



**UNIVERSITE KASDI MERBAH-OUARGLA**

**Faculté des sciences de la nature et de la vie**

**Département des Sciences Agronomiques**



**THESE**

**Pour l'Obtention du Diplôme de Doctorat ès Sciences**

**Spécialité : Sciences Agronomiques**

**Option : Sciences du sol**

**Analyse de l'organisation spatiale des sols dans un paysage aride par l'approche  
de la cartographie numérique du sol**

**Présentée et soutenue publiquement**

**Par**

**ASSAMI TAREK**

**Le 28/06/2020**

**Devant le jury composé de :**

|                                |                                         |                           |
|--------------------------------|-----------------------------------------|---------------------------|
| <b>M. CHELOUFI Hamid</b>       | <b>Professeur Université de Ouargla</b> | <b>Président</b>          |
| <b>M. HAMDI-AISSA Baelhadj</b> | <b>Professeur Université de Ouargla</b> | <b>Directeur de thèse</b> |
| <b>M. ABDESSELAM Salah</b>     | <b>Professeur Université de Batna</b>   | <b>Examineur</b>          |
| <b>M. BENMESSAOUD Hassen</b>   | <b>Professeur Université de Batna</b>   | <b>Examineur</b>          |
| <b>M. MOSTEPHAOUI Tewfik</b>   | <b>M.R.A. CRSTRA, Biskra</b>            | <b>Examineur</b>          |
| <b>M. BERKAL Ismaïel</b>       | <b>M.C.A. Université de Ouargla</b>     | <b>Examineur</b>          |

**ANNEE UNIVERSITAIRE : 2019-2020**

*Dédicaces*

*À la Mémoire de mes parents*

*À ma Famille*

### ***Remerciements***

Je tiens à exprimer mes sincères remerciements à mon directeur de thèse M. **HAMDIAISSA Baelhadj**, professeur à l'université de Ouargla, pour avoir accepté de diriger ce travail, pour sa disponibilité et ses directives qui m'ont permis de mener à terme ce travail.

Mes remerciements sont adressés à M. **CHELOUFI Hamid** professeur à l'université de Ouargla, qui m'a fait l'honneur de présider le jury. Je remercie également les membres du jury qui ont bien voulu examiner la thèse : M. **ABDESSELAM Salah** professeur à l'université de Batna, M. **BENMESSAOUD Hassen** professeur à l'université de Batna, M. **BERKAL Ismaïel** maître de conférences « A » à l'université de Ouargla et M. **MOSTEPHAOUI Tewfik** maître de recherche « A » et directeur de la division de recherche "gestion et valorisation de la ressource en sols en régions arides" au CRSTRA.

Ce travail a été domicilié au laboratoire de Biogéochimie des milieux désertiques de l'université de Ouargla. Je remercie tous ses membres en particulier le directeur du laboratoire **Pr. HADJ MAHAMMED Mahfoud**.

Mes vifs remerciements vont à mes collègues du CRSTRA, en particulier ceux du laboratoire central et du laboratoire de géomatique. Enfin, je tiens à remercier le personnel de la bibliothèque de l'Agence Nationale des Ressources Hydrauliques d'Alger de m'avoir fourni les documents relatifs à la zone d'étude et aussi les agriculteurs que j'ai rencontré sur le terrain.

## **Résumé**

Les sols constituent une ressource essentielle, rendant de nombreux services du point de vue sécurité alimentaire et environnementale. Aussi, pour une meilleure analyse de la variation spatiale des sols, la cartographie numérique du sol en particulier les modèles d'apprentissage machine offrent des nouvelles perspectives pour la cartographie des pédopaysage arides en Algérie, où le taux de couverture par les cartes pédologiques est faible.

Dans cette étude, nous avons comparé six algorithmes d'apprentissage machine pour modéliser et cartographier les sols de Zeb El Gherbi, dans la Wilaya de Biskra. Un total de 331 observations ponctuelles du sol classées selon la base de référence mondiale pour les ressources en sols et des covariables optimales ont été utilisées comme données d'entrée dans les modèles. La carte finale a été validée à partir de 135 données ponctuelles de sol.

Nos résultats ont montré que la variation des sols a été modélisée par une approche déterministe, selon le concept sol-paysage (SCORPAN). La forêt aléatoire a été le modèle le plus précis. Par ailleurs, le relief à travers l'élévation ainsi que le matériau parental, représenté par la covariable indice de taille de grain, ont été les principaux facteurs déterminants la distribution spatiale des sols à l'échelle du paysage. La précision du modèle forêt aléatoire a augmenté avec le niveau taxonomique le plus détaillé, c'est-à-dire les groupes de sol de référence avec les principaux qualificatifs. Les Haplic Petric Gypsisols, les Gleyic Solonchaks et les Haplic Gypsisols ont été les plus fréquents, couvrant 67,46% de la superficie totale de la région d'étude. Finalement, la méthodologie développée a permis de modéliser la relation entre les sols et les facteurs de pédogénèse, ce qui peut contribuer à accélérer la cartographie des sols dans les régions arides de l'Algérie.

## **Mots clés**

Cartographie numérique du sol, apprentissage machine, base de référence mondiale, paysage aride, Zeb El Gherbi, facteurs de pédogénèse, modélisation.

## **Abstract**

The soils are an essential resource, providing many services in terms of food and environmental security. Also, for a better analysis of spatial variation of soils, digital soil mapping, in particular, machine learning models offer a new perspective for mapping arid soil-landscapes in Algeria, where the coverage rate of soil maps is low.

In this study, we compared six machine learning algorithms to model and map soils in Zeb El Gherbi, Biskra Wilaya. A total of 331 soil point observations classified according to the world reference base for soil resources and optimal covariates were used as input data in the models. The final map was validated from data of 135 soil points.

Our results showed that soil variation was modelled by a deterministic approach, according to the soil-landscape concept (SCORPAN). Random forest was the most accurate model. Besides, the relief through elevation and parent material, represented by grain size index covariate, were the most important factors determining the spatial distribution of soils at the landscape scale. The accuracy of the random forest model increased with the most detailed taxonomic level, i.e. the reference soil groups with principal qualifiers. Haplic Petric Gypsisols, Gleyic Solonchaks and Haplic Gypsisols were the most common soils, representing 67.46% of the total area of the study region. Lastly, the methodology developed allowed modelling the relationship between soils and pedogenesis factors, which can help to accelerate soil mapping in Algeria's arid regions.

## **Keywords**

Digital soil mapping, machine learning, world reference base, arid landscape, Zeb El Gherbi, pedogenesis factors, modelling.

## ملخص

تعتبر التربة موردا أساسيا من حيث تقديمها للعديد من الخدمات وبالأخص في مجال الأمن الغذائي والبيئي. ولأجل الحصول على تحليل أفضل للتنظيم المكاني للتربة، تقدم تقنيات رسم الخرائط الرقمية وبالأخص التلقين الآلي مجالات جديدة لرسم خرائط المناظر الطبيعية للتربة القاحلة في الجزائر، حيث أن معدل التغطية لخرائط التربة منخفض بها.

تأسيسا على ما سبق قمنا في هذه الدراسة بمقارنة ست خوارزميات للتلقين الآلي لنمذجة ورسم خرائط تربة الزاب الغربي في ولاية بسكرة، حيث تم استخدام 331 نقطة من التربة وفقا للقاعدة المرجعية العالمية لموارد التربة، واستخدمت المتغيرات المشتركة المثلى كمدخلات للنماذج، وتم التحقق من صحة الخريطة النهائية انطلاقا من بيانات 135 نقطة من التربة.

أظهرت النتائج المتوصل إليها أنه على نطاق المناظر الطبيعية تم صياغة التغيرات في التربة وفقا للمقاربة الحتمية حسب مفهوم تضاريس-التربة (SCORPAN). وكانت الغابة العشوائية النموذج الأكثر دقة، من ناحية أخرى كانت التضاريس ممثلة في الارتفاع وتشكيل التربة السطحية الممثلة في مؤشر حجم الحبوب أهم العوامل الرئيسية في التوزيع المكاني للتربة، بالإضافة إلى ذلك كانت دقة نموذج الغابات العشوائية أعلى مع المستوى التصنيفي الأكثر تفصيلا، أي مجموعات التربة المرجعية مع التصفيات الرئيسية. وتعتبر أنواع التربة: جيبسيسول بيتريك هابليك، سولونساكز جلايك وجيبسيسول هابليك الأكثر شيوعا، حيث غطت 67.46٪ من إجمالي مساحة منطقة الدراسة.

في الأخير يمكن القول أن المنهجية المقدمة سمحت بنمذجة العلاقة بين التربة وعوامل تكوينها، الأمر الذي يؤدي إلى تسريع رسم خرائط التربة للمناطق القاحلة في الجزائر.

## الكلمات المفتاحية

رسم الخرائط الرقمية للتربة، التلقين الآلي، القاعدة المرجعية العالمية لموارد التربة، المناظر الطبيعية القاحلة، الزاب الغربي، عوامل تكوين التربة، النمذجة.

## **LISTE DES ABREVIATIONS**

ANRH : Agence Nationale des Ressources Hydrauliques

ASTER GDEM: Advanced Spaceborne Thermal Emission and Reflection Global Digital Elevation Model

BCT : Bagged Classification Tree

CNS : Cartographie Numérique du Sol

COST: Cosine of the Sun Zenith Angle

CPCS : Commission de Pédologie et de Cartographie des Sols

DPAT : Direction de la Planification et de l'Aménagement du Territoire

DEMRH : Direction des Etudes de Milieu et de la Recherche Hydraulique

GSI : Grain Size Index

MLP : Multilayer-Perceptron Neural Network

MNA : Modèle Numérique d'Altitude

Mrvbf: Multi-resolution index of valley bottom flatness

NNET: Single-Hidden-layer Neural Networks

ONM : Office National de la Météorologie

PA: Producer's Accuracy

RF: Random Forest

RFE: Recursive Feature Elimination

RSGs: Reference Soil Groups

SRTM: Shuttle Radar Topography Mission

SVML: Linear support vector machines

SVMR: Radial-basis Support Vector Machines

OA: Overall Accuracy

TCILow: Terrain Classification Index for Low lands

TRI: Terrain Ruggedness Index

TWI: Topographic Wetness Index

UA: User's Accuracy

WRB: World Reference Base

## LISTE DES FIGURES

|                                                                                                                                                            |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1. Fonctionnement des modèles d'apprentissage machine .....                                                                                         | 34 |
| Figure 2. Localisation de la région d'étude (a) et répartition des données ponctuelles de sol (b)<br>.....                                                 | 42 |
| Figure 3. Diagramme pluvio-thermique et bioclimatique .....                                                                                                | 45 |
| Figure 4. Carte géologique de la région de Zeb El Gherbi (Extrait de la carte géologique de Biskra<br>au 1/200 000. ANRH 1980).....                        | 46 |
| Figure 5. Le réseau hydrographique traversant la région d'étude délimitée en rouge .....                                                                   | 49 |
| Figure 6. Les indices spectraux calculés à partir de l'imagerie Landsat 5 TM. rat_7_1(a),<br>sali_indx (b), gsi_indx (c) et gyps_indx (d).....             | 54 |
| Figure 7. Les indices de courbure calculés à partir du MNA. Profil de courbure (a) et Plan de<br>courbure (b).....                                         | 55 |
| Figure 8. Les dérivés de terrain calculés à partir du MNA. Élévation (a), Pente (b), pente du bassin<br>versant (c) et Mrvbf (d) .....                     | 57 |
| Figure 9. Les dérivés de terrain calculés à partir du MNA. TRI (a), TCILow (b) et TWI (c)<br>.....                                                         | 58 |
| Figure 10. Histogrammes des données de l'élévation avant (a), et après transformation en log-<br>normale (b).....                                          | 59 |
| Figure 11. Le diagramme Q-Q de l'élévation avant (a) et après transformation en log-normale (b)<br>.....                                                   | 60 |
| Figure 12. Représentation d'un semi-variogramme théorique .....                                                                                            | 61 |
| Figure 13. Illustration des algorithmes de réseaux de neurones artificiels. Le perceptron multicouche<br>(a) et le réseau à couche unique cachée (b) ..... | 65 |
| Figure 14. Mode de fonctionnement de l'arbre de décision .....                                                                                             | 66 |
| Figure 15. Architecture de l'algorithme de forêt aléatoire .....                                                                                           | 68 |
| Figure 16. Principe de l'algorithme machines à vecteur de support .....                                                                                    | 70 |

|                                                                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 17. Représentation spatiale de la tendance de l'élévation .....                                                                                                 | 78 |
| Figure 18. Exemples de dérivés MNA avant (a et b) et après l'application du filtre spatial (a.1 et b.1).....                                                           | 80 |
| Figure 19. Les valeurs moyennes de kappa pour les deux niveaux taxonomiques de la WRB .....                                                                            | 81 |
| Figure 20. Importance des covariables utilisées par la forêt aléatoire au niveau des groupes de sols de référence avec qualificatifs principaux .....                  | 85 |
| Figure 21. Distribution des groupes de sols de référence avec qualificatifs principaux en fonction de l'élévation (a), indice de taille de grain (b) et Mrvbf (c)..... | 85 |
| Figure 22. Carte des classes de sol de Zeb El Gherbi modélisée par la forêt aléatoire au niveau des groupes de référence du sol avec qualificatifs principaux.....     | 86 |
| Figure 23. Carte de l'indice de confusion de Zeb El Gherbi au niveau des groupes de référence du sol avec qualificatifs principaux.....                                | 88 |

## LISTE DES TABLEAUX

|                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------|----|
| Tableau 1. Données climatiques de la station de Biskra durant la période de 1998 jusqu'à 2016 (Source ONM).....                   | 44 |
| Tableau 2. Données agricoles de la région de Zeb El Gherbi (DPAT, 2015) .....                                                     | 50 |
| Tableau 3. Nombre et fréquence des données de sol pour les deux niveaux taxonomiques de la base de référence mondiale (WRB) ..... | 77 |
| Tableau 4. Tendence globale et transformation des attributs du terrain.....                                                       | 78 |
| Tableau 5. Propriétés des semi-variogrammes utilisés.....                                                                         | 79 |
| Tableau 6. Ajustement de Bonferonni au niveau des groupes de sols de référence avec qualificatifs principaux .....                | 82 |
| Tableau 7. Ajustement de Bonferonni au niveau des groupes de sols de référence (RSGs) .....                                       | 82 |
| Tableau 8. Comparaison de la performance entre l'ensemble des covariables et celles sélectionnée par RFE.....                     | 83 |
| Tableau 9. Covariables sélectionnées par la fonction d'élimination de variables par récursion (RFE) .....                         | 84 |
| Tableau 10. Superficie couverte par les 13 classes de sol prédite par le modèle forêt aléatoire .....                             | 88 |
| Tableau 11. Matrice de confusion des 13 classes de sol utilisant des données de validation externe (N = 135).....                 | 89 |

# TABLE DES MATIÈRES

|                             |    |
|-----------------------------|----|
| INTRODUCTION GENERALE ..... | 16 |
|-----------------------------|----|

