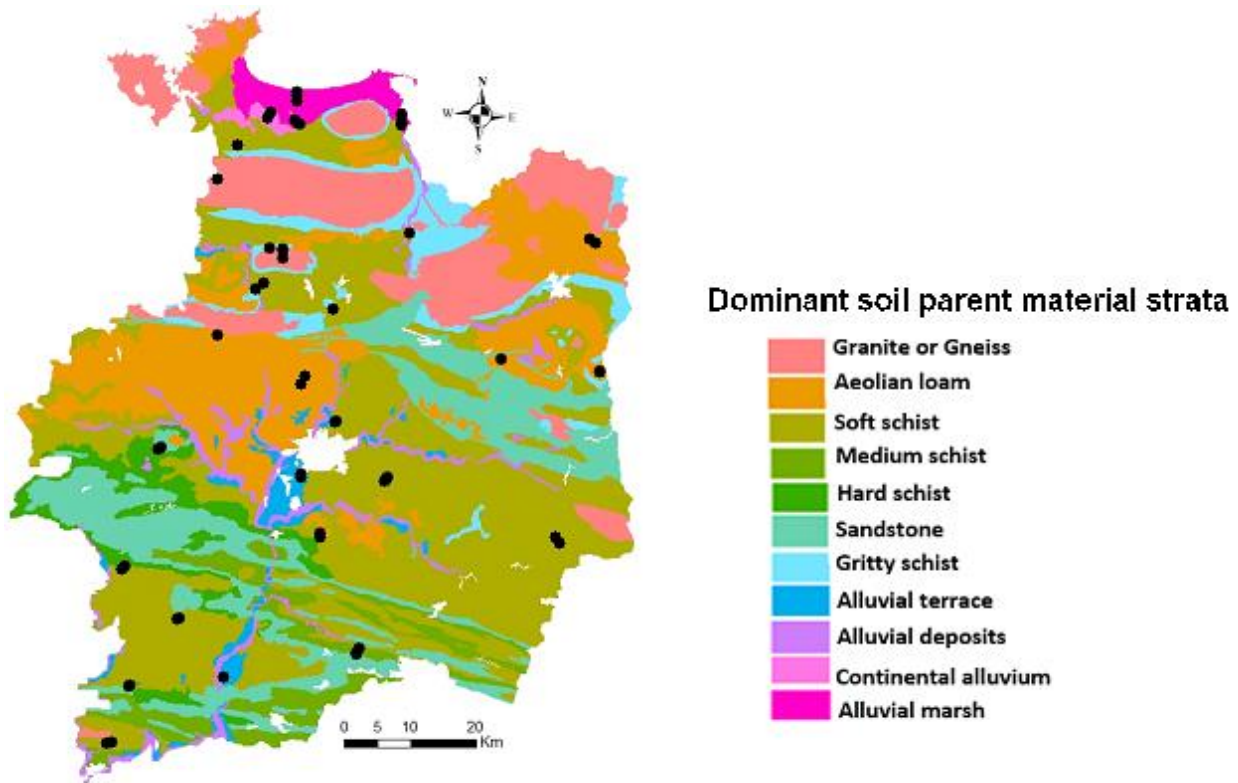


Figure 6.2 : Location of the 64 soil sites selected by stratified simple random sampling based on dominant soil parent strata (Ellili et al., 2019).

Additional soil characteristics required to parameterise STICS were derived either from tables in the STICS manual (Baumont et al., 2010) or measurements performed in the study area. Soil physical characteristics, such as water content at field capacity, permanent wilting point and bulk density of

fine earth, were estimated using local pedotransfer functions based on analyses of more than 500 representative soil samples (Table 6.1).

The thickness of the biologically active layer for mineralisation and carbon dynamics was set to 30 cm. The maximum rooting depth, which depends strongly on soil type, was set to the depth of the beginning of the C horizon for shallow soils and 100 cm for soils deeper than 100 cm (Table 6.1). For each soil layer, water infiltration rate was set to a default value of 50 mm per day.

Table 6.1 : Characteristics of six soil profiles in the dataset of 64 sampling points based on soil analysis and estimates used as input parameters for the STICS model.

^a Soil type		G30a		L61a		N86b		O22a		T121b		Vcm2c	
Depth		0-30	30-40	0-30	90-100	0-30	0-25	40-48	0-30	50-70	0-32	75-100	
^b Soil													
Clay	g kg ⁻¹	16.50		13.4		23.4	21.3		13.60		37.4		
Organic C	g kg ⁻¹	0.15		0.19		0.26	0.21		0.11		0.11		
Organic N	g kg ⁻¹	11.03		10.10		9.84	10.7		10.78		11.02		
pH (water)		6.04		6.51		5.72	6.34		5.55		7.36		
W _{FC}	g kg ⁻¹	23.42	17.2	29.15	22.87	29.91	29.83	25.13	22.66	20.09	45.45	44.31	
W _{WP}	g kg ⁻¹	12.05	8.1	12.3	10.57	16.54	15.48	13.64	10.26	10.00	28.74	27.20	
Bulk density	g cm ⁻³	1.27	1.43	1.24	1.49	1.24	1.24	1.34	1.33	1.46	1.24	1.24	
Stones	% vol	21.46	37	2.16	5.11	36.34	18.3	37.22	3.06	15.01	0.00	1.14	
^c STICS parameters													
Profhum	cm	30		30		30	30		30		30		
Obstarac	cm	40		100		30	48		70		100		

^aSoil types: G30a = Dystric Cambisol developed from granite with a depth of 30 cm; L61a = Endostagnic Cambisol on deep loam with a depth of 150 cm; N86b = Leptosol developed from soft schist with a depth of 30 cm; O22a = Cambisol developed from medium schist with a depth of 48 cm; T121b = Cambisol overlying alluvial terrace with a depth of 70 cm; and Vcm2c = Histosol developed on marsh with a depth of 220 cm

^bSoil analyses: W_{FC} = water content at field capacity, W_{WP} = water content at permanent wilting point.

^cSTICS parameters: profhum = biologically active soil depth for mineralisation; obstarac = maximum rooting depth.

6.2.3.3. Crop management

Crop management was simulated based on the dominant practices performed by conventional farmers in the Ille-et-Vilaine department. Requirements of the European Union Nitrates Directive, revised in 2014 (EC, 2013), were considered to define the amounts of organic and mineral N inputs. Over the baseline period, the same maize-wheat-catch crop rotation was simulated with fixed annual sowing and soil-tillage dates and organic and mineral fertilisation schedule. The crop rotation simulated lasts two years and has three successive crops (Figure 6.3):

- The grain maize crop was grown from 10 April to 30 October (harvest) in the first year of each rotation. Before sowing maize, soils were ploughed to ca. 25 cm and 35 t ha⁻¹ of pig slurry was applied. A semi-early maize variety (DK 250) was sown on 15 April with a planting density of 10 plants m⁻². To avoid delayed germination due to potential drought conditions in the topsoil, 15 mm of water was provided by irrigation at sowing. Maize was fertilised with 100 kg ha⁻¹ of ammonium nitrate 30 days after planting. After harvest, stubble and roots were left on the field as crop residues, and the aboveground biomass was exported.
- The winter wheat crop was grown after the maize crop, from 1 November of the first year of the rotation to 31 July (harvest) of the second year. As simulations were run sequentially over the baseline period, the final crop residues and the soil water content of the maize crop simulation were used to initialize the carbon, N and water balances of the wheat crop simulation. Before sowing wheat, soils were ploughed to a depth of 25 cm then sown at a density of 311 plants m⁻². The crop was fertilised with 40 t ha⁻¹ of pig slurry on 15 March, followed 70 kg ha⁻¹ of ammonium of nitrate on 15 May.
- The catch crop (white mustard) was grown between the winter and spring crops, as required by regulations. The cover crop was sown on 5 September and ploughed into the soil on 3 March as crop residues, one month before sowing the following maize crop.

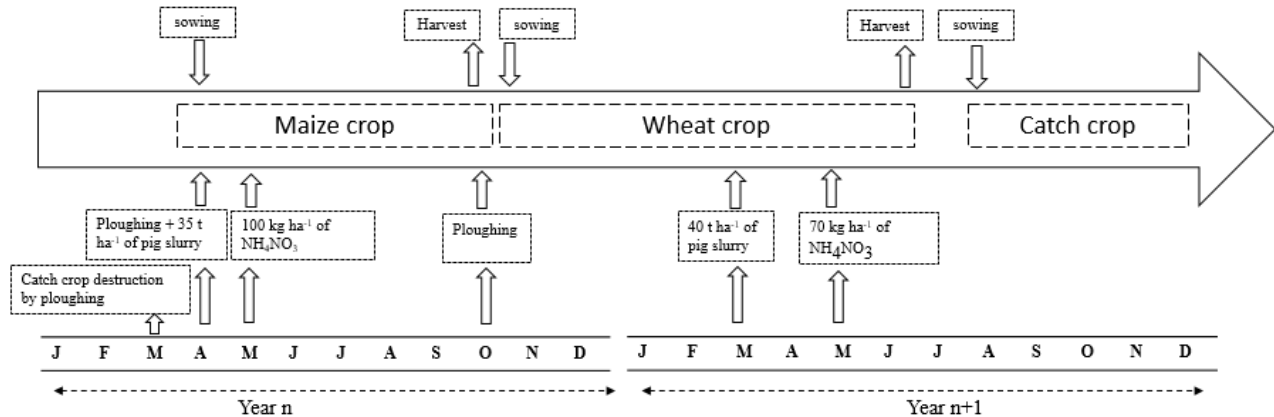


Figure 6.3 : The crop rotation simulated over the baseline period (maize-wheat-catch crop).

6.2.4. Simulation procedure

The crop rotation was simulated with the latest version of STICS (v9.0) and its parameter sets. Over the study period (1988-2018), soil parameters remained static except for carbon, N and water in soil layers, under the influence of dynamic climatic conditions and crop growth. Because the three cropping periods of the rotation were run sequentially, simulation n was initialised with results of simulation $n-1$. By combining soil characteristics of the 64 sites sampled, and crop rotations identified over the 30 year-period, 3565 STICS simulations were performed by using a Java package developed in this study to automate simulations. STICS provides more than 200 outputs, some of which describe daily values of soil properties, such as the water content of soil layers, available water, N mineralisation, N transfer, and water and carbon dynamics. Of these, nine daily outputs were selected to calculate SES indicators: crop transpiration, crop yield, water drainage, N leaching, plant N needs, amount of N mineralised from humus, N mineralised from organic residues, amount of organic N added to soil, and amount of SOC in humified organic matter.

6.2.5. Assessment of SES indicators

Six SES were assessed: (1) climate regulation via carbon sequestration, (2) water quality regulation, (3) N to plant provision, (4) water to plant provision, (5) plant biomass provision and (6) groundwater recharge. Furthermore, both inter-annual and overall SES variabilities were analysed based on

biophysical SES indicators. All indicators used to infer SES were chosen based on the available literature.

We developed two equations for each SES indicator. The first calculated the SES for each calendar year, while was the mean of these annual values over the baseline period (1988-2018).

- **Climate regulation (CS)** was defined as the ability of soils to sequester organic carbon. To quantify CS, we used the annual change in SOC stock from 0-30 cm as the indicator (Eq 6.1):

$$CS_j = CStock_{i+1} - CStock_i \quad (\text{Eq 6.1})$$

$$\overline{CS} = \frac{1}{n} \sum_{j=1}^n CS_j \quad (\text{Eq 6.2})$$

where $CStock_i$ is SOC stock from 0-30 cm of year i and n is the number of years considered ($n=30$, 1988-2018).

\overline{CS} is the annual average of CS over the baseline period.

- **Water quality (WQ) regulation** describes the soil's ability to store N and decrease N leaching. It was calculated as a ratio of the mean annual amount of N not leached to the total amount of N in the soil and supplied by fertilisation (Eq. 6.3):

$$WQ_j = \frac{1}{365} \sum_{i=1}^{365} \left(1 - \frac{N \text{ leaching}_i}{N \text{ min}_i + azomes_{(i-1)} + anit_i} \right) \quad (\text{Eq 6.3})$$

$$\overline{WQ} = \frac{1}{n} \sum_{j=1}^n WQ_j \quad (\text{Eq 6.4})$$

where i and j are respectively, the day and year considered, $Nleaching_i$ is the amount of N leached daily from the soil, $Nmin_i$ is the amount of mineral N produced daily by mineralisation of soil organic matter, crop residues and organic fertilisers, $anit_i$ is the amount of fertiliser N added daily to the crop, and $azomes_{(i-1)}$ is the amount of N in the soil profile on day $i-1$.

\overline{WQ} is the annual average of WQ_j indicator over the baseline period.

WQ_j ranges from 0 (poor water regulation service) to 1 (high water regulation service).

- **N to plant provision (NP)** describes the soil's ability to meet plant N needs during the crop rotation without any mineral fertilisation. We chose a normalised indicator that ranges from 0-1. First, we calculated the ratio of available mineralised N, which can be easily taken up by roots, to plant N needs from sowing to harvest. Based on STICS outputs, this SES indicator was calculated using Eq. 6.5:

$$NP_j = \frac{1}{365} \sum_{i=1}^{365} \min\left(1, \frac{Nh_i + Nr_i + azomes^* - Nleaching_i}{N_{plant_i}}\right) \quad (\text{Eq 6.5})$$

$$\overline{NP} = \frac{1}{n} \sum_{j=1}^n NP_j \quad (\text{Eq 6.6})$$

where Nh_i is the amount of N mineralised daily from humus, Nr_i is the amount of N mineralised daily from organic residues (crop residues and organic fertilisers), $azomes^*$ is the amount of N in the soil profile on the sowing date of each year and N_{plant_i} is daily plant N needs.

\overline{NP} is the annual average of NP_j over the baseline period.

The ratio equals 1 if all plant needs are met, exceeds 1 if available N exceeds plant needs and is less than one if available N is insufficient. NP equals the ratio limited to a maximum value of 1; thus, excess N available for plants was not considered. Also, as the catch crop was completely returned to the soil and not exported, it was not considered in SES assessment.

- **Water to plant provision (WP)** describes the soil's ability to store and provide water to plants without irrigation. It depends strongly on available water capacity, rainfall pattern and crop phenological stage. We adopted EFESE-EA (2017)'s indicator: mean annual recovery rate of water plant needs during the crop rotation. Consequently, we used plant transpiration rates over the rotation of spring and winter crops (Eq 6.7):

$$WP_j = \sum_{i=\text{sowing date crop 1}}^{\text{harvest date crop 1}} ep_i + \sum_{k=\text{sowing date crop 2}}^{\text{end of calendar year}=365} ep_k \quad (\text{Eq 6.7})$$

$$\overline{WP} = \frac{1}{n} \sum_{j=1}^n WP_j \quad (\text{Eq 6.8})$$

where ep_i is daily crop1 (maize) transpiration and ep_k is daily crop2 (wheat) transpiration.

\overline{WP} is the annual average of WP_j over the baseline period.

- **Plant biomass provision (YE)** describes the soil's ability to produce plant biomass by photosynthesis. It represents the annual yield of cash crops and expressed in energy units ($\text{kcal ha}^{-1} \text{ year}^{-1}$). Hence, like WP and NP, YE was calculated by excluding the catch crop. We converted dry matter yields of maize and wheat into energy according to FAO specifications (FAO, 2003) using Eq 6.9:

$$YE_j = \frac{1}{n} \sum_{i=1}^n B_i * k \quad (\text{Eq 6.9})$$

$$\overline{YE} = \frac{1}{n} \sum_{j=1}^n YE_j \quad (\text{Eq 6.10})$$

where B_i is the dry matter of harvested biomass and k is a conversion coefficient to express biomass in terms of energy (3.34 kcal g⁻¹ for maize and 3.19 kcal g⁻¹ for wheat).

\overline{YE} is the annual average of YE_j over the baseline period.

- **Groundwater recharge (GW)** describes the soil's ability to recharge groundwater. It considers water that infiltrates into the soil and percolates into the groundwater. GW is calculated as the sum of daily water drainage from the soil, which is assumed to reach the groundwater, over one calendar year (Eq 6.11):

$$GW_j = \sum_{i=1}^{365} D_i \quad (\text{Eq 6.11})$$

$$\overline{GW} = \frac{1}{n} \sum_{j=1}^n GW_j \quad (\text{Eq 6.12})$$

where D_i is the amount of water drained daily from the soil (mm).

\overline{GW} is the annual average of GW_j over the baseline period.

6.2.6. Statistical analysis

Simulation outputs characterised spatio-temporal variability in SES indicators at 64 sites that covered the primary variability of the study area. To analyse inter-annual variability in SES, SES relationships and the soil properties that influenced them, several statistical analyses were performed.

First, we calculated coefficients of variance (CV) of the inter-annual SES values to compare distributions of SES with different units. It measures the variability of each indicator by dividing its mean value by its standard deviation. Formally, it is calculated using Eq. 6.13, in which M and S are the mean and standard deviation of each SES indicator over the baseline period, respectively.

$$CV = \frac{M \times 100}{S} \quad (\text{Eq 6.13})$$

Second, we used a bagplot representation to visualise the relationship between each pair of SES indicators. The bagplot is a generation of the univariate boxplot (Tukey, 1975) to bivariate data. It depicts the location (position of median depth), spread (bag area), correlation skewness (bag shape) and outliers of the dataset (Rousseeuw et al., 1999). We generated bagplots using the bagplot function of the aplpack package of R software (R Core Team, 2019).

Third, we calculated a Pearson correlation matrix between each pair of mean SES indicators and soil determinants to demonstrate the degree of interactions among variables and select those exerting the most influence. To visualize the correlation matrix, we generated a network chart using the `qgraph` package of R software. To remove edges that were not significant, we used the Bonferroni threshold (Curtin and Shulz., 1998).

Fourth, we performed a normalised PCA to assess gradients that separated sites according to SES indicator means and their associated standard deviations over the baseline period. Continuous soil properties such as SOC content, available water capacity, maximum rooting depth, soil pH and clay content were used as supplemental variables. These soil properties were selected because they are the ones used most frequently in SES studies, as reported by Greiner et al. (2017) and Dominati et al. (2013).

Fifth, to identify linear relationships between each SES metric and soil properties, stepwise multiple-linear regression was performed. We used the Akaike Information Criterion (AICc) to select the best models. The statistical significance of model parameters was estimated by analysis of deviance (validated by F tests) between a null model (no effect) and a model that included one or more explanatory variables. Statistics were calculated and graphical outputs were generated using the `ade4` package (Dray and Dufour, 2007) of R software.

Finally, we performed multivariate analysis of variance (MANOVA) to test differences in means of SES indicators between two or more groups of soil properties. The resulting variances were similarly tested against F statistics approximated by Wilks' lambda. For the MANOVA, we transformed continuous soil properties into categorical properties. We defined two classes for pH (<6, >6), three classes for clay content (< 20%, 20-30%, > 30%), two classes for soil depth (shallow (< 60 cm) and deep (>60 cm)) and three classes for soil drainage (well drained, moderately drained and poorly drained). Finally, based on field pedological description of soil profiles, we defined 10 classes for parent materials and seven for soil types.

6.3. Results

6.3.1. Inter-annual means of SES over the study area

All SES indicator means over the study area varied greatly, except for NP (Table 6.2). For instance, YE varied from 1-3 10^7 kcal ha⁻¹ year⁻¹, WP varied from 110.4-286.9 mm year⁻¹ (median = 192.6 mm year⁻¹) and GW varied from 103.3-282.5 mm year⁻¹ (median = 186.4 mm year⁻¹). For regulation SES, mean WQ was high and ranged from 0.69-0.99. CS varied more, from -773.7 to 914.7 kg ha⁻¹ year⁻¹. In contrast, NP varied little, ranging from 0.84-1.00 (median = 0.99).

Table 6.2 : Descriptive statistics of mean values of soil ecosystem services (SES) over the baseline period estimated at the 64 sites (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision).

SES	Min	Median	Mean	Max
WP (mm year ⁻¹)	110.4	192.6	193.6	286.9
GW (mm year ⁻¹)	103.3	186.4	182.5	282.5
NP	0.84	0.99	0.98	1.00
YE 10^7 (kcal ha ⁻¹ year ⁻¹)	1.00	1.85	1.96	3.03
CS (kg ha ⁻¹ year ⁻¹)	-773.7	139.5	126.5	914.7
WQ	0.69	0.86	0.86	0.99

6.3.2. Inter-annual temporal variability in SES indicators

For two example contrasting soils studied – shallow and sandy (Leptosol) vs. deep and loamy (Endostagnic Cambisol) – inter-annual variability (CVs) in SES was high: 17-36% and 7-65%, respectively (Figure 6.4). Over the baseline period, wheat and maize yields were nearly twice as high for the deep loamy soil as those for the shallow sandy soil. Moreover, YE was relatively constant over time for the shallow soil regardless of the crop (0.9-1.5 10^7 kcal ha⁻¹ year⁻¹). In contrast, YE for the deep soil had more inter-annual variability for maize (1.7-4.2 10^7 kcal ha⁻¹ year⁻¹) than for wheat. WP was highest for the deep loamy soil, exceeding 300 mm year⁻¹ in 1993, 1997, 2007 and 2015, but did not exceed 200 mm year⁻¹ for the shallow soil in the same years. As the winter and spring crops had similar trends in WP for both soils, we can conclude that both crops had similar water needs over their phenological stages. In contrast, GW was highest for the shallow sandy soil (>400 mm year⁻¹ in

1999) but lowest for the deep loamy soil ($<50 \text{ mm year}^{-1}$). Furthermore, GW for both soils displayed cyclical and unstable behaviour that was lower and less variable for wheat over the baseline period. NP usually remained high over the study period for the deep loamy soil, varying from 0.7-1.0 for maize and reaching its maximum value (1.0) for wheat. In contrast, NP for the shallow soil had high inter-annual variability, varying from 0.38-1.0 for maize and 0.5-1.0 for wheat.

For regulation SES, both soils had high WQ, which usually ranged from 0.8-1.0 for the deep loamy soil. For instance, it reached its maximum value (1.0) in 1988, 1999, 2004 and 2017 and was nearly 0.9 in 1990, 1994, 1999, 2008 and 2018, meaning that at least 10% of the N provided was leached from the soil. However, the shallow soil had large inter-annual variability over the baseline period, ranging from 0.4-0.9. Interestingly, for the deep loamy soil, WQ was similar for both crops over time. In contrast, WQ dynamics for the shallow sandy soil were similar to those for the deep loamy soil for wheat but differed greatly for maize.

Finally, CS was higher for the shallow soil than for the deep soil, regardless of the crop. For the shallow soil, carbon sequestration was predicted for all years, except for wheat in 2014 and 2018, when CO_2 losses reached nearly $180 \text{ kg ha}^{-1} \text{ year}^{-1}$. Meanwhile, CS for the deep loamy soil was marked over the baseline period by an alternating cycle of years of net soil carbon storage and loss. The wettest years, particularly 1997, 2000, 2014 and 2015, with total rainfall greater than 800 mm year^{-1} , had high CO_2 losses.

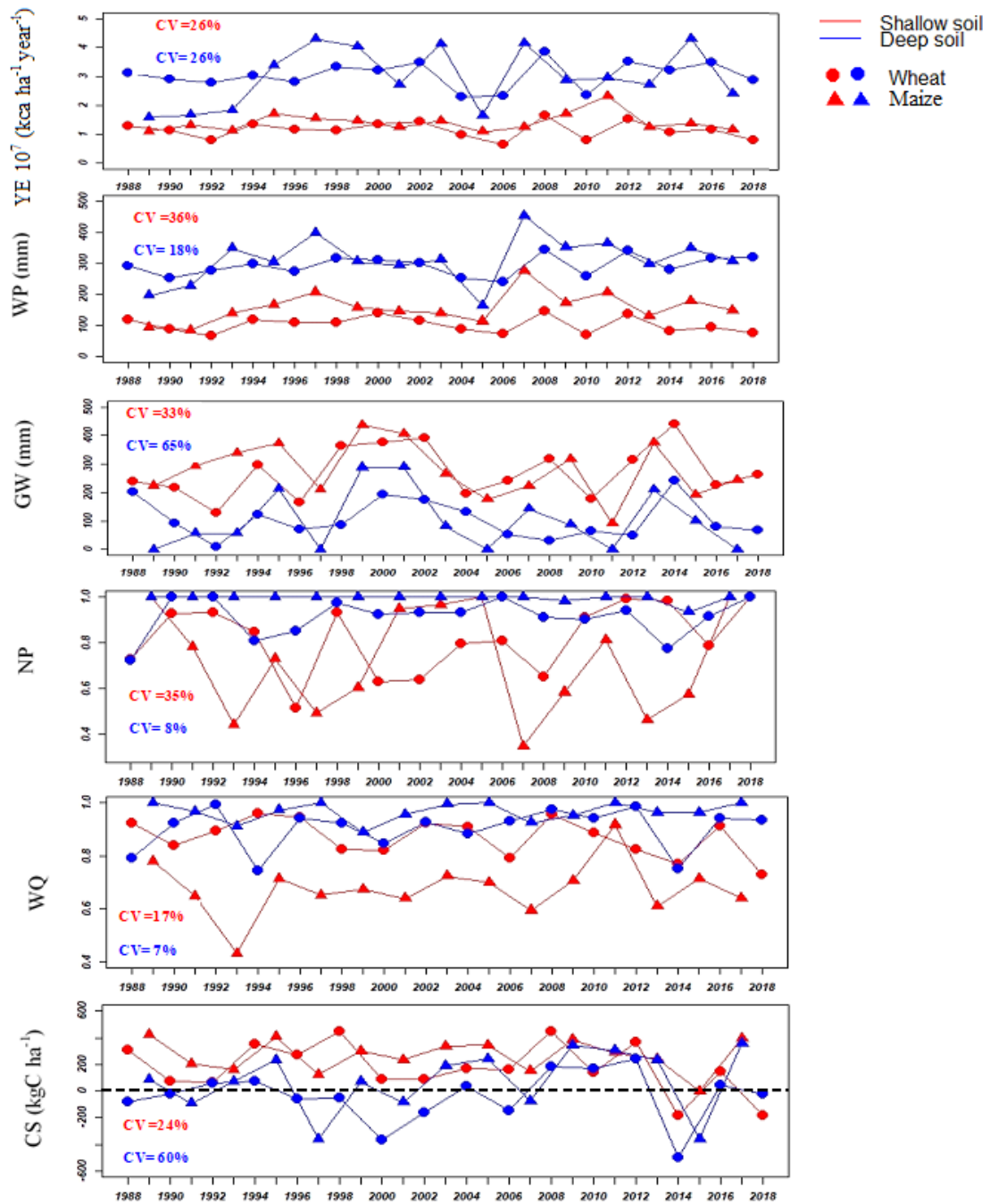


Figure 6.4 : Annual soil ecosystem service values for maize and wheat crops for two contrasting soils over the baseline period (red: coarse-textured Leptosol overlying gritty schist with a depth of 30 cm; blue: Endostagnic Cambisol on deep loam with a depth of 150 cm).