## PREMIERE PARTIE : REVUE DE LITTÉRATURE SUR LE SOL ET SPATIALISATION

|                                                                                         |           |
|-----------------------------------------------------------------------------------------|-----------|
| <b>1.1. Introduction .....</b>                                                          | <b>21</b> |
| <b>1.2. La variabilité spatiale des sols .....</b>                                      | <b>21</b> |
| 1.2.1 Approche déterministe .....                                                       | 21        |
| 1.2.2 Approche stochastique .....                                                       | 22        |
| 1.2.3 Approche hybride.....                                                             | 23        |
| <b>1.3. Les modèles d'organisation des sols dans le paysage.....</b>                    | <b>23</b> |
| 1.3.1 Concept de voisinage .....                                                        | 23        |
| 1.3.2 Concept de sol-paysage (Pédopaysage).....                                         | 24        |
| <b>1.4. Le rôle de l'échelle dans l'étude de la variabilité spatiale des sols .....</b> | <b>25</b> |
| <b>1.5. Les facteurs de formation des sols (facteurs de pédogenèse).....</b>            | <b>26</b> |
| 1.5.1 Le climat.....                                                                    | 26        |
| 1.5.2 Le relief .....                                                                   | 27        |
| 1.5.3 Le matériau parental.....                                                         | 28        |
| 1.5.4 Les organismes vivants.....                                                       | 29        |
| 1.5.5 L'âge .....                                                                       | 30        |
| <b>1.6. Représentation spatiale des sols.....</b>                                       | <b>30</b> |
| 1.6.1 La cartographie conventionnelle.....                                              | 30        |
| 1.6.2 La cartographie numérique du sol (CNS).....                                       | 31        |
| 1.6.3 Les méthodes de cartographie numérique du sol .....                               | 32        |

|                                                                |           |
|----------------------------------------------------------------|-----------|
| 1.6.3.1. Les méthodes supervisées.....                         | 32        |
| 1.6.3.2. Les méthodes non supervisées.....                     | 35        |
| 1.6.3.3. Les méthodes géostatistiques.....                     | 35        |
| 1.6.4. Pertinence de l'approche multi-échelle dans la CNS..... | 36        |
| <b>1.7. L'incertitude en cartographie.....</b>                 | <b>39</b> |
| 1.7.1. Les sources de l'incertitude.....                       | 39        |
| 1.7.2. La quantification de l'incertitude.....                 | 40        |

## **DEUXIEME PARTIE : MATERIELS ET METHODES**

|                                                          |           |
|----------------------------------------------------------|-----------|
| <b>2.1. Caractéristiques générales de la région.....</b> | <b>42</b> |
| 2.1.1. Situation géographique .....                      | 42        |
| 2.1.2. Le climat.....                                    | 43        |
| 2.1.2.1. La pluviosité .....                             | 43        |
| 2.1.2.2. La température .....                            | 43        |
| 2.1.2.3. Le vent.....                                    | 43        |
| 2.1.2.4. Synthèse climatique.....                        | 45        |
| 2.1.3. La géologie.....                                  | 45        |
| 2.1.4. La géomorphologie .....                           | 46        |
| 2.1.4.1. Les djebels .....                               | 47        |
| 2.1.4.2. Les glaci.....                                  | 47        |
| 2.1.4.3. La plaine et les terrasses alluviales.....      | 47        |
| 2.1.5. Les eaux souterraines .....                       | 47        |
| 2.1.5.1. La nappe phréatique .....                       | 47        |
| 2.1.5.2. La nappe albienne.....                          | 48        |

|                                                                                                   |           |
|---------------------------------------------------------------------------------------------------|-----------|
| 2.1.5.3. <i>La nappe calcaire</i> .....                                                           | 48        |
| 2.1.5.4. <i>La nappe du Mio-Pliocène</i> .....                                                    | 48        |
| 2.1.6. <i>Le réseau hydrographique</i> .....                                                      | 48        |
| 2.1.7. <i>La couverture végétale</i> .....                                                        | 49        |
| 2.1.8. <i>L'occupation du sol</i> .....                                                           | 49        |
| <b>2.2. Collecte et préparation des données des sols</b> .....                                    | <b>50</b> |
| 2.2.1. <i>Utilisation de la base de référence mondiale pour les ressources en Sols (WRB)</i> .... | 51        |
| <b>2.3. Facteurs environnementaux</b> .....                                                       | <b>52</b> |
| 2.3.1. <i>Collecte des covariables</i> .....                                                      | 52        |
| 2.3.1.1. <i>Images satellitaires</i> .....                                                        | 52        |
| 2.3.1.2. <i>Modèle numérique d'altitude (MNA)</i> .....                                           | 54        |
| 2.3.2. <i>Intégration de l'approche multi-échelle dans la CNS</i> .....                           | 58        |
| 2.3.2.1. <i>Exploration des données</i> .....                                                     | 59        |
| 2.3.2.2. <i>Analyse de la structure spatiale</i> .....                                            | 60        |
| 2.3.3. <i>Sélection des covariables optimales</i> .....                                           | 62        |
| <b>2.4. Modèles d'apprentissage machine</b> .....                                                 | <b>63</b> |
| 2.4.1. <i>Réseaux de neurones artificiels (ANN)</i> .....                                         | 63        |
| 2.4.2. <i>Arbres de décision</i> .....                                                            | 65        |
| 2.4.2.1. <i>Forêt aléatoire (RF)</i> .....                                                        | 66        |
| 2.4.2.2. <i>Bagged classification tree (BCT)</i> .....                                            | 68        |
| 2.4.3. <i>Machines à vecteur de support (SVM)</i> .....                                           | 69        |
| <b>2.5. Évaluation métrique des modèles</b> .....                                                 | <b>70</b> |
| <b>2.6. Signification statistique des différences observées entre les modèles</b> .....           | <b>71</b> |

|                                                   |           |
|---------------------------------------------------|-----------|
| <b>2.7. Évaluation des cartes modélisées.....</b> | <b>72</b> |
| <b>2.8. Estimation de l'incertitude .....</b>     | <b>72</b> |

## **TROISIEME PARTIE : RESULTATS ET DISCUSSION**

|                                                                                               |            |
|-----------------------------------------------------------------------------------------------|------------|
| <b>3.1. Résultats .....</b>                                                                   | <b>74</b>  |
| <i>3.1.1. Classes de sol .....</i>                                                            | <i>74</i>  |
| <i>3.1.2. Analyse multi-échelle des facteurs environnementaux .....</i>                       | <i>77</i>  |
| <i>3.1.3. Comparaison des modèles.....</i>                                                    | <i>80</i>  |
| <i>3.1.4. Sélection des principaux facteurs environnementaux.....</i>                         | <i>83</i>  |
| <i>3.1.5. Qualité de la carte modélisée.....</i>                                              | <i>86</i>  |
| <b>3.2. Discussion.....</b>                                                                   | <b>90</b>  |
| <i>3.2.1. Modèle optimal.....</i>                                                             | <i>90</i>  |
| <i>3.2.2. Variables environnementales contrôlant la distribution des classes de sol .....</i> | <i>90</i>  |
| <i>3.2.3. Estimation des classes de sol.....</i>                                              | <i>92</i>  |
| <i>3.2.4. Analyse spatiale des sols.....</i>                                                  | <i>95</i>  |
| <i>3.2.5. Importance de l'approche multi-échelle dans la modélisation des sols.....</i>       | <i>97</i>  |
| <i>3.2.6. Prédiction de l'incertitude .....</i>                                               | <i>98</i>  |
| <b>CONCLUSION GENERALE .....</b>                                                              | <b>100</b> |
| <b>RÉFÉRENCES BIBLIOGRAPHIQUES .....</b>                                                      | <b>103</b> |
| <b>ANNEXES .....</b>                                                                          | <b>117</b> |

## **INTRODUCTION GENERALE**

Au regard des enjeux environnementaux et de la sécurité alimentaire d'une population mondiale croissante, les régions arides, qui représentent 12,1% de la superficie mondiale (CNULD, 2011), subissent une pression environnementale et démographique pouvant avoir un impact négatif sur l'équilibre des terres. L'Algérie n'échappe pas à ces pressions qui s'exercent aussi sur les régions désertiques, qui représentent 90% du territoire nationale (FAO, 2005). De plus, face au défi de la sécurité alimentaire, l'Algérie mène une politique d'extension des terres agricoles au Sahara. Cependant, une meilleure gestion des territoires sahariens exige une connaissance de la nature et de l'organisation spatiale des sols. Il faut savoir que le sol présente une hétérogénéité à l'échelle du paysage et que la compréhension de sa distribution ne peut se réaliser sans étudier son organisation et sa relation avec les facteurs de pédogénèse (climat, matériau parental, relief, organismes vivants, temps,...), qui représentent les variables externes responsables de la formation des sols.

Les informations sur les sols des régions désertiques d'Algérie sont rares et le taux de couverture par les cartes pédologiques est extrêmement faible. Dans ces régions, le peu de travaux de cartographie réalisés ont été effectués à l'aide de la méthode conventionnelle, qui a été depuis longtemps un outil de premier choix pour obtenir de l'information sur la couverture pédologique. La carte conventionnelle des sols est élaborée à partir d'un modèle qualitatif, qui met en évidence la relation sol-paysage en se basant sur le travail de terrain et sur l'utilisation des cartes topographiques et des photographies aériennes. Actuellement, en Algérie les anciennes cartes pédologiques ne sont pas mises à jour et la production de nouvelles cartes est longue et coûteuse, en particulier lorsqu'on travaille sur de grandes superficies (Lagacherie et McBratney, 2006).

En réponse à ces contraintes qui affectent la carte pédologique conventionnelle, une nouvelle approche de cartographie quantitative initialement inspirée des travaux de Jenny (1941) a été

améliorée par McBratney et al. (2003) et qu'on définit par le terme de cartographie numérique du sol (CNS). Selon Minasny et al. (2013), la cartographie numérique du sol est une expression utilisée pour décrire des approches visant à cartographier les propriétés ou les classes de sols à l'aide de techniques numériques, en prenant en compte les divers aspects des caractères des sols dans un paysage.

L'usage de la télédétection, notamment à l'échelle régionale, apporte une dimension plus large à la cartographie des pédopaysages. Cela est vrai dans les régions arides où l'environnement est propice à l'application de la télédétection à cause de la faible couverture végétale et nuageuse. Les travaux de télédétection consacrés à la cartographie des pédopaysages arides ont été axés sur l'étude des états de surface des sols (Messedi et Delaître, 2007 ; Breunig et al., 2008) et sur la salinisation des sols (Ding et al., 2011 ; Peng et al., 2019).

Bien que les techniques de cartographie numérique du sol soient opérationnelles depuis quelques décennies dans de nombreux pays à travers le monde, en particulier les plus développés. En Algérie, les études qui utilisent ces techniques à l'échelle des paysages arides sont rares (Hamdi-Aïssa et Girard, 2000; Mostephaoui et al., 2013).

Plusieurs méthodes quantitatives, sous forme de modèles statistiques et mathématiques, sont utilisées dans la cartographie numérique du sol. Parmi ces méthodes, l'apprentissage machine est largement appliqué pour prédire les classes de sols (Hengl et al., 2017; Massawe et al., 2018; Ramcharan et al., 2018). Néanmoins, peu de chercheurs ont comparé plusieurs techniques d'apprentissage machine dans la même zone d'étude (Brungard et al., 2015 ; Heung et al., 2016; Levi, 2017; Taghizadeh-Mehrjardi et al., 2015).

Sous l'hypothèse que la couverture pédologique dans le paysage aride n'est pas déterminée par le hasard et qu'une relation quantitative est possible entre le sol et les facteurs contribuant à sa formation, nous avons voulu répondre aux questions suivantes ;

- Comment les sols varient et quels sont les modèles qui décrivent le mieux l'organisation spatiale des sols dans un paysage aride?
- Quels sont les facteurs de pédogénèse les plus déterminants dans cette distribution spatiale ?

Pour cela, l'idée principale est de tester les techniques de cartographie numérique du sol, notamment les modèles d'apprentissage machine dans un paysage aride, où nous avons supposé que l'organisation spatiale des sols est contrôlée par des facteurs de pédogénèse. Il est donc possible de modéliser et de cartographier les sols en formalisant le concept de sol-paysage par des modèles d'apprentissage machine.

C'est dans ce contexte qu'a été inscrite cette thèse en lien avec un projet de recherche à impact socio-économique intitulé : «Valorisation des données pédologiques des oasis du Bas Sahara » *Val-Péd-Oasis*. Ce projet est mené par le laboratoire de biogéochimie des milieux désertique (université de Ouargla) et a été financé par la Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT).

Les objectifs de cette thèse sont :

- (a) comparer les différentes méthodes d'apprentissage machine et identifier le modèle optimal pour décrire et cartographier la distribution des sols en utilisant la classification de base de référence mondiale (IUSS Working Group WRB, 2014) ;
- (b) identifier les principaux facteurs environnementaux, qui contrôlent la distribution spatiale des sols.

Ce travail se compose de trois parties :

- La première partie est consacrée à une synthèse bibliographique sur la variabilité spatiale des sols à l'échelle du paysage et à la cartographie des sols.
- Dans la deuxième partie, nous décrivons la zone d'étude, les données et les modèles d'apprentissage machine utilisés.
- La troisième partie est axée sur les résultats obtenus, ainsi que sur la discussion et l'interprétation des résultats.

**PREMIERE PARTIE : REVUE DE LITTÉRATURE SUR LE SOL ET  
SPATIALISATION**

## **1.1. Introduction**

Plusieurs études ont mis en évidence les différents aspects de la variabilité spatiale des sols et de leurs propriétés (Ettema et Wardle, 2002; Heuvelink et Webster, 2001; Lin et al., 2005). Ainsi, il semble que la très grande variation naturelle, peut parfois entraver ou empêcher la quantification de la variabilité des sols. En effet, Webster (2000) souligne que la complexité de la relation entre les sols et les facteurs de formation, la variation du sol en fonction du temps, et le caractère incomplet des connaissances sur la distribution du sol donnent l'impression que la variation du sol à certaines échelles est aléatoire. Alors que, les processus physico-chimiques qui aboutissent à la formation des sols obéissent à des lois déterministes pouvant être décrite par un langage mathématique.

## **1.2. La variabilité spatiale des sols**

La littérature scientifique décrit la variabilité spatiale des sols dans le paysage selon trois approches qui sont :

### *1.2.1. Approche déterministe*

Depuis longtemps, l'approche déterministe domine la pensée scientifique pour expliquer des phénomènes tel que la loi de la gravité, la relativité de l'espace temps et même l'évolution des espèces vivantes. Le principe du déterminisme nie toute existence du hasard dans un milieu physique parfaitement ordonné, il se base sur le principe que la succession des événements est due à l'effet de causalité. La pédologie n'échappe pas à cette conception d'expliquer la variabilité spatiale des sols selon le concept de causalité, elle a été imprégnée depuis longtemps par le principe du déterminisme. En effet, la recherche d'un mécanisme explicative de la dépendance des sols envers leur milieu a été présentée la première fois par Dokuchaev qui a observé un changement des sols en fonction du climat lors de ses voyages en Russie, il a établi un schéma liant le sol aux facteurs de l'environnement. Jenny (1941) a été parmi les premiers à appréhender le déterminisme spatial des sols de manière quantitative. Il a pris en

considération les facteurs de formation des sols comme des variables explicatives dans la modélisation de la distribution spatiale des sols. Depuis lors, de nombreux travaux dans le monde reflètent l'approche déterministe, qui lie le sol aux différents facteurs de l'environnement, au moyen de modèles mathématiques de plus en plus nombreux et complexes favorisés par la puissance des ordinateurs et la disponibilité des bases de données sur les sols. Par ailleurs, le principe du déterminisme est à l'origine de la pensée génétique, qui a fortement influencé la classification des sols, cette dernière reste une référence essentielle de la cartographie des types de sols (Pedro, 1989)

### *1.2.2. Approche stochastique*

Dans une nature très complexe, l'approche déterministe montre une certaine limite quand il s'agit de trouver la relation cause et effet. Selon, Walter (2002) les principaux inconvénients de l'approche déterministe sont :

- les déterminants des sols sont nombreux et interagissent entre eux, ce qui rend très difficile d'élucider les processus qui ont conduit à la formation des sols;
- la faible densité d'échantillonnage des sols conduit à une variabilité spatiale importante d'un site à un autre.