6.3.3. Spatial variability of the annual estimates of SES indicators

Although cropping practices and weather data were the same, each SES indicator varied greatly among soil types (Figure 6.5). For instance, YE differed greatly among soils. Mean (± 1 standard deviation) YE (maize and wheat combined) reached $2.63 \pm 0.60 \cdot 10^7 \text{ kcal ha}^{-1} \text{ year}^{-1}$ (Table 6.3) for a Luvisol developed on Aeolian loam soil (L1C1) but did not exceed $1.40 \pm 0.36 \cdot 10^7 \text{ kcal ha}^{-1} \text{ year}^{-1}$ (Table 3) for a sandy Leptosol (L4C1). Likewise, WP was higher for fine-textured soils ($289 \pm 56 \text{ mm year}^{-1}$ for R17b) but lowest for coarse-textured soils ($130 \pm 48 \text{ mm year}^{-1}$ for R0R5).

GW was low ($122 \pm 88 \text{ mm year}^{-1}$) for a Luvisol (L0C1) in a lowland with high available water storage capacity (200 mm for a depth of 100 cm) (Table 6.3). GW reached $283 \pm 90 \text{ mm year}^{-1}$ for a Leptosol (R0R5) with a low available water storage capacity ($<36 \text{ mm}$) (Table 6.3). Overall, GW depended strongly on soil depth and thus on available water storage capacity. It was high for shallow sandy soils and low for deeper loamy soils.

Mean NP values were high, ranging from 0.75 (T8a) to 1.0 (Vm8Th1) (Table 6.3), indicating that at least 75% of plant N needs were covered. As the same amount of fertiliser was applied each year over the baseline period, NP was usually high regardless of the soil type. In general, no N stress was predicted during the phenological crop stages, and all crop N needs were satisfied by mineralisation of crop residues. More importantly, only deep soils had the highest NP (1.0). Shallow soils had lower NP.

Mean WQ was high, even though the rotation had a constant annual N supply. WQ ranged from 0.70 ± 0.16 (T0B5) to 0.98 ± 0.02 (Vm3Th1, Table 6.3). Furthermore, WQ differed significantly among soil types, being highest for deep soils with a fine texture, such as an Endostagnic Cambisol with a loamy texture (L4C1: 0.94 ± 0.05 , Table 6.3). In contrast, WQ was relatively low for shallow soils, such as a Leptosol developed on granite with a coarse texture (T0B5: 0.70 ± 0.16 , Table 3). Overall, WQ had the lowest mean values for fine-textured soils in depressions and the highest mean values for coarse-textured soils on coastal plains and for soils rich in organic matter.

For CS, situations of carbon storage or loss were identified. As soils were ploughed twice per year to a depth of 25 cm, both deep and shallow soils had negative carbon balances. For 21% of the sties, mean annual loss rate varied from nearly -773 to -3 kg ha^{-1} (Figure 6.5). In general, negative balance was due to mineralisation, which released SOC as CO_2 into the atmosphere. In comparison, 79% of

the sites had positive carbon balances, with mean annual increase of 1.42-914 kg ha⁻¹ (Figure 6.5). Interestingly, shallow soils had higher mean CS than deep soils.

Overall, except for NP, all SES indicators varied greatly among sites and reached their maximum value at least once. More importantly, soils with the highest YE had the highest WP, lowest GW, maximum NP (1.0) and highest WQ. In addition, except for CS, the standard deviation of all SES varied greatly among sites. Hence, soil diversity in the landscape, which is related to the nature of the geological substrate and other pedogenetic factors, was the main driver of SES provision.

Table 6.3 : Mean (and standard deviation) of soil ecosystem service values over the baseline period for 16 sites. These sites were selected by ranking the 64 sites by increasing yield then retaining every fourth site (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision).

Soil	WP (mm)	GW (mm)	NP	YE (10 ⁷ kcal)	CS (kg ha ⁻¹)	WQ
R0R5	130 (48)	283 (90)	0.99 (0.05)	1 (0.3)	-7.76 (150)	0.88 (0.05)
N0R5	142 (48)	256 (91)	0.84 (0.24)	1.4 (0.36)	187 (188)	0.78 (0.14)
T0B5	150 (54)	263 (93)	0.75 (0.27)	1.5 (0.42)	232 (219)	0.70 (0.16)
Vm8Th1	199 (52)	158 (84)	1(0)	1.57 (0.48)	-406 (299)	0.98 (0.02)
T4B4	172 (55)	210 (93)	0.91 (0.13)	1.62 (0.42)	126 (206)	0.87 (0.08)
R5B3	182 (84)	176 (58)	0.97 (0.06)	1.65 (0.66)	318 (262)	0.80 (0.08)
V5V4	187 (56)	298 (94)	0.96 (0.07)	1.74 (0.45)	-267 (307)	0.81 (0.10)
N0B3	200 (56)	212 (95)	0.92 (0.12)	1.77 (0.45)	2.67 (229)	0.80 (0.0)
U5C1	208 (55)	203 (94)	0.93 (0.11)	1.87 (0.05)	142 (234)	0.81 (0.09)
L1C1	221 (50)	203 (98)	0.97 (0.05)	1.95 (0.48)	203 (211)	0.83 (0.09)
T3B3	220 (58)	187 (93)	0.86 (0.18)	2 (0.56)	443 (211)	0.83 (0.10)
Vm3Th1	247 (55)	109 (82)	0.96 (0.07)	2.1 (0.6)	213 (178)	0.98 (0.02)
Vm0Th1	254 (56)	154 (89)	0.98 (0.05)	2.38 (0.61)	-170 (222)	0.86 (0.08)
Vm0Th1	264 (56)	152 (91)	0.97 (0.04)	2.45 (0.58)	1.42 (278)	0.87 (0.09)
L4C1	275 (56)	127 (83)	0.92 (0.1)	2.63 (0.6)	402 (225)	0.94 (0.05)
L0C1	289 (56)	122 (88)	0.93 (0.1)	2.81 (0.7)	186 (202)	0.91 (0.07)

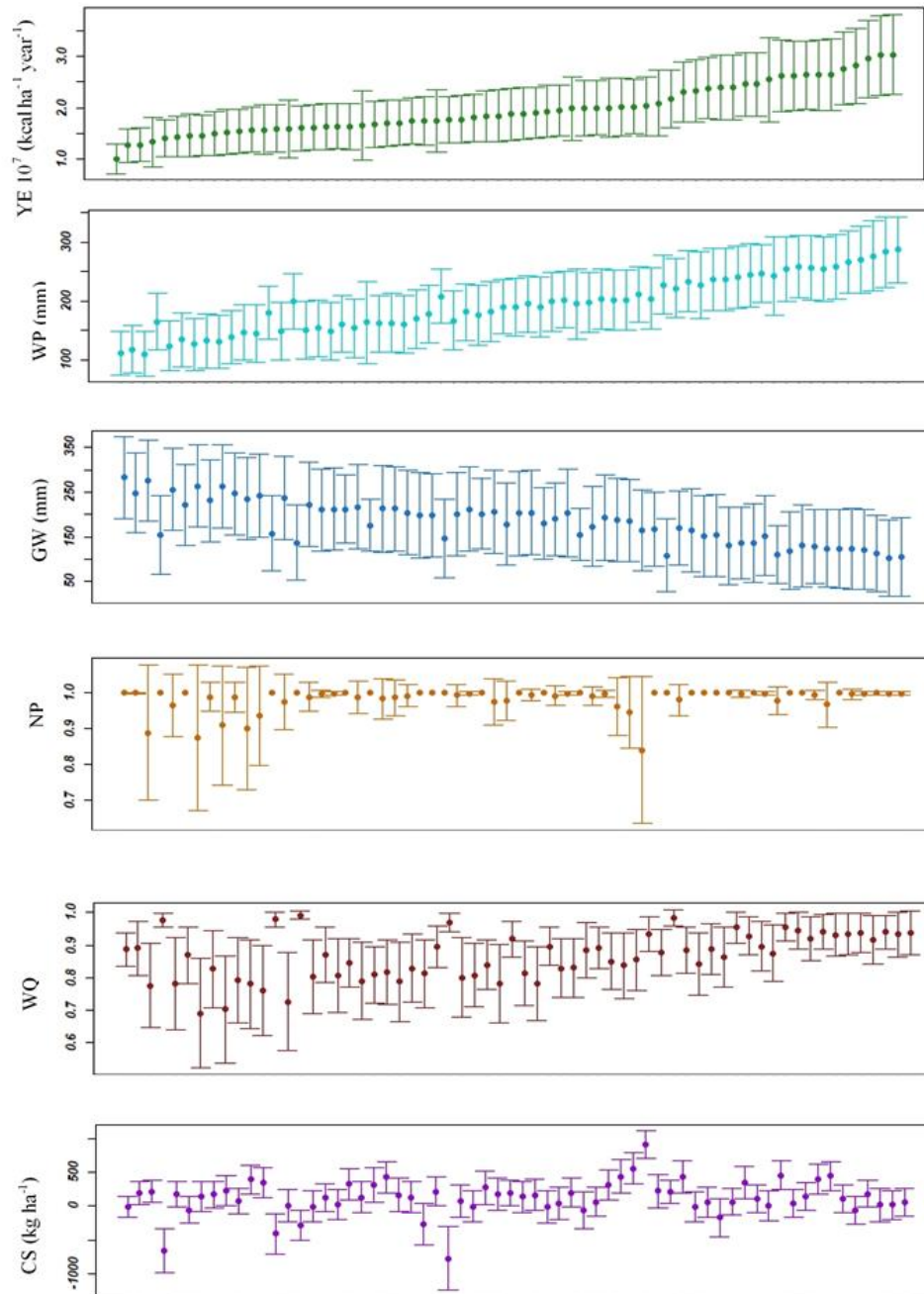


Figure 6.5 : Mean soil ecosystem service indicators of the 64 sites studied, ranked by increasing plant biomass provision (YE). Error bars represent the standard deviation.

6.3.4. Interrelation between SES

Most bagplots of each pair of SES indicators means for the 64 sites over 30 years had roughly the same area, so their datasets had a similar spread (Figure 6.6). However, the bags had different orientations: some sloped upward (positive correlation) (YE-WP, Figure 6.6a; WQ-WP, Figure 6.6f), whereas others sloped downward (GW-YE, Figure 6.6b). All bagplots were relatively circular and slightly skewed because their medians lay near the centre of each bag. Finally, all datasets were medium-tailed based on the size of the “loop” (the boundary separating inliers from outliers) and the absence of outliers.

Overall, linear correlations produced results similar to those revealed by the bagplots. WP had the highest positive correlation, with YE ($r=0.97$), followed by WQ ($r=0.66$), and the highest negative correlation, with GW ($r=-0.93$) (Table 4). YE had a high negative correlation with GW ($r=-0.86$) and a high positive correlation with WQ ($r=0.52$). Overall, the correlation matrix highlighted significant correlations between all SES, except for NP and CS, which were negative ($r=-0.48$). The network chart of the SES highlighted the main relationships between them (Figure 6.7). Overall, YE was strongly synergic with WP and WQ but showed a trade-off with GW. Likewise, WQ had strong synergies with WP and NP. Significant trade-offs were observed among many pairs of SES, such as GW and WP, GW and WQ, CS and NP, and NP and GW.

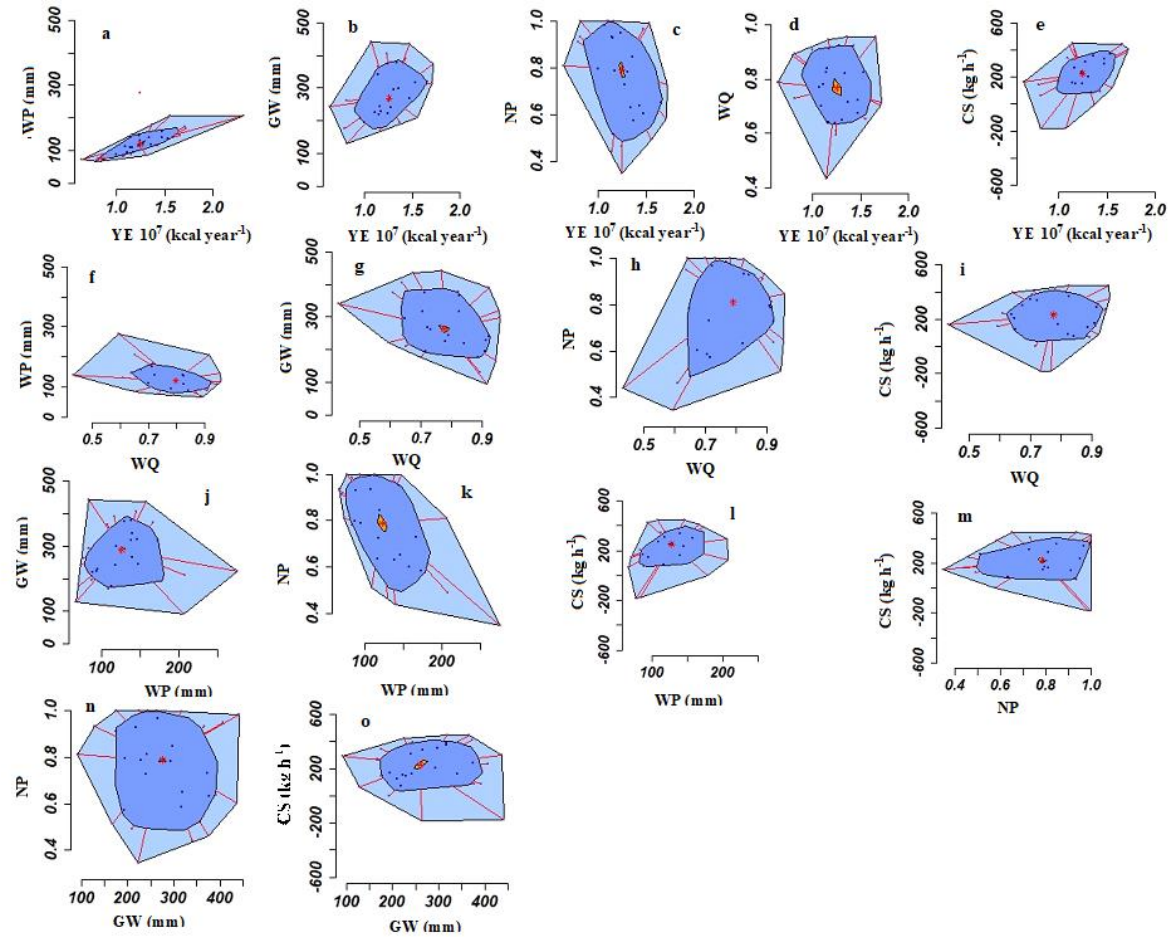


Figure 6.6 : Bagplots of pairs of soil ecosystem service indicators. Darker shading identifies the bag that contains 50% of observations, while lighter coloured shading identifies the boundary separating inliers from outliers (the “loop”). Asterisks identify the point with highest half space depth (median depth) with its 95% confidence region.

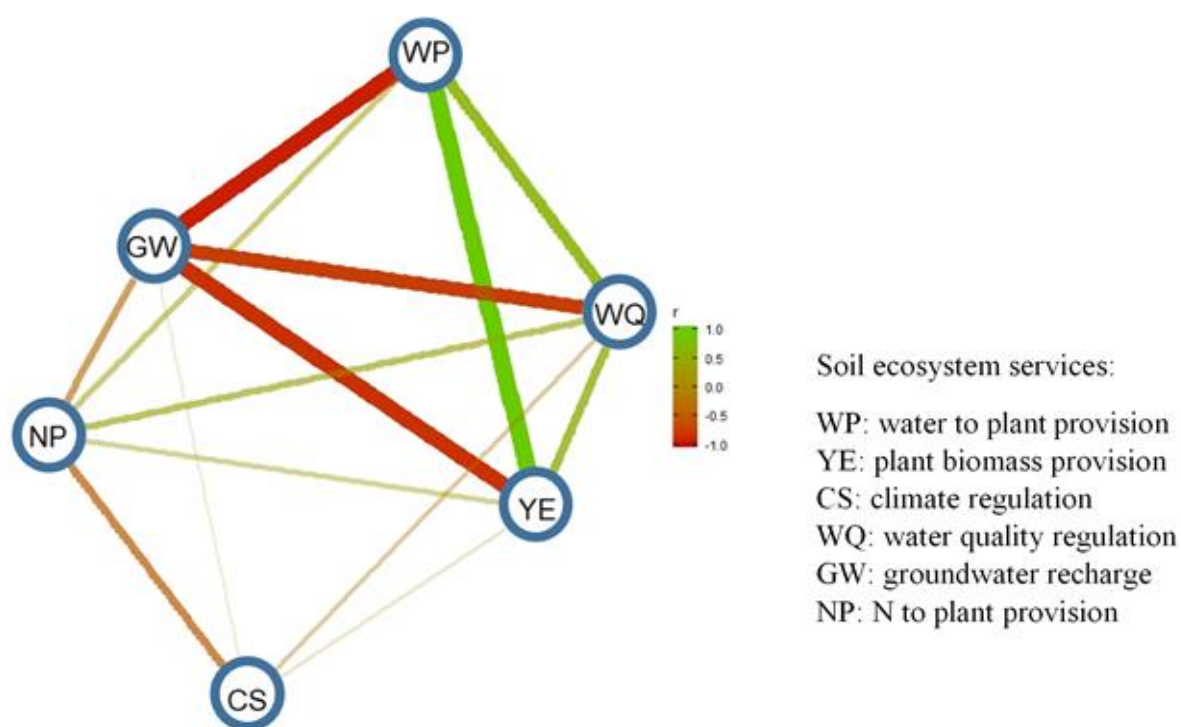


Figure 6.7 : Network chart showing interrelations among soil ecosystem service indicators based on their mean values over the baseline period. Only significant correlations are shown. Line width indicates correlation intensity.

Table 6.4 : Pearson correlation coefficients (r) between soil ecosystem service (SES) indicators and soil properties. (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision; AW: available water capacity; Obs: maximum rooting depth; pH: soil pH of the topsoil; CC: soil organic carbon content of the topsoil)

SES	WP	GW	NP	YE	CS	WQ	AW	Cly	pH	CC
WP	1									
GW	-0.93***	1								
NP	0.31	-0.39	1							
YE	0.97***	-0.86***	0.25	1						
CS	0.02	0.11	-0.48**	0.13	1					
WQ	0.66***	-0.81***	0.4	0.52***	-0.23	1				
AW	0.98***	-0.94***	0.33	0.92***	-0.04	0.73***	1			
Cly	0.31	0.03	0.18	-0.47***	-0.51**	0.37	-0.17	1		
pH	0.27	-0.35	0.55**	0.17	-0.71***	0.37	0.34	0.42*	1	
CC	0.33	0.16	0.05	-0.45*	-0.66***	0.23	-0.26	0.61***	0.1	1
Obs	0.86***	-0.85***	0.38	0.73***	-0.16	0.68***	0.88***	-0.01	0.1	-0.12

Asterisks denote significance (t-test) (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$);

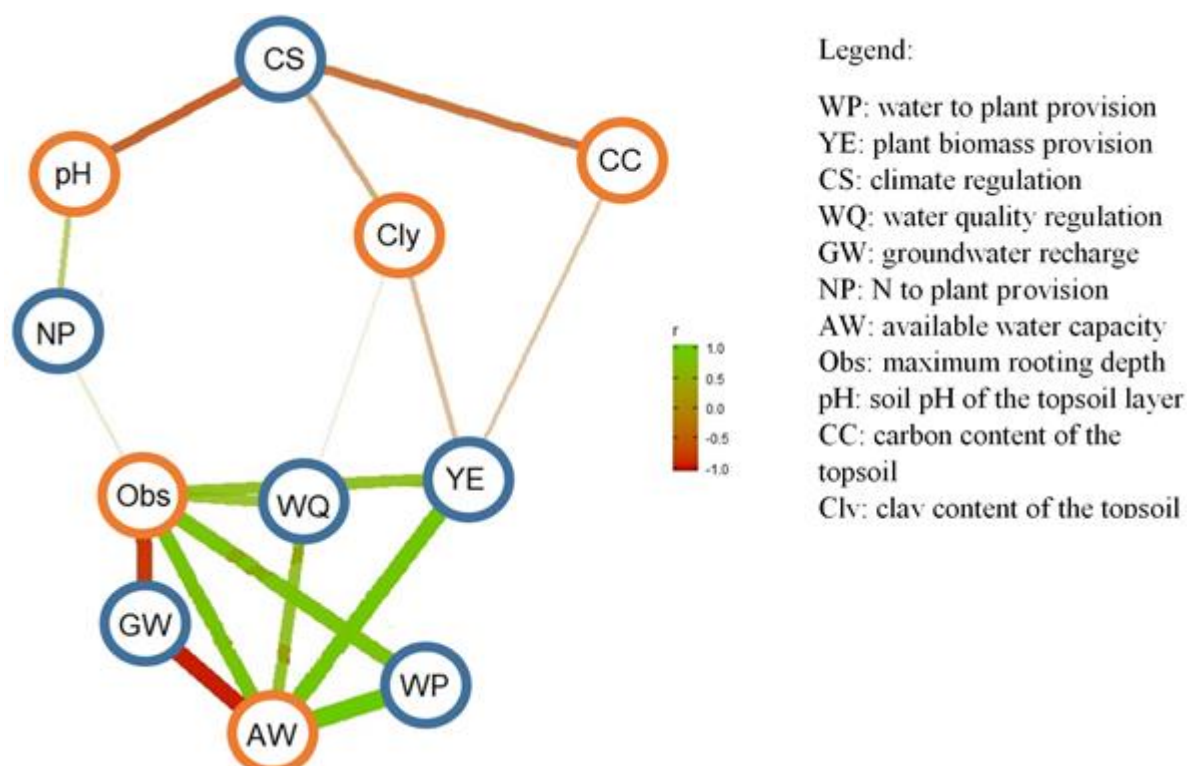


Figure 6.8 : Network chart showing interrelations between the SES indicator means values over the baseline period (blue circles) and soil properties (orange circles). Only significant correlations are shown. Line width indicates correlation intensity.

6.3.5. Interrelation between SES and soil properties

In general, WP was strongly and positively correlated with some soil properties, such as available water storage capacity ($r=0.98$) and maximum rooting depth ($r=0.86$) (Table 6.4). Similarly, YE had high positive correlation with available water storage capacity ($r=0.92$) and maximum rooting depth ($r=0.73$) and significant negative correlation with clay content ($r=-0.47$) and SOC content ($r=-0.45$). However, CS was strongly correlated only with physicochemical soil properties such as SOC content ($r=-0.66$), pH (-0.71) and clay content ($r=-0.51$). Visual inspection of this network chart (Figure 6.8) highlighted the predominant role of soil depth via maximum rooting depth and available water storage capacity. These variables were strongly correlated with WP, YE, WQ and GW. More importantly, GW was negatively correlated with WP, WQ, YE (Figure 6.7) and soil depth variables (Figure 6.8). In contrast, NP and CS were correlated only with physicochemical soil properties. Thus, soil depth and soil hydrodynamic features were the main drivers of GW, WQ, WP and YE, whereas physicochemical soil properties were the main drivers of NP and CS.

6.3.6. Multivariate analysis of interrelations between SES and soil properties

Projecting all variables onto the PCA factorial plan, the first two PCs explained nearly 70% of the total variance, of which PC1 explained 50.5% (Figure 6.9a). PC1 contrasted YE, WP, WQ and GW, whereas PC2 contrasted CS and NP. More importantly, WP and GW were positively correlated with their standard deviations, whereas NP, CS and WQ were negatively correlated with their standard deviations. Projection of supplemental soil properties on PC1 highlighted that water holding capacity and soil depth were strongly correlated with YE, WP and GW. Furthermore, pH, clay content and SOC content were strongly correlated with NP and CS. Thus, depth and texture soil characteristics influenced water-related SES, while physicochemical soil properties influenced carbon and N balances.

By adding barycentre ellipses ($\alpha = 5\%$) of parent material and soil type groups to the PCA results, we distinguished seven ellipses for parent material (Figure 6.9b) and four ellipses for soil type (Figure 6.9c). Overall, the PCA accurately separated the sites according to their soil type, particularly Histosols, Luvisols, Cambisols and Leptosols. Luvisols were associated with high YE and WP. Histosols were associated with high WQ and NP. Cambisols had high CS, while Leptosols had the highest GW.

Regarding parent material, clusters of alluvial terrace, gritty schist, sandstone and granite overlapped substantially, whereas those of Aeolian loam and marsh were widely separated. Aeolian loam parent material was associated with YE and WP, marsh was associated with NP and WQ, and the other clusters were associated with CS and GW. Hence, shallow sandy soils had the highest GW and CS, deep loamy soils with high soil water storage capacity had the highest YE and WP and the relatively hydromorphic Histosols had the highest WQ and NP.