Face à une impression d'une complexité dans la distribution spatiale des sols, les méthodes stochastiques, qui prennent en considération l'aspect aléatoire des phénomènes, sont utilisées dans la modélisation de la distribution spatiale des sols et de leurs propriétés. Ainsi, les incertitudes associées à la cartographie sont généralement modélisées à l'aide de méthodes stochastiques. Par exemple, Park et al. (2007) et Li et Zhang. (2007) ont prédit les types de sol, en utilisant la chaîne de Markov. En plus, les méthodes géostatistiques permettent de prendre en considération la composante stochastique en supposant que la distribution spatiale des sols et de leurs propriétés est une réalisation d'un processus aléatoire.

### *1.2.3. Approche hybride*

L'approche hybride est basée sur la combinaison de la composante déterministe et stochastique, qui est utilisée pour expliquer l'aspect aléatoire de la variation spatiale des résidus du modèle (McBratney et al., 2000). Cette approche offre une grande puissance de prédiction spatiale, en particulier pour les études régionales et de bassin versant (Minasny et McBratney, 2016).

## **1.3. Les modèles d'organisation des sols dans le paysage**

### *1.3.1. Concept de voisinage*

La couverture pédologique est une composante majeure de l'écosystème terrestre montrant une structure dans sa distribution spatiale qui dépend d'une part, des facteurs de pédogénèse et d'autre part de la relation d'un type de sol par rapport à ses voisins. Pedro (1989) explique que dans certaines situations, lorsque la trajectoire gravitaire de l'eau peut être verticale et latérale, le sol ne peut être considéré en lui-même, par référence à la seule superposition verticale, il se définit toujours par rapport à ses voisins. De même, Lagacherie (2002) note que les lois de voisinage permettent d'estimer sur un site dépourvu d'observation du sol, une distribution de probabilité à partir d'un ensemble de sites voisins sur lesquels l'unité du sol est connue. Quantifier ces lois de voisinage implique une analyse comparative des valeurs que prennent les sites voisins et de vérifier leur proximité spatiale, on parle dans ce cas d'auto-corrélation spatiale. Parmi les méthodes permettant de déterminer la loi du voisinage, on peut citer :

- ➔ La méthode des plus proches voisins : elle consiste à sélectionner les points voisins en fonction de leur proximité du point d'étude. On définit un nombre de points les plus proches « k » qu'on désignera comme voisins.
- ➔ La distance d'auto-corrélation : elle permette de quantifier la corrélation spatiale et donc le voisinage entre les points. Les méthodes de krigeage utilisent le

semi-variogramme pour calculer la portée, qui est défini comme la distance à partir de laquelle la relation entre les points devient non structurée.

### 1.3.2. Concept de sol-paysage (Pédopaysage)

En science du sol, le concept du paysage indique que le sol est un milieu organisé dans les trois dimensions de l'espace et dans la dimension temporelle, et montre que les propriétés de sol et les processus pédologiques ne peuvent être compris lorsqu'ils sont isolés de leur contexte spatial et/ou temporel (Lacoste, 2012). L'objectif de la modélisation du pédopaysage est de comprendre la relation entre la distribution spatiale des sols et le paysage. Cette relation est traduite par un modèle qui peut être simplement qualitative (modèle mental) ou quantitative sous forme de modèles mathématiques. La modélisation sol-paysage tente d'intégrer les sols et les facteurs de formation du sol, qui sont le matériau parental, le relief, la végétation, l'âge, l'occupation du sol et les activités humaines. Hudson (1992) considère le concept sol-paysage comme la base scientifique fondamentale de l'étude des sols. Jenny (1941) a évalué la composante déterministe de la variabilité spatiale des sols à travers le paysage de manière quantitative. Il a pris en considération les facteurs de pédogenèse à savoir : le climat, les organismes vivants, le relief, le matériau parental et l'âge comme des variables explicatives de la distribution des sols. En plus, il a formalisé la relation entre les facteurs de pédogenèse et le sol sous forme d'équation :

$$S = f(cl, o, r, p, t) \quad (1)$$

Où S représente les sols ou ses attributs, (cl) climat, (o) organismes vivants, (r) relief, (p) matériau parental, (t) l'âge ou le temps.

Ensuite, cette formule a été améliorée par McBratney et al. (2003) en ajoutant la dimension spatiale et le sol comme des facteurs prédictifs. Le modèle conceptuel établi par McBratney et al. (2003) est appelé SCORPAN. Ce concept est utilisé pour quantifier les classes et/ou les propriétés du sol à partir des observations empiriques et de la connaissance des variables

environnementales. SCORPAN facilite également l'estimation de l'erreur ou de l'incertitude de la modélisation spatiale des classes ou des propriétés des sols (Soil Science Division Staff, 2017). Sept facteurs de formation du sol sont identifiés dans le modèle SCORPAN

$$S=f(s, c, o, r, p, a, n) \quad (2)$$

Où S, représente les classes ou les propriétés de sol, f : représente le modèle, autre propriétés du sol à un point donné (s), climat (c), organismes (o), reliefs (r), matériau parental (p), l'âge ou le temps (a) et la position spatiale (n).

#### **1.4. Le rôle de l'échelle dans l'étude de la variabilité spatiale des sols**

La notion d'échelle en science du sol est complexe, elle ne traduit pas simplement un rapport entre une distance mesurée sur la carte et la même distance sur le terrain (Legros, 1996), mais elle se réfère généralement à la dimension spatiale ou temporelle d'un phénomène. L'interaction entre la variabilité spatiale et le facteur d'échelle dépendent de la propriété du sol en question et les facteurs de formation des sols régissant le changement spatial. La variabilité spatiale peut être sur un pédon ou un segment limité de la topographie. Dans d'autres conditions, elle peut se produire sur des grandes distances correspondant aux unités cartographiques et géomorphologiques. À chaque niveau organisationnel, les critères qui caractérisent la variabilité spatiale des sols changent. Par exemple, l'hydromorphie analysée à l'échelle du pédon prend en considération la porosité, la présence ou l'absence d'un horizon imperméable et la texture, par contre le même phénomène étudié à l'échelle du paysage prend en considération d'autres facteurs tel que le relief, la géologie et la végétation.

De nombreux travaux ont étudié la relation entre l'échelle et la variabilité spatiale des sols. Par exemple, Lin et al. (2005) concluent, qu'il peut y avoir des variations importantes dans les propriétés du sol, comme la profondeur de l'horizon A et le pH à l'échelle métrique. Sur des échelles similaires, Franklin et Mills (2003) indiquent que l'hétérogénéité spatiale des propriétés du sol affecte la variabilité structurale de la communauté microbienne du sol. À

l'échelle du paysage, les facteurs de formation des sols influencent la pédogenèse à différentes échelles, de sorte que la distribution spatiale des sols et de leurs propriétés peuvent varier à différentes échelles et dans différentes directions (Kerry et Oliver, 2011).

### **1.5. Les facteurs de formation des sols (facteurs de pédogenèse)**

#### *1.5.1. Le climat*

La pluie et la température jouent un rôle important dans la formation des sols. En effet, les précipitations influent sur les réactions d'altération des minéraux et sur l'activité biologique dans le sol à travers la quantité d'eau stockée. D'un autre côté, l'eau de pluie permet le transport des matériaux en solution et en suspension dans le sol. Quant à la température, elle influence également les réactions chimiques et biologiques dans le sol. Ainsi, les réactions chimiques augmentent avec la hausse de la température. Par exemple, dans les climats arides, le gypse est plus susceptible de précipiter dans les sols. Alors que dans les régions froides, les basses températures ont un effet inhibiteur sur l'activité microbienne, qui entraîne une diminution de la vitesse de décomposition de la matière organique et son accumulation dans le sol.

Par ailleurs, le vent joue un rôle important dans l'érosion et l'accumulation du matériau parental, il influence la formation des dunes de sable dans les régions désertiques. Le rayonnement solaire est un facteur climatique important car, d'une part, il favorise la photosynthèse qui produit le carbone organique, et d'autre part il est une source de flux thermique dans le sol.

McBratney et al. (2003) indiquent que seulement 5% des études utilisent les facteurs climatiques comme covariables dans la cartographie numérique du sol. Cela est dû à la résolution spatiale grossière des données climatiques, qui sont généralement fournies sous forme raster avec une résolution d'un kilomètre. Seuls les projets de cartographie numérique du sol qui couvrent de vastes zones, c'est-à-dire à l'échelle nationale ou continentale, peuvent

utiliser ces données. Par exemple, Barthold et al. (2013) ont cartographié les types de sols de la steppe de Mongolie en se basant sur la température et les précipitations comme variables de modélisation. De même, Hengl et al. (2017) ont produit la carte numérique mondiale des classes de sols et de leurs propriétés (carbone organique, densité apparente, capacité d'échange cationique, pH, fractions de texture du sol et éléments grossiers) en utilisant des données de précipitation sous forme raster avec une résolution d'un kilomètre.

#### *1.5.2. Le relief*

Le relief est un facteur déterminant dans l'organisation spatiale des sols, quel que soit l'échelle spatiale (Jenny, 1941). Le relief joue un rôle important dans la circulation de l'eau dans le paysage, influençant ainsi le développement du sol (Moore et al., 1993). Les sols varient de façon continue le long d'une pente et constituent ce qu'on appelle les chaînes du sol. Ainsi, à travers les différents processus pédologique, il existe un lien génétique entre les sols du haut de la pente et ceux du bas de la pente. A titre d'exemple, sous l'influence de la pente, l'érosion produit des sols superficiels en amont, par contre en aval les sols sont plus profonds. Les sels solubles sont plus susceptibles de s'accumuler au bas de la pente, ce qui donne des sols plus salés. L'entraînement des particules fines sous l'effet de la pente produit un appauvrissement en éléments fins des sols en amont, alors que les sols situés en aval présentent une quantité importante d'argiles et de limons.

Les variables qui décrivent le relief, et qu'on désigne par le terme « attributs du terrain » peuvent être facilement extraites d'un modèle numérique d'altitude (MNA) disponible en format raster. Le MNA correspond à une grille composée d'un ensemble de cellules appelées pixels, chaque pixel se réfère à une valeur d'élévation déterminée. La résolution spatiale de MNA dépend de la taille des pixels qui varie de 30 m jusqu'à 90 m pour le SRTM (shuttle radar topography mission). Les données du relief sont utilisées dans 80% des études de

cartographie numérique du sol (McBratney et al., 2003). Il existe deux types d'attributs du terrain dérivés du MNA :

- ➔ Les dérivés primaires qui sont calculés directement à partir de données d'élévation, telles que la pente et l'aspect.
- ➔ Les dérivés secondaires qui sont des combinaisons des dérivés primaires, tel que les indices de courbure, qui décrivent la forme du terrain. Un autre dérivé secondaire couramment utilisée est l'indice topographique d'humidité, qui décrit la saturation potentielle en eau des sols (Moore et al., 1991).

### *1.5.3. Le matériau parental*

La structure à partir duquel les sols se forment est définie comme le matériau parental, ainsi les formations superficielles telles que les alluvions et les colluvions peuvent constituer le matériau d'origine pour de nombreux sols. Le matériau parental représente la composante géochimique du sol. En effet, les minéraux qui composent le matériau parental servent d'éléments nutritifs pour les plantes et traduit la fertilité du sol. Par exemple, les sols formés sur du calcaire sont enrichies en carbonates de calcium, alors que les sols acides formés sur du granite ont généralement une faible base cationique et une très faible quantité de silice. La dureté et la composition granulométrique du matériau parental influencent la vitesse de l'altération, cette dernière est plus importante sur des matériaux perméables et riches en minéraux altérables (Gaddas, 2001). La texture du sol est souvent liée au matériau parental, la texture grossière du sol provient des matériaux, qui ont des grains de quartz tels que les grès et les granites. Tandis que, la texture fine résulte de sédiments riches en argile et en limon, comme le schiste argileux.

Le matériau parental est considéré comme un facteur important dans l'étude des sols et de leurs propriétés (Gray et al., 2016). Cependant, 25% des publications sur la cartographie

numérique du sol utilisent le matériau parental comme variable explicative (McBratney et al., 2003). Cette faible utilisation est due, d'une part à la difficulté de l'observation directe depuis la surface terrestre (Zhang et al., 2017), et d'autre part à la disponibilité de l'information sur le matériau parental qu'à des échelles cartographiques grossières (Grunwald, 2009). L'information sur le matériau parental est généralement fournie par des cartes géologiques et des images satellitaires, qui fournies des informations sur la minéralogie et/ou sur les formations superficielles (Zhang et al., 2017).

#### *1.5.4. Les organismes vivants*

Les plantes modifient l'environnement du sol dans lequel elles se développent. Les éléments que les plantes incorporent dans leur biomasse sont ajoutés au sol, de sorte que l'horizon superficiel du sol s'enrichit en carbone organique, azote et autres éléments nutritifs. Aussi, le type de végétation peut être un indicateur des différents changements édaphiques du sol. Par exemple, les plantes halophytes (*Salsola vermiculata* et *Suaeda mollis*) indiquent le degré de salinité des sols. Par ailleurs, les effets de la faune et des microorganismes peuvent être observés sur les sols développés sous la forêt et ceux développés sous la steppe. Une grande partie de la matière organique des sols forestiers, qui provient du feuillage des arbres est transformée en différents types d'humus qui constituent la litière. Dans les sols des steppes, les racines et le couvert végétal de surface sont mélangés, grâce à la faune et aux microorganismes, avec la partie minérale et créent l'horizon A épais et sombre.

Dans la cartographie numérique du sol, l'information sur les organismes vivants provient essentiellement à partir de la couverture végétale, qui est fournie par des cartes de végétation et/ou des données satellitaires. La mesure de la végétation par télédétection repose sur la réflectance de l'énergie électromagnétique par les plantes. L'indice de différence normalisée de végétation (NDVI) est souvent utilisé pour représenter la relation entre la végétation et les sols (Mulder et al., 2011).

#### *1.5.5. L'âge*

Tous les facteurs de formation du sol doivent interagir pendant un certain temps pour produire le sol. En effet, la formation du sol est un processus très lent qui nécessite des milliers, voire des millions d'années. L'âge, en tant que facteur de formation du sol, s'est avéré difficile à caractériser, en dépit des techniques de mesure existantes tel que la datation avec du  $^{14}\text{C}$ . Ces méthodes mesurent le temps de manière ponctuelle, ce qui ne permet pas de produire des couvertures complètes d'une région comme la télédétection (McBratney et al., 2003). Les informations spatiales disponibles sur le temps ou l'âge de formation du sol sont en partie fournies par des cartes stratigraphiques, géologiques et géomorphologiques.

### **1.6. Représentation spatiale des sols**

La cartographie des sols consiste à délimiter les corps naturels des sols, à classer et regrouper les sols délimités en unités cartographiques et à collecter des informations sur les propriétés des sols afin d'interpréter et de représenter la distribution spatiale des sols sur une carte (Soil Science Division Staff, 2017).

#### *1.6.1. La cartographie conventionnelle*

La cartographie conventionnelle représente la variabilité spatiale de la couverture pédologique, elle repose en grande partie sur l'application manuelle des connaissances d'experts à travers le paradigme sol-paysage (Hudson, 1992). Dans la démarche de la cartographie classique des sols, l'information sur les différents facteurs de pédogenèse extraite des photographies aériennes, des cartes géologiques et topographiques est associée pour établir des hypothèses sur les règles de la distribution spatiale des sols.