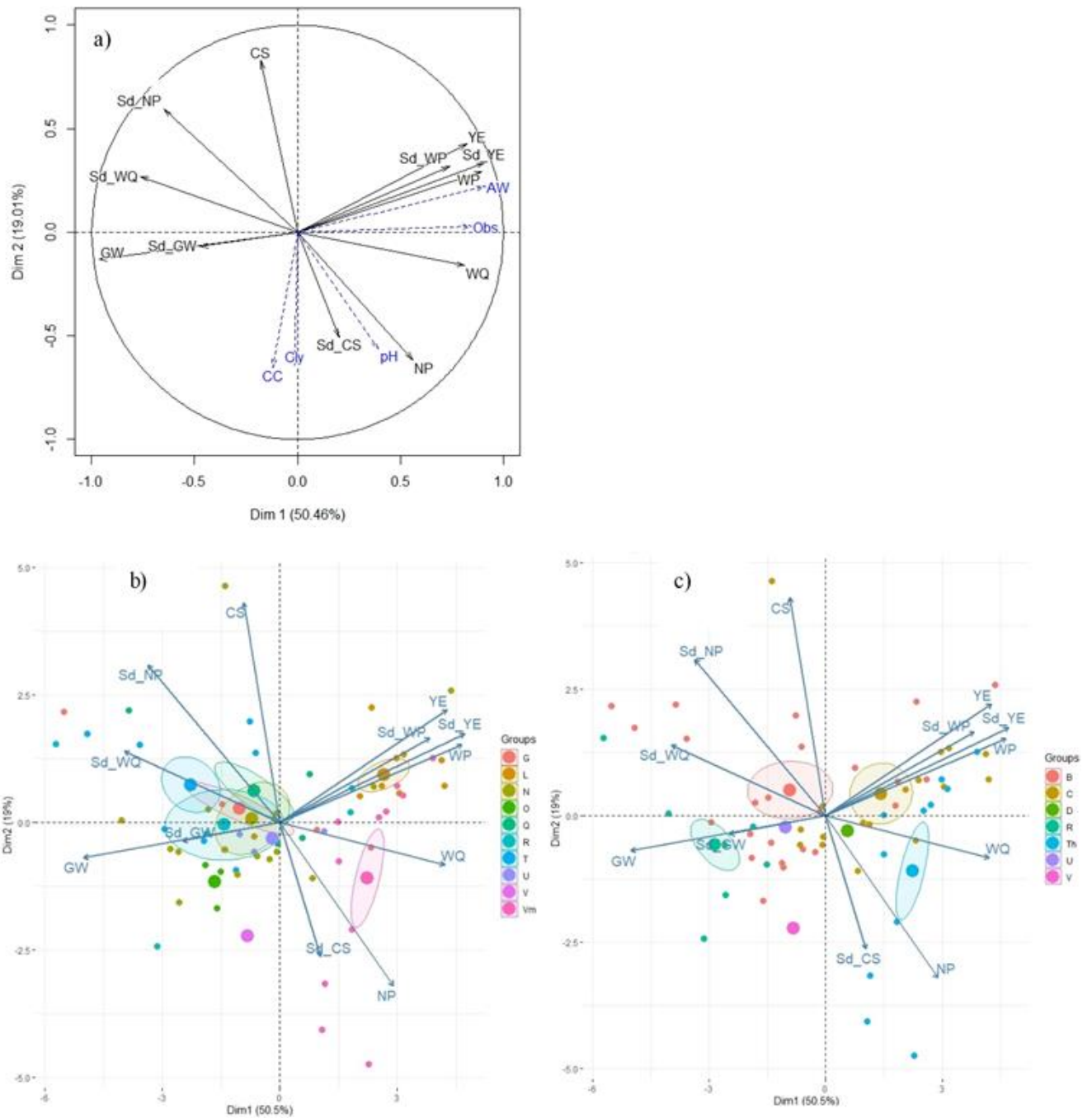


Figure 6.9 : a) Plots of SES indicator means with their associated standard deviations and some soil properties as supplemental variables on the two first PCA principal components. b) Biplot of PCA variables and the 64 sites according to parent material ($\alpha=5\%$) (G: granite, L: Aeolian loam, N: soft schist, O: medium schist, R: gritty schist, T: alluvial terrace, U: colluvial deposits, V: alluvial deposits, Vm: marsh) c) Biplot of PCA variables and the 64 sites according to soil type ($\alpha=5\%$) (B: Cambisol, C: Endostagnic Cambisol, D: Stagnic Alluvisol, R: Leptosol, Th: Histosol, U: Colluvic Cambisol, V: Fluvisol).

Stepwise multiple-linear regression for each of the six SES indicators highlighted the large contribution of inherent and manageable soil properties to explaining the indicators. In general, all SES regression equations had high adjusted R^2 values (>0.80), except for NP, for which it was low (0.30) (Table 6.5). All SES indicator regressions were significant at $p < 0.001$.

As expected, variability in SES indicators was driven mainly by soil characteristics. The MANOVA (Table 6.6) demonstrated that SES provision differed significantly between the three classes of clay content, the two classes of soil pH, the two classes of soil depth, the three classes of soil drainage, the 10 classes of soil parent material and the seven classes of soil type. Therefore, SES provision differed for each soil property class, and the pure effect of soil determinants was significant regardless of the SES.

Table 6.5 : Statistics of stepwise multiple linear regression models predicting means soil ecosystem service (SES) values over the baseline period as a function of soil properties. (WP: water to plant provision; YE: Plant biomass provision; CS climate regulation; WQ: water quality regulation; GW: groundwater recharge; NP: N to plant provision; AW: available water capacity; Obs: maximum rooting depth; pH: soil pH of the topsoil; CC: soil organic carbon content of the topsoil; Cly: clay content of the topsoil).

SES	Soil properties selected in the model	Adjusted R^2	F-statistic	Significance
WP	AW, clay, soil depth class	0.97	849	< 0.001
GW	AW, clay, obs, soil drainage, soil depth class	0.91	119	< 0.001
NP	pH	0.30	26	< 0.001
YE	AW, clay, obs	0.94	362	< 0.001
WQ	AW, pH, CC, clay, obs, soil depth class	0.82	49.6	< 0.001
CS	pH, CC, clay, soil depth class	0.87	110	< 0.001

Table 6.6 : Results of a Multivariate Analysis of Variance (MANOVA) model comparing clay, soil pH, parent material, soil type, soil depth and soil drainage effects on mean values of soil ecosystem services over the baseline period. The F-statistic, approximated from Wilks' lambda, denotes the importance of model terms in explaining SES.

Term	F-statistic	Significance
Clay	0.22	< 0.001
Soil pH	0.57	< 0.001
Parent material	0.08	< 0.001
Soil type	0.15	< 0.001
Soil depth	0.52	< 0.001
Soil drainage	0.26	< 0.001

6.4. Discussion

6.4.1. Comparing SES evaluations with a focus on the modelling approach limitations

Six SES indicators were estimated using the STICS model, and then interrelations between SES and between SES and soil properties were investigated. This study contributes to an emerging literature that attempts to assess multiple SES and their interrelations at a large scale. Analysing how SES provision varies with pedogenetic variability over a long temporal scale distinguishes our approach from other studies, such as those of Grimaldi et al. (2014) and Calzolari et al. (2016). These studies considered SES as static and assessed them using a set of indicators expressed at a given point in time.

As mentioned by Beaudoin et al. (2008), STICS simulates crop production without any significant temporal drift. For instance, the mean yield of winter wheat among the 64 sites was predicted at 7.6 and 6.3 t ha⁻¹ in 2014 and 2018, respectively, similar to those reported in agriculture surveys in the Agreste database for the Ille-et-Vilaine department (7.5 and 6.8 t ha⁻¹, respectively). Likewise, the mean yield of grain maize was predicted at 8.4 and 8.6 t ha⁻¹ in 2015 and 2017, respectively, similar to those in the Agreste database (ca. 8.5 q ha⁻¹ in both years). Even though STICS does not consider many factors, such as pests and diseases, it predicted biomass production over the baseline period well. When present, these factors can reduce crop production greatly (Meynard et al., 2002; Beaudoin et al., 2008; Fraga et al., 2018).

Our results showed that GW depended strongly on soil depth and available water storage capacity. Shallow soils with coarse texture had the highest GW, whereas deep soils with a medium texture had the lowest GW. GW was estimated at 298 (±94) or 122 (±88) mm for soils with a water storage capacity <40 mm or 120 mm, respectively (Table 6.3) The EFESE-EA (2017) study also demonstrated this trend, highlighting that GW increased as water storage capacity decreased, and vice versa. In their study, GW reached 430 mm or did not exceed a mean of 300 mm for soils with a water storage capacity <40 or >120 mm, respectively. The difference in GW reported by these studies may have several explanations. First, EFESE-EA STICS simulations were based on pedoclimatic units whose soil features were available at 1:1,000,000 scale. Hence, the relative accuracy of soil data, particularly those of stony materials, can influence GW predictions significantly. Second, hydrodynamic soil properties such as W_{FC} , W_{WP} and bulk density were estimated using different regional and national pedotransfer soil functions. Finally, EFESE-EA

predictions were based on a hydrological year time step (1 September to 31 August), whereas our predictions were based on a calendar-year time step (1 January to 31 December). In addition, some limitations of STICS should be noted. STICS is a field-scale model that simulates only water flows at a single point, rather than within an area. Therefore, it does not consider lateral spatial interactions, though it can consider deep percolation.

WP was directly proportional to water storage capacity. Indeed, mean WP was ca. 130 mm year⁻¹ for shallow sandy soils but greater than 200 mm year⁻¹ for deep loamy soils. This result agrees with those of the EFESE-EA, which reported similar WPs strongly correlated with water storage capacity. In general, both soil texture and structure influence pore-size distribution: medium- and fine-textured soils have high total water storage capacity due to high porosity and the ideal combination of meso- and micro-porosity (O'Geen, 2012). Likewise, Calzolari et al. (2016) demonstrated that fine-textured soils had a high water storage capacity, whereas coarse-textured soils had a low water storage capacity.

As WQ is an N-based indicator, WQ and NP had similar responses to pedogenetic variability. Both indicators exceeded a mean of 70%. They had high variability for shallow soils (low YE and high GW) and low variability for deep soils (high YE and low GW). This behaviour was related strongly to inherent soil properties (e.g. soil texture, soil depth), agricultural practices (e.g. mineral fertilisation, soil tillage) and rainfall pattern. Due to their large pore space, deep soils with less sand and more silt, clay or organic matter retained more N, which corroborates results of Powell-Graines and Gaines (1994), who studied the relationship between soil texture and N leaching. Furthermore, fine-textured soils can retain more N than coarse-textured soils, which directly influences N availability for plants: more N is available for plants in deep fine-textured soils than in shallow coarse-textured soils. Our results agreed those of the EFESE-EA (2017), which highlighted high N retention capacity of soils in Brittany. Furthermore, Van Wijnen et al. (2012) assessed natural decrease in pollutants using indicators of N-related processes (e.g. soil pH, potential mineralisation rate of organic N, potential carbon mineralisation rate, functional microbial activity). They found that natural decrease in pollutants is high in arable fields on clay soils and low in most arable fields on sandy soils. The same trend was also confirmed by Dominati et al. (2013) and Calzolari et al. (2016). However, most studies like ours assess WQ as the amount of N released into groundwater after passing through the soil profile. In fact, soils also filter organic matter, phosphorus (P) and pathogens, which can differ in form, stability and solubility and thus in the probability that they will be retained or released. In our context, the agricultural

practices used involve adding exogenous organic matter and mineral fertilisation. Therefore, assessing WQ as a function of P and dissolved organic carbon (DOC) leaching would provide an informative three-dimensional indicator (N, P and DOC).

CS reflects carbon flows that influence SOC stock from 0-30 cm. Processes involved in carbon balance include net primary production, incorporation of crop residues, organic matter decomposition, erosion and DOC leaching. The net balance among these processes determines whether a soil loses or accumulates carbon. CS is provided when soil carbon accumulates (positive carbon net balance), but loss of soil carbon is considered degradation (Dominati et al., 2013). In the present study, situations of carbon accumulation or loss were identified, but no trends were identified as a function of pedogenesis. This can be explained mainly by the complexity of carbon sequestration, which STICS failed to simulate accurately. Carbon can be sequestered through temporal changes in land use and agricultural practices (Ellili et al., 2019).

6.4.2. Explaining and interpreting SES interrelations

According to Qin et al. (2015) and Vallet et al. (2018), few studies have attempted to assess SES interrelations. In general, trade-off, synergy and neutral interactions between SES depend strongly on agro-pedoclimatic conditions. Our results showed strong synergy for the pairs WQ-NP, WQ-WP, YE-WQ and YE-WP, which means that both SES decrease or increase together. As expected, YE is synergic with WP, which is consistent with the literature. Many researchers (e.g. Rodney-Bennett and Harms, 2011) observed a positive correlation between crop yield and crop transpiration. Water is a transporting agent for all plant physiological processes, and any water deficit can impair plant metabolism and consequently decrease crop yield. It was not surprising to find synergy between WQ and NP. As NP is based on soil N mineralisation, increasing N availability for plants can increase biomass production and thus plant N uptake, which reduces N leaching from the soil. Likewise, both water and N requirements increase when crop yield increases, which may explain the synergy between WP and NP. Calzolari et al. (2016) found a similar relationship between YE and WQ when assessing multiple indicators. Other studies, such as that of Jopke et al. (2015), found a trade-off between crop production and water regulation ($r=-0.23$) at the European scale, albeit without distinguishing crop areas from forest or grassland. It is notable that the pairs CS-YE, CS-GW and CS-WQ had non-significant relationships (r values close to 0). Jopke et al. (2015) also studied multiple SES, observing a neutral relationship between crop production and carbon storage, as well as between water provision and air quality regulation,

with no significant correlation ($r = -0.06$ and -0.01 , respectively). Egoh et al. (2002) reported a weak correlation between primary productivity and carbon storage ($r = -0.14$), carbon storage and water supply ($r = 0.08$) and carbon storage and soil water retention ($r = -0.17$). In contrast, Raudsepp-Hearne et al. (2010), found a strong trade-off between crop production and carbon sequestration ($r = -0.88$). The differences among results can be explained by the simulation model we used to quantify SES: the absence of model validation for this set of SES may limit the confidence with which we can assess SES interrelations (Bagstad et al., 2013; Nemeč and Raudsepp-Hearne, 2012).

The pairs GW-WP, GW-WQ and GW-YE displayed trade-off relationships. Overall, a decrease in GW implied a decrease in the amount of N that can reach the groundwater and thus an increase in WQ. Hence, as in former studies (e.g., United Nations, 2005), we identified a trade-off between provisioning SES (GW) and regulation SES (WQ). However, the trade-off between GW and WP and between GW and YE may be explained by temporal and geographic scale effects. In our context, the spring crop (maize) had high water requirements, particularly during the flowering period, which coincided with moderate rainfall amounts. The winter crop did not have this problem. Hence, an important caveat to our analysis is that temporal correlation was not considered when interpreting SES interrelations. Many researchers have used the “change over time” approach (Tomscha and Gergel, 2016) or “correlation analysis between the amounts of changes in SES” (Zheng et al., 2014, Qin et al., 2015) to detect changes in SES relationships over time.

6.4.3. Explaining and interpreting interrelations between SES and soil properties

Clearly, soil properties have a strong influence on SES provision. Globally, soils differ at regional scales (Vincent et al., 2018; Ellili et al., 2019) down to the local scale of soil catenas (Ellili et al., 2019), at which landscape ecology is studied most. We considered easily accessible inherent soil properties such as soil parent material, soil type, soil depth and maximum rooting depth, as well as other properties measured in the shallowest layer, such as SOC content, texture and pH. The correlation network chart indicated that maximum rooting depth and soil water storage capacity, which integrates effects of soil texture, structure and depth, were the main drivers of YE, WP, GW and WQ. These results are consistent with the literature. Dominati (2013) clearly stated that available water capacity is an essential determinant of food provision supported by natural capital

stocks. The distribution of pore sizes influences the supply of gases and nutrients to plant roots. In addition, available water capacity is a key factor not only for plant development but also for flow regulation, because it influences draining of excess water. Dominati (2013) also pointed out that filtering of nutrients or contaminants and groundwater recharge depend strongly on soil depth, texture and structure. Many researchers (Egoh et al 2008, Grimadi et al 2014, Clazolarli et al 2016, Clech'h et al., 2016, Greiner et al., 2017, Drobnik et al., 2018) argue that inherent soil properties, particularly soil structure, available water capacity and depth are the main drivers of provision of YE, WP, GW and WQ.

In the present study, YE and carbon content had a negative correlation ($r = -0.47$). The same result was reported by Raudsepp-Hearne et al. (2010) ($r = -0.48$) and Jopke et al. (2015). We also found a negative correlation between CS and pH, which is a strong driver of soil microbial diversity and strongly influences the microbial carbon cycle. It can regulate processes that influence carbon dynamics, particularly mineralisation, assimilation and accumulation of organic matter inputs by soil microbial communities. Many studies, such as that of Malik et al. (2018), identified that soil pH negatively influences SOC stock. In contrast to the review of Singh et al. (2017), we also found a negative correlation between CS and clay content. In general, carbon mineralisation is lower in fine-textured soils than in coarse-textured soils (Hassink, 1992). As SOC is less accessible in fine-textured soils, the latter have a greater ability to protect SOC against microbial attack than coarse-textured soils (Hassink, 1997). However, carbon sequestration relies on complex interactions among soil properties, landscape management and environmental conditions that are complicated to simulate simultaneously.

As seen, the results of this study were generally similar to those of related studies. However, as correlation coefficients are the only measure available with which to compare results, we cannot yet compare the additional information obtained with the PCA, MANOVA, regression-based methods or bagplots. More importantly, all statistical analyses were complementary and confirmed the strong influence of soil properties in explaining SES variances, which was consistent with Grimaldi et al. (2014)'s results. Surprisingly, for all SES except NP, more than 87% of the total variance of SES could be explained using only inherent soil properties. The MANOVA revealed that the bundle of SES differed significantly among soil-property groups ($p < 0.001$). The PCA, a co-inertia analysis, which is a more flexible multivariate method, confirmed that soil properties were the key drivers of SES provision and provided additional information about the relationships between SES and their associated standard deviations. Hence,

as mentioned by Mouchet et al. (2014), PCA, bagplots, MANOVA, regression-based methods and correlation matrices are effective approaches to analyse SES interrelations, while network analysis is a powerful way to visualize their relationships. Although the present study attempted to capture temporal variability by averaging SES indicators over 30 years, it did not address the temporal dimension sufficiently.

6.5. Conclusion

We developed a methodological framework for assessing SES and analysing interactions among them. We provided multiple SES indicators predicted by the STICS crop model, which used input data describing soils, weather and agricultural practices over 30 years. Our results showed that landscape-scale pedogenetic variability, as well as within-soil variability, even though it was briefly investigated, were the key drivers of SES provision. We revealed that deep soil with good hydrodynamic properties produced the most biomass and satisfied plant water requirements the most due to its high water holding capacity. In contrast, shallow soils with a coarse texture had the highest groundwater recharge but lower water-quality buffering. Most of our results were consistent with the literature, specifically with EFESE-EA (2017) results, regardless of the data used. In our context, YE-WP, YE-WQ and YE-NP displayed synergies, whereas GW-WQ, GW-WP and GW-YE displayed trade-offs. Like other researchers (e.g. Jopke et al. (2015)), we failed to demonstrate the well-known trade-off between CS and YE, possibly due to the scale effect and the complexity of the carbon sequestration process. Overall, the spatial pattern of SES provision and interactions among SES should be studied in the future to provide maps for decision support for both stakeholders and decision makers.

Conclusion du chapitre 6

Dans ce chapitre nous avons cherché, dans un premier temps à évaluer les services écosystémiques des sols agricoles à l'échelle parcellaire en considérant une échelle temporelle moyenne de 30 ans. Cette étude a essentiellement porté sur l'évaluation biophysique de six services écosystémiques, à savoir la régulation du climat ; la régulation de la qualité de l'eau ; la recharge de la nappe souterraine ; la fourniture d'eau à la plante ; la fourniture d'azote à la plante ; et la production de biomasse agricole. L'évaluation biophysique a été faite à l'aide du modèle de culture STICS en intégrant une grande variabilité pédologique et simulant des pratiques agricoles et des itinéraires techniques couramment mis en œuvre à l'échelle de la zone d'étude. Les données pédologiques mobilisées sont issues d'observations et de mesures de nombreuses propriétés du sol faites sur 64 sites choisis pour couvrir une diversité importante des sols agricoles, notamment de par leur matériau parental d'origine. Sur l'ensemble des sites sélectionnés, la variabilité interannuelle ainsi que la variabilité globale sur toute la période d'étude ont été caractérisées en simulant l'évolution dynamique des processus du sol. Dans un second temps, nous avons cherché à exploiter les évaluations réalisées pour caractériser les interrelations entre les services écosystémiques, mais aussi pour identifier les déterminants pédologiques des services écosystémiques instruits. De manière globale, l'approche adoptée est novatrice car elle promeut la prise en compte de l'information pédologique dans le processus d'évaluation des services écosystémiques des sols, l'analyse conjointe des services instruits et leurs interrelations, et l'identification de leurs déterminants pédologiques.

En synthèse, nous avons constaté que la variabilité des services écosystémiques des sols présente une double composante de variabilité intra-annuelle d'origine climatique et de variabilité pédologique. En effet, les services de production de biomasse agricole et de fourniture d'eau à la plante atteignent des niveaux importants pour les années humides et les sols ayant des caractéristiques agronomiques favorables telles qu'une porosité et une profondeur de sol importantes. En revanche, les services de régulation de la qualité de l'eau et la fourniture d'azote à la plante révèlent des niveaux importants indépendamment du type de sol et de ses propriétés. Globalement, les évaluations biophysiques établies sont en cohérentes, avec les travaux de l'EFESE-EA, notamment vis-à-vis de la fourniture de services écosystémiques en fonction de la variabilité des sols. Par exemple, les sols dont la capacité de rétention en eau est limitée montrent des niveaux importants du service de recharge de la nappe souterraine.

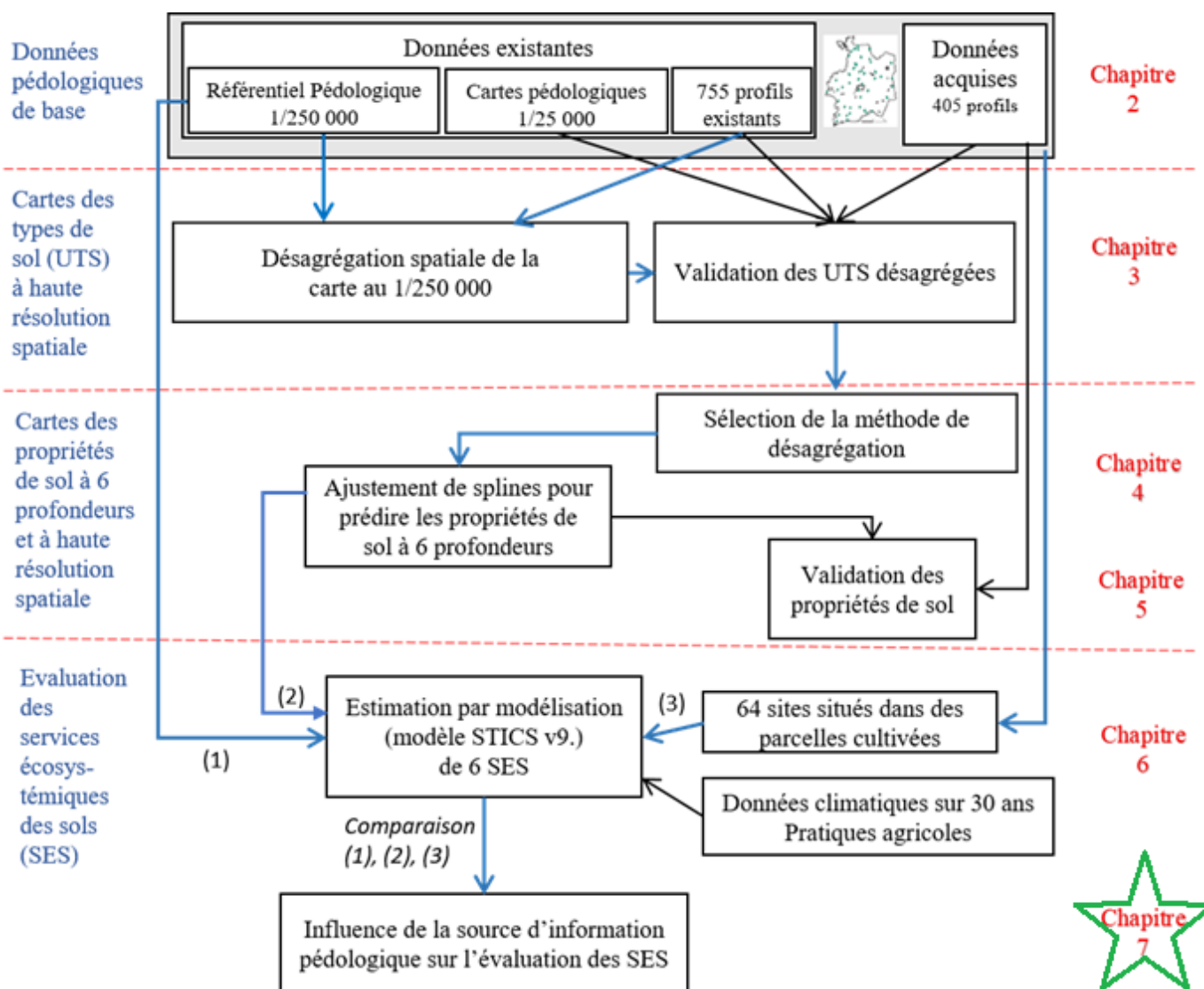


Figure 7.0 Schéma de la structure générale de la thèse explicitant les données mobilisées et les différentes étapes de désagrégation spatiale de l'information pédologique existante en vue de l'évaluation des services écosystémiques des sols par modélisation dynamique des sols.

Chapitre 7 :
Sensibilité de l'évaluation des
services écosystémiques
rendus par les sols à la source
de l'information pédologique

Sensibilité de l'évaluation des services écosystémiques rendus par les sols à la source de l'information pédologique

Abstract

Soil is a natural capital that provides several ecosystem services ensuring human well-being and sustainable socioeconomic development. The scarcity of soil information is the main shortcoming to assess soil ecosystem services (SESs). At regional level, soil information is available at coarse spatial resolution (1:250,000 in Brittany, France) and represented as a tessellation of polygon defining Soil Map Units (SMU), where each SMU can include one or several Soil Type Units (STU) with given proportions derived from expert knowledge. Spatial disaggregation of complex SMU appears to be a relevant approach to enhance soil spatial resolution. The aim of this study was to test the hypothesis that using disaggregated soil information rather than legacy soil data 1:250, 000 leads to a more accurate SESs assessment. Four SESs (climate regulation, crop production, water quality regulation and soil fertility maintain) were assessed using the STICS soil crop model mobilizing successively three different sources of soil information: (i) legacy soil properties maps 1:250,000, (ii) disaggregated soil properties maps at 50 m resolution and (iii) field measurements as validation dataset. Over an area of 6775 km², 64 sites located in croplands were sampled according to the standard *GlobalSoilMap* depths (0-5cm, 5-15cm, 15-30cm, 30-60cm, 60-100cm and 100-200cm) to determine soil properties: soil organic carbon content, 5 particle-size fractions, coarse fragments, CEC, and pH. Conventional soil map 1:250,000 was previously disaggregated using DSMART algorithm with soil landscape relationships at 50 m resolution (Vincent et al., 2018). Soil properties were then predicted at the *GlobalSoilMap* standard depths by fitting spline functions. For legacy soil maps, for each SMU and soil property, the value of the dominant STU was allocated to the whole SMU. 8263 STICS simulations were computed over a period of 30 years. Six SESs indicators were estimated from STICS outputs to evaluate SES supplies. The main results show that 1) SES provision depends on inherent soil properties and is marginally sensitive to the source of soil information considered. 2) Disaggregating conventional soil improves the accuracy of the soil properties prediction at regional scale and therefore partially compensates the lack of soil information for SES assessment over large areas. Overall, the results of this work allowed evaluating the impact of pedological information quality on the SESs assessment using available tools and models.

7.1. Introduction

Dans les chapitres précédents, nous avons cherché à améliorer la mise à disposition de l'information spatialisée sur les sols et à quantifier leurs incertitudes. Pour cela, nous avons mis en œuvre une approche de désagrégation spatiale du RRP de Bretagne, en utilisant l'algorithme DSMART avec des règles *a priori* sur les lois de distribution des sols spécifiques au contexte breton. Les cartes désagrégées, à une résolution spatiale de 50 m, ont été ensuite mobilisées pour caractériser les distributions spatiales des propriétés de sol à six profondeurs standards selon les spécifications du consortium *GlobalSoilMap* (Chapitre 4). Un jeu de données indépendant, collecté selon un échantillonnage probabiliste, a été utilisé pour quantifier la qualité prédictive des cartes pédologiques délivrées (Chapitre 5). Tout ce travail préliminaire avait pour objectif principal d'améliorer la disponibilité de l'information pédologique existante pour promouvoir son intégration et sa prise en compte dans des procédures d'évaluation des services écosystémiques des sols. La majorité des synthèses bibliographiques, telles que celle de Adhikari et Hartemink (2016), expliquent l'absence d'une considération formelle de l'information pédologique dans l'évaluation biophysique des services écosystémiques des sols par sa faible disponibilité de manière exhaustive sur les sites d'étude, mais aussi par sa faible qualité. Cependant, aucune étude n'a cherché à évaluer l'effet de la qualité de l'information pédologique sur l'évaluation des services écosystémiques des sols. Cela est généralement justifié en considérant qu'une information pédologique non précise ne permet pas de délivrer des quantifications biophysiques cohérentes et surtout fiables des services instruits. Ainsi, l'étude de l'effet de la source de l'information sur les sols sur l'évaluation des services écosystémiques est nécessaire pour identifier les intérêts opérationnels des bases de données pédologiques existantes et proposer des pistes d'amélioration pour générer des évaluations cohérentes avec les caractéristiques du milieu et les problématiques traitées.

L'objectif de cette étude est d'étudier l'effet de la source d'information pédologique sur l'évaluation de plusieurs services écosystémiques (deux services de régulation et quatre services d'approvisionnement) évalués par une approche de modélisation. Pour l'essentiel, trois sources d'information pédologiques ont été mobilisées : (i) l'information contenue dans la base de données du RRP en considérant les caractéristiques de l'UTS dominante de chaque UCS ; (ii) les propriétés de sols dérivées des cartes désagrégées ; (iii) et les propriétés de sol issues des observations de terrain et analyses faites en 64 sites répartis sur tout le département d'Ille-et-Vilaine et situés dans des parcelles agricoles.

7.2. Matériel et méthodes

7.2.1. Données d'entrée pour la modélisation mise en œuvre

Comme dans le chapitre 5, 64 sites situés dans des parcelles agricoles en Ille-et-Vilaine, ont été retenus pour évaluer l'effet de l'information pédologique sur la quantification biophysique des services écosystémiques des sols. Pour cela, nous avons opté pour une approche de modélisation fondée sur l'utilisation du modèle de culture STICS (Brisson et al., 2008) dans sa version 9.0, en considérant une série temporelle de 30 ans. Trois types de données sont nécessaires pour alimenter le modèle STICS :

- Les données climatiques : celles-ci portent sur des données journalières de 6 variables climatiques, à savoir les précipitations, la vitesse du vent, les températures maximale et minimale, l'humidité relative de l'air et le rayonnement global. Toutes ces variables sont exprimées au pas journalier sur toute la période du 1^{er} janvier 1988 au 31 décembre 2018. Les données climatiques sont extraites de la base de données CLIMATIK de l'Inra (Inra, 2019) et correspondent à la station météorologique de Saint Jacques-de-la-Lande, située près de Rennes. Les caractéristiques générales de ces données climatiques ont été présentées dans le chapitre 6.
- Les données sur les itinéraires techniques agricoles : comme pour les données climatiques, les données sur les itinéraires techniques mobilisées dans cette étude, correspondent à celles utilisées dans le chapitre 6. Pour le jeu de données dont nous disposons, l'information sur les pratiques agricoles n'est pas renseignée. De ce fait, nous avons décidé de simuler des pratiques agricoles correspondant aux pratiques agricoles dominantes des agriculteurs en Ille-et-Vilaine en termes de rotation culturale (maïs-blé-interculture), de travail du sol, de calendriers des travaux et de fertilisations organique et minérale. Les pratiques sont donc fixées homogènes pour l'ensemble des sites étudiés.
- Les données pédologiques : trois sources d'information pédologique ont été mobilisées. La première source concerne les données d'observation et les analyses issues de la campagne de terrain portant sur les 64 sites en milieu agricole (chapitre 6). La deuxième source de données correspond aux données de propriétés du sol dérivées des cartes de sols désagrégées (chapitre 4). Ces données sont issues d'une désagrégation spatiale du RRP de Bretagne en utilisant l'algorithme DSMART et des règles *a priori* sur les lois de distribution des sols en Bretagne. Le couplage de l'ajustement des fonctions splines aux caractéristiques des UTS et les cartes caractérisant la distribution spatiale des UTS en Bretagne, a permis de délivrer des cartes caractérisant la distribution spatiale des propriétés de sol à l'échelle régionale. La dernière source de données est issue d'une simple extraction des données pédologiques fournies dans le RRP de Bretagne. Nous

admettons que l'UTS dominante de chaque UCS est représentative de la variabilité pédologique de l'UCS dans laquelle elle se trouve. De ce fait, un jeu de données avec les caractéristiques pédologiques de l'UTS dominante de chaque UCS a été construit pour caractériser l'information brute de la base sémantique Donesol 3.

En synthèse, chaque site est caractérisé par trois jeux de données sur les propriétés des sols issues des trois sources d'information.

7.2.2. La procédure de simulation STICS

Après avoir rassemblé les données d'entrée du modèle de culture STICS, nous avons croisé toutes les données sur les pratiques agricoles, le climat et les sols, pour les exprimer au format requis par le modèle. Les données sur les itinéraires techniques et le climat sont identiques pour tous les sites, mais l'information sur le sol varie en fonction des trois sources de données. Le croisement des données pédoclimatiques permet de construire des unités de simulation de base (USM), qui correspondent à des parcelles homogènes en termes de sol, de climat et de culture. Plusieurs USM ont été définies et enchainées pour simuler les rotations de culture couvrant l'ensemble de la série temporelle (1988 – 2018).

Dans le cas des données sur le sol, en plus de la teneur en argile, du pH et de la teneur en azote organique, STICS nécessite des données caractérisant les propriétés hydrodynamiques des sols à savoir la teneur en eau à la capacité au champ, la teneur en eau au point de flétrissement permanent, mais aussi la densité apparente de chaque couche de sol. Or, ces données ne sont pas mesurées sur le terrain, ni renseignées dans la base de données Donesol 3. De ce fait, nous avons utilisé des fonctions de pédotransfert régionales (Remy et al, 2015) pour les caractériser vis-à-vis des trois jeux de données de sol mobilisés.

Par ailleurs, afin d'automatiser le lancement séquentiel des simulations STICS, nous avons développé un programme JAVA qui permet d'enchaîner de façon séquentielle sur la période étudiée (1988-2018) les USM créées avec chaque jeu de données pédologiques. Le croisement, au niveau des 64 sites d'étude, de l'information pédologique, des pratiques agricoles et des données climatiques génère 3565 USM pour les données d'observations et les données désagrégées, mais uniquement 307 USM pour les données brutes du RRP de Bretagne, plusieurs sites pouvant être caractérisés par la même UTS dominante. Au final, un total de 7437 USM a été simulé en considérant les trois sources de données pédologiques.

7.2.3. Evaluation des services écosystémiques des sols

Six services écosystémiques des sols ont été évalués par le biais de six indicateurs, choisis en fonction de leur pertinence, mais aussi en fonction des sorties de modèle disponibles. Ces services portent sur la contribution du sol à la régulation du climat (CS), à la régulation de la qualité de

l'eau (WQ), à la fourniture d'azote à la plante (NP), à la fourniture d'eau à la plante (WP), à la production de biomasse (YE) et à la recharge de la nappe souterraine (GW). La définition précise des indicateurs associés à ces services est détaillée dans le §2.5 du chapitre 6. Ces six indicateurs sont évalués pour chaque année calendaire : la variation annuelle du stock de carbone dans le sol sur 30 cm, la quantité d'eau drainée au-delà du système racinaire de la culture, la transpiration des plantes, le rapport entre la quantité d'azote lixiviée et l'azote retenu dans la porosité matricielle du sol et finalement l'efficacité de couverture des besoins azotés des plantes.

Tous ces indicateurs sont définis et calculés en fonction des variables caractérisant les processus dynamiques des sols issus du modèle STICS v9.0, comme expliqué dans le chapitre précédent. Cependant, on ne considère dans ce chapitre que la valeur moyenne interannuelle des indicateurs, qui intègre la variabilité temporelle couvrant la période 1988-2018.

7.3. Résultats

7.3.1. Distribution des SES moyens sur 30 ans selon la source d'information pédologique

Pour permettre une comparaison visuelle, nous avons utilisé des représentations en boîtes à moustache des indicateurs calculés en mettant côte à côte, pour chaque indicateur, la distribution des moyennes sur 30 ans, calculées à partir des trois sources d'information pédologique et pour les 64 sites retenus (Figure 7.1). Cette présentation en boîtes à moustache permet de caractériser la variabilité globale des indicateurs associés aux services écosystémiques instruits. En moyenne, les services écosystémiques évalués à partir des données brutes fournies dans le RRP de Bretagne et les données désagrégées sont similaires sans aucune différence significative de moyennes mise en évidence par un test de Tukey au niveau de signification de 5%. Par exemple, la médiane du service de fourniture d'eau à plante est de 230 mm/an en mobilisant les données brutes du RRP de Bretagne et est estimé à 228 mm/an avec les données désagrégées. Néanmoins, les évaluations faites à partir des données désagrégées apparaissent plus variables entre les sites que celles faites à partir des données brutes du RRP. Cette variabilité est plus accentuée pour le service de fourniture d'eau à la plante (WP), l'énergie contenue dans la biomasse récoltée (YE) et la recharge de la nappe souterraine (GW). Dans le cas des services liés à l'azote notamment la fourniture d'azote à la plante (NP) et la régulation de la qualité de l'eau (WQ), les évaluations réalisées à partir des données brutes du RRP de Bretagne sont légèrement plus variables que les évaluations réalisées à partir des données désagrégées.

Par comparaison, les services écosystémiques évalués à partir des observations de terrain apparaissent en moyenne significativement différents pour les services de fourniture d'eau à la plante (WP), d'énergie de la biomasse récoltée (YE), de régulation de la qualité de l'eau (WQ) et

de recharge de la nappe souterraine (QW). Par exemple, les moyennes sur 30 ans du service de fourniture d'eau à la plante ainsi que la recharge de la nappe souterraine couvrent une large gamme de valeurs qui oscillent entre 120 mm/an et 295 mm/an. Quant aux valeurs médianes, on voit clairement que le service de fourniture d'eau à la plante, la production d'énergie (biomasse agricole) ainsi que la régulation de la qualité de l'eau présentent des médianes plus faibles dans le cas des données observées, que les évaluations effectuées à partir des deux autres sources. Cependant, le service de recharge de la nappe révèle une valeur médiane plus importante que pour les deux autres évaluations. En ce qui concerne le service de fourniture d'azote à la plante (NP) et la régulation du climat via la séquestration de carbone dans le sol (CS), les trois sources de données pédologiques fournissent des distributions similaires (la médiane varie entre 150 et 200 kg/ha/an), avec une variabilité plus importante qui caractérise le service de fourniture d'azote aux plantes dans le cas des données observées.

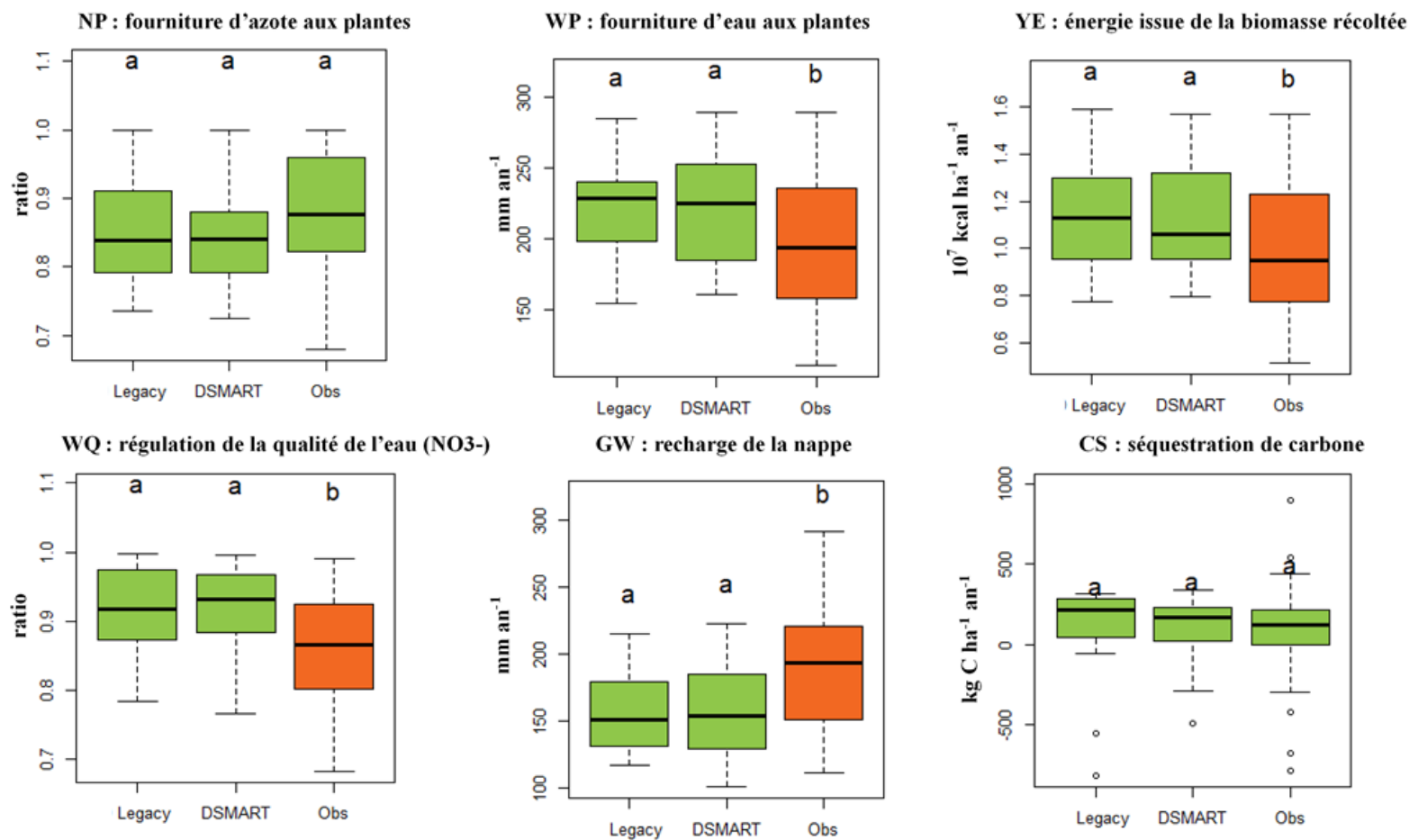


Figure 7.1 : Statistiques non paramétriques en 64 sites d'Ille-et-Vilaine des services écosystémiques des sols moyens sur 30 ans, en fonction de la source de l'information pédologique : (i) Legacy correspond à l'utilisation de l'UTS dominante du RRP ; (ii) DSMART aux données issues des cartes désagrégées avec le modèle DSMART ; (iii) Obs aux données observées et analysées en chacun des sites. Les lettres représentent la significativité du Test de Tukey, la même lettre indique des moyennes non significativement différentes au seuil de 0.05.

7.3.2. Comparaison des SES évalués à partir des données de terrain et à partir des données cartographiques

La figure 7.2 compare pour chaque service la valeur estimée en chaque point à partir des données du RRP ou à partir des cartes désagrégées, en fonction de la valeur estimée à partir des données de terrain. Cette comparaison montre que la distribution de nuages de points varie en fonction des services instruits. Par exemple, le service de régulation de la qualité de l'eau (WQ) présente un nuage de points très dispersé alors que le service de production de biomasse agricole (YE) présente un nuage de points concentré entre 0,5 et 1,5 10⁷ kcal/ ha/ an. De plus, les trois services associés à la fourniture d'eau à la plante (WP), la production de biomasse agricole (YE) et la régulation de la qualité de l'eau (WQ), sont généralement surestimés par les deux sources de données mobilisant le RRP de Bretagne. Le service de recharge de la nappe souterraine (GW) est quant à lui globalement sous-estimé avec ces deux mêmes sources de données pédologiques. Par contraste, le service de régulation du climat (CS) ainsi que celui de fourniture d'azote à la plante (NP) ne présentent pas de biais systématique de sur- ou sous-estimation.

Par ailleurs, la comparaison des coefficients de détermination (R^2), met en évidence une légère amélioration de la prédiction des services en utilisant les données désagrégées. Par exemple, pour le service de recharge de la nappe souterraine (GW), le R^2 est estimé à 0.09 avec les données brutes du RRP de Bretagne et atteint 0,33 avec les données désagrégées. Le même constat s'applique également aux autres services, mais avec des pourcentages d'amélioration moindres.

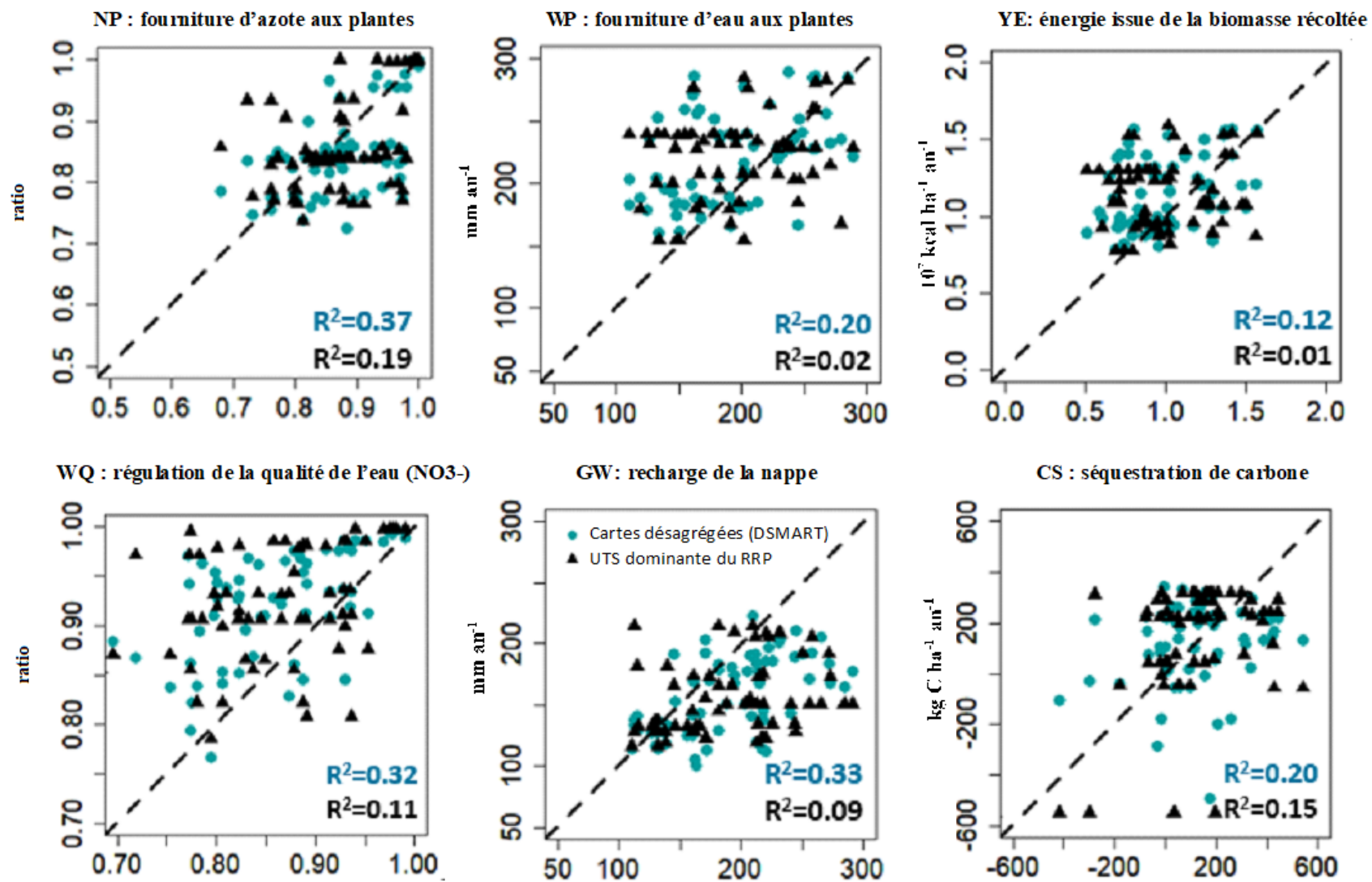


Figure 7.2 : Représentation en chacun des 64 sites de l'évaluation des services écosystémiques moyens sur 30 ans à partir des données cartographiques (RRP : triangle noir ; cartes désagrégées : cercle bleu) en fonction de celle issue des données d'observation en abscisse.

7.3.3. Analyse en composantes principales (ACP) des services écosystémiques

Une analyse en composantes principales normalisée (ACP) a été effectuée afin d'étudier les relations entre les services écosystémiques instruits sur les trois jeux de données (Figure 7.3), en considérant toujours les valeurs moyennes des indicateurs, calculées sur l'ensemble de la période d'étude (30 ans). Pour tester la significativité des différences entre les trois sources de données, nous avons opté pour une option d'habillage des individus (les 64 sites) selon les trois modalités décrivant la source des données (données d'observations, données désagrégées et données brutes du RRP de Bretagne). Par la suite, nous avons construit le barycentre de chaque modalité ainsi que son intervalle de confiance ($\alpha=5\%$). La sélection des axes de l'ACP a été basée sur deux critères empiriques à savoir le critère du coude associé à l'observation de l'éboullis des valeurs propres et le critère Kaiser-Guttman (Kaiser, 1991) qui préconise de ne retenir que les axes avec des valeurs propres supérieures à 1 dans le cas d'une ACP normalisée.

L'observation de l'éboullis des valeurs propres ainsi que l'application de la règle de Kaiser-Guttman conduisent à retenir les deux premières composantes principales qui résument plus de 87 % de l'inertie totale (Figure 7.3.a). Le premier axe factoriel explique 57% de la variabilité totale de l'information. Il discrimine nettement les quatre services associés à la production de biomasse agricole, à la régulation de la qualité de l'eau, à la fourniture d'azote à la plante et à la fourniture d'eau à la plante d'une part et au service de recharge de la nappe et à la régulation du climat d'autre part. Quant à la deuxième composante, celle-ci explique environ 30% de la variabilité totale et discrimine les services de production agricole, de fourniture d'eau à la plante et de régulation du climat des services de régulation d'eau, de fourniture d'azote à la plante et de recharge de la nappe.

Par ailleurs, le service de fourniture d'eau à la plante et celui de production de biomasse agricole sont fortement corrélés. De même, les services associés à l'azote notamment celui de régulation de la qualité de l'eau ainsi que celui de fourniture d'azote à la plante sont fortement corrélés et varient dans le même sens. D'un autre côté, le service de régulation du climat et celui de la fourniture d'azote à la plante sont strictement opposés et donc fortement corrélés négativement. L'analyse du nuage de points associé à la projection des individus (64 sites) sur le premier plan factoriel, met en évidence une dispersion importante des évaluations réalisées (Figure 7.3.b). Les points représentatifs des évaluations faites à partir des données désagrégées, ainsi que les données brutes du RRP de Bretagne, sont principalement situés dans la partie supérieure du plan factoriel.

Ils sont ainsi associés à des niveaux importants de service de production de biomasse agricole, de fourniture d'eau à plante et de régulation du climat via la séquestration de carbone. Les points représentatifs des évaluations faites à partir des données d'observation ne présentent pas une structure particulière et couvrent une large gamme de situations.

L'ajout du barycentre de chaque modalité décrivant la source des données, montre le chevauchement des cercles de confiance des barycentres des données désagrégées et les données brutes du RRP, alors que le cercle de confiance du barycentre des données d'observation se distingue nettement des autres. Cela confirme encore une fois la ressemblance forte entre les évaluations faites à partir des données désagrégées et les données brutes du RRP de Bretagne.