En cartographie conventionnelle, le pédologue construit un modèle qualitatif (Hudson, 1992; Miller et Schaetzl, 2014), qui met en évidence la relation sol-paysage en se basant sur le travail de terrain et l'utilisation des photos aériennes sur lesquelles le sol est découpé sous forme de polygones (Yang et al., 2011). Depuis longtemps, la carte conventionnelle des sols

est restée un outil de premier choix pour obtenir une information sur la couverture pédologique. Néanmoins, elle présente un caractère statique, une imprécision, et elle découpe la couverture pédologique en polygones qui ne peuvent être combinés avec les formes raster, tels que le modèle numérique d'altitude et les images satellitaires (Hartemink et al., 2010). En plus, la carte conventionnelle des sols présente des difficultés à aborder la variabilité spatiale des sols d'une manière cohérente et précise (Brevik et al., 2016). Le modèle qualitatif qui aboutit à la création de la carte pédologique n'est pas clairement décrit (Jafari et al., 2012).

#### *1.6.2. La cartographie numérique du sol (CNS)*

La cartographie numérique du sol est définie comme la création et la diffusion de systèmes d'information spatiale sur les sols par l'utilisation des méthodes d'observation sur le terrain et en laboratoire, associées à des systèmes d'inférence spatiale et non spatiale des sols (Lagacherie et McBratney, 2006). Les progrès de la cartographie numérique du sol sont le fruit de plusieurs facteurs tels que : la disponibilité de données spatiales numériques (MNA, images satellitaires), la disponibilité de la puissance de calcul pour traiter de grandes données, le développement d'outils d'exploration de données et les progrès des systèmes d'information géographique, les applications des modèles statistiques et géostatistiques ainsi qu'une nouvelle génération de scientifiques des sols qui sont intéressés par l'analyse spatiale des sols (Minasny et McBratney, 2016).

La cartographie numérique du sol est objective, systématique, reproductible, actualisable et représente une expression optimale d'une compréhension statistiquement validée des relations sol-environnement en fonction des données actuellement disponibles (Hengl, 2018). Néanmoins, elle est basée sur l'utilisation d'une grande quantité de données sur le sol, qui nécessite des dépenses importantes sur le terrain en temps et en argent.

Les différences entre la cartographie conventionnelle des sols et la cartographie numérique du sol se rapportent principalement au degré d'utilisation des méthodes statistiques robustes pour

élaborer des modèles de prévision et le calcul de degré de précision de la carte à travers l'élaboration de la carte de l'incertitude (Minasny et McBratney, 2016).

En Algérie, la cartographie numérique du sol à l'échelle du paysage est peu utilisée. La plupart des études trouvées sont basées sur l'utilisation des images satellitaires, comme source d'information du couvert végétal, des états de surface du sol et des formations superficielles du sol. Douaoui et al. (2006) ont cartographié la salinité superficielle du sol de la plaine de Chlef en utilisant d'une part, les données de télédétection combinées avec les données de salinité, et d'autre part le modèle de régression-krigeage. Dans les régions du Sahara, la cartographie numérique du sol a été utilisée pour la cartographie des sols en appliquant la méthode de maximum de vraisemblance aux images satellitaires dans la région de Ouargla (Hamdi-Aïssa et Girard, 2000) et dans la région de Biskra (Mostephaoui et al., 2013 ; Assami et al., 2014).

### *1.6.3. Les méthodes de cartographie numérique du sol*

#### *1.6.3.1. Les méthodes supervisées*

Les méthodes supervisées sont basées sur une connaissance a priori de la relation sol-paysage. Pour ces méthodes, les valeurs utilisées comme données d'apprentissage, c'est-à-dire la variable d'entrée et la variable cible, sont connues. Ainsi, l'algorithme d'apprentissage supervisé apprend à partir des données de formation pour ajuster le modèle, une fois que le modèle devient performant, il peut être extrapolé à des endroits où la valeur à prédire est inconnue.

Parmi les méthodes supervisées, nous avons les méthodes d'apprentissage machine (machine learning) (Figure 1), qui ont pris de l'ampleur ces dernières années en raison, d'une part, de l'augmentation de la puissance de traitement des ordinateurs et, d'autre part, des sources des données telles que les images satellitaires, les appareils de mesures sur le terrain qui se basent pour la plupart sur des approches géophysiques (induction électromagnétique, conductivité

électrique, domaine de réflectométrie du temps, etc..) et le modèle numérique d'altitude. L'apprentissage machine est basé sur des méthodes statistiques, qui traduites sous forme de codes informatique permettent de construire des modèles quantitatifs aussi précis que possible. Les données sont essentielles pour l'apprentissage machine et doivent être de bonne qualité, c'est-à-dire nous devons éliminer toutes les informations erronées et/ou sans intérêt pour la modélisation, cette procédure est appelée nettoyage des données. Les modèles d'apprentissage machine utilisent une grande quantité de données (Somarathna et al., 2017).

Les modèles d'apprentissage machine sont caractérisés par deux types de variables :

- Les variables explicatives : appelées aussi prédicteurs, covariables ou variables prédictives, sont des variables à partir desquelles le modèle est construit. Pour l'étude des sols, ces prédicteurs correspondent généralement à des facteurs de formation des sols tels que le relief et la lithologie.
- Une variable cible : qui représente la variable à modéliser, dans notre cas, il s'agit des classes des sols.

La théorie de l'apprentissage machine est basée sur un ensemble de données de  $n$  points, chacun étant décrit par des covariables, fournies par  $X_i$ , qui sont modélisées par une fonction  $F(x)$  pour nous donner la valeur cible  $y$ . Pour chaque fonction proposée, nous devons évaluer le fonctionnement en examinant la valeur du risque prévu. Ceci est fourni par une fonction de perte. Ainsi, nous comparons différentes fonctions et recherchons celles qui nous donnent le risque minimum attendu. Parfois, l'algorithme d'apprentissage machine tend à minimiser sa fonction de perte en sur-ajustant les données. C'est pourquoi, après avoir appris une fonction basée sur les données du jeu d'apprentissage, cette fonction est validée sur un jeu de données de test, qui ne figurait pas dans l'ensemble d'apprentissage.

Selon Hastie et al. (2009), le but de l'apprentissage machine est de trouver une approximation optimale de la fonction  $F(x)$ , qui permet d'établir la relation prédictive entre les covariables

d'entrée et les résultats souhaités. L'apprentissage machine comprend des modèles linéaires (régression linéaire simple et multiple, analyse discriminante) et non linéaires (réseaux de neurones, forêt aléatoire). D'après Brungard et al. (2015), les modèles d'apprentissage machine peuvent être divisés en trois groupes : simple, modéré et complexe. Cette différence est due à la capacité d'interprétation des résultats des modèles et au nombre de paramètres de chaque modèle.

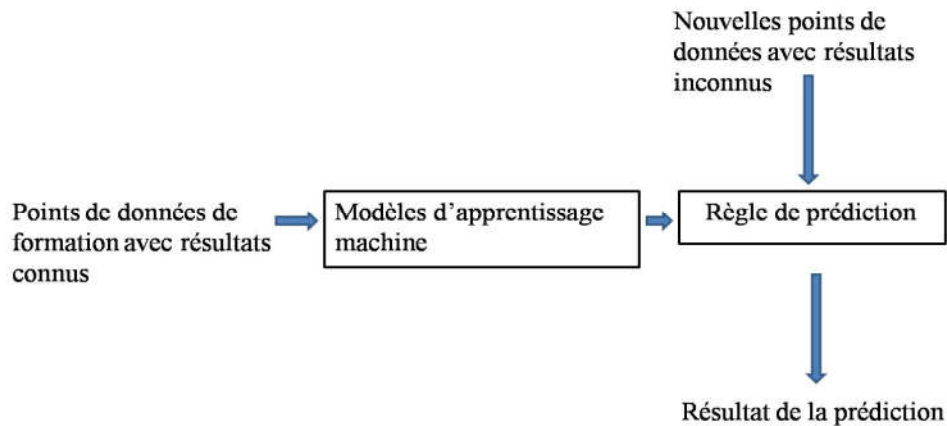


Figure 1. Fonctionnement des modèles d'apprentissage machine.

Pour l'analyse spatiale des sols, les différentes méthodes d'apprentissage machine utilisées sont classées en fonction du type de données. Dans le cas des données continues telles que la salinité du sol, le taux d'argile et le stock de carbone du sol, les méthodes de régression sont les plus utilisées. Pour les données discrètes (catégorielles) telles que les classes de sols, les classes texturale et les classes de drainage, les méthodes de classification sont les plus utilisées.

#### *1.6.3.2. Les méthodes non supervisées*

L'apprentissage non supervisé n'exige pas des données de formation, nous n'avons pas de connaissance a priori de nos résultats. L'algorithme découvre la fonction inhérente aux données et regroupe les pixels ayant des caractéristiques similaires en classes uniques basées sur des critères déterminés statistiquement (Duda et al., 2001). La définition des classes résultantes dépend uniquement des données représentant les covariables SCORPAN et de quelques paramètres définis au moment de l'exécution de la classification (Soil Science Division Staff, 2017). McBratney et al. (2003) décrivent la classification non supervisée « comme une première option pour réaliser une carte numérique du sol, lorsque nous n'avons pas une connaissance de la relation sol-paysage. Une fois que la carte des classes des sols a été établie sur les sites d'observation, des méthodes de classification supervisées devraient être utilisées pour ajuster le modèle et ensuite effectuer des prédictions à partir d'autres endroits où aucune propriété du sol n'a été observée ». Les méthodes ISODATA (Iterative Self-Organizing Data Analysis Technique) et k-means sont les algorithmes de classification non supervisés les plus couramment utilisés (Soil Science Division Staff, 2017).

#### *1.6.3.3. Les méthodes géostatistiques*

La géostatistique est basée sur la variable régionalisée, qui est une réalisation présentant une structure spatiale. La recherche d'éventuelle structure des données s'effectue par la reconnaissance de la structure spatiale à partir du semi-variogramme. Tandis que, l'estimation des valeurs dans des points non échantillonnés est effectuée par des méthodes d'interpolation telles que le krigeage. Les méthodes d'interpolation sont fondées sur des données spatialement localisées et sur l'hypothèse selon laquelle les valeurs des données les plus proches sont plus similaires que celles des données plus éloignées (Soil Science Division Staff, 2017). La majorité des méthodes géostatistiques utilisées en sciences du sol sont testées sur des données

continues (pH, la salinité, la granulométrie et le carbone organique, etc.....). Les méthodes de krigeage les plus simples sont le krigeage ordinaire et le krigeage universel. Cependant, il existe des variantes du krigeage, qui incorporent à la fois des composantes déterministes et stochastiques comme le cokrigeage et la régression-krigeage.

#### *1.6.4. Pertinence de l'approche multi-échelle dans la CNS*

Dans la cartographie numérique du sol, l'échelle spatiale se compose de trois entités (Malone et al., 2013) qui sont :

- ✓ L'extension : il s'agit de l'étendue de la surface ou de la couverture du domaine cartographique tel qu'une carte du monde, un pays, une région, ou une parcelle.
- ✓ La résolution : est la taille de l'espacement du pixel du raster. Une image raster composée de pixels ayant chacun une superficie de 10 sur 10 mètres au sol a une résolution de 10 mètres. plus la résolution est grande, plus l'image raster est précise
- ✓ Le support : est l'entité sur laquelle l'information est définie, il peut être un volume, une surface ou un point. Par exemple, un profil du sol correspond à un support représenté par un point.

L'importance de l'échelle spatiale dans la cartographie numérique du sol a pris de l'ampleur, en particulier avec l'introduction du modèle numérique d'altitude (MNA), qui est considéré comme une source d'information sur la topographie. Selon Cavazzi et al. (2013), l'échelle du MNA façonne le résultat final de l'analyse du CNS, indiquant qu'une meilleure compréhension et connaissance quantitative de l'échelle aide à améliorer la modélisation du sol. Toutefois, le choix de l'échelle du MNA qui doit être intégrée dans l'analyse du CNS reste posé. Les études menées sur l'importance de l'échelle spatiale dans la modélisation des sols indiquent, d'une part, que l'utilisation du MNA avec une résolution très fine ne semble pas améliorer la qualité de la modélisation des pédopaysages (Thompson et al., 2001) puisque les attributs du terrain présentent des informations très détaillées, ce qui introduit des erreurs

et donnent un modèle imprécis (Cavazi et al., 2013; Roecker et al., 2010), et d'autre part, l'augmentation de la résolution du MNA détériore la qualité de la modélisation des sols (Chaplot et al., 2000). Une approche alternative est proposée pour améliorer la qualité du modèle, qui consiste à modifier la taille de la fenêtre du MNA en appliquant un filtre spatial de convolution, qui fait intervenir le voisinage du pixel. Par exemple, Smith et al. (2006) suggèrent qu'au lieu d'utiliser des données du MNA avec des résolutions fines, il est plus intéressant de faire correspondre les caractéristiques du terrain issu du MNA en utilisant la taille de la fenêtre adéquate. Roecker et Thompson (2010) montrent qu'en variant la taille de la fenêtre du MNA de manière qu'elle prend en considération l'échelle du processus du sol, tout en maintenant une résolution constante, à un impact plus important que l'augmentation de la résolution des pixels, qui présente un risque de perte d'information. De même, Zhu et al. (2008) indiquent que la résolution du MNA n'est pas aussi importante que la taille de la fenêtre dans le calcul des dérivés de terrain du MNA. Maynard et Johnson (2014) concluent que la corrélation optimale entre les propriétés du sol et les attributs du terrain peut être réalisée en utilisant un MNA à résolution modérée avec un ajustement de la taille du voisinage pour correspondre à l'échelle de la propriété à modéliser. D'un autre côté, Shi et al. (2012) ont appliqué deux types de voisinage de forme circulaire et quadratique pour calculer la taille de la fenêtre sur la variable pente, ils concluent que la forme circulaire est plus performante.

Les facteurs de pédogenèse en tant que forces agissantes sur la formation des sols montrent une variabilité locale et régionale au niveau des paysages. De plus, les interactions des facteurs de pédogenèse, qui se manifestent à plusieurs échelles influencent la genèse et la dépendance spatiale du sol. Behrens et al. (2010) expliquent que lors de l'extrapolation des classes des sols ou des propriétés des sols par les méthodes de CNS, il est important d'intégrer les informations sur les facteurs de formation du sol non seulement au point où la classe du

sol et les valeurs de propriétés du sol sont mesurées, mais aussi dans un contexte spatial plus large. En effet, de nombreuses études montrent que l'intégration du contexte spatial à différentes échelles (l'aspect multi-échelle) en combinant des données topographiques dérivées de filtres spatiaux avec des données topographiques brutes peut améliorer la précision globale de la modélisation des sols et de leurs propriétés (Behrens et al., 2010; Grinand et al., 2008; Moran et Bui, 2002).