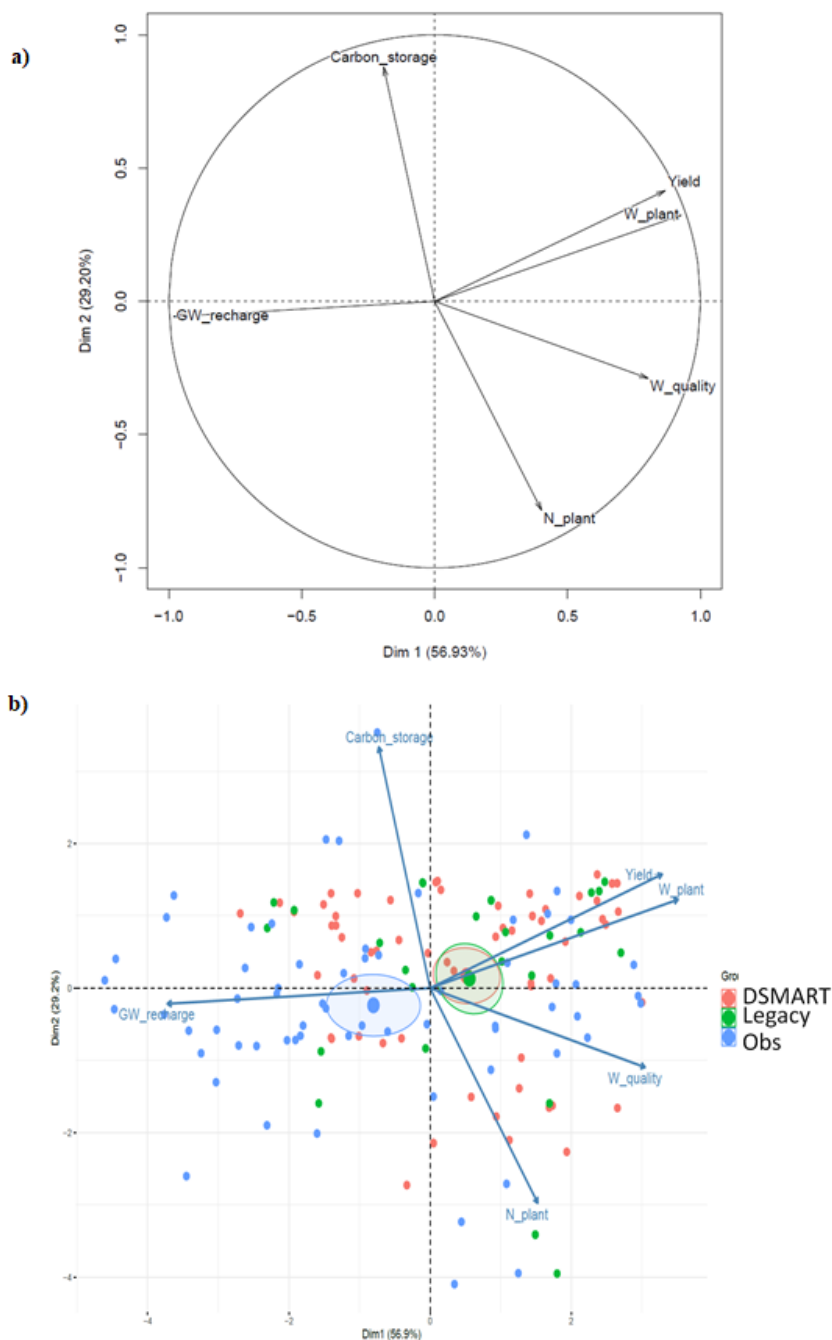


Figure 7.3 : Représentation du premier plan factoriel d'une analyse en composantes principales regroupant les individus des trois jeux de données mobilisés pour l'évaluation des services écosystémiques : a) projection des variables et b) projection des individus (64 sites x3) selon l'origine des données : Legacy correspond à l'utilisation de l'UTS dominante du RRP ; DSMART aux données issues des cartes désagrégées avec le modèle DSMART ; Obs aux données observées et analysées en chacun des sites.

7.4. Discussion et conclusion

L'utilisation d'une information sur les sols issue des cartes désagrégées conduit à une évaluation des services écosystémiques moyens sur 30 ans très similaire à celle obtenue en ne considérant que le type de sol dominant renseigné dans le RRP : aucune différence significative des moyennes n'a été mise en évidence (Figure 7.1) et les barycentres des deux sources de données sont superposées sur le plan factoriel de l'ACP regroupant les six services considérés (Figure 7.3b). Néanmoins, en comparant l'évaluation des services issues des données cartographiques avec celle obtenue par les données de terrain (Figure 7.2), on constate une légère amélioration de l'estimation des services lors de l'utilisation des données désagrégées.

Cette similitude est expliquée par plusieurs éléments notamment la qualité de la base de données d'origine, mais aussi l'approche de désagrégation mobilisée pour améliorer la résolution de l'information pédologique contenue dans le RRP de Bretagne, et finalement l'approche de modélisation mise en œuvre.

7.4.1. La qualité de la base de données Donesol du RRP de Bretagne

Bien qu'elle soit globalement très bien renseignée pour l'ensemble des propriétés de sol (Cf. chapitre 2), la base de données du RRP de Bretagne présente certaines limites. Ces limites expliquent une grande partie des similitudes constatées entre les évaluations biophysiques faites à partir des données désagrégées et les données brutes du RRP de Bretagne, mais aussi les différences constatées entre les évaluations réalisées à partir des données d'observations et celles mobilisant les données du RRP de Bretagne. Comme expliqué dans le chapitre 2 de la thèse (matériels et méthodes), les données pédologiques qui alimentent la base de données sémantique du RRP de Bretagne sont issues de la synthèse des analyses de profils réalisées pour un même type de sol, mais intègrent une expertise pédologique complémentaire. De plus, les profils pédologiques analysés n'ont pas été sélectionnés selon un échantillonnage probabiliste. Ils sont pour la plupart issus d'études antérieures au RRP (533 profils), et l'emplacement des 179 profils réalisés spécifiquement pour le RRP a été décidé par les experts-cartographes sur la base d'une stratégie d'échantillonnage déterministe, visant à sélectionner des profils supposés représentatifs des types de sol. De ce fait, la non-homogénéité des stratégies d'échantillonnage adoptées pour collecter les profils a pu biaiser les paramètres statistiques (mode, minimum, maximum)

renseignés dans la base de données pour chaque propriété de sol considérée, sans que ce biais ait pu être quantifié.

Un autre aspect est lié à la prise en compte pour caractériser les UTS, de profils pédologiques décrits et analysés au cours d'études pédologiques réalisées entre 1980 et 2012 (Lemercier et al., 2010). Cette étendue temporelle assez large peut là encore biaiser les estimations issues du RRP, notamment pour la teneur en carbone organique et le pH des horizons, ces deux propriétés étant sensibles à l'évolution des pratiques agricoles, notamment en matière de fertilisation, de rotations culturales, de travail du sol, mais aussi pour le pH en matière de chaulage des terres agricoles. Or, les études existantes fondées sur la compilation des analyses de terre effectuées par les agriculteurs (Walter et al, 1995 ; Lemercier et al., 2006 ; Saby et al., 2014) suggèrent une baisse des teneurs en matière organique des sols bretons depuis les années 1980, notamment dans les zones à teneurs initialement fortes, et au contraire une augmentation du pH des terres cultivées liée à des pratiques de chaulage. Il y a donc pu y avoir des évolutions temporelles de certaines propriétés des sols, qui là encore peuvent être la source de biais de l'information du RRP.

Par ailleurs, l'expertise pédologique incorporée dans la base de données relationnelle du RRP de Bretagne peut générer également des erreurs de caractérisation des UTS. Cette expertise a pour objectif principal de garantir une certaine cohérence entre les propriétés de sol renseignées et les processus pédologiques définissant une UTS donnée. Par exemple, elle permet de renseigner des teneurs en argile plus importantes dans les strates de profondeur pour caractériser un processus de lessivage d'argile pour des sols rattachés aux Néoluvisols ou aux Luvisols. Du fait qu'elle repose sur des règles expertes, mais non-normalisées, l'expertise pédologique peut générer des erreurs dans le renseignement des caractéristiques de certaines UTS, surtout en cas d'analyses manquantes. Elle impacte donc la fiabilité et la justesse des données contenues dans la base de données sémantique du RRP de Bretagne, sans que cet effet puisse être quantifié autrement que par la comparaison à des données indépendantes de validation, comme nous l'avons fait dans le chapitre 5.

7.4.2. Les limites de la méthode de désagrégation des cartes

La méthode de désagrégation que nous avons mise en oeuvre peut également générer des biais dans les prédictions de données pédologiques. Cela s'explique notamment par les règles de relation sol-paysage implémentées pour retranscrire les relations complexes entre l'organisation spatiale des sols et les caractéristiques du milieu. En effet, l'algorithme DSMART se base

principalement sur les proportions surfaciques des UTS qui composent les UCS. Or, ces proportions sont renseignées par expertise en fonction des fosses et des profils pédologiques réalisés, mais aussi en fonction de la connaissance *a priori* du milieu. Un autre aspect, lié aux règles d'apprentissage définies par le modèle pour prédire les UTS, explique également une partie des incertitudes associées aux données désagrégées. En effet, nous avons montré dans le chapitre 3 que la prise en compte de règles de distribution spatiale des sols permettait d'améliorer la représentation spatiale des types de sol dans les paysages. Ces règles n'obéissent pas à des lois physiques, mais se fondent plutôt sur des fréquences d'occurrence des différents types de sols en fonction des contextes géologiques, géomorphologiques et topographiques. Ces fréquences d'occurrence sont généralement établies à partir d'un nombre limité d'études et sont difficilement appréhendables de façon précise via quelques déterminants de l'organisation spatiale des sols à savoir le matériau parental, la position dans le versant et les conditions d'hydromorphie des sols. Dans le cas des cartes désagrégées, les propriétés de sol ont été prédites à partir des trois premières UTS prédites par désagrégation et des probabilités d'occurrence associées. En analysant les résultats, il apparaît la première UTS prédite par désagrégation, ayant la probabilité d'occurrence la plus élevée, correspond très souvent à l'UTS dominante de chaque UCS renseignée dans le RRP. Par conséquent, l'estimation d'une propriété du sol en un pixel est majoritairement influencée par les caractéristiques pédologiques de l'UTS dominante de chaque UCS. Dans l'ensemble, la forte similitude entre les données désagrégées et les données brutes du RRP est à l'origine des similitudes constatées entre les évaluations biophysiques des services écosystémiques faites à partir des données désagrégées et à partir de celles considérant l'UTS dominante de chaque UCS.

L'analyse du degré de linéarité entre les évaluations effectuées montre une amélioration marginale de corrélation en utilisant les données pédologiques désagrégées. Cela est notamment expliqué par l'utilisation des mêmes caractéristiques de l'UTS dominante de chaque UCS pour les points proches. En effet, vu les larges étendues spatiales des UCS, il y a une forte probabilité que des points proches les uns des autres se situent dans la même UCS et donc présentent les mêmes niveaux de fourniture de services. Cela explique également l'alignement horizontal des prédictions faites à partir des données brutes de RRP de Bretagne (Figure 7.2).

7.4.3. L'approche d'évaluation des services écosystémiques mise en œuvre.

La démarche retenue pour l'évaluation de services écosystémiques repose sur les sorties du modèle dynamique de culture STICS, simulant les flux hydriques, de carbone et d'azote dans le système sol-plante-atmosphère. Pour ce faire, STICS nécessite des données d'entrée qui caractérisent les conditions climatiques, les pratiques agricoles ainsi que les propriétés de sol à savoir la teneur en argile en surface, le rapport C/N, le pH et des propriétés hydrodynamiques des sols, mais aussi l'épaisseur et la teneur en éléments grossiers des différentes couches de sol. Certaines de ces propriétés, notamment les propriétés hydrodynamiques (teneur en eau à la capacité au champ, teneur en eau au point de flétrissement permanent) et la densité apparente des couches des sols, ont été estimées par le biais de fonctions de pédotransfert locales (Remy et al., 2014). Dans sa thèse, C. Le Bas (2016) montre que le facteur prépondérant dans l'impact des incertitudes liées aux données sols sur les sorties du modèle STICS, est lié au dimensionnement du réservoir utile en eau des sols. L'importance de l'estimation du réservoir utile est également apparue dans le chapitre 6 (§3.4), puisque cette grandeur est très corrélée aux variations des six services écosystémiques étudiés. Cela montre l'importance d'une validation approfondie des fonctions de pédotransfert utilisées, qui n'a pu être menée dans le cadre de cette thèse. Dans ce contexte, l'incertitude voire le degré de lissage des données pédologiques incorporées dans le modèle vis-à-vis des trois sources de données est difficilement quantifiable.

L'autre choix important tient dans la définition même des indicateurs de services construits à partir des sorties du modèle STICS. Dans l'ensemble, nous avons trouvé une tendance globale cohérente avec celle de l'étude nationale EFESE-EA (Tibi et Théron, 2018) notamment au regard des services de recharge de la nappe souterraine, de fourniture d'eau à la plante et de régulation de la qualité de l'eau. Cependant, ce dernier service reste fortement dépendant des scénarios de pratiques agricoles simulées. En outre, au regard des valeurs médianes, le service de fourniture d'eau à la plante et celui de production de biomasse agricole sont plus importants avec les données issues du RRP de Bretagne (données brutes et données désagrégées) qu'en considérant les données d'observées. Quant au service de recharge de la nappe souterraine, la médiane des évaluations faites à partir des données d'observations est au contraire plus importante que celle faite à partir des données brutes du RRP de Bretagne et des données désagrégées. Cela est notamment expliqué par deux effets majeurs, à savoir la surestimation quasi-systématique de la profondeur des sols par le RRP et également par les données désagrégées, et aussi par la sous-

estimation de la charge en éléments grossiers dans ces mêmes bases cartographiques. Ces deux sources de biais influencent fortement le réservoir utile en eau pour la plante et expliquent les écarts entre les services estimés à partir des observations de terrain et ceux estimés à partir des deux sources cartographiques (Figure 7.2).

Un autre constat très intéressant qui découle de cette étude est que les relations entre les services écosystémiques, identifiées sur l'ACP, sont conservées même après la constitution d'un jeu de données rassemblant les trois jeux de données. En effet, la comparaison des résultats de l'ACP faite sur les données d'observations uniquement et les résultats de l'ACP mobilisant simultanément les trois jeux de données, révèle des projections similaires des variables (services écosystémiques) sur le premier plan factoriel avec des inerties totales légèrement différentes (chapitre 6, Figure 6.9a). Cela montre clairement que les interrelations entre services écosystémiques ne dépendent que faiblement de la source d'information sur les sols mobilisée ; ce sont le mode de représentation des processus dans le modèle STICS et les données climatiques et agronomiques, qui structurent en premier lieu les corrélations entre services. La source d'information sur les sols module la valeur de chacun des services écosystémiques, mais influence marginalement leurs interactions.

Chapitre 8 :

Discussion et conclusion générale

Discussion et conclusion générale

Ce travail de thèse a porté sur le développement et la mise en œuvre d'une méthodologie d'amélioration de l'information pédologique contenue dans le RRP de Bretagne par des techniques de cartographie numérique des sols, en vue d'estimer les services écosystémiques des sols cultivés. Cette amélioration concerne la résolution spatiale, mais aussi changement de support de l'information selon la profondeur (strates dans le RRP versus couches standard *GlobalSoilMap*). La méthodologie élaborée permet d'une part d'appréhender la distribution spatiale des propriétés des sols à l'échelle régionale (chapitre 3 et 4), et d'autre part d'évaluer la qualité prédictive des cartes pédologiques produites (chapitre 5). Cette dernière constitue une étape incontournable pour caractériser le niveau d'incertitude qui peut être incorporé dans des outils de gestion agro-environnementale et plus précisément des outils d'aide à la décision. Les résultats de valorisation des données pédologiques issues du RRP de Bretagne ont été ultérieurement mobilisés dans des processus d'évaluation de plusieurs services écosystémiques jugés comme emblématiques dans le contexte agricole breton. Ce volet a permis d'analyser l'apport du concept de services écosystémiques dans la caractérisation de l'évolution dynamique des processus biophysiques des sols, mais aussi de déterminer les facteurs de contrôle de la fourniture des services écosystémiques instruits (chapitre 6). Un autre aspect novateur de ce travail de recherche est associé à l'évaluation de l'effet de la source de l'information pédologique sur l'évaluation des services écosystémiques (chapitre 7). Cela consiste à comparer les évaluations biophysiques des services écosystémiques faites à partir de plusieurs sources de données, qui se différencient notamment par leurs origines, mais aussi par leurs échelles de restitution cartographique. A notre connaissance, cette question n'a jamais été abordée dans la littérature scientifique, malgré son importance dans l'optique de promouvoir la prise en compte de l'information pédologique dans les processus d'évaluation des services écosystémiques des sols.

Tous ces éléments viennent renforcer la question de l'utilisation des cartes pédologiques et la pertinence potentielle de leurs usages. Ce travail de recherche a donc été l'occasion de mettre en avant deux points principaux qu'il s'agit de discuter :

- L'apport de la cartographie numérique et l'aptitude des bases de données actuelles au 250 000^{ème} pour la compréhension de l'organisation tridimensionnelle des sols, ainsi que pour la modélisation de leur fonctionnement ;
- L'évaluation biophysique des services écosystémiques des sols cultivés.

8.1. L'amélioration de la prédiction spatiale des types de sol et de leurs propriétés

8.1.1. La méthodologie de désagrégation spatiale

Pour améliorer l'information pédologique existante, la démarche choisie dans la thèse a été de s'appuyer sur les bases de données existantes pour caractériser ensuite plus finement la répartition spatiale des sols à l'échelle régionale de la Bretagne. Pour cela, nous avons opté pour une approche de désagrégation spatiale, prometteuse en matière de prédiction des types de sols sur des vastes territoires surtout quand les données pédologiques sont disponibles de manière exhaustive, mais à une échelle de restitution cartographique relativement grossière et donc insuffisante pour résoudre les problématiques agro-environnementales (Odgers et al., 2014).

Le sol est un objet tridimensionnel dont l'organisation spatiale varie en fonction des différents éléments du paysage (Jenny, 1941), à savoir la topographie, le matériau parental, l'occupation du sol, la structure du paysage, les pratiques agricoles et le climat. Le couplage de méthodes statistiques a permis en premier lieu la cartographie prédictive fine de la répartition spatiale des sols et en deuxième lieu la cartographie prédictive en trois dimensions de plusieurs propriétés du sol nécessaires à la caractérisation des sols, mais aussi à la modélisation dynamique des processus des sols.

En ce qui concerne la cartographie prédictive des types de sol (Unités Typologiques de Sol – UTS) à l'échelle régionale, nous avons retenu le département d'Ille-et-Vilaine pour réaliser et tester les développements méthodologiques. Pour l'essentiel, trois approches mobilisant l'algorithme DSMART ont été testées pour désagréger la carte de pédopaysages : DSMART classique, DSMART avec 755 profils pédologiques supplémentaires et DSMART avec des règles sur les lois de distribution des sols en contexte breton. Les résultats obtenus montrent la forte similitude, en terme de performance de prédiction des types de sols, des trois approches précédemment citées. Les prédictions des UTS sont aussi assorties de leurs probabilités d'occurrence en chaque pixel de 50 m de côté et des indices de confusion entre la première et la deuxième UTS prédites. Dans l'ensemble, les trois cartes pédologiques montrent clairement la prédominance de l'UTS dominante (ayant la proportion surfacique la plus importante dans chaque UCS). Néanmoins, la carte des types de sols issue de l'implémentation des règles décrivant les relations sol-paysage en contexte local, présente un niveau de détails plus important notamment pour les sols développés dans les alluvions, de part et d'autre des cours d'eau (Fluviosol). Le même constat est observé mais à un degré moindre avec l'ajout de 755 profils de sol supplémentaires dans la procédure de désagrégation. Plus encore, une désagrégation intra UCS

plus marquée caractérise les cartes pédologiques générées par les deux versions modifiées de l'algorithme DSMART, ce que l'on peut traduire par une augmentation de la représentation de la variabilité des sols.

La sélection de la meilleure approche de désagrégation de la carte des UCS de l'Ille-et-Vilaine est basée sur les résultats de validation des cartes des UTS prédites. Cette dernière consiste à mesurer la performance du processus de désagrégation en mesurant le degré de concordance entre des observations faites sur le terrain et leurs prédictions associées. Les approches de validation appliquées dans le cadre de la thèse sont de trois niveaux : validation pixel à pixel, validation par voisinage (3*3 pixels) et validation sémantique qui consiste à réduire le nombre d'UTS présentes en regroupant les UTS similaires par expertise. On observe une légère amélioration des résultats de validation via l'implémentation des règles *a priori* sur les lois de distribution des sols en contexte breton. Par exemple, au regard des trois cartes d'UTS les plus probables et l'approche de validation sémantique, la concordance entre les observations et les prédictions s'élève à 28,5 % pour DSMART avec règles sols-paysages contre 25,3 % pour DSMART classique et 21 % pour DSMART avec 755 profils de sol supplémentaires. Ces indicateurs de validation sont calculés à partir des 405 profils décrits et échantillonnés spécifiquement dans le cadre de la thèse. Nous avons retenu la méthode de désagrégation mobilisant l'algorithme DSMART et des règles sol-paysage, inhérentes au contexte local, pour désagréger le RRP de Bretagne à une résolution spatiale de 50 m (Vincent et al., 2018). En effet, en plus de fournir les meilleurs résultats de validation des cartes des trois UTS les plus probables, cette approche est celle qui montre le niveau de désagrégation des UCS le plus marqué (diversité des UTS prédites au sein des UCS) et elle restitue la distribution spatiale des types de sol autour du réseau hydrographique.

Bien que l'approche de désagrégation spatiale adoptée montre des avancées pour assurer la cohérence de l'organisation horizontale et verticale de la couverture pédologique, elle présente néanmoins certaines limites. Celles-ci sont notamment associées aux limites des règles d'apprentissage, mais aussi aux règles *a priori* sur les lois de distribution des sols spécifiques au contexte breton, implémentées dans l'algorithme DSMART par Vincent et al. (2018).

En ce qui concerne la méthode d'apprentissage implémentée par défaut dans DSMART, l'algorithme C4.5 d'arbre de régression est purement statistique et dépourvu d'expertise pédologique. Les méthodes d'apprentissage constituent le fondement des modèles empiriques permettant de relier les types de sol à leurs facteurs de variation préalablement définis. La majeure partie du jeu de calibration du modèle est issue d'un échantillonnage virtuel et aléatoire au sein des UCS. Cette stratégie d'échantillonnage permet de tirer un nombre fixe et important de points virtuels (30 par polygone), supposés suffisants pour représenter la variabilité de l'ensemble de l'UCS et ainsi permettre de cartographier l'hétérogénéité présente sur le territoire d'étude. Elle

ne tient donc pas compte des étendues spatiales des UCS, du fait que l'échantillonnage ne soit pas proportionnel à leurs superficies.

Dans le cas de la procédure DSMART classique, la méthode d'allocation aléatoire pondérée introduit également un certain biais dans le processus de désagrégation spatiale. Cette méthode d'allocation induit que chaque UTS au sein d'une UCS a été pondérée avec sa proportion supposée d'occurrence dans l'UCS en question. L'inconvénient majeur est que ce mode d'allocation ne tient pas compte des relations sol paysage. Une UTS donnée possède donc la même probabilité d'être prédite à n'importe quel point de l'UCS, même à des endroits incohérents d'un point de vue pédologique. Par exemple, un Colluviosol (sol caractéristique des milieux et bas pentes) pourra être alloué à un échantillon virtuel sur une zone complètement plane ou en haut de pente. Plus encore, l'exécution de l'algorithme DSMART repose sur l'hypothèse très forte que les pourcentages d'UTS au sein des UCS sont correctement estimés dans la base de données au 250 000^{ème}. Autant ce pourcentage peut être considéré comme valide lorsque le pédologue cartographe dispose d'une densité d'informations importante (par exemple lorsque les UCS délimitées au 250 000^{ème} correspondent à une synthèse de cartes plus détaillées à grande échelle), autant il est de plus en plus empreint de subjectivité lorsque la cartographie a été établie d'une manière descendante et globale (c'est-à-dire en délimitant de grands ensembles physiographiques puis en les caractérisant avec une faible densité de sondages). De fait, les erreurs de prédiction générées par l'algorithme DSMART trouvent en partie leur origine dans la qualité de la base de données de départ et notamment par la justesse avec laquelle les proportions surfaciques des UTS au sein des UCS ont été renseignées.

Pour pallier les limites inhérentes à l'algorithme DSMART, il est intéressant d'utiliser une allocation aléatoire stratifiée. Cette approche permet de guider l'allocation des UTS aux échantillons virtuels générés par DSMART, en prenant en compte la topographie, le matériau parental et la végétation, principaux facteurs de la distribution des sols dans un contexte climatique donné (Hearing et al., 2012). Holmes et al. (2015) et Nauman et Thompson (2014) ont supposé que l'ajout de règles de distribution des sols pourrait substantiellement améliorer la qualité des prédictions. Le chapitre 3 de la thèse a notamment confirmé que l'implémentation des règles *a priori* sur les lois de distribution des sols dans un paysage breton permet d'obtenir une carte des sols plus cohérente avec les caractéristiques du milieu.

La prise en compte des règles sol-paysage dans le processus de désagrégation du RRP de Bretagne a permis d'améliorer de façon substantielle la représentation de la répartition spatiale des sols à l'échelle régionale. Pour l'essentiel, ces règles reposent sur l'expertise du pédologue cartographe et sur sa connaissance *a priori* du milieu. Pour cela, Vincent et al. (2018) ont développé des règles dites « expertes » mobilisant quelques facteurs déterminants de l'organisation spatiale des sols en

Bretagne à savoir le matériau parental, la position topographique et les conditions d'hydromorphie des sols. Ces règles sont extrapolées d'une manière systématique à l'échelle de la Bretagne entière sans tenir compte de la complexité pédologique locale qui se développe dans des conditions pédogénétiques rares et peu fréquentes à l'échelle régionale. De plus, les règles implémentées sont strictes, homogènes et représentatives des lois les plus fréquentes sur l'ensemble du territoire d'étude. Or, les lois d'organisation des sols dans le paysage ne sont pas inscrites dans des lois physiques incontestables. Elles ne sont donc pas des lois universelles mais se fondent plutôt sur des fréquences d'occurrence des différents types de sols en fonction des contextes géologiques, géomorphologiques et topographiques. Ces fréquences d'occurrence sont généralement établies à partir d'un nombre limité d'études et sont difficilement appréhendables de façon précise via quelques déterminants de l'organisation spatiale des sols à savoir ceux établis par Vincent et al. (2018).

8.1.2. Perspectives d'amélioration de l'approche de désagrégation spatiale

Plusieurs pistes peuvent être investiguées pour améliorer la procédure de désagrégation spatiale du RRP de Bretagne. Pour l'essentiel, quatre voies majeures sont proposées :

- Dans un premier temps, il s'agirait d'améliorer le jeu de covariables SCORPAN mobilisé. En effet, des covariables environnementales supplémentaires devraient être implémentées dans le modèle afin de mieux caractériser la variabilité pédologique au sein de la zone d'étude. Pour cela, il existe des méthodes d'optimisation du choix des covariables environnementales selon des critères empiriques et statistiques. Ces méthodes permettent d'identifier un jeu de covariables induisant les meilleures prédictions pour une zone d'étude donnée. Elles permettent aussi d'éliminer les covariables exprimant des informations redondantes dans le modèle ou nuisant à la qualité des inférences spatiales. Concernant les nouvelles covariables environnementales, il pourrait s'agir de données décrivant plus finement les variabilités spatiales et interannuelles de l'occupation des sols. En effet, une représentation plus réaliste de l'occupation des sols et des rotations des cultures, même en considérant des échelles temporelles moyennes, permettrait de retranscrire la complexité paysagère. Plus encore, l'intégration des éléments linéaires du paysage (haies, talus, fossés, bord de champs) comme des covariables environnementales permet dans le contexte armoricain d'améliorer les prédictions des types de sols et de leurs propriétés (Follain et al., 2007 ; Chaplot et al., 2004 ; Lacoste et al., 2014). Par exemple, la prise en compte de la position des haies permet de distinguer les sols développés de part et d'autre de ces structures linéaires en fonction de leur position en

amont ou en aval et de les différencier par un épaissement des horizons organo-minéraux en amont des haies en fonction de la distance à celles-ci. Il existe des inventaires des réseaux bocagers réalisés par photo-interprétation et contrôle de terrain, mais actuellement ces inventaires ne sont pas exhaustifs à l'échelle régionale.

- Dans un second temps, il s'agirait d'améliorer les règles décrivant *a priori* les lois de distribution des sols implémentées dans le processus de désagrégation. Il s'agirait de collecter un jeu de données supplémentaires et non biaisées couvrant l'ensemble de l'hétérogénéité pédologique, géomorphologique, paysagère et environnementale de l'ensemble de la zone d'étude, puis d'utiliser des méthodes d'intelligence artificielle pour développer des lois décrivant plus précisément les relations sol-paysage. Ces méthodes innovantes permettraient d'explorer voire de déterminer une grande partie des lois à l'origine de l'organisation spatiale des sols dans un contexte donné.
- Dans un troisième temps, simplifier l'information à désagréger par l'algorithme DSMART pourrait être aussi une solution envisageable pour améliorer les prédictions finales. En effet, il est possible de regrouper les UTS similaires d'un point de vue des processus pédologiques et/ou selon des critères d'aptitude agronomique en utilisant les typologies agronomiques des sols comme par exemple la typologie Typterres développée dans le cadre du RMT Sols et Territoires (www.sols-et-territoire.org). Cela simplifierait considérablement la mise en œuvre du processus de désagrégation et fournirait aussi des cartes opérationnelles et facilement appréhendables par les utilisateurs, qui ne sont forcément pas pédologues.
- Enfin, complexifier les règles, en les adaptant en fonction de la qualité de la représentation exhaustive du matériau parental ou de l'hydromorphie qui permettent d'allouer l'UTS à un point donné. Par exemple, si on sait que la prédiction du matériau parental n'est pas correcte dans une certaine zone, il serait judicieux de ne pas utiliser cette covariable dans la procédure.

8.1.3. La cartographie prédictive des propriétés des sols à l'échelle régionale

Pour réaliser une cartographie prédictive des propriétés des sols à l'échelle régionale, nous avons combiné les cartes de distribution spatiale des trois UTS les plus probables, leurs probabilités d'occurrence (chapitre 3) et aussi les caractéristiques des UTS exprimées en six intervalles de profondeur de sol. Cette dernière approche est assurée par l'ajustement de fonctions splines aux données strates des UTS. L'utilisation des fonctions de conservation de masse (splines) pour harmoniser l'information pédologique est recommandée par de nombreuses études scientifiques

(Malone et al., 2011 ; Odgers et al., 2012) et notamment par le consortium *GlobalSoilMap* dans son guide méthodologique (Arrouays et al., 2014). Dans le chapitre 4 de la thèse, nous avons pu montrer la performance des fonctions splines dans la conservation des interrelations entre les propriétés des sols. Ces interrelations sont représentatives de certains processus pédologiques spécifiques tels que le lessivage d'argile en profondeur. Les fonctions splines assurent ainsi une certaine cohérence de l'organisation verticale de la couverture pédologique via la prédiction simultanée de plusieurs propriétés des sols. Pour atténuer au maximum les effets de bord (sous-estimation ou surestimation des variables du sol sur les strates définissant les bornes des UTS), la première et la dernière strate de chaque UTS ont été systématiquement subdivisées toutes les deux en deux strates virtuelles, tel que proposé par Odgers et al. (2012) et Vaysse (2015). Ce remodelage de l'information pédologique permet de générer des prédictions respectant les spécifications du consortium *GlobalSoilMap* et constituent une étape indispensable pour appréhender la distribution spatiale des sols à l'échelle régionale.

Par ailleurs, les résultats issus de l'ajustement des fonctions splines ainsi que les prédictions spatialisées des types de sols ont été utilisés pour estimer, pour chaque propriété, une moyenne pondérée par les probabilités d'occurrence des trois UTS les plus probables. Cela a permis de caractériser explicitement les distributions spatiales des propriétés des sols pour chaque pixel de 50 m de côté. Dans l'ensemble, la méthodologie implémentée représente une première tentative pour entreprendre une cartographie numérique des propriétés des sols sur l'ensemble du territoire breton de façon systématique. Les cartes des propriétés des sols obtenues sont cohérentes avec les caractéristiques géo-environnementales de la région Bretagne et surtout cohérentes entre elles. Bien que les prédictions des propriétés des sols constituent un support pertinent pour répondre à la demande croissante en données spatialisées sur les sols et notamment renforcer la prise en compte de l'information pédologique dans le développement territorial, elles présentent néanmoins quelques verrous méthodologiques. En effet, nous avons essayé de réduire au maximum le biais induit par l'ajustement des fonctions splines, mais cela n'a pas été quantifié d'une manière précise. Une analyse de propagation des incertitudes liées à l'ajustement des fonctions splines est nécessaire pour identifier le niveau d'incertitude propre aux fonctions ajustées. L'apport de la création de deux strates virtuelles n'a été quantifié que d'une façon visuelle sans réelle analyse quantitative.

Par ailleurs, dans les chapitres 6 et 7, afin de modéliser le fonctionnement dynamique des sols cultivés, nous avons appliqué trois fonctions de pédotransfert pour prédire la teneur en eau à la capacité au champ, la teneur en eau au point de flétrissement permanent des plantes et la densité apparente des horizons de sol. Bien que ces fonctions aient été construites à partir des données acquises sur des échantillons de sol représentatifs de la variabilité pédologique de la Bretagne

entière, celles-ci n'ont pas encore été publiées dans la littérature internationale. Ainsi, le degré de lissage de l'information pédologique mobilisées par ces fonctions ou encore leurs performances n'ont pas été quantifiées dans cette étude.

Perspectives d'amélioration de la cartographie numérique des propriétés des sols

Dans le cadre de la thèse, la distribution spatiale des propriétés des sols a été appréhendée à partir des données surfaciques des UTS et des strates renseignées dans la base de données Donesol 3 associée au RRP de Bretagne. Pour ce faire, la carte des pédopaysages de Bretagne a tout d'abord été désagrégée, et des cartes de propriétés des sols en ont ensuite été dérivées. Cependant, il existe d'autres approches de cartographie numérique des sols qui pourraient aisément être appliquées dans notre contexte. Ces méthodes permettent d'aboutir à une cartographie prédictive des propriétés des sols sans avoir la nécessité de désagréger la carte des UCS pour délimiter spatialement les UTS. Ces approches ont été par exemple appliquées dans la région de Languedoc-Roussillon par Vaysse et al. (2014) ou à l'échelle de la France entière par Chen (2019). Ces auteurs se sont appuyés sur des modèles d'apprentissage comme des forêts d'arbres décisionnels, de régression-krigeage et sur des simulations stochastiques pour caractériser la distribution des propriétés des sols en utilisant les profils pédologiques répertoriés dans les bases de données Donesol ou décrits dans le Réseau de Mesure de la Qualité des Sols. Les prédictions ont été assorties de leurs incertitudes déterminées par l'algorithme de forêts de régressions quantiles (Quantile Regression Forest). Plus encore, Angelini et al. (2017) ont utilisé une approche novatrice dans le champ de la cartographie numérique des sols basée sur le Structural Equation Modelling (Grace, 2006). La mise en œuvre de ce modèle permet de prédire de nombreuses propriétés des sols simultanément, assurant ainsi la cohérence de l'organisation verticale de la couverture pédologique. Toutefois, comme pour la désagrégation spatiale, même les approches alternatives mobilisant des données ponctuelles n'assurent pas une prédiction dynamique des propriétés des sols.

8.1.4. La validation des cartes désagrégées

Afin de caractériser la qualité prédictive des cartes pédologiques, nous avons développé une méthodologie originale qui s'appuie sur la validation des cartes de propriétés des sols, mais aussi des critères cartographiques à savoir le matériau parental, la classe d'hydromorphie, le type de développement du solum et la classe de profondeur. Pour ce faire, nous avons collecté un jeu de données indépendant, sans biais, selon un échantillonnage aléatoire stratifié selon le matériau parental dominant des UCS, à l'échelle du département d'Ille-et-Vilaine. La démarche générale

comprend deux volets principaux. Le premier consiste à confronter les prédictions quantitatives aux données mesurées, à partir des analyses physico-chimiques réalisées sur les échantillons de sols collectés, en calculant des paramètres statistiques à savoir des coefficients de détermination, des erreurs moyennes et des erreurs quadratiques moyennes. Le deuxième volet consiste à calculer des indices de pureté partielle vis-à-vis des critères cartographiques précédemment cités, des indices de puretés moyennes et des indices de puretés strictes à différentes échelles spatiales : le transect linéaire, la strate et le département d'Ille-et-Vilaine entier.

Les résultats de validation mettent en évidence une variabilité très importante des indices de pureté qui se décline en trois niveaux :

- Une variabilité inter-critères cartographiques. En effet, la pureté partielle du matériau parental atteint des valeurs de pureté très importantes pour l'ensemble des strates identifiées. Elle est maximale (100%) pour la strate des alluvions continentales et la strate des terrasses alluviales et est estimée à 63 % pour la strate de schistes gréseux. Le type développement de profil présente des indices de pureté nettement plus faibles que ceux du matériau parental. Par exemple, il est estimé à 17 % pour la strate des terrasses alluviales et à 33% pour la strate des alluvions.
- Une variabilité inter-cartes désagrégées. Pour l'ensemble des indices de pureté, la carte de la première UTS prédite se distingue par les valeurs les plus élevées, suivie par la carte de la deuxième et de la troisième UTS prédites. Par exemple, au regard de la pureté stricte, la première UTS atteint 34% à l'échelle départementale, contre 12% pour la deuxième UTS prédite et uniquement 5% pour la troisième UTS prédite. La même tendance caractérise aussi les valeurs de pureté moyenne mais avec des valeurs nettement plus élevées. Ces résultats montrent que la probabilité d'occurrence estimée par la méthode de désagrégation est un bon indicateur de la pureté des types de sol telle qu'estimée à partir d'une campagne de validation indépendante.
- Une variabilité inter-échelles. Les indices de puretés calculés varient aussi en fonction du support de validation considéré, à savoir celui d'un transect disposé sur un versant, celui des grands ensembles de matériau parental ou encore de la zone d'étude entière. Bien que la qualité de prédiction soit indépendante de la position dans le versant (bas, milieu ou haut de versant), on observe une amélioration des indices de pureté en considérant une échelle plus importante. Ce constat est également confirmé dans des travaux antérieurs de Bishop et al. (2015) dans un contexte australien.

Quant aux propriétés analytiques des sols, on obtient des performances de prédiction du même ordre que celles décrites par Vaysse et al., 2014 en région Languedoc-Roussillon. Par rapport à des travaux internationaux, nos performances de validation sont relativement meilleures que

celles obtenues en Australie par Odgers et al. (2015) qui ont suivi une démarche analogue de désagrégation des types de sol suivie d'une prédiction spatiale des propriétés des sols.

Perspectives d'amélioration de la validation des cartes pédologiques

Dans la littérature scientifique, la question de la validation des cartes pédologiques est souvent abordée sous l'angle du type de sol et des propriétés analytiques des sols. Un nombre limité d'études a porté sur la validation des critères cartographiques (Walter, 1990) comme cela a été entrepris dans le cadre de cette thèse. Cela est justifié par le coût élevé mais aussi la lourdeur de la collecte des données supplémentaires. Généralement, ce sont des approches de validation interne (Minasny et al., 2013), mobilisant une grande partie du jeu de données ayant déjà servi pour calibrer de modèle, qui sont appliquées pour appréhender la qualité de prédiction. D'un point de vue statistique, un biais est introduit et difficilement quantifiable, dû à la double utilisation des données d'origine pour calibrer et valider les modèles.

Pour éviter ces risques, Brus et al. (2011) recommandent d'utiliser un jeu de données de validation indépendant et surtout probabiliste. Toutefois, la densité d'échantillonnage reste à déterminer en fonction de la variabilité pédologique des sites d'étude et des ressources financières et de l'expertise pédologique disponibles.

Bien que la stratégie de validation mise œuvre dans les travaux de la thèse soit originale et permette de lever les verrous majeurs existants, il existe néanmoins quelques pistes qui permettraient d'améliorer l'interprétation voire l'opérationnalité des cartes désagrégées. Le défi majeur concernant la validation des prédictions est la prise en compte de la distance taxonomique entre les types de sol (Girard, 1983 ; Carre et Jacobson, 2009). Même s'il existe 171 UTS distinctes en Ille-et-Vilaine au sein du RRP au 250 000^{ème}, il est possible que des UTS différentes soient proches sur le plan taxonomique et se retrouvent donc dans des conditions environnementales similaires. Cela complique la détermination des probabilités de prédictions par de simples règles d'apprentissage et par conséquent, biaise les résultats de validation. La prise en compte de la distance taxonomique entre les différentes UTS permettrait non seulement d'améliorer les résultats de validation, notamment pour les puretés strictes des UTS probables, mais aussi de réduire l'indice de confusion élevé à certains endroits de la zone d'étude (Figure 3.5). D'un point de vue fonctionnel, certaines confusions entre UTS peuvent avoir des conséquences importantes, d'autres être beaucoup moins préjudiciables. Par exemple, pour de nombreuses applications agro-environnementales, la confusion d'un Néoluvisol avec un Rankosol sera plus sévère que sa confusion avec un Brunisol. La notion de distance taxonomique permettrait de tenir compte de la sévérité de la confusion, et ainsi de relativiser les erreurs de prédiction. Cela permettrait

également de cibler des secteurs ou des contextes pédologiques prioritaires pour densifier les observations et ainsi réduire les erreurs de prédiction.

8.1.5. Les incertitudes associées aux prédictions spatialisées sur les sols et la qualité de la base de données du RRP de Bretagne

Pour respecter les spécifications du consortium *GlobalSoilMap*, les prédictions pédologiques doivent être assorties de leurs incertitudes, et plus précisément de leurs intervalles de confiance. Dans ce travail de recherche, les prédictions pédologiques ont été dérivées à partir de la base de données Donesol 3 du RRP de Bretagne en mobilisant les valeurs modales des strates constituant les UTS. Vu l'incomplétude de la base de données vis-à-vis des bornes de distribution associées aux valeurs modales des propriétés des sols, les intervalles de confiance des prédictions n'ont pas été déterminés. L'action prioritaire pour remédier à cette situation est de renseigner les valeurs manquantes des bornes des propriétés des sols. Cela pourrait être réalisé par expertise pédologique et/ou par l'utilisation de modèles statistiques. Ces derniers mobilisent des règles d'apprentissage permettant de prédire les valeurs minimales et maximales des propriétés des sols en fonction des valeurs modales et d'une combinaison de variables bien renseignées voire optimisées dans la base de données Donesol.

8.2. L'évaluation biophysique des services écosystémiques des sols cultivés

Les chapitres 6 et 7 ont porté sur l'évaluation biophysique de six services écosystémiques des sols (deux services de régulation et quatre services d'approvisionnement) en s'appuyant sur la modélisation dynamique du fonctionnement des sols. Au total, 64 sites localisés dans des parcelles cultivées ont été sélectionnés pour évaluer les services écosystémiques ciblés en utilisant le modèle de culture STICS. Pour alimenter ce modèle, nous avons utilisé les données météorologiques des 30 dernières années, un scénario de pratiques agricoles homogènes pour l'ensemble des simulations et finalement les caractéristiques des sols décrivant les 64 solums choisis. En fonction des variables de sorties du modèle STICS, un indicateur par service a été calculé afin de caractériser leur variabilité interannuelle, mais aussi leur variabilité moyenne sur 30 ans, en chacun de ces 64 sites. Par la suite, nous avons analysé les interrelations entre les services étudiés et nous avons pu identifier les déterminants pédologiques les plus en lien avec la variabilité de fourniture des services écosystémiques étudiés.

Les résultats mettent en avant une grande variabilité des services rendus, en lien avec la variabilité pédologique de la zone d'étude. Par exemple, les sols profonds ayant une texture équilibrée enregistrent des niveaux importants pour tous les services instruits, sauf pour celui de recharge de la nappe souterraine. Au contraire, les sols peu profonds ayant une texture grossière enregistrent des niveaux de fourniture relativement faibles des services écosystémiques sauf pour le service de recharge de la nappe souterraine qui se distingue par des niveaux bien élevés. D'un autre côté, l'analyse de la variabilité interannuelle de l'ensemble des services étudiés pour deux types de sol assez contrastés, montre encore une fois une variabilité importante des services et qui s'avère plus importante pour les cultures de printemps que pour les cultures d'hiver.

L'analyse des corrélations statistiques entre les différents services permet de caractériser les interrelations induites. Dans l'ensemble, les services écosystémiques étudiés sont fortement corrélés entre eux à l'exception des services de régulation du climat et de fourniture de l'azote à la plante qui présentent de faibles corrélations avec le reste des autres services. En outre, la caractérisation des interrelations entre les services écosystémiques et quelques propriétés intrinsèques et extrinsèques des sols, permet d'identifier les facteurs de contrôle des services instruits. Il s'avère que la profondeur, les variables texturales et le réservoir utile en eau des sols sont fortement liés aux services écosystémiques associés à la dynamique de l'eau dans le système sol-plante-climat. En parallèle, les propriétés physico-chimiques du sol, notamment le pH, la teneur en carbone organique et la teneur en argile, sont fortement corrélées avec le service de régulation du climat et à un degré moindre, avec la fourniture d'azote à la plante.

Le groupe d'experts de l'EFESE-EA (2017) a identifié comme piste pertinente, pour améliorer les évaluations des services écosystémiques des sols, l'utilisation de données pédologiques plus précises, notamment les données contenues dans les bases de données relationnelles des RRP régionaux qui couvriront très prochainement l'intégralité du territoire français. Pour vérifier la pertinence de cette piste et évaluer l'effet de la qualité de l'information pédologique mobilisée sur l'estimation des services écosystémiques des sols, nous avons comparé les évaluations faites sur les 64 sites à partir de trois sources d'information pédologique : i) les caractéristiques de l'UTS dominante de chaque UCS du RRP; ii) les données désagrégées du RRP ; iii) les données issues des observations sur le terrain. Globalement, il n'y a pas d'effet lié à la source de l'information pédologique pour les services de fourniture d'azote à la plante et de régulation du climat via la séquestration de carbone dans les sols. Les autres services affichent des distributions similaires qu'ils soient évalués à partir des données de l'UTS dominante de l'UCS du RRP ou des données désagrégées, mais se différencient significativement des évaluations réalisées à partir des données collectées au champ. La comparaison des évaluations faites à partir des données du RRP de Bretagne ou à partir des cartes désagrégées avec celles issues des données de terrain, met

en évidence une légère, mais assez variable, amélioration en fonction du service étudié via l'utilisation des données désagrégées pour tous les services ciblés.

8.2.1. Avantages et limites d'une approche d'évaluation des SES par modélisation dynamique

Pour intégrer à la fois la variabilité des pratiques agricoles et celles des caractéristiques du milieu, l'utilisation d'un modèle dynamique tel que le modèle de culture STICS présente des avantages en matière de caractérisation du fonctionnement des compartiments sol-plante-climat pour simuler l'évolution dynamique de plusieurs processus des sols, au prix néanmoins d'une configuration difficile des paramètres d'entrée du modèle. Le modèle STICS présente l'avantage d'une organisation modulaire permettant de simuler de nombreux systèmes de cultures selon différents contextes de gestion, mais aussi d'intégrer une large variabilité pédologique et climatique. Il a ainsi été sélectionné dans le cadre de l'expertise EFESE-EA pour évaluer les services écosystémiques des sols cultivés en grande culture à l'échelle de la France entière (Sbiti et Therond, 2018). Cependant, comme toute approche retranscrivant les relations complexes des interactions sol-plante-climat, le modèle STICS comporte certaines limites. Celles-ci sont notamment liées aux besoins importants en données d'entrée et au temps nécessaire pour la préparation des fichiers d'entrée, mais aussi en temps d'apprentissage du modélisateur en fonction de son expérience. Une phase d'optimisation des paramètres STICS est nécessaire pour ajuster au mieux certains d'entre eux. Cette phase consiste à fixer quelques paramètres mesurés directement sur le terrain et à faire ensuite converger le modèle pour optimiser d'autres paramètres préalablement définis par l'utilisateur selon des critères empiriques. De plus, le modèle de culture STICS est un modèle unidimensionnel (1-D), qui s'intéresse aux processus de transfert verticaux et ignore de ce fait les écoulements latéraux. Cet aspect peut certainement introduire un biais conséquent dans le calcul des flux d'entrée-sortie du modèle et donc du bilan environnemental. De plus, STICS nécessite des données ponctuelles, qui sont supposées être homogènes à l'échelle parcellaire. Cela peut limiter la généralité d'application de ce modèle pour définir des stratégies de gestion agro-environnementale sur de vastes territoires du fait de la complexité de la modélisation à mettre en oeuvre.

8.2.2. Vers une approche hybride d'évaluation des services écosystémiques des sols

Pour pallier les limites de la modélisation dynamique des processus et développer une approche pertinente permettant d'évaluer plus facilement les services écosystémiques, il apparaît intéressant de s'appuyer sur une approche hybride couplant la modélisation et l'expertise. Cela

se traduit par le développement d'indicateurs simples, mais pertinents vis-à-vis de leurs aptitudes à appréhender les services écosystémiques des sols. Cette approche repose notamment sur la modélisation mécaniste lorsque les processus biophysiques conditionnant la fourniture des services écosystémiques sont bien maîtrisés par les modèles existants et le recours à une typologie experte permettant d'identifier les proxys pertinents pour appréhender les services dont on ne maîtrise pas encore la modélisation de leurs processus, à l'instar de la fourniture d'habitats pour la biodiversité, la régulation des gaz à effet de serre, etc).

Pour la mise en place d'une telle approche hybride, trois voies majeures complémentaires sont possibles (Figure 8.1) :

- La modélisation d'un ensemble de scénarios potentiels sur un territoire donné. Cela nécessite de faire un inventaire de l'ensemble des systèmes de cultures et de la variabilité pédologique et climatique d'un territoire. Chacune des combinaisons identifiées fait l'objet d'une modélisation dynamique qui permet de quantifier différents services écosystémiques jugés d'intérêt au sein de ce territoire. Par la suite, les services écosystémiques peuvent être déduits par une simple identification de la situation actuelle et l'application du principe d'homologie. Dans l'ensemble, pour cette voie, trois phases essentielles se distinguent : 1) une phase d'apprentissage par la modélisation des différentes combinaisons systèmes de culture x climat x sol existantes sur un territoire ou devant être évaluées ; 2) une phase de validation des résultats de modélisation pour caractériser l'aptitude du modèle à retranscrire les relations complexes entre les composantes du système sol-plante-climat et donc la fiabilité des évaluations biophysiques et finalement 3) une phase d'application dans une nouvelle situation, par la reconnaissance de la combinaison la plus proche qui sert de référence pour l'estimation des services écosystémiques. Cette approche est proposée par Fossey et al. (soumis) pour évaluer les services écosystémiques des sols cultivés au sein de la métropole de Nantes - Saint-Nazaire, en considérant les systèmes de culture actuels, mais également des systèmes de culture dont l'adoption est envisagée et dont on cherche à évaluer l'intérêt.
- La modélisation mécaniste des processus des sols en intégrant la variabilité pédologique et climatique couvrant celle du territoire d'étude. Les services écosystémiques sont en premier lieu évalués par modélisation dynamique pour ensuite être mis en relation avec les variables pédo-environnementales. Cette étape permet d'identifier les principaux facteurs de contrôle des services modélisés et par conséquent d'identifier les proxys conditionnant la fourniture des services écosystémiques. Dans notre travail, les résultats du chapitre 6 de la thèse mettent en évidence des corrélations très importantes entre les services écosystémiques instruits et certaines propriétés pédologiques notamment le

réservoir utile en eau, les variables texturales, la profondeur du sol et la teneur en carbone organique, avec des coefficients de corrélations supérieurs à 70% pour la plupart des services étudiés. Ces fortes corrélations suggèrent que ces propriétés du sol pourraient être mobilisées dans la discrimination des sols selon leurs aptitudes à fournir des services écosystémiques. L'utilisation de proxies, préalablement sélectionnés par une approche de modélisation, permettrait de caractériser la multifonctionnalité des sols. En effet, plus le nombre de fonctions remplies par un sol donné est important, plus son indice de polyvalence d'usage est important et par conséquent plus son aspect multifonctionnel est intéressant.

- La mobilisation d'une expertise transdisciplinaire pour identifier les proxies pertinents pour l'évaluation des services écosystémiques. L'idée générale consiste à confier à des experts de différents domaines tels que l'analyse environnementale, l'agronomie et la pédologie pour identifier une liste non exhaustive de proxies à mobiliser dans l'évaluation des services écosystémiques. La mutualisation des apports des proxies permettrait une analyse multidimensionnelle des services ciblés.