Globalement, deux démarches permettent d'intégrer le contexte spatial en considérant l'aspect multi-échelle de l'organisation des sols dans le paysage :

1. Le changement de la résolution du pixel : cette approche peut entraîner des erreurs dans le modèle lorsque la résolution du pixel du MNA est augmentée (Behrens et al., 2010). Bien que cette approche permette de calculer des attributs du terrain à l'échelle spatiale à laquelle les propriétés du sol varient, le changement de la résolution du pixel peut affecter de manière négative la modélisation de la relation sol-paysage.
2. Le changement de la taille de la fenêtre : par l'utilisation d'un filtre spatial de convolution qui permet de prendre en considération la relation spatiale des pixels voisins (opération de voisinage). Le filtre spatial consiste à appliquer une fenêtre flottante sur le raster en remplaçant la valeur de chaque pixel du raster d'origine par le résultat d'une transformation sur le pixel et ses voisins, le filtre est défini par la taille et le type de la fenêtre flottante. Une taille optimale de la fenêtre est recherchée, elle correspond à la distance entre les pixels qui permette d'une part de connaître la longueur à partir de laquelle il n'y a aucune relation spatiale entre les pixels, et d'autre part de conserver les détails significatifs sur le terrain. Cette distance est déterminée soit par la méthode géostatistique en se basant sur l'analyse du semi-variogramme local dont la portée correspond à la taille de la fenêtre (Moran et Bui, 2002), soit par itération en utilisant les attributs du terrain avec des fenêtres de différentes tailles

(Grinand et al., 2008 ; Behrens et al., 2010). Cette dernière méthode reste une approche simple puisque la détermination de la taille de la fenêtre ne repose sur aucune condition, contrairement à l'approche utilisant le semi-variogramme, qui permet de mesurer la variabilité spatiale des covariables et d'ajuster la taille de la fenêtre (Moran et Bui, 2002; Grinand et al., 2008).

### **1.7. L'incertitude en cartographie**

La cartographie des sols consiste à extrapoler l'information dans des zones où aucune mesure du sol n'est observée. Cela peut produire des erreurs causées par l'incertitude des véritables classes du sol dans ces endroits non observés. L'incertitude est due à la complexité de la variabilité spatiale des sols. Ainsi, l'évaluation de l'incertitude en cartographie est importante pour la prise de décision et les analyses de risque.

#### *1.7.1. Les sources de l'incertitude*

Il existe de nombreuses sources d'incertitude dans la cartographie numérique du sol :

- ✓ Les estimations sur le terrain peuvent être subjectives lorsqu'il s'agit d'estimer, par exemple, la couleur ou la texture du sol. De même, un mauvais échantillonnage sur le terrain peut biaiser les mesures effectuées sur le sol. Par ailleurs, les pédologues peuvent être en désaccord sur la taxonomie des sols lors de la classification ;
- ✓ Une lecture et/ou une manipulation incorrecte par le personnel du laboratoire entraîne des erreurs dans la mesure des propriétés du sol. De plus, un mauvais étalonnage de l'équipement de mesure a une incidence sur les résultats des mesures des propriétés du sol ;
- ✓ L'absence ou la mauvaise localisation des coordonnées géographiques du solum contribue également à accroître l'incertitude ;

- ✓ Les variables explicatives peuvent également comporter des incertitudes. Par exemple, un modèle numérique d'altitude, qui est une approximation de l'élévation réelle comporte une incertitude verticale liée au type du MNA (SRTM, ASTER GDEM, ect...). Il en résulte un effet de propagation de l'incertitude vers les attributs du terrain extraits du MNA, tels que la pente ou l'indice d'humidité topographique ;
- ✓ Même si nous savons que la distribution spatiale des sols dans le paysage peut être formalisée selon des modèles mathématiques et mise en œuvre dans un langage informatique, il reste que les modèles sont des représentations simplifiées d'un monde réel beaucoup plus complexe, par conséquent ces modèles introduisent aussi des incertitudes.

#### *1.7.2. La quantification de l'incertitude*

Dans la cartographie numérique du sol, l'incertitude peut être quantifiée par :

1- La probabilité de distribution des classes de sol, qui indique la similarité de l'occurrence des classes de sol dans chaque pixel du raster, c'est le cas de l'indice de confusion (Burrough et al., 1997).

2- Les intervalles de prédiction qui indiquent la plage de valeurs dans laquelle la valeur réelle des propriétés des sols est susceptible de se produire (Malone et al., 2011). La cartographie numérique du sol utilise le plus souvent des intervalles de 90%, qui indiquent la plage de valeurs dans laquelle une nouvelle mesure sera trouvée 9 fois sur 10 (Soil Science Division Staff. 2017). Les intervalles de l'incertitude sont le plus souvent montrés sous forme de cartes affichés côte à côte, et qui représente l'intervalle de modélisation inférieur à 90%, la moyenne et l'intervalle de modélisation supérieur à 90%.

## **DEUXIEME PARTIE : MATERIELS ET METHODES**

## 2.1. Caractéristiques générales de la région d'étude

### 2.1.1 Situation géographique

La région d'étude est localisée dans le Zeb El Gherbi à 25 km de la ville de Biskra et au sud-est de l'Algérie (Figures 2a et 2b). Elle est située entre les longitudes 5°16'E et 5°30'E et les latitudes 34°46'N et 34°38'N, et couvre une superficie d'environ de 22573.35 ha. La région est délimitée au nord par le mont de Zeb, au sud par l'oued Djedi, le canal de M'lili forme la frontière orientale et à l'ouest par l'oued Fellag et l'oued El Ouzene. Le choix de la zone est lié à la disponibilité d'une quantité importante de données pédologiques et à la présence de sols représentatifs des régions arides.

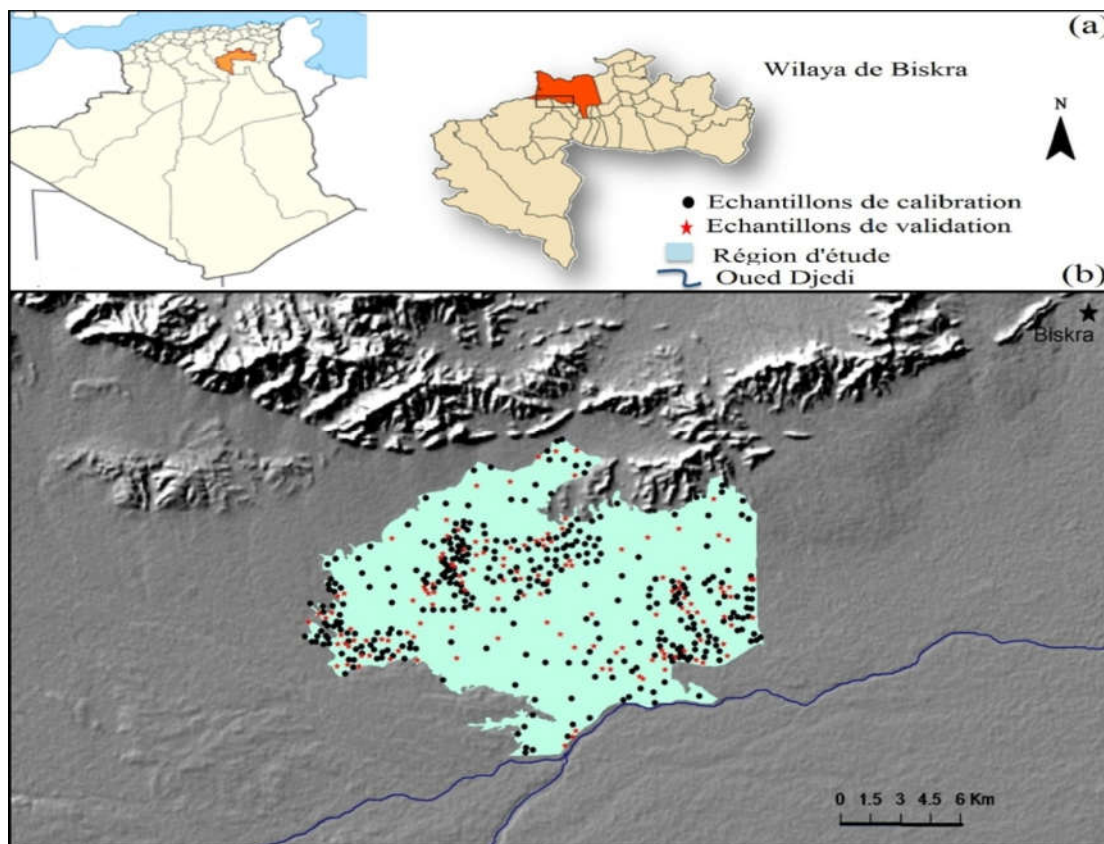


Figure 2. Localisation de la région d'étude (a) et répartition des données ponctuelles de sol (b).

### *2.1.2. Le climat*

#### *2.1.2.1. La pluviosité*

Dans la région, les pluies sont insuffisantes, irrégulières et inégalement réparties tout au long de l'année. Les précipitations ont atteint une moyenne de 130 mm entre 1998 et 2016. L'année 2004 a été la plus pluvieuse avec une moyenne de 297 mm. Tandis que, l'année 2002 a été la moins pluvieuse avec une moyenne de 49 mm (Tableau 1). La pluie est pratiquement nulle pendant la période estivale, c'est-à-dire juin, juillet et août.

#### *2.1.2.2. La température*

Le tableau 1 montre que la température moyenne est de 22,9°C entre 1998 et 2016. La température la plus élevée est de 24,7°C enregistrée en 2002. Par contre, la température la plus basse est de 22,2 °C enregistrée dans l'année 1998 et 2004. Les températures maximales moyennes sont enregistrées durant les mois de juillet et août, tandis que les températures minimales sont enregistrées pendant la saison hivernale.

#### *2.1.2.3. Le vent*

Les vents froids et humides du nord-ouest dominant pendant la période hivernale. Alors, qu'en été les vents secs et chauds du sud sont plus fréquents. Le tableau 1 indique que le vent est relativement fréquent durant toute l'année, avec une vitesse annuelle moyenne de 4,2 m/s pendant la période de 1998 jusqu'à 2016.

Tableau 1. Données climatiques de la station de Biskra durant la période de 1998 jusqu'à 2016 (Source ONM).

| <b>Année</b>                | <b>Pluviométrie cumulée (mm)</b> | <b>Température moyenne (°C)</b> | <b>Vitesse moyenne du vent (m/s)</b> |
|-----------------------------|----------------------------------|---------------------------------|--------------------------------------|
| 1998                        | 92                               | 22.2                            | 5.1                                  |
| 1999                        | 191                              | 23.2                            | 5.2                                  |
| 2000                        | 64                               | 22.5                            | 4.8                                  |
| 2001                        | 88                               | 22.9                            | 5.3                                  |
| 2002                        | 49                               | 24.7                            | 4.9                                  |
| 2003                        | 125                              | 23.8                            | 4.2                                  |
| 2004                        | 297                              | 22.2                            | 3.9                                  |
| 2005                        | 60                               | 22.5                            | 3.8                                  |
| 2006                        | 173                              | 22.8                            | 4.3                                  |
| 2007                        | 79                               | 22.9                            | 4.7                                  |
| 2008                        | 119                              | 22.7                            | 4.2                                  |
| 2009                        | 171                              | 22.8                            | 3.9                                  |
| 2010                        | 188                              | 22.5                            | 3.9                                  |
| 2011                        | 252                              | 22.5                            | 3.5                                  |
| 2012                        | 56                               | 23.2                            | 3.2                                  |
| 2013                        | 144                              | 22.7                            | 3.8                                  |
| 2014                        | 70                               | 23.4                            | 3.6                                  |
| 2015                        | 105                              | 22.8                            | 4                                    |
| 2016                        | 156                              | 23.5                            | 3.9                                  |
| <b>Pluviométrie Moyenne</b> | 130                              | 22.9                            | 4.2                                  |

#### 2.1.2.4. Synthèse climatique

À partir des données climatiques de la région, nous avons établi une synthèse climatique, qui est représentée par le diagramme pluvio-thermique et bioclimatique réalisé par Pouget (1980). Ce diagramme est une adaptation du quotient d'Emberger pour l'Algérie. La figure 3 indique que la région de Biskra est caractérisée par un étage bioclimatique aride inférieur à hiver doux.

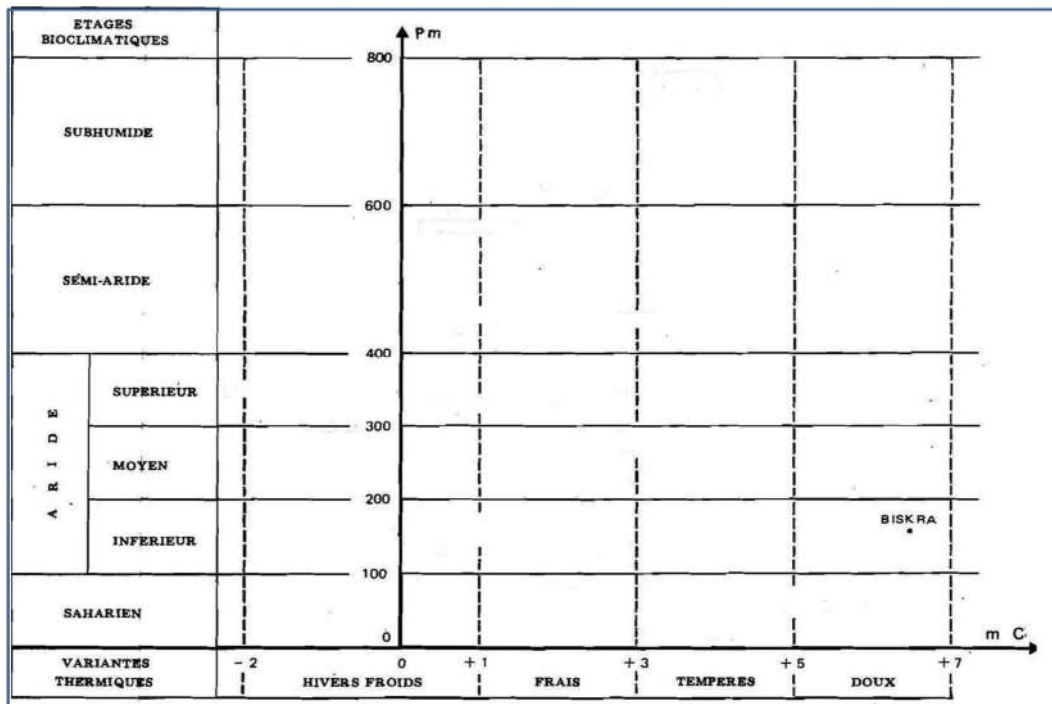


Figure 3. Diagramme pluvio-thermique et bioclimatique.

#### 2.1.3. La géologie

D'un point de vue géologique et lithologique (Figure 4), la région d'étude est caractérisée par:

- ✓ Le Secondaire : localisé essentiellement sur le mont de Zeb, il est constitué d'affleurements de roche calcaire et de calcaire marneux de la période Maestrichtien-Campanien et Santonien-Coniacien.

- ✓ Le Tertiaire : au nord, nous avons des calcaires massifs à silex de l'Eocène inférieur. Dans le sud nous avons observé sur une butte témoin, la présence de couches de sable rouge et de gypse du Pliocène supérieur.
- ✓ Le Quaternaire : est le plus dominant dans la région. Au nord, il est constitué par des éboulis sur les versants et par la présence d'une croûte continue calcaire-gypseuse. Au centre et au sud, le substrat est dominé par des alluvions qui contiennent des matériaux sableux.