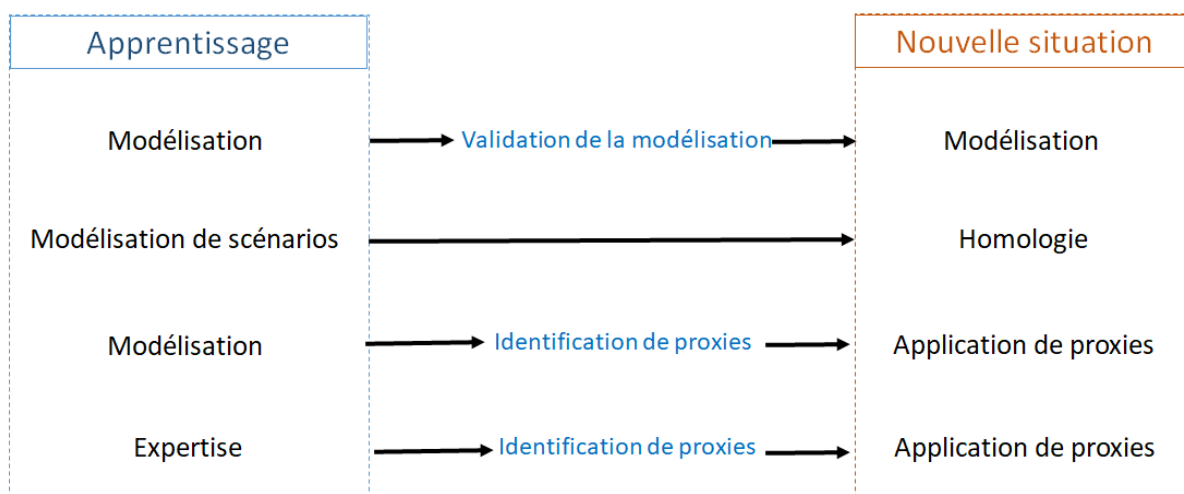


Figure 8.1 : Schéma général de différentes approches d'évaluation des services écosystémiques des sols

Références bibliographiques

- Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services — A global review. *Geoderma* 262, 101–111. <https://doi.org/10.1016/j.geoderma.2015.08.009>
- Agence européenne de l'environnement, 2013. Towards a Common International Classification of Ecosystem services (CICES), Version 4.3 [WWW Document]. URL <http://cices.eu/>
- Albizua, A., Williams, A., Hedlund, K., Pascual, U., 2015. Crop rotations including ley and manure can promote ecosystem services in conventional farming systems. *Applied Soil Ecology* 95, 54–61. <https://doi.org/10.1016/j.apsoil.2015.06.003>
- Andrew, M.E., Wulder, M.A., Nelson, T.A., Coops, N.C., 2015. Spatial data, analysis approaches, and information needs for spatial ecosystem service assessments: a review. *GIScience & Remote Sensing* 52, 344–373. <https://doi.org/10.1080/15481603.2015.1033809>
- Andrew, S.S., Karlen, D. L., Camardella, C.A., 2004. The Soil Management Assessment Framework: A quantitative Soil Quality Evaluation Method. *Soil Science of America Journal*, 68.
- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014. GlobalSoilMap: toward a fine-resolution global grid of soil properties. In: Sparks, D.L. (Ed.), *Advances in Agronomy* vol 125. Elsevier Academic Press Inc, San Diego, pp. 93–134.
- Beckett, P.H.T., Webster, R., 1971. Soil variability: A review. *Soils and Fertilizers*. 34, 1: 1-15
- Bagstad, K.J., Semmens, D.J., Winthrop, R., 2013. Comparing approaches to spatially explicit ecosystem service modeling: A case study from the San Pedro River, Arizona. *Ecosystem Services* 5, 40–50. <https://doi.org/10.1016/j.ecoser.2013.07.007>
- Baize, D., 2000. Guide des analyses en pédologie, INRA. Seconde édition revue et augmentée. pp 257.
- Baize, D., Girard, M.C., 2008. Référentiel pédologique 2008. Association française pour l'étude du sol
- Ballabio, C., Panagos, P., Montanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* 261, 110–123. <https://doi.org/10.1016/j.geoderma.2015.07.006>
- Barrios, E., 2007. Soil biota, ecosystem services and land productivity. *Ecological Economics* 64, 269–285. <https://doi.org/10.1016/j.ecolecon.2007.03.004>
- Baumont R, Dulphy JP, Sauvant D, Tran G, Meschy F, Aufrère J, Peyraud JL, Champciaux P. Les tables de la valeur des aliments (Livre rouge). In : Alimentation des bovins, ovins et caprins. Besoins des animaux – Valeurs des aliments. Tables INRA 2010 (édition remaniée), J Agabriel (coord), pp.185-279
- Baveye, P.C., Baveye, J., Gowdy, J., 2016. Soil “Ecosystem” Services and Natural Capital: Critical Appraisal of Research on Uncertain Ground. *Frontiers in Environmental Science* 4. <https://doi.org/10.3389/fenvs.2016.00041>

- Beaudoin, N., Launay, M., Sauboua, E., Ponsardin, G., Mary, B., 2008. Evaluation of the soil crop model STICS over 8 years against the “on farm” database of Bruyères catchment. *European Journal of Agronomy* 29, 46–57. <https://doi.org/10.1016/j.eja.2008.03.001>
- Bennett, D.R., Harms, T.E., 2011. Crop Yield and Water Requirement Relationships for Major Irrigated Crops in Southern Alberta. *Canadian Water Resources Journal* 36, 159–170. <https://doi.org/10.4296/cwrj3602853>
- Bennett, E.M., Peterson, G.D., Gordon, L.J., 2009. Understanding relationships among multiple ecosystem services: Relationships among multiple ecosystem services. *Ecology Letters* 12, 1394–1404. <https://doi.org/10.1111/j.1461-0248.2009.01387>
- Beroutchachvili, N., Bertrand, G., 1978. Le géosystème ou «système territorial naturel». *Revue géographique des Pyrénées et du Sud-Ouest* 49, 167–180. <https://doi.org/10.3406/rgpso.1978.3548>
- Berthier, L., Le Bris, A.-L., Lemerrier, B., Walter, C., 2013a. Organisation des sols des Côtes d'Armor. Version 1.1. Programme Sols de Bretagne, p. 246.
- Berthier, L., Le Bris, A.-L., Lemerrier, B., Walter, C., 2013b. Organisation des sols du Finistère. Version 1.1. Programme Sols de Bretagne, p. 226
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin* 24, 43–69. <https://doi.org/10.1080/02626667909491834>
- Bishop, T.F.A., Horta, A., Karunaratne, S.B., 2015. Validation of digital soil maps at different spatial supports. *Geoderma* 241–242, 238–249. <https://doi.org/10.1016/j.geoderma.2014.11.026>
- Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91, 27–45. [https://doi.org/10.1016/S0016-7061\(99\)00003-8](https://doi.org/10.1016/S0016-7061(99)00003-8)
- Biswas, A., Zhang, Y., 2018. Sampling Designs for Validating Digital Soil Maps: A Review. *Pedosphere* 28, 1–15. [https://doi.org/10.1016/S1002-0160\(18\)60001-3](https://doi.org/10.1016/S1002-0160(18)60001-3)
- Blum, W.E.H., 2005. Functions of Soil for Society and the Environment. *Reviews in Environmental Science and Bio/Technology* 4, 75–79. <https://doi.org/10.1007/s11157-005-2236>
- Boudes, P., 2008. L'environnement, domaine sociologique La sociologie française au risque de l'environnement 537.
- Bouma, J., 2010. Implications of the knowledge paradox for soil science *Adv. Agron.*, 106 pp. 143–171.
- Bouma, J., 2014. Soil science contributions towards Sustainable Development Goals and their implementation: linking soil functions with ecosystem services. *Journal of Plant Nutrition and Soil Science* 177, 111–120. <https://doi.org/10.1002/jpln.201300646>
- BRGM, 2009. <http://sigesbre.brgm.fr/Histoire-geologique-de-la-Bretagne-59.html>
- Breiman, L., Friedman, J., Stone, C., Olsen, R., 1984. *Classification and regression Trees*. Boca Raton, CA, USA.
- Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussièrre, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model stics. *European Journal of Agronomy* 18, 309–332. [https://doi.org/10.1016/S1161-0301\(02\)00110](https://doi.org/10.1016/S1161-0301(02)00110)
- Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2008. Conceptual basis, formalizations and parametrization of the STICS crop model, *Editions Quae*.
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* 62, 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364>

- Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. *Soil Research* 37, 495. <https://doi.org/10.1071/S98047>
- Bui, E.N., Moran, C.J., 2001. 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](https://doi.org/10.1016/S0016-7061(01)00070-2)
- Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L., 2018. Soil quality – A critical review. *Soil Biology and Biochemistry* 120, 105–125. <https://doi.org/10.1016/j.soilbio.2018.01.030>
- Burrough, P.A., van Gaans, P.F.M., Hootsmans, R., 1997. Continuous classification in soil survey: spatial correlation, confusion and boundaries. *Geoderma* 77, 115–135. [https://doi.org/10.1016/S0016-7061\(97\)00018-9](https://doi.org/10.1016/S0016-7061(97)00018-9)
- Busch, M., La Notte, A., Laporte, V., Erhard, M., 2012. Potentials of quantitative and qualitative approaches to assessing ecosystem services. *Ecological Indicators* 21, 89–103. <https://doi.org/10.1016/j.ecolind.2011.11.010>
- Calzolari, C., Ungaro, F., Filippi, N., Guermandi, M., Malucelli, F., Marchi, N., Staffilani, F., Tarocco, P., 2016. A methodological framework to assess the multiple contributions of soils to ecosystem services delivery at regional scale. *Geoderma* 261, 190–203. <https://doi.org/10.1016/j.geoderma.2015.07.013>
- Cardona, A., 2012. L'introduction de la notion de “ services écosystémiques ”: pour un nouveau regard sur le sol. In 6ème Journées de Recherches En Sciences Sociales, (SFER), 14p.
- Chaney, N.W., Wood, E.F., McBratney, A.B., Hempel, J.W., Nauman, T.W., Brungard, C.W., Odgers, N.P., 2016. 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>
- Chaplot, V., Bouahom, B., Valentin, C., 2010. Soil organic carbon stocks in Laos: spatial variations and controlling factors. *Glob. Chang. Biol.* 16, 1380–1393.
- Chatfield, C., 1995. Problem Solving. A statistician's guide, second edition. New York: Chapman and Hall/CRC. <https://doi.org/10.1201/b15238>
- Chen, S., 2019 - Cartographie à haute résolution de propriétés du sol à l'échelle de la France métropolitaine : application au carbone organique du sol et à ses potentiels additionnels de séquestration et de stockage. Thèse d'Agrocampus Ouest sous le sceau de l'Université Bretagne-Loire, 175 p.
- Chen, S., Richer-de-Forges, A.C., Saby, N.P.A., Martin, M.P., Walter, C., Arrouays, D., 2018. 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>
- CICES, 2013. CICES V4.3 (January 2013). <http://cices.eu/>.
- Le Clec'h, S., 2016 – Spatialisisation des services écosystémiques en contexte de front pionnier amazonien : analyse critique d'un outil de la gestion agroenvironnementale. Thèse De l'Université de Rennes 2 sous le sceau de l'Université Bretagne-Loire, 410 p.
- Le Clec'h, S., Oszwald, J., Decaens, T., Desjardins, T., Dufour, S., Grimaldi, M., Jegou, N., Lavelle, P., 2016. Mapping multiple ecosystem services indicators: Toward an objective-oriented approach. *Ecological Indicators* 69, 508–521. <https://doi.org/10.1016/j.ecolind.2016.05.021>
- Cook, S., Corner, R., Groves, P., Grealish, G., 1996. Use of airborne gamma radiometric data for soil mapping. *Soil Research* 34, 183. <https://doi.org/10.1071/SR9960183>
- Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., V. O'Neill, R., J. Paruelo, R. G. Raskin, P. Sutton, and M. van den Belt., 1997. The value of the world's ecosystem services and natural capital. *Ecological Economics*, 25:3615, 1997.
- Curtin, F., Schulz, P., 1998. Multiple correlations and bonferroni's correction. *Biological Psychiatry* 44, 775–777. [https://doi.org/10.1016/S0006-3223\(98\)00043-2](https://doi.org/10.1016/S0006-3223(98)00043-2)

- Crutzen, P.J., Stoermer, E.F., "The 'Anthropocene' ", 2000. Global Change. IGBP Newsletter, 41, p. 17-18.
- Daily, G., 1997. Nature's Services: Societal Dependence on Natural Ecosystem. Island Press. Washington DC
- de Groot, R.S., Alkemade, R., Braat, L., Hein, L., Willemsen, L., 2010. Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity* 7, 260–272. <https://doi.org/10.1016/j.ecocom.2009.10.006>
- de Groot, R.S., Wilson, M.A., Boumans, R.M.J., 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics* 41, 393–408. [https://doi.org/10.1016/S0921-8009\(02\)00089-7](https://doi.org/10.1016/S0921-8009(02)00089-7)
- de Groot, R., et al. (2018). "RE: There is more to Nature's Contributions to People than Ecosystem Services – A response to de Groot et al." Current Opinion in *Environmental Sustainability* 359.
- De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. Sampling for Natural Resource Monitoring. *Springer*, Berlin.
- De Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016. Farm-scale soil carbon auditing. *Geoderma* 265, 120–130. <https://doi.org/10.1016/j.geoderma.2015.11.010>
- Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R.T., Molnár, Z., Hill, R., Chan, K.M.A., Baste, I.A., Brauman, K.A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P.W., van Oudenhoven, A.P.E., van der Plaats, F., Schröter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C.A., Hewitt, C.L., Keune, H., Lindley, S., Shirayama, Y., 2018. Assessing nature's contributions to people. *Science* 359, 270–272. <https://doi.org/10.1126/science.aap8826>
- Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecological Economics* 69, 1858–1868. <https://doi.org/10.1016/j.ecolecon.2010.05.002>
- Dominati, E.J., n.d. Natural capital and ecosystem services of soils 11.
- Dominati, E.J., Mackay, A., Lynch, B., Heath, N., Millner, I., 2014. An ecosystem services approach to the quantification of shallow mass movement erosion and the value of soil conservation practices. *Ecosystem Services* 9, 204–215. <https://doi.org/10.1016/j.ecoser.2014.06.006>
- Doran, J.W., 2002. Soil health and global sustainability: translating science into practice 9.
- Doran, J.W., Safley, M., 1997. Defining and assessing soil health and sustainable productivity. In: Pankhurst, C., Doube, B.M., Gupta, V.V.S.R. (Eds.), *Biological Indicators of Soil Health*. CAB International, Wallingford, Oxon, UK, pp. 1–28.
- Doran, J.W., Zeiss, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. *Applied Soil Ecology* 15, 3–11. [https://doi.org/10.1016/S0929-1393\(00\)00067-6](https://doi.org/10.1016/S0929-1393(00)00067-6)
- Dray S, Dufour AB, 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22(4):1–20
- Drobnik, T., Greiner, L., Keller, A., Grêt-Regamey, A., 2018. Soil quality indicators – From soil functions to ecosystem services. *Ecological Indicators* 94, 151–169. <https://doi.org/10.1016/j.ecolind.2018.06.052>
- Dymond, J.R., Ausseil, A.-G.E., Ekanayake, J.C., Kirschbaum, M.U.F., 2012. Tradeoffs between soil, water, and carbon – A national scale analysis from New Zealand. *Journal of Environmental Management* 95, 124–131. <https://doi.org/10.1016/j.jenvman.2011.09.019>
- European Commission. (2006). Proposal for a Directive of the European Parliament and the Council Establishing a Framework for the Protection of Soil Amending Directive 2004-35-.
- Egoh, B., Drakou, E.G., Dunbar, M.B., Maes, J., Willemsen, L., 2012. Indicators for mapping ecosystem services: a review. European Commission, Joint Research Centre (JRC), Ispra.

- Egoh, B., Reyers, B., Rouget, M., Richardson, D.M., Le Maitre, D.C., van Jaarsveld, A.S., 2008. Mapping ecosystem services for planning and management. *Agriculture, Ecosystems & Environment* 127, 135–140. <https://doi.org/10.1016/j.agee.2008.03.013>
- Eigenbrod, F., Armsworth, P.R., Anderson, B.J., Heinemeyer, A., Gillings, S., Roy, D.B., Thomas, C.D., Gaston, K.J., 2010. The impact of proxy-based methods on mapping the distribution of ecosystem services. *Journal of Applied Ecology* 47, 377–385. <https://doi.org/10.1111/j.1365-2664.2010.01777>
- Ellili Bargaoui, Y., Walter, C., Michot, D., Saby, N.P.A., Vincent, S., Lemerrier, B., 2019. Validation of digital maps derived from spatial disaggregation of legacy soil maps. *Geoderma* 356, 113907. <https://doi.org/10.1016/j.geoderma.2019.113907>
- Ellili, Y., Walter, C., Michot, D., Pichelin, P., Lemerrier, B., 2019. 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>
- European commission, Joint research centre, Institute for environment and sustainability (Ispra, I., Maes, J., Paracchini, M.L., Zulian, G.B., 2011. A European assessment of the provision of ecosystem services: Towards an atlas of ecosystem services. Publications Office of the European Union, Luxembourg.
- ESRI, 2012. ArcMap 10.1. Environmental Systems Resource Institute, Redlands, California
- Faber, J.H., van Wensem, J., 2012. Elaborations on the use of the ecosystem services concept for application in ecological risk assessment for soils. *Sci. Total Environ.* 415, 3-8.
- FAO, 2003. Evaluation du bilan en éléments nutritifs du sol. Approches et méthodologies. FAO Fertilizer and Plant Nutrition Bulletin 14, Rome.
- Fisher, B., Edwards, D.P., Larsen, T.H., Ansell, F.A., Hsu, W.W., Roberts, C.S., Wilcove, D.S., 2011a. Cost-effective conservation: calculating biodiversity and logging trade-offs in Southeast Asia: Cost-effective conservation in logged forests. *Conservation Letters* 4, 443–450. <https://doi.org/10.1111/j.1755-263X.2011.00198>
- Fisher, B., Turner, R.K., Burgess, N.D., Swetnam, R.D., Green, J., Green, R.E., Kajembe, G., Kulindwa, K., Lewis, S.L., Marchant, R., Marshall, A.R., Madoffe, S., Munishi, P.K.T., Morse-Jones, S., Mwakalila, S., Paavola, J., Naidoo, R., Ricketts, T., Rouget, M., Willcock, S., White, S., Balmford, A., 2011b. Measuring, modeling and mapping ecosystem services in the Eastern Arc Mountains of Tanzania. *Progress in Physical Geography: Earth and Environment* 35, 595–611. <https://doi.org/10.1177/0309133311422968>
- Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. *Ecological Economics* 68, 643–653. <https://doi.org/10.1016/j.ecolecon.2008.09.014>
- Fossey, M., Bustany, C., Angers, D., Cudennec, C., Durand, P., Gascuel, C., Jaffrezic, A., Pérès, G., Bois, S., Warot, G., Walter, C. Can soil ecosystem services be considered in territorial planning? Article soumis à Frontiers.
- Fraga, H., García de Cortázar Aauri, I., Santos, J.A., 2018. Viticultural irrigation demands under climate change scenarios in Portugal. *Agricultural Water Management* 196, 66–74. <https://doi.org/10.1016/j.agwat.2017.10.023>
- Fu, Q., Li, B., Yang, L., Wu, Z., Zhang, X., 2015. Ecosystem Services Evaluation and Its Spatial Characteristics in Central Asia's Arid Regions: A Case Study in Altay Prefecture, China. *Sustainability* 7, 8335–8353. <https://doi.org/10.3390/su7078335>
- Garcia, D, C, A, I., 2006 – Adaptation du modèle STICSa la vigne (vitis vinifera L.). Utilisation dans le cadre d'une étude d'impact du changement climatique à l'échelle de la France. Thèse de Montpellier SupAgro sous le sceau de l'école doctorale BSIAE, 292 p.
- Gaines, T. P., Gaines, S. T., 1994. Soil texture effect on nitrate leaching in soil percolates, *Communications in Soil Science and Plant Analysis*, 25:13-14, 2561-2570, <https://doi.org/10.1080/00103629409369207>

- Garrigues, E., Corson, M.S., Angers, D.A., van der Werf, H.M.G., Walter, C., 2012. Soil quality in Life Cycle Assessment: Towards development of an indicator. *Ecological Indicators* 18, 434–442. <https://doi.org/10.1016/j.ecolind.2011.12.014>
- Grace, J.B., 2006. Structural Equation Modeling and Natural Systems Cambridge University Press, Cambridge, UK.
- Graux, A.-I., Gaurut, M., Agabriel, J., Baumont, R., Delagarde, R., Delaby, L., Soussana, J.-F., 2011. Development of the Pasture Simulation Model for assessing livestock production under climate change. *Agriculture, Ecosystems & Environment* 144, 69–91. <https://doi.org/10.1016/j.agee.2011.07.001>
- Greiner, L., Keller, A., Grêt-Regamey, A., Papritz, A., 2017. Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. *Land Use Policy* 69, 224–237. <https://doi.org/10.1016/j.landusepol.2017.06.025>
- Grimaldi, M., Oszwald, J., Dolédec, S., Hurtado, M. del P., de Souza Miranda, I., Arnauld de Sartre, X., Assis, W.S. de, Castañeda, E., Desjardins, T., Dubs, F., Guevara, E., Gond, V., Lima, T.T.S., Marichal, R., Michelotti, F., Mitja, D., Noronha, N.C., Delgado Oliveira, M.N., Ramirez, B., Rodriguez, G., Sarrazin, M., Silva, M.L. da, Costa, L.G.S., Souza, S.L. de, Veiga, I., Velasquez, E., Lavelle, P., 2014. Ecosystem services of regulation and support in Amazonian pioneer fronts: searching for landscape drivers. *Landscape Ecology* 29, 311–328. <https://doi.org/10.1007/s10980-013-9981-y>
- Guo, L.B., Gifford, R.M., 2002. Soil carbon stocks and land use change: a meta analysis. *Global Change Biology* 8, 345–360. <https://doi.org/10.1046/j.1354-1013.2002.00486>
- Guo, Z., Xiao, X., Li, D., 2000. An assessment of ecosystem services: water flow regulation and hydroelectric power production. *Ecol. Appl.* 10 (3), 925–936.
- FAOSTAT, 2004. Food balance sheets [WWW Document]. Food and Agriculture Organization of the United Nations. URL <http://www.fao.org/faostat/en/>.
- Haines-Young, R., Potschin, M., 2010. The links between biodiversity, ecosystem services and human well-being, in: Raffaelli, D.G., Frid, C.L.J. (Eds.), *Ecosystem Ecology*. Cambridge University Press, Cambridge, pp. 110–139. <https://doi.org/10.1017/CBO9780511750458.007>
- Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., Schröder, B., 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. *Geoderma* 185–186, 37–47. <https://doi.org/10.1016/j.geoderma.2012.04.001>
- Hassink, J., 1992. Effects of soil texture and structure on carbon and nitrogen mineralization in grassland soils. *Biology and Fertility of Soils* 14, 126–134. <https://doi.org/10.1007/BF00336262>
- Hassink, J., 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant and Soil* 191, 77–87.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. *Springer Series in Statistics*
- Hartemink, A., 2015. The definition of soil since the early 1800's. *Adv. Agron.*, 137, pp. 73–126
- Hauck, J., Albert, C., Fürst, C., Geneletti, D., La Rosa, D., Lorz, C., Spyra, M., 2016. Developing and applying ecosystem service indicators in decision-support at various scales. *Ecological Indicators* 61, 1–5. <https://doi.org/10.1016/j.ecolind.2015.09.037>
- Haygarth, P.M., Ritz, K., 2009. The future of soils and land use in the UK: Soil systems for the provision of land-based ecosystem services. *Land Use Policy* 26, S187–S197. <https://doi.org/10.1016/j.landusepol.2009.09.016>
- Hengl, T., Kempen, B., Heuvelink, G., Malone, B., Hannes, R., 2013. GSIF: Global Soil Information Facilities. R package version 0.3–2.
- Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. 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>

- Hewitt, A., Dominati, E., Webb, T., Cuthill, T., 2015. Soil natural capital quantification by the stock adequacy method. *Geoderma* 241–242, 107–114. <https://doi.org/10.1016/j.geoderma.2014.11.014>
- Holmes, K.W., Griffin, E.A., Odgers, N.P., 2015. 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>
- IGN, 2008. BD ALTI®. <http://www.ign.fr>.
- INRA Infosol, 2014. Donesol Version 3.4.3. Dictionnaire de données
- IPBES. 2018. Catalogue of Assessments on Biodiversity and Ecosystem Service <http://catalog.ipbes.net>
- IUSS Working Group WRB: World reference base for soil resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome, 116 pp.
- Jenny, H., 1941. Factors of Soil Formation: a System of Quantitative Pedology. McGraw-Hill, New York and London.
- Jobbagy, E.G., Jackson, R.B., 2000. The Vertical Distribution of Soil Organic Carbon and Its Relation to Climate and Vegetation. *Ecological Applications* 10, 423. <https://doi.org/10.2307/2641104>
- Jopke, C., Kreyling, J., Maes, J., Koellner, T., 2015. Interactions among ecosystem services across Europe: Bagplots and cumulative correlation coefficients reveal synergies, trade-offs, and regional patterns. *Ecological Indicators* 49, 46–52. <https://doi.org/10.1016/j.ecolind.2014.09.037>
- Karlen, D.L., Ditzler, C.A., Andrews, S.S., 2003. Soil quality: why and how? *Geoderma* 114, 145–156. [https://doi.org/10.1016/S0016-7061\(03\)00039-9](https://doi.org/10.1016/S0016-7061(03)00039-9)
- Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. 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>
- Kerry, R., Goovaerts, P., Rawlins, B.G., Marchant, B.P., 2012. 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>
- Keskin, H., Grunwald, S., 2018. Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma* 326, 22–41. <https://doi.org/10.1016/j.geoderma.2018.04.004>
- Kirschbaum, M.U.F., n.d. Will changes in soil organic carbon act as a positive or negative feedback on global warming, 32.
- Lacoste, M., Lemerrier, B., Walter, C., 2011. 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>
- Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. 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>
- Lal, R., 2004. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* 304, 1623–1627. <https://doi.org/10.1126/science.1097396>
- Larondelle, N., Haase, D., Kabisch, N., 2014. Mapping the diversity of regulating ecosystem services in European cities. *Global Environmental Change* 26, 119–129. <https://doi.org/10.1016/j.gloenvcha.2014.04.008>
- Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.-P., 2006a. Soil invertebrates and ecosystem services. *European Journal of Soil Biology* 42, S3–S15. <https://doi.org/10.1016/j.ejsobi.2006.10.002>
- Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.-P., 2006b. Soil invertebrates and ecosystem services. *European Journal of Soil Biology* 42, S3–S15. <https://doi.org/10.1016/j.ejsobi.2006.10.002>

- Lavelle, P., Rodríguez, N., Arguello, O., Bernal, J., Botero, C., Chaparro, P., Gómez, Y., Gutiérrez, A., Hurtado, M. del P., Loaiza, S., Pullido, S.X., Rodríguez, E., Sanabria, C., Velásquez, E., Fonte, S.J., 2014. Soil ecosystem services and land use in the rapidly changing Orinoco River Basin of Colombia. *Agriculture, Ecosystems & Environment* 185, 106–117. <https://doi.org/10.1016/j.agee.2013.12.020>
- Layke, C., Mapendembe, A., Brown, C., Walpole, M., Winn, J., 2012. Indicators from the global and sub-global Millennium Ecosystem Assessments: An analysis and next steps. *Ecological Indicators* 17, 77–87. <https://doi.org/10.1016/j.ecolind.2011.04.025>
- Le Bas, C. 2016 – Impact des incertitudes des paramétrages sols estimés à partir de la Base de Données Géographique des Sols de France sur les sorties du modèle de culture STICS. Approche méthodologique. Thèse d'AgroParisTech sous le sceau de l'école doctorale ABIES, 447 p.
- Le Bris, A.-L., Berthier, L., Lemerrier, B., Walter, C., 2013a. Organisation des sols d'Ille-et-Vilaine. Version 1.1. Programme Sols de Bretagne, p. 266.
- Le Bris, A.-L., Berthier, L., Lemerrier, B., Walter, C., 2013b. Organisation des sols du Morbihan. Version 1.1. Programme Sols de Bretagne, p. 258.
- Lee, H., Lautenbach, S., 2016. A quantitative review of relationships between ecosystem services. *Ecological Indicators* 66, 340–351. <https://doi.org/10.1016/j.ecolind.2016.02.004>
- Le Du-Blayo, L., Corpetti, T., Gouery, P., Bourget, E., 2008. 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.
- Lemaire G., Gastal F., Plénet D., 1997. Dynamics of N uptake and N distribution in plant canopies. Use of crop N status index in crop modelling. In: Diagnostic Procedures for Crop N Management. Eds G Lemaire and I G Burns, pp 15-29. INRA, Paris
- Lemerrier, B., Lacoste, M., Loum, M., Berthier, L., Le Bris, A.L., Walter, C., 2013. 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
- Lemerrier, B., Lacoste, M., Loum, M., Walter, C., 2012. 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>
- Lemerrier, B., Walter, C., Saby, N., Schvartz, C., Arrouays, D., Follain, S. 2006 - Suivi des teneurs en carbone organique et en phosphore extractible dans les sols agricoles de trois régions françaises. Analyse à partir de la Base de Données des Analyses de Terre. *Étude et Gestion des Sols*, vol 13 (3), 165-180.
- Liu, F., Geng, X., Zhu, A., -xing, Fraser, W., Song, X., Zhang, G., 2016. Soil polygon disaggregation through similarity-based prediction with legacy pedons. *Journal of Arid Land* 8, 760–772. <https://doi.org/10.1007/s40333-016-0087-7>
- Lü, Y., Fu, B., Feng, X., Zeng, Y., Liu, Y., Chang, R., Sun, G., Wu, B., 2012. A Policy-Driven Large Scale Ecological Restoration: Quantifying Ecosystem Services Changes in the Loess Plateau of China. *PLoS ONE* 7, e31782. <https://doi.org/10.1371/journal.pone.0031782>
- Machado, I.R., Giasson, E., Campos, A.R., Costa, J.J.F., Silva, E.B. da, Bonfatti, B.R., 2018. 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/10.1590/18069657rbcsc20170193>
- Maes, J., Liqueste, C., Teller, A., Erhard, M., Paracchini, M.L., Barredo, J.I., Grizzetti, B., Cardoso, A., Somma, F., Petersen, J.-E., Meiner, A., Gelabert, E.R., Zal, N., Kristensen, P., Bastrup-Birk, A., Biala, K., Piroddi, C., Egoh, B., Degeorges, P., Fiorina, C., Santos-Martín, F., Naruševičius, V., Verboven, J., Pereira, H.M., Bengtsson, J., Gocheva, K., Marta-Pedroso, C., Snäll, T., Estreguil, C., San-Miguel-Ayanz, J., Pérez-Soba, M., Grêt-Regamey, A., Lillebø, A.I., Malak, D.A., Condé, S., Moen, J., Czúcz, B., Drakou, E.G., Zulian, G., Lavelle,

- C., 2016. An indicator framework for assessing ecosystem services in support of the EU Biodiversity Strategy to 2020. *Ecosystem Services* 17, 14–23. <https://doi.org/10.1016/j.ecoser.2015.10.023>
- Maes, J., Paracchini, M.L., Zulian, G., Dunbar, M.B., Alkemade, R., 2012. Synergies and trade-offs between ecosystem service supply, biodiversity, and habitat conservation status in Europe. *Biological Conservation* 155, 1–12. <https://doi.org/10.1016/j.biocon.2012.06.016>
- Malik, A.A., Puissant, J., Buckeridge, K.M., Goodall, T., Jehmlich, N., Chowdhury, S., Gweon, H.S., Peyton, J.M., Mason, K.E., van Agtmaal, M., Blaud, A., Clark, I.M., Whitaker, J., Pywell, R.F., Ostle, N., Gleixner, G., Griffiths, R.I., 2018. Land use driven change in soil pH affects microbial carbon cycling processes. *Nature Communications* 9. <https://doi.org/10.1038/s41467-018-05980-1>
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154, 138–152. <https://doi.org/10.1016/j.geoderma.2009.10.007>
- Maris, V., 2014. Nature à vendre : les limites des services écosystémiques. *Sciences en questions. Quae.ed.*
- Marsman, B.A., De Gruijter, J.J., 1986. Quality of soil maps: a comparison of soil survey methods in a sandy area, Soil survey papers. Soil Survey Inst, Wageningen.
- McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203–213. <https://doi.org/10.1016/j.geoderma.2013.08.013>
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Merot, Ph., Ezzahar, B., Walter, C., Arousseau, P., 1995. Mapping waterlogging of soils using digital terrain models. *Hydrological Processes* 9, 27–34. <https://doi.org/10.1002/hyp.3360090104>
- Meynard, J.-M., Cerf, M., Guichard, L., Jeuffroy, M.-H., Makowski, D., 2002. Which decision support tools for the environmental management of nitrogen? *Agronomie* 22, 817–829. <https://doi.org/10.1051/agro:2002064>
- Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington, DC. 86 p.
- Minasny, B., McBratney, A.B., 2010. 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
- 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>
- Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. In: *Advances in Agronomy*. Elsevier, pp. 1–47.
- Møller, A.B., Malone, B., Odgers, N.P., Beucher, A., Iversen, B.V., Greve, M.H., Minasny, B., 2019. Improved disaggregation of conventional soil maps. *Geoderma* 341, 148–160. <https://doi.org/10.1016/j.geoderma.2019.01.038>
- Mouchet, M.A., Lamarque, P., Martín-López, B., Crouzat, E., Gos, P., Byczek, C., Lavorel, S., 2014. An interdisciplinary methodological guide for quantifying associations between ecosystem services. *Global Environmental Change* 28, 298–308. <https://doi.org/10.1016/j.gloenvcha.2014.07.012>
- Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. 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>

- Nahlik, A.M., Kentula, M.E., Fennessy, M.S., Landers, D.H., 2012. Where is the consensus? A proposed foundation for moving ecosystem service concepts into practice. *Ecological Economics* 77, 27–35. <https://doi.org/10.1016/j.ecolecon.2012.01.001>
- Naidoo, R., Ricketts, T.H., 2006. Mapping the Economic Costs and Benefits of Conservation. *PLoS Biology* 4, e360. <https://doi.org/10.1371/journal.pbio.0040360>
- Nauman, T.W., Thompson, J.A., 2014. 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>
- Nauman, T.W., Thompson, J.A., Rasmussen, C., 2014. 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>
- Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, Dr., Chan, K.M., Daily, G.C., Goldstein, J., Kareiva, P.M., Lonsdorf, E., Naidoo, R., Ricketts, T.H., Shaw, Mr., 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontiers in Ecology and the Environment* 7, 4–11. <https://doi.org/10.1890/080023>
- Nemec, K.T., Raudsepp-Hearne, C., 2013. The use of geographic information systems to map and assess ecosystem services. *Biodiversity and Conservation* 22, 1–15. <https://doi.org/10.1007/s10531-012-0406-z>
- Nowosad, J., Stepinski, T.F., 2018. Spatial association between regionalizations using the information-theoretical V –measure. <https://doi.org/10.1080/13658816.2018.1511794>.
- OECD, 2008. Strategic environmental assessment and ecosystem services [WWW Document]. URL <http://www.oecd.org/dac/environment-development/41882953.pdf>
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* 67, 215–226.
- Odgers, N., McBratney, A., Minasny, B., Sun, W., Clifford, D., 2014. Dsmart: An algorithm to spatially disaggregate soil map units, in: Arrouays, D., McKenzie, N., Hempel, J., de Forges, A., McBratney, Alex (Eds.), *GlobalSoilMap*. CRC Press, pp. 261–266. <https://doi.org/10.1201/b16500-49>
- Odgers, N.P., Holmes, K.W., Griffin, T., Liddicoat, C., 2015a. Derivation of soil-attribute estimations from legacy soil maps. *Soil Research* 53, 881. <https://doi.org/10.1071/SR14274>
- Odgers, N.P., Libohova, Z., Thompson, J.A., 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma* 189–190, 153–163. <https://doi.org/10.1016/j.geoderma.2012.05.026>
- Odgers, N.P., McBratney, A.B., Minasny, B., 2015b. 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>
- O’Geen, A.T., 2012. Soil Water Dynamics. *Nature Education Knowledge* 3 (6):12
- Palm, C., Sanchez, P., Ahamed, S., Awiti, A., 2007. Soils: A Contemporary Perspective. *Annual Review of Environment and Resources* 32, 99–129. <https://doi.org/10.1146/annurev.energy.31.020105.100307>
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691
- Powell Gaines, T., Gaines, S.T. 1994. Soil Texture Effect on Nitrate Leaching in Soil Percolates. Tifton Physical Soil Testing Laboratory, Tifton, Georgia. Published in COMMON. SOIL SCI. PLANT ANAL., 25-1314, 2561-2570.
- Ponce-Hernandez, R., Marriott, F.H.C., Beckett, P.H.T., 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. *Journal of Soil Science* 37, 455–467. <https://doi.org/10.1111/j.1365-2389.1986.tb00377.x>

- Prado, R.B., Fidalgo, E.C.C., Monteiro, J.M.G., Schuler, A.E., Vezzani, F.M., Garcia, J.R., Oliveira, A.P. de, Viana, J.H.M., Pedreira, B. da C.C.G., Mendes, I. de C., Reatto, A., Parron, L.M., Clemente, E. de P., Donagemma, G.K., Turetta, A.P.D., Simões, M., 2016. Current overview and potential applications of the soil ecosystem services approach in Brazil. *Pesquisa Agropecuária Brasileira* 51, 1021–1038. <https://doi.org/10.1590/s0100-204x2016000900002>
- Qin, K., Li, J., Yang, X., 2015. Trade-Off and Synergy among Ecosystem Services in the Guanzhong-Tianshui Economic Region of China. *International Journal of Environmental Research and Public Health* 12, 14094–14113. <https://doi.org/10.3390/ijerph121114094>
- Quinlan, J.R., 1993. C4.5: Programs for Machine Learning, 1. Morgan Kaufmann Publishers
- Rapport scientifique de l'étude réalisé par l'INRA-volet « écosystèmes agricoles » de l'Evaluation Française des Ecosystèmes et des Services Ecosystémiques (EFESE), 2017. 970 p.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R.Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0. <http://www.R-project.org>.
- Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J.A.M., Scholten, T., 2014. Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma* 226–227, 140–150. <https://doi.org/10.1016/j.geoderma.2014.02.002>
- Raudsepp-Hearne, C., Peterson, G.D., Bennett, E.M., 2010. Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. *Proceedings of the National Academy of Sciences* 107, 5242–5247. <https://doi.org/10.1073/pnas.0907284107>
- Remy A., Lemercier B, Berthier L, Michot D., Walter C. 2015. Etablissement de fonctions de pédotransfert des propriétés hydrodynamiques des sols en Bretagne à partir de mesures physiques au laboratoire. Rapport interne du programme Sols de Bretagne. 36 pages
- Reyers, B., O'Farrell, P.J., Cowling, R.M., Egoh, B.N., Le Maitre, D.C., Vlok, J.H.J., 2009. Ecosystem Services, Land-Cover Change, and Stakeholders: Finding a Sustainable Foothold for a Semiarid Biodiversity Hotspot. *Ecology and Society* 14. <https://doi.org/10.5751/ES-02867-140138>
- Rivière, J.M., Tico, S., Dupont, C., 1992. Méthode Tarière Massif Armoricaïn. Caractérisation des sols, Rennes : INRA Editions, p. 20.
- Robinson, D.A., Emmett, B.A., Reynolds, B., Rowe, E.C., Spurgeon, D., Keith, A.M., Lebron, I., Hockley, N., n.d. Soil Natural Capital and Ecosystem Service Delivery in a *World of Global Soil Change* 28.
- Robinson, D.A., Fraser, I., Dominati, E.J., Davíðsdóttir, B., Jónsson, J.O.G., Jones, L., Jones, S.B., Tuller, M., Lebron, I., Bristow, K.L., Souza, D.M., Banwart, S., Clothier, B.E., 2014. On the Value of Soil Resources in the Context of Natural Capital and Ecosystem Service Delivery. *Soil Science Society of America Journal* 78, 685. <https://doi.org/10.2136/sssaj2014.01.0017>
- Robinson, D.A., Hockley, N., Dominati, E., Lebron, I., Scow, K.M., Reynolds, B., Emmett, B.A., Keith, A.M., de Jonge, L.W., Schjønning, P., Moldrup, P., Jones, S.B., Tuller, M., 2012. Natural Capital, Ecosystem Services, and Soil Change: Why Soil Science Must Embrace an Ecosystems Approach. *Vadose Zone Journal* 11, 0. <https://doi.org/10.2136/vzj2011.0051>
- Rosenberg, A., Hirschberg, J., 2007: 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.
- Russell, J.S., Moore, A.W., 1968. Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. *Transactions of the 9th International Congress of Soil Science* 4, 205–213.
- Rousseeuw, P. J., Ruts, I., Tukey, J. W., 1999. Statistical Computing and Graphics: The Bagplot: A bivariate Boxplot. *The American Statistician*, 382–387.
- Rutgers, M., van Wijnen, H.J., Schouten, A.J., Mulder, C., Kuiten, A.M.P., Brussaard, L., Breure, A.M., 2012. A method to assess ecosystem services developed from soil attributes with

- stakeholders and data of four arable farms. *Science of The Total Environment* 415, 39–48. <https://doi.org/10.1016/j.scitotenv.2011.04.041>
- Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A., Üveges, J.B., Freudenschuß, A., Simota, C., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology* 14, 2432–2442. <https://doi.org/10.1111/j.1365-2486.2008.01658.x>
- Saby, N.P.A., Lemerrier, B., Arrouays, D., Lemenager, S., Louis, B.P., Millet, Paroissein, F., Schellenberger, E., Squvidant, H., Swiderski, C., Toutain, C., Walter, C., Bardy, M., 2014. Le programme Base de Données des Analyses de Terre (BDAT) : Bilan de 20 ans de collecte de résultats d'analyses. *Etude et Gestion des Sols*, vol.21, 141-150
- Sandhu, H.S., Wratten, S.D., Cullen, R., 2010a. Organic agriculture and ecosystem services. *Environmental Science & Policy* 13, 1–7. <https://doi.org/10.1016/j.envsci.2009.11.002>
- Sandhu, H.S., Wratten, S.D., Cullen, R., 2010b. The role of supporting ecosystem services in conventional and organic arable farmland. *Ecological Complexity* 7, 302–310. <https://doi.org/10.1016/j.ecocom.2010.04.006>
- Schulte, R.P.O., Creamer, R.E., Donnellan, T., Farrelly, N., Fealy, R., O'Donoghue, C., O'hUallachain, D., 2014. Functional land management: A framework for managing soil-based ecosystem services for the sustainable intensification of agriculture. *Environmental Science & Policy* 38, 45–58. <https://doi.org/10.1016/j.envsci.2013.10.002>
- Schwilch, G., Bernet, L., Fleskens, L., Giannakis, E., Leventon, J., Marañón, T., Mills, J., Short, C., Stolte, J., van Delden, H., Verzandvoort, S., 2016. Operationalizing ecosystem services for the mitigation of soil threats: A proposed framework. *Ecological Indicators* 67, 586–597. <https://doi.org/10.1016/j.ecolind.2016.03.016>
- Scull, P., Franklin, J., Chadwick, O.A., 2005. 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>
- Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S., Schmidt, S., 2011. A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead: Priorities for ecosystem service studies. *Journal of Applied Ecology* 48, 630–636. <https://doi.org/10.1111/j.1365-2664.2010.01952.x>
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Sharp, R., Tallis, H., Ricketts, T., Guerry, A., Wood, S., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M., Mandle, L., Hamel, P., Vogl, A. 2014. InVEST User's Guide.
- Sherrouse, B., Semmens, D. 2010. Social values for ecosystem services (solves). Using gis to include social values
- Silva, F.A.M., Naudin, K., Corbeels, M., Scopel, E., Affholder, F., 2019. Impact of conservation agriculture on the agronomic and environmental performances of maize cropping under contrasting climatic conditions of the Brazilian Cerrado. *Field Crops Research* 230, 72–83. <https://doi.org/10.1016/j.fcr.2018.10.009>
- Singh, M., n.d. Stabilization of Soil Organic Carbon as Influenced by Clay Mineralogy 53.
- Smith, V.H., Schindler, D.W. 2009. Eutrophication science: where do we go from here ? *Trends in Ecology Evolution* 24(4) : 201-207. <https://doi.org/10.1016/j.tree.2008.11.009>.
- Soil Science Division Staff. *Soil Survey Manual; USDA Handbook 18*; Ditzler, C., Sche_e, K., Monger, H.C. 2017. Eds.; Government Printing Office: Washington, DC, USA

- Soltani, I., Fouad, Y., Michot, D., Bréger, P., Dubois, R., Cudennec, C., 2019. A near infrared index to assess effects of soil texture and organic carbon content on soil water content: Effect of texture and organic carbon on soil water with NIR. *European Journal of Soil Science* 70, 151–161. <https://doi.org/10.1111/ejss.12725>
- Squidant, H., 1994. MNTSURF: Logiciel de traitement des modèles numériques de terrain. 36 ENSAR, Rennes, France.
- Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V. de R. de, Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment* 164, 80–99. <https://doi.org/10.1016/j.agee.2012.10.001>
- Stoorvogel, J.J., Bakkenes, M., Brink, B.J.E., Temme, A.J.A.M., 2017a. To What Extent Did We Change Our Soils? A Global Comparison of Natural and Current Conditions. *Land Degradation & Development* 28, 1982–1991. <https://doi.org/10.1002/ldr.2721>
- Stoorvogel, J.J., Bakkenes, M., Temme, A.J.A.M., Batjes, N.H., Brink, B.J.E., 2017b. S-World: A Global Soil Map for Environmental Modelling. *Land Degradation & Development* 28, 22–33. <https://doi.org/10.1002/ldr.2656>
- Styc, Q., Lagacherie, P., 2019. What is the Best Inference Trajectory for Mapping Soil Functions: An Example of Mapping Soil Available Water Capacity over Languedoc Roussillon (France). *Soil Systems* 3, 34. <https://doi.org/10.3390/soilsystems3020034>
- Swinton, S.M., Lupi, F., Robertson, G.P., Hamilton, S.K., 2007. Ecosystem services and agriculture: Cultivating agricultural ecosystems for diverse benefits. *Ecological Economics* 64, 245–252. <https://doi.org/10.1016/j.ecolecon.2007.09.020>
- Tibi A., Théron O. 2018. Services écosystémiques fournis par les espaces agricoles. Evaluer et caractériser. Editions Quae, collection Matière à débattre et décider, 188 p.
- Tomscha, S.A., Gergel, S.E., 2016. Ecosystem service trade-offs and synergies misunderstood without landscape history. *Ecology and Society* 21. <https://doi.org/10.5751/ES-08345-210143>
- United Nations, 2015. Sustainable Development Goals. <https://sustainabledevelopment.un.org>.
- Vallet, A., Locatelli, B., Levrel, H., Wunder, S., Seppelt, R., Scholes, R.J., Oszwald, J., 2018. Relationships Between Ecosystem Services: Comparing Methods for Assessing Tradeoffs and Synergies. *Ecological Economics* 150, 96–106. <https://doi.org/10.1016/j.ecolecon.2018.04.002>
- van Wijk, M.T., Klapwijk, C.J., Rosenstock, T.S., van Asten, P.J.A., Thornton, P.K., Giller, K.E., 2016. Methods for Environment: Productivity Trade-Off Analysis in Agricultural Systems, in: Rosenstock, T.S., Rufino, M.C., Butterbach-Bahl, K., Wollenberg, L., Richards, M. (Eds.), *Methods for Measuring Greenhouse Gas Balances and Evaluating Mitigation Options in Smallholder Agriculture*. Springer International Publishing, Cham, pp. 189–198. https://doi.org/10.1007/978-3-319-29794-1_10
- van Wijnen, H.J., Rutgers, M., Schouten, A.J., Mulder, C., de Zwart, D., Breure, A.M., 2012. How to calculate the spatial distribution of ecosystem services — Natural attenuation as example from The Netherlands. *Science of The Total Environment* 415, 49–55. <https://doi.org/10.1016/j.scitotenv.2011.05.058>
- Vaysse, K. 2015– Application et développement de méthodes de cartographie numérique des propriétés des sols à l'échelle régionale : cas du Languedoc-Roussillon. Thèse Montpellier SupAgro sous le sceau de l'école doctorale SIBAGHE, 190 p.
- Vaysse, K., Arrouays, D., McKenzie, N.J., Coste, S., Lagacherie, P., 2014. Estimation of GlobalSoilMap.net grids cells from legacy soil data at the regional scale in Southern France, in: *GlobalSoilMap: Basis of the Global Spatial Soil Information System*. CRC Press, pp. 133–138.

- Vaysse, K., Lagacherie, P., 2015. 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>
- Veronesi, F., Corstanje, R., Mayr, T., 2012. Mapping soil compaction in 3D with depth functions. *Soil and Tillage Research* 124, 111–118. <https://doi.org/10.1016/j.still.2012.05.009>
- Villa, F., Ceroni, M., Bagstad, K., Johnson, G., Krivov, S. 2009. ARIES (Artificial Intelligence for Ecosystem Services): A new tool for ecosystem services assessment, planning, and valuation. In 11th annual BIOECON conference on economic instruments to enhance the conservation and sustainable use of biodiversity, conference proceedings. Venice, Italy.
- Vincent, S., Lemerrier, B., Berthier, L., Walter, C., 2018. 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>
- Vrebos, D., Staes, J., Vandenbroucke, T., D'Haeyer, T., Johnston, R., Muhumuza, M., Kasabeke, C., Meire, P., 2015. Mapping ecosystem service flows with land cover scoring maps for data-scarce regions. *Ecosystem Services* 13, 28–40. <https://doi.org/10.1016/j.ecoser.2014.11.005>
- Wallace, K.J., 2007. Classification of ecosystem services: Problems and solutions. *Biological Conservation* 139, 235–246. <https://doi.org/10.1016/j.biocon.2007.07.015>
- Walter, C., 1991. Estimation de propriétés du sol et quantification de leur variabilité à moyenne échelle: cartographie pédologique et géostatistique dans le sud de l'Île et Vilaine (France). PhD Thesis. Ecole nationale supérieure agronomique de Rennes
- Walter, C., Bouedo, T., Arousseau, P. 1995 - Cartographie communale des teneurs en matière organique des sols bretons et analyse de leur évolution temporelle de 1980 à 1995. Rapport final. Conseil Régional de Bretagne - Agence Loire-Bretagne, 31 p
- Webster, R. and Oliver, M. 2007. *Geostatistics for Environmental Scientists*. John Wiley & Sons, New York. <http://dx.doi.org/10.1002/9780470517277>
- Wilding, L.P., Jones, R.B., Schafer, G.M., 1965. Variation of Soil Morphological Properties within Miami, Celina, and Crosby Mapping Units in West-Central Ohio. *Soil Science Society of America Journal* 29, 711. <https://doi.org/10.2136/sssaj1965.03615995002900060033x>
- Williams, A., Hedlund, K., 2013. Indicators of soil ecosystem services in conventional and organic arable fields along a gradient of landscape heterogeneity in southern Sweden. *Applied Soil Ecology* 65, 1–7. <https://doi.org/10.1016/j.apsoil.2012.12.019>
- Wischmeier, W.H. et Smith, D.D., 1978. Predicting Rainfall Erosion Losses. Handbook 537. US Department of Agriculture, Agricultural Research Service, Washington, DC.
- Wu, J., Feng, Z., Gao, Y., Peng, J., 2013. Hotspot and relationship identification in multiple landscape services: A case study on an area with intensive human activities. *Ecological Indicators* 29, 529–537. <https://doi.org/10.1016/j.ecolind.2013.01.037>
- Yang, L., Jiao, Y., Fahmy, S., Zhu, A.-X., Hann, S., Burt, J.E., Qi, F., 2011. Updating Conventional Soil Maps through Digital Soil Mapping. *Soil Science Society of America Journal* 75, 1044. <https://doi.org/10.2136/sssaj2010.0002>
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Finke, P., 2017. Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology* 285, 186–204. <https://doi.org/10.1016/j.geomorph.2017.02.015>
- Zhang, W., Ricketts, T.H., Kremen, C., Carney, K., Swinton, S.M., 2007. Ecosystem services and dis-services to agriculture. *Ecological Economics* 64, 253–260. <https://doi.org/10.1016/j.ecolecon.2007.02.024>
- Zheng, Z., Fu, B., Hu, H., Sun, G., 2014. A method to identify the variable ecosystem services relationship across time: a case study on Yanhe Basin, China. *Landscape Ecology* 29, 1689–1696. <https://doi.org/10.1007/s10980-014-0088-x>