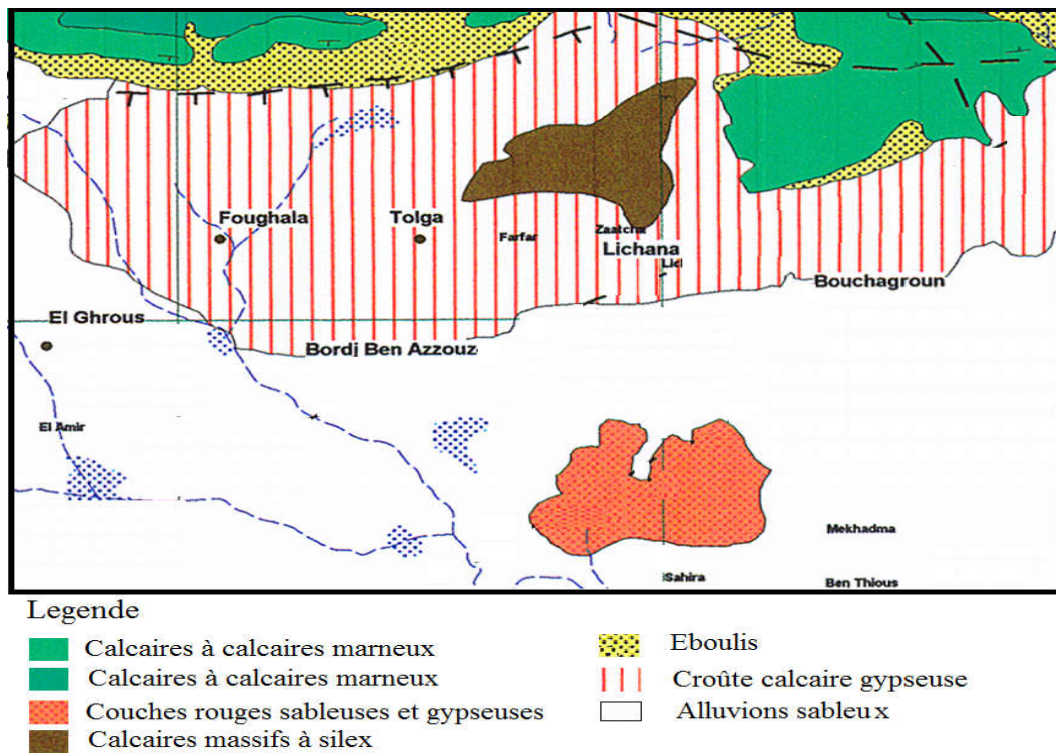


Figure 4. Carte géologique de la région de Zeb El Gherbi (Extrait de la carte géologique de Biskra au 1/200 000. ANRH 1980).

#### 2.1.4. La géomorphologie

La région d'étude est caractérisée par une altitude qui varie entre 81 et 196 m, elle est constituée par un ensemble morphologique formé par :

#### *2.1.4.1. Les djebels*

Situés au nord de la région d'étude, la hauteur des djebels varie de 735 m pour le djebel Aroussine, 714 m pour le djebel Gouara, 596 m pour le djebel Oum Djenib, et 357 m pour le djebel Mennchar.

#### *2.1.4.2. Les glacis*

Au nord, le glacis se présente comme une large surface concave, qui prend naissance en bas du mont de Zeb et s'étend jusqu'à la plaine alluviale. Le glacis est formé pendant le quaternaire et constitué d'une pente douce, qui décroît vers le sud. Au sud, quelques collines à encroûtement gypseux émergent et constituent des buttes témoins caractérisées par un encroûtement avec une couverture caillouteuse. Ces buttes témoins sont les indicateurs d'un long processus d'érosion que la région a traversé.

#### *2.1.4.3. La plaine et les terrasses alluviales*

Dans l'ensemble, c'est le plat qui caractérise la plaine avec une très légère pente suivant la direction nord-sud. La plaine occupe le fond d'un bassin de sédimentation où s'accumulent les matériaux alluvionnaires. Tout au long du quaternaire récent, l'oued Djedi a déposé ses alluvions sous forme de terrasses, dont certaines ont disparu sous l'effet de l'érosion hydrique provoquée par l'oued pendant son débordement. Les quelques traces de terrasses qui restent ont montré une alternance de niveaux d'argile et de sable.

#### *2.1.5. Les eaux souterraines*

##### *2.1.5.1. La nappe phréatique*

La nappe est présente dans des accumulations alluvionnaires, où elle est parfois peu profonde, entre 0 à 10 m. Elle est alimentée par les pluies, les eaux des oueds et l'irrigation. Du point de vue de la qualité, l'eau de la nappe est médiocre avec un degré élevé de salinité.

#### *2.1.5.2. La nappe albiennne*

Cette nappe est rarement utilisée dans la région, elle est située dans de grandes profondeurs allant de 1700 à 2600 mètres, et se caractérise par une température très élevée.

#### *2.1.5.3. La nappe calcaire*

Appelée aussi « nappe de Tolga », elle est la plus exploitée en raison de sa profondeur relativement faible allant de 150 à 200 m. La nappe calcaire contient une grande réserve en eaux liée, d'une part à l'état de fissuration de la roche qui permet le captage des eaux superficielles, et d'autre part à son alimentation en eaux à partir de l'atlas Saharien. Actuellement, la nappe subit une baisse en raison de la surexploitation.

#### *2.1.5.4. La nappe du Mio-Pliocène*

Cette nappe a une extension considérable, sa profondeur varie entre 90 et 250 mètres. Quant à son épaisseur, elle reste faible sur les versants et augmente au niveau de la plaine où elle atteint 100 mètres.

#### *2.1.6. Le réseau hydrographique*

Les nombreux oueds qui traversent la région d'étude sont marqués par un écoulement épisodique, qui ne peut se produire qu'en cas de précipitations. La nature des formations superficielles combinée à un relief légèrement ondulé ont contribué à la formation d'un réseau hydrographique diffus (Figure 5). Ce type de réseau est marqué par des incisions en forme de ravins, qui s'exercent généralement sur le matériel meuble du glacis. Dans le sud de la région, l'oued Djedi coule périodiquement pendant la saison des pluies.

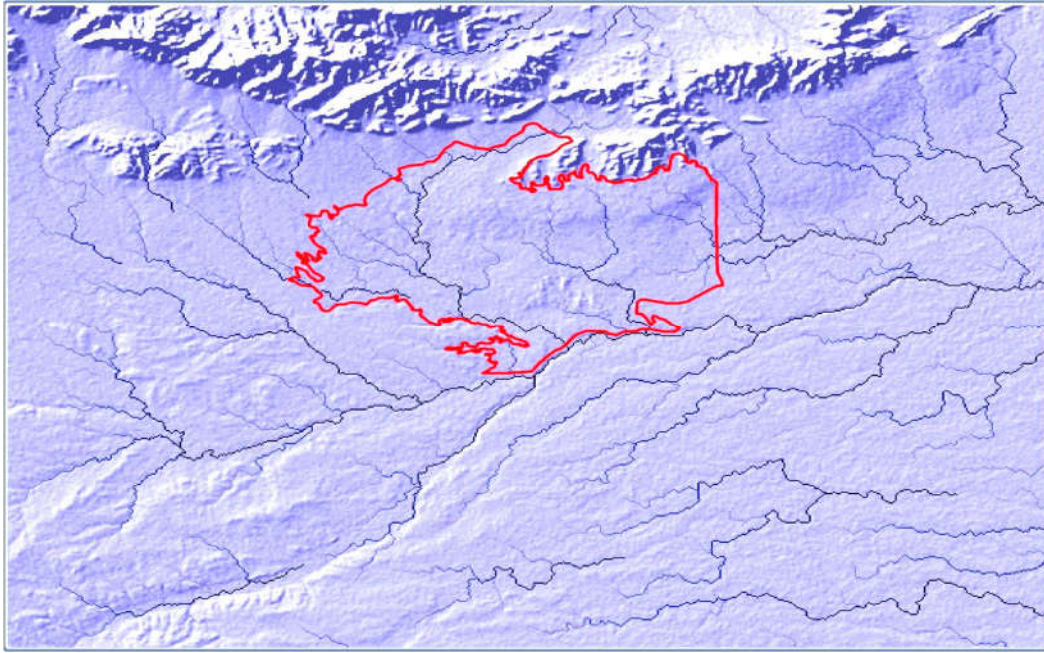


Figure 5. Le réseau hydrographique traversant la région d'étude délimitée en rouge.

#### 2.1.7. La couverture végétale

La couverture végétale est peu dense avec une dominance des plantes à strate herbacée. Au niveau des glaciers et des collines encroûtées, nous avons observé des plantes caractéristiques des sols gypseux (*Astragalus armatus*, *Limoniastrum guyonianum*, *Zygophyllum album*). Dans les zones plates, nous avons une association de plantes spécifiques à l'écosystème salin (*Suaeda Mollis*, *Salsola Vermiculata* et *Atriplex Halimus*) et hydromorphe (*Tamarix Articulata*).

#### 2.1.8. L'occupation du sol

La zone est dominée par des oasis, où le palmier dattier de type Deglet Nour occupe la première place. Les oasis de la région sont caractérisées par une répartition spatiale en deux lignes (Aidaoui, 1994). Une première ligne est constituée de cinq oasis allant de Bouchagroun à Laghroun au pied du massif du Zeb et une deuxième ligne est constituée des oasis de Mekhadama, Shira et Lioua qui sont installées sur des anciennes terrasses de l'oued Djedi. La

superficie agricole utile de toute la région de Zeb El Gherbi est de 25 750 hectares (Tableau 2). En plus de la phéoniculture, d'autres types de cultures sont valorisés et concernent essentiellement :

- La culture céréalière dont la superficie est de 580 hectares est basée sur l'orge et le blé ;
- La culture maraîchère d'une superficie de 2845 hectares, qui comprend les cultures en serre et en plein champ (Tableau 2).

Tableau 2. Données agricoles de la région de Zeb El Gherbi (DPAT, 2015).

| Communes          | Surface agricole utile (SAU) en hectare | Nombre de palmiers dattiers | Céréalicultures en hectare | Maraichers en hectares | Autres cultures |
|-------------------|-----------------------------------------|-----------------------------|----------------------------|------------------------|-----------------|
| Tolga             | 4384                                    | 301830                      | 235                        | 673                    | 27              |
| Bou- Chagroun     | 1845                                    | 110040                      | 0                          | 182                    | 0               |
| Bordj Ben- Azzouz | 1666                                    | 143020                      | 0                          | 31                     | 0               |
| Lichana           | 2122                                    | 140200                      | 0                          | 459                    | 0               |
| Foughala          | 1799                                    | 118700                      | 0                          | 168                    | 0               |
| Laghrou           | 7030                                    | 177160                      | 25                         | 816.38                 | 33.35           |
| Mekhadma          | 2409                                    | 204377                      | 140                        | 537                    | 35              |
| Lioua             | 4495                                    | 247012                      | 180                        | 795                    | 65              |
| Total             | 25750                                   | 1442339                     | 580                        | 2845                   | 127             |

## 2.2. Collecte et préparation des données des sols

Plusieurs études agro-pédologiques ont été réalisées dans la région de Zeb El Gherbi, qui est considérée comme une zone à fort potentiel agricole en particulier pour le palmier dattier. Nous avons recueilli et organisé toutes les données sur le sol (observations à la tarière et

solums) obtenues à partir des observations effectuées sur le terrain et des études pédologiques réalisées dans la région entre 1972 et 2006 (ANRH, 2006 ; DEMRH, 1972, 1973). Nous avons transformé une partie des données géoréférencées de l'ancienne projection Lambert Nord Algérie vers le système de référence Mercator Transversal Universel (WGS 84 UTM 31N). Au total, 466 données sur le sol ont été utilisées et exportées sous forme de fichier texte vers le logiciel R (R Development Core Team, 2016). Les données ont été divisées en deux ensembles à l'aide d'un échantillonnage stratifié aléatoire, soit 331 échantillons pour la calibration et 135 pour la validation externe (Figure 2b).

### *2.2.1. Utilisation de la base de référence mondiale pour les ressources en sols (WRB)*

Alors qu'en France, l'ancienne classification française des sols (CPCS, 1967) a été remplacée par le référentiel pédologique (Baize et Girard 2009). En Algérie, la CPCS (1967) est restée à ce jour la référence pour la légende des cartes de sol à toutes les échelles. En effet, dans la plupart des études pédologiques réalisées en Algérie les unités cartographiques sont rattachées au CPCS (1967). Cette dernière est une classification hiérarchique qui définit différentes unités taxonomiques (niveau taxonomique) allant du niveau de la classe jusqu'au niveau du sous groupe pour les unités majeures de classification des sols et pour les unités mineurs de la famille jusqu'à la phase, séparant ainsi de façon arbitraire des individus sols, qui sont caractérisés par une continuité spatiale. Cette rigidité de la classification CPCS est un inconvénient par rapport à la base de référence mondiale pour les ressources en sols (WRB ; IUSS Working Group WRB, 2014), qui définit plusieurs intergrades entre les unités principales et réduit le nombre de taxonomie à deux niveaux (Duchaufour 1998). En outre, l'absence de données mesurables a conduit à délaisser la CPCS au profit de la soil taxonomy et de la classification des sols WRB, qui sont basés sur des données pédologiques mesurables, qui peuvent être traitées et analysées statistiquement (Legros, 1996). Dans cette étude, nous avons choisi la WRB pour la classification des sols. Elle est conçue comme un moyen facile

de communication entre les scientifiques pour identifier, caractériser et nommer les principaux types de sols. En plus, elle est considérée comme un outil offrant une meilleure corrélation entre les systèmes nationaux et permet d'agir en tant que facteur commun par lequel les systèmes nationaux peuvent être comparés.

Dans cette étude, les données ponctuelles du sol ont été converties de l'ancienne classification française des sols (CPCS, 1967) vers la WRB (IUSS Working Group WRB, 2014) en utilisant la description morphologique et analytique des profils représentatifs de chaque classe de sol. Les sols ont été classés à deux niveaux :

- Le premier niveau a correspondu aux groupes de sols de référence (RSGs; reference soil groups) ;
- Le deuxième niveau a correspondu aux groupes de sols de référence avec qualificatifs principaux, qui décrivent en détail les propriétés du RSGs en ajoutant un ensemble de qualificatifs basés sur les horizons, les propriétés et les matériaux de diagnostic. À ce niveau, la classification des sols a été basée sur certaines propriétés pertinentes qui ont expliqué l'organisation des sols dans le paysage telles que : la présence d'encroûtement gypseux, la présence de signes d'hydromorphie, la présence de charge caillouteuse.

## **2.3. Facteurs environnementaux**

### *2.3.1. Collecte des covariables*

#### *2.3.1.1. Images satellitaires*

Pour montrer le maximum d'information sur les sols, une image Landsat 5 TM datée de Juin 1984 a été utilisée. L'image acquise est constituée de six bandes de résolution de 30 m, qui ont été téléchargées sous forme de fichiers TIFF. À l'origine, les bandes spectrales sont formées de nombre numérique qui varie de 0 à 255. Le logiciel SAGA geographic information system software (Conrad et al., 2015) a été utilisé pour la correction

atmosphérique de l'image, nous avons utilisé le modèle de correction COST (Cosine of the Sun Zenith Angle) pour corriger la réflectance atmosphérique (Chávez, 1996). La correction COST utilise le cosinus de l'angle d'élévation solaire de l'image Landsat, pour corriger l'angle solaire. Un facteur « tau » a été utilisé dans le modèle de normalisation pour corriger l'angle du capteur par rapport à la surface du sol.

Quatre indices spectraux ont été calculés à l'aide du logiciel SAGA GIS, incluant:

- ✓ L'indice de salinité 3 (salinity index 3 ; Figure 6b) : a été utilisé pour identifier la salinité de la surface du sol (Douaoui et al., 2006).

$$\text{sali\_indx} = \sqrt{G^2 + R^2} \quad (3)$$

Où, G= bande 2 et R=bande 3

- ✓ L'indice gypsique (gypsic index ; Figure 6d) : est basé sur un rapport de différence normalisé. Il a été utilisé pour identifier le gypse de la surface du sol (Nield et al., 2007).

$$\text{gyps\_indx} = (\text{bande 5} - \text{bande 7}) / (\text{bande 5} + \text{bande 7}) \quad (4)$$

- ✓ Le rapport de la bande 1 sur la bande 7 (Figure 6a) : a été utilisé pour identifier la salinité de la surface du sol.

$$\text{rat\_7\_1} = \text{bande 1} / \text{bande 7} \quad (5)$$

- ✓ L'indice de taille de grain (GSI; gsi\_indx; grain size index ; Figure 6c) : a été développé à partir de la réflectance spectrale de la surface du sol et corrélé avec la teneur en sable fin (Xiao et al., 2006).

$$\text{gsi\_indx} = (R - B) / (R + B + G) \quad (6)$$

Où, B= bande 1, G=bande 2 et R= bande 3

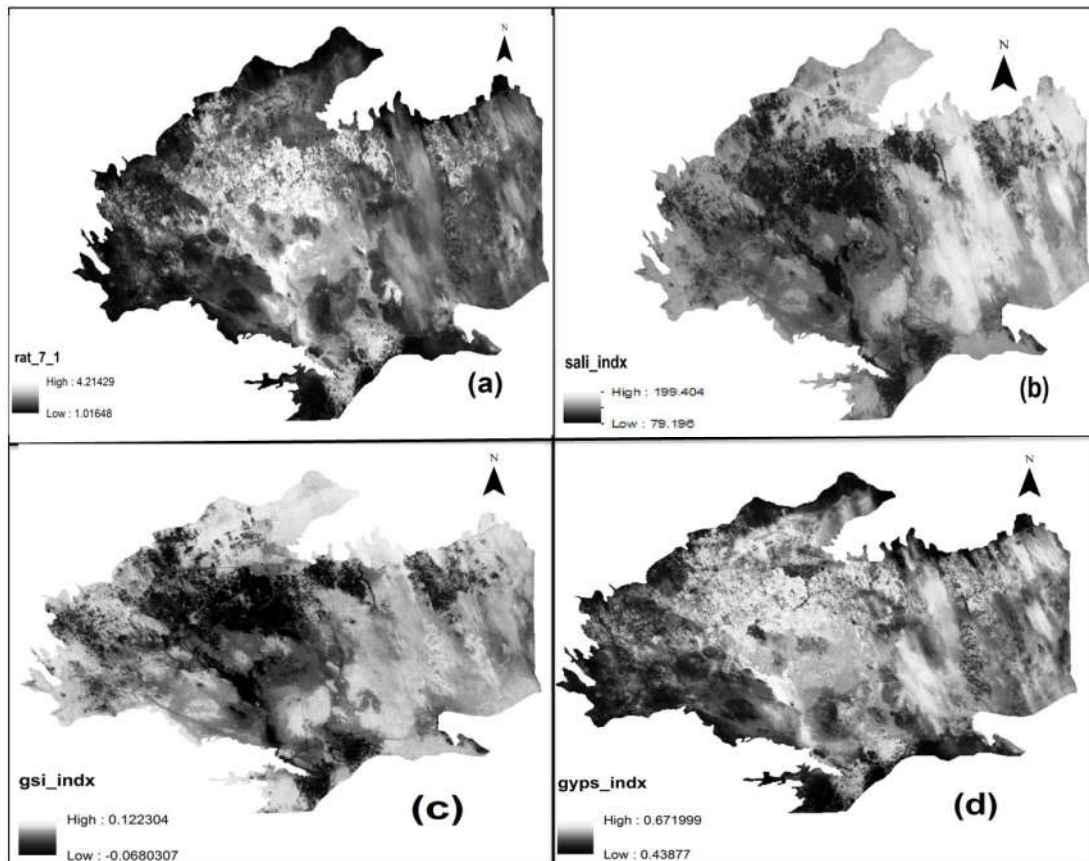


Figure 6. Les indices spectraux calculés à partir de l'imagerie Landsat 5 TM. rat\_7\_1(a), sali\_indx (b), gsi\_indx (c) et gyps\_indx (d).

#### 2.3.1.2. Modèle numérique d'altitude (MNA)

Un modèle numérique d'altitude de résolution de 90 m (SRTM; shuttle radar topography mission elevation data at 3 arc-seconds for global coverage) a été acquis pour générer les attributs du terrain représentant le relief. Le MNA a été obtenu en format TIFF. Le prétraitement a été effectué grâce au logiciel SAGA GIS. Il a consisté à l'élimination des erreurs causées par les artefacts et à remplir le vide créé dans le MNA original en utilisant l'outil « fill sinks » installé dans SAGA GIS (Planchon et Darboux, 2002). Ensuite, le MNA de résolution initiale de 90 m a été ré-échantillonné à la même résolution que l'image Landsat

5 TM, soit 30 m. Pour cela, nous avons utilisé la méthode d'interpolation bilinéaire qui, pour déterminer le pixel ré-échantillonné, prend une moyenne pondérée en fonction de la proximité des quatre pixels les plus proches. Neuf attributs du terrain ont été extraits du MNA à l'aide du SAGA GIS comprenant:

- ✓ Les indices de courbure: mesure la forme ou la courbure du relief, deux types de courbure ont été utilisés dans cette étude :
  - Le profil de courbure représente la courbure de la surface dans la direction de la pente la plus raide (Figure 7a). Le profil de courbure affecte la vitesse d'écoulement de l'eau de drainage en surface et influence le processus d'érosion et de dépôt. Dans les endroits convexes, l'érosion prévaut, alors que dans les zones concaves, c'est le processus de dépôt qui domine.
  - Le plan de courbure: représente la courbure dans un plan horizontal (Figure 7b). Il a été utilisé pour différencier entre les crêtes et les vallées. Le plan de courbure est positif pour les cellules aux contours convexes, et négatif pour les cellules aux contours concaves.

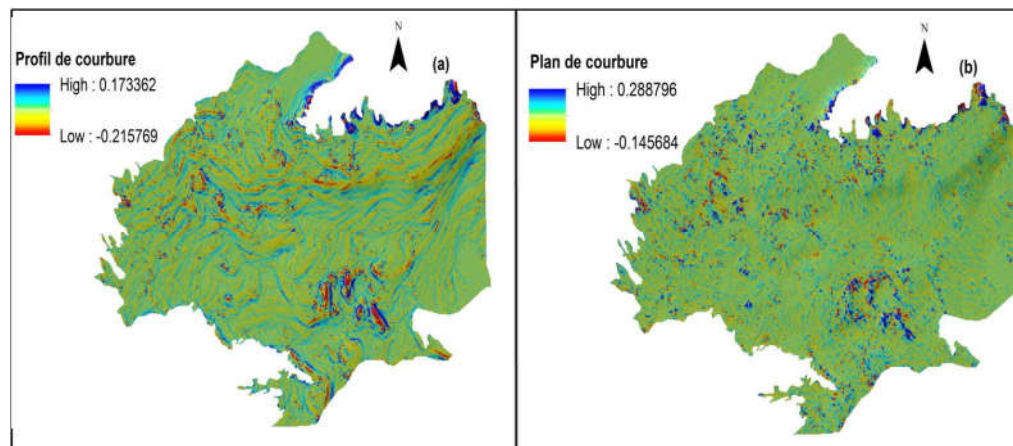


Figure 7. Les indices de courbure calculés à partir du MNA. Profil de courbure (a) et Plan de courbure (b).

- ✓ L'élévation (elevation) : elle représente les valeurs brutes de MNA et correspond à la hauteur du terrain perpendiculaire au-dessus de la surface de référence qui est généralement la surface de la mer. L'élévation a été utilisée pour expliquer le flux de transfert de matière (eaux et matière solides) à travers le paysage entre les positions élevées et basses (Figure 8a).
- ✓ La pente (slope) : est une dérivée directe de l'élévation (dérivée primaire) calculée en pourcentage. Elle permet de différencier les sols en fonction de leur position le long de la pente. En plus, elle contrôle le flux de l'eau et de l'humidité dans le sol (Figure 8b).
- ✓ La pente du bassin versant (Catchment slope) : est définie comme la pente moyenne au-dessus du tracé de l'écoulement (Figure. 8c). Elle a été utilisée pour différencier entre les lieux d'accumulation et l'enlèvement des sédiments (Evans et Hartemink, 2014).
- ✓ L'indice de planéité du fond de vallée à résolutions multiples (Mrvbf; multi-resolution index of valley bottom flatness) : est un attribut du terrain qui identifie les fonds des vallées plates (Gallant et Dowling, 2003). Une valeur élevée identifie les zones de fond de vallée, qui représente des zones de dépôt. Le Mrvbf est calculé à différentes échelles, grâce à un algorithme appliqué sur le MNA (Figure 8d).

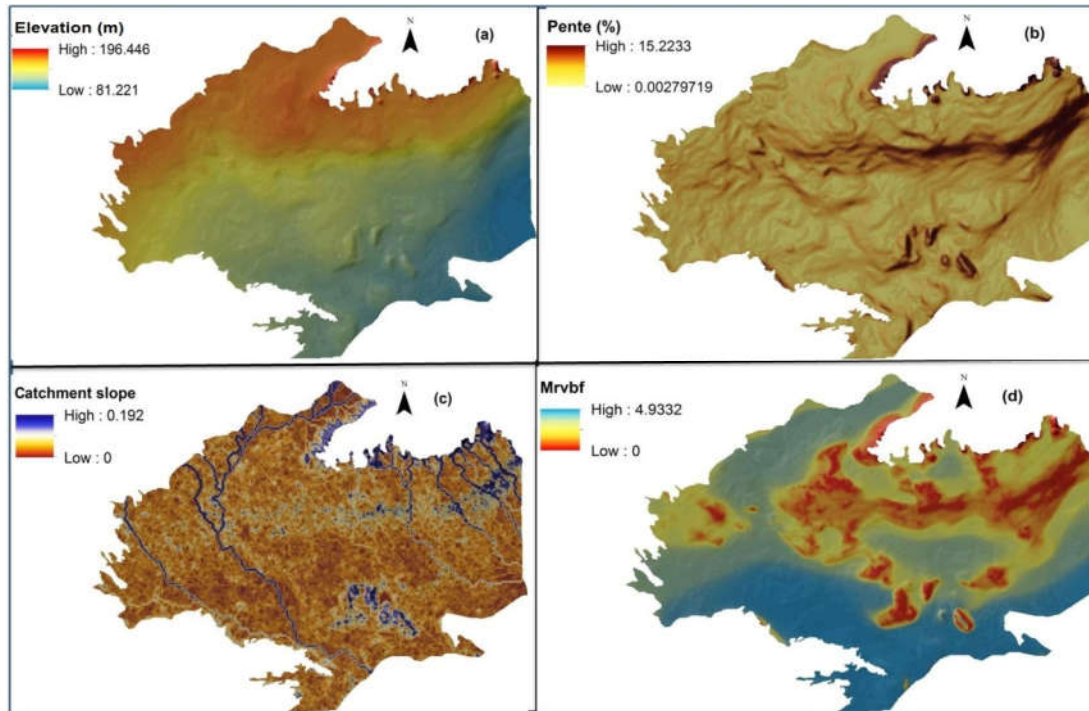


Figure 8. Les dérivés de terrain calculés à partir du MNA. Élévation (a), Pente (b), Pente du bassin versant (c) et Mrvbf (d).

- ✓ L'indice de rugosité du terrain (TRI; Terrain ruggedness index ; Figure 9a) : a permis de mesurer l'hétérogénéité de la topographie sur le terrain (Riley et al., 1999).
- ✓ L'indice de classification du terrain pour les basses terres (TCILow ; Terrain classification index for low lands ; Figure 9b) : ce paramètre représente plusieurs processus qui forment les caractéristiques morphologiques du terrain et montre beaucoup de détails du terrain, en particulier dans les régions à faible pente (Bock et al., 2007).
- ✓ L'indice topographique d'humidité (TWI ; Topographic wetness index ; Figure 9c) : décrit la tendance d'un point à accumuler l'eau (Moore et al., 1991). Une valeur élevée de cet indice indique une accumulation importante de l'eau dans le sol. Il est défini comme suit:

$$TWI = \ln(A/B) \quad (7)$$

Où,  $\ln$  représente la fonction logarithmique, A représente la surface drainée, et B la pente locale en pourcentage.

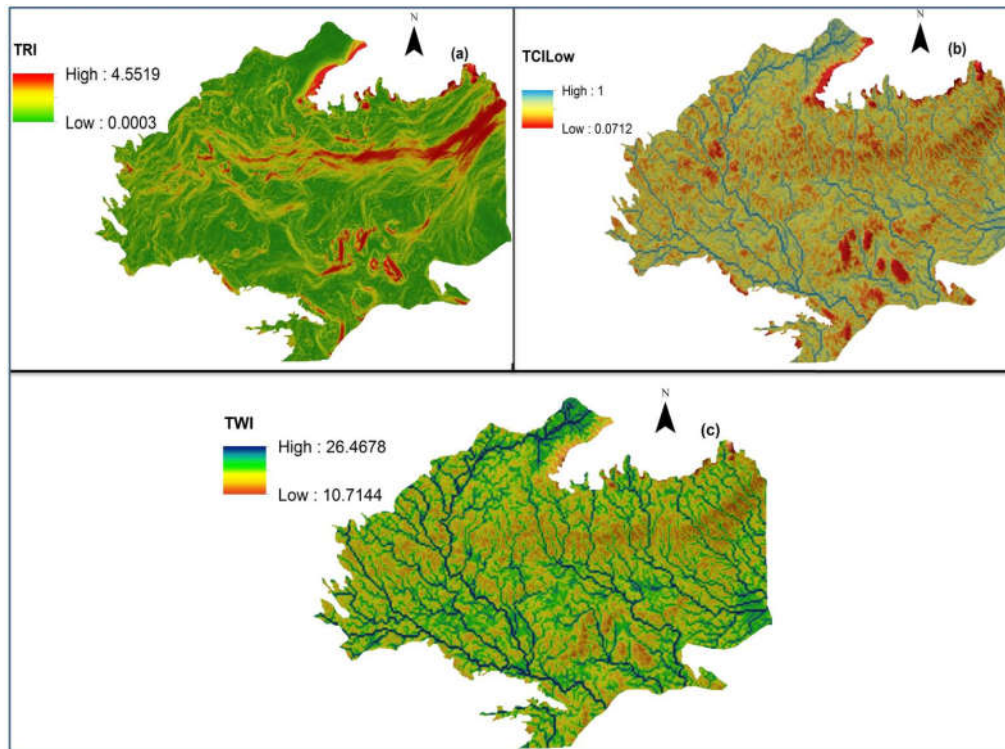


Figure 9. Les dérivés de terrain calculés à partir du MNA. TRI (a), TCILow (b) et TWI (c).

### 2.3.2. Intégration de l'approche multi-échelle dans la CNS

L'intégration du contexte spatial à différentes échelles a été basée sur le traitement des pixels des indices topographiques issus du MNA, en prenant en compte non seulement l'information brute contenue dans le pixel mais aussi l'information dans le voisinage du pixel à traiter. Pour cela, nous avons appliqué aux indices topographiques bruts des filtres spatiaux qui effectuent une opération de voisinage avec une fenêtre de type moyen et de forme circulaire considérée comme un moyen efficace de calculer les différentes échelles des variables explicatives (Grinand et al., 2008), en utilisant l'outil statistiques focales d'ArcGIS 10.5 (ESRI, 2016).

Pour calculer la taille de la fenêtre, nous avons utilisé la méthode du semi-variogramme appliquée par Moran et Bui (2000). Cette méthode permet de calculer la portée du semi-variogramme local du krigeage ordinaire dont la valeur correspond au rayon du cercle. La variable élévation a été utilisée comme exemple pour démontrer les différentes étapes de traitement pour le calcul de la portée en se basant sur l'outil géostatistique d'ArcGis 10.5 :

### 2.3.2.1. Exploration des données

L'exploration des données a consisté à analyser la distribution et la tendance spatiale de l'élévation.

#### ➤ La distribution spatiale

Les valeurs de l'élévation n'ont pas présenté une distribution normale (Figure 10.a). L'objectif est d'obtenir une distribution normale des données, dans ce cas, la transformation en log-normale a donné les meilleurs résultats (Figure 10.b).

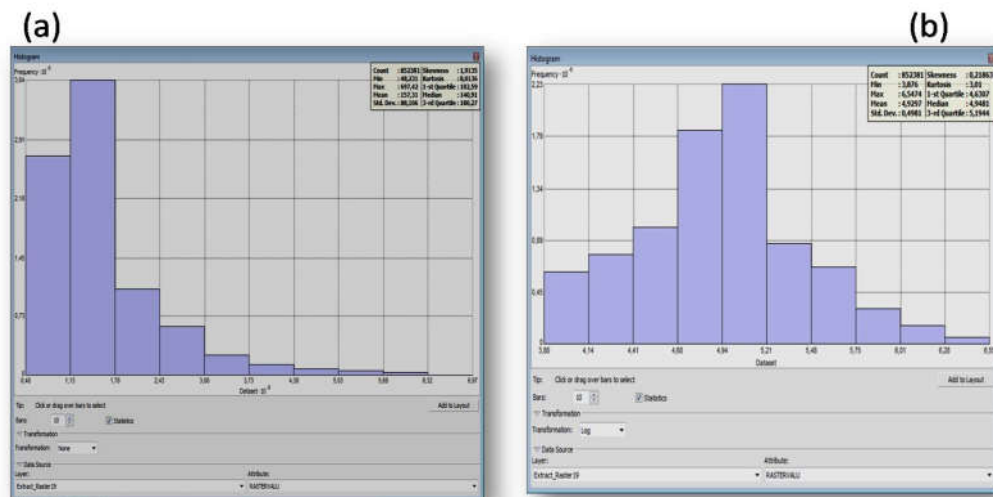


Figure 10. Histogrammes des données de l'élévation avant (a) et après transformation en log-normale (b).

Une autre façon de tester la normalité a été l'utilisation du diagramme quantile-quantile (Q-Q) qui permet d'évaluer la pertinence de l'ajustement linéaire de la distribution des données à un modèle normal standard (Figure 11.a). Après la transformation en log-normale, nous avons remarqué une relation linéaire entre les données et le modèle normal standard (Figure 11.b).

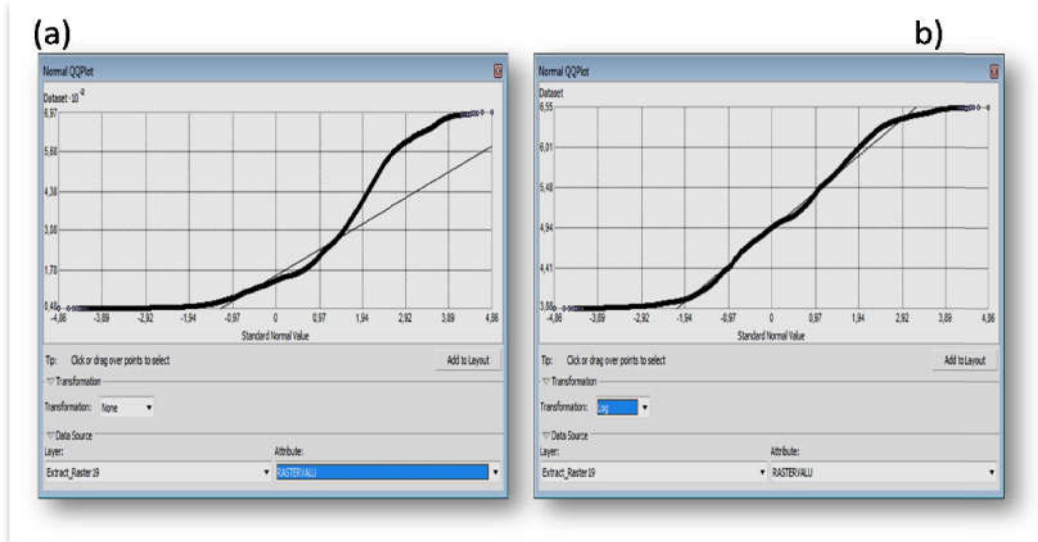


Figure 11. Le diagramme Q-Q de l'élévation avant (a) et après transformation en log-normale (b).

### ➤ La tendance spatiale

L'analyse de la tendance spatiale permet d'identifier la présence ou l'absence d'une tendance globale dans les données. Si une tendance existe, cela implique la présence d'une composante déterministe qui peut être représentée par une formule mathématique de type polynômiale d'ordres 1, 2 ou 3. Dans ce cas, il est nécessaire d'éliminer la tendance afin de satisfaire les hypothèses de stationnarité des données, ce qui conduit généralement à une analyse plus précise.

#### 2.3.2.2. Analyse de la structure spatiale

Après la transformation et l'élimination de la tendance spatiale des données, nous avons déterminé la structure spatiale des données grâce au semi-variogramme, qui est un outil

d'analyse de la dépendance des données. La figure 12 présente un modèle du semi-variogramme théorique, où la semi-variance  $\gamma(h)$  est fonction de l'intervalle d'échantillonnage (h) et dont l'équation est donnée par :  $\gamma(h) = 1/2 \text{ var } (Z(x) - Z(x+h))$  (8)

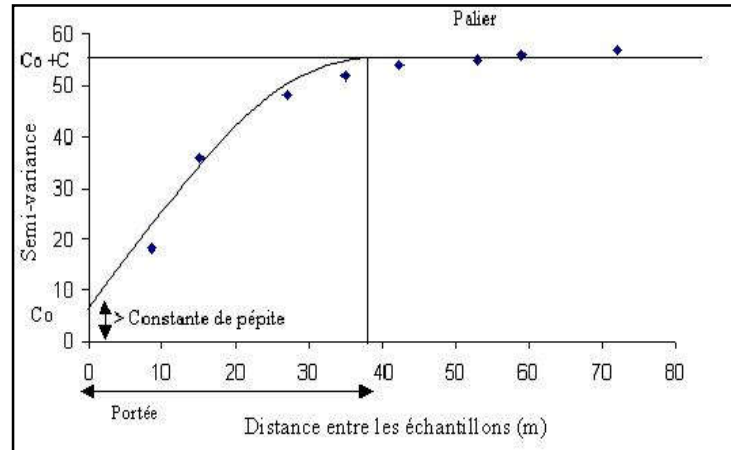


Figure 12. Représentation d'un semi-variogramme théorique.

L'ajustement d'un modèle théorique (type sphérique, type exponentielle, type gaussien, etc....) au semi-variogramme expérimental a pour objectif de rendre compte des caractéristiques qui sont :

- ✓ L'effet de pépité, qui est une discontinuité attribuée à des erreurs de mesure sur les valeurs et/ou à l'existence d'une micro-régionalisation non estimée à cause d'un pas d'échantillonnage trop grand (Figure 12).
- ✓ Le palier : le comportement à l'infini du semi-variogramme peut être délimité ou non autour d'une valeur appelée le palier, qui est égale à la variance sur l'ensemble des valeurs. D'une autre manière, la valeur où le semi-variogramme atteint la portée est appelée le palier (Figure 12).
- ✓ La portée : est définie comme étant la distance à partir de laquelle le semi-variogramme devient constant, elle correspond à la distance d'autocorrélation (Figure 12). Les emplacements des échantillons séparés par une distance inférieure ou égale à

la portée sont spatialement dépendants, tandis que les emplacements séparés par une distance supérieure à la portée sont spatialement indépendants. Parfois, les modèles du semi-variogramme changent non seulement avec la distance mais aussi avec la direction, dans ce cas nous sommes en présence d'anisotropie. Ainsi, le palier est atteint plus rapidement dans certaines directions que dans d'autres. La distance la plus longue pour atteindre le palier est appelée la portée majeure et la distance la plus courte pour atteindre le palier est appelée la portée mineure.

L'évaluation du semi-variogramme a été réalisée par une validation croisée, qui représente des tests statistiques permettant de choisir le type du semi-variogramme d'ajustement le plus performant.

### *2.3.3. Sélection des covariables optimales*

Dans la cartographie numérique du sol, la sélection des covariables optimales est une étape, qui consiste à choisir les variables pertinentes pour la modélisation. La sélection de covariables optimales permette de réduire d'une part, la durée de traitement des algorithmes lors de la modélisation, ce qui augmente la performance des algorithmes (Heung et al., 2017; Taghizadeh-Mehrjardi et al., 2015), et d'autre part d'éliminer les variables explicatives qui présentent une information redondante. Il existe de nombreuses techniques de sélection de covariables, nous citons les plus utilisées :

- ✓ La sélection ascendante (forward selection): les variables sont ajoutées une à une dans le modèle et à chaque itération on mesure si l'erreur décroît. Le processus s'arrête lorsqu'un critère de performance tel que la valeur de l'erreur ou de la précision est satisfaisante.
- ✓ La sélection descendante (backward selection) : dans ce cas, nous avons un ensemble initial de prédicteurs et à chaque itération une variable explicative est supprimée, de manière que le modèle devient plus robuste. La fonction d'élimination de variables par

réursion (recursive feature elimination; RFE) fait partie de cette technique. Elle permet de d'obtenir un groupe optimal de prédicteurs déterminé par la mesure d'un indicateur de performance (Guyon et al., 2002; Kuhn et Johnson, 2013). La méthode RFE a été appliquée en cartographie numérique du sol pour la modélisation des classes de sols (Brungard et al. 2015; Levi, 2017) et des propriétés de sols (Ballabio, 2009; Jeong et al., 2017).

Après que les covariables ont été centrées et mises à l'échelle, nous avons utilisé la méthode RFE pour sélectionner un sous ensemble de covariables les plus performantes. La fonction RFE est disponible dans le paquet caret (Kuhn, 2008) installé dans le logiciel R. Le modèle forêt aléatoire a été appliqué comme fonction de contrôle avec 5 répétitions de la validation croisée avec chacune 10-folds (Levi, 2017).

#### **2.4. Modèles d'apprentissage machine**

Nous avons testé six modèles complexes identifiés comme les plus précis par Brungard et al. (2015). Ces modèles sont installés dans le paquet caret sous R, incluant Bagged classification tree (BCT), la forêt aléatoire (RF ; Random forest), les machines à vecteur de support linéaire (SVML ; Linear support vector machines), les machines à vecteur de support à base radiale (SVMR ; Radial-basis support vector machines), le réseau de neurones à couche unique cachée (NNET ; Single-hidden-layer neural network) et le réseau de neurones perceptron multicouche (MLP ; Multi-layer perceptron neural network). Nous les décrivons tour à tour.

##### *2.4.1. Réseaux de neurones artificiels (ANN)*

Le réseau de neurones artificiels (Artificial Neural Networks) est un classificateur inspiré par le fonctionnement des réseaux neuronaux dans le cerveau humain. C'est un algorithme puissant adapté aux problèmes non linéaires (Kuhn et Johnson, 2013). La structure d'un ANN se compose d'un élément de base appelé neurone qui est interconnecté et organisé sous la forme d'une ou plusieurs couches de neurones. Chaque neurone d'entrée qui représente les

prédicteurs se voit attribuer une pondération basée sur sa valeur relative par rapport aux autres neurones d'entrée (Behrens et al., 2005). Ces neurones d'entrée sont reliés à une ou plusieurs couches cachées qui traitent les données pour en extraire des relations significatives. Ensuite, les couches cachées sont liées à une sortie qui représente les variables à prédire (couche de sortie). Dans cette étude, deux modèles d'ANN ont été appliqués :

- Le perceptron multicouche (MLP) : composée de plusieurs couches de réseaux de neurones cachés et interconnectés, au sein desquelles l'information circule de la couche d'entrée vers la couche de sortie (Figure 13.a).
- Le réseau à couche unique cachée (NNET) : c'est le réseau le plus simple, qui contient une seule couche cachée (Figure 13.b).

Le réseau de neurones artificiels a été utilisé dans la cartographie numérique du sol pour modéliser les variables catégorielles tels que les classes de sols (Chagas et al. 2017; Pásztor et al. 2018) et les classes texturales (Wu et al. 2018). En plus, il a été utilisé pour modéliser des variables continues de sols, comme le stock de carbone organique (Taghizadeh-Mehrjardi et al., 2016).

Le réseau de neurones artificiels possède des avantages, telles qu'une grande adaptation aux problèmes de classification, une grande capacité d'apprentissage et de modélisation des relations non linéaires et aucune restriction sur les variables d'entrée, comme la normalité des variables continues. Par contre, ses inconvénients sont l'intensité élevée en calcul, surtout si le réseau est complexe, la sensibilité aux problèmes de sur-apprentissage et les résultats difficiles à interpréter.

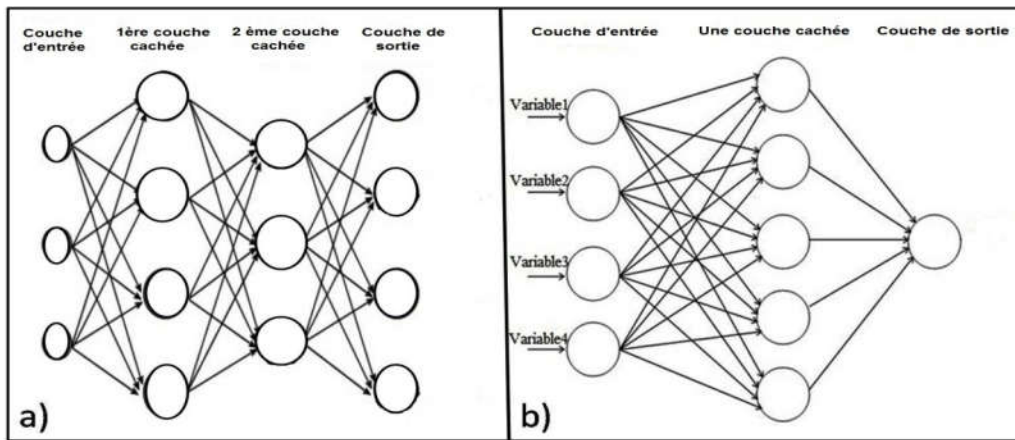


Figure 13. Illustration des algorithmes de réseaux de neurones artificiels. Le perceptron multicouche (a) et le réseau à couche unique cachée (b).

#### 2.4.2. Arbres de décision

L'arbre de décision est un type d'algorithme d'apprentissage supervisé sous la forme d'un arbre utilisé principalement pour les problèmes de classification. Dans l'arbre de décision, l'échantillon est divisé en deux ou plusieurs ensembles homogènes de décision fondés sur les attributs les plus significatifs (Figure 14). L'arborescence est caractérisée par des nœuds représentant les variables explicatives, tandis que les branches représentent la valeur de chaque attribut testé et les feuilles représentent les classes prédites. Les arbres de décision présentent deux caractéristiques :

- L'induction d'arborescence, qui consiste à prendre en considération les variables explicatives sur lesquelles il est préférable de fractionner le jeu de données jusqu'à ce que toutes les instances de formation soient classées. Lors de la construction de l'arbre, l'objectif est de choisir les attributs qui créent les nœuds les plus purs possibles. Cette pureté est généralement mesurée par des indices tels que l'indice de Gini et le gain d'information.
- L'élitage des arbres, qui consiste à élaguer la structure trop complexe et difficile à interpréter d'un arbre de décision, afin de réduire le sur-apprentissage. Ce dernier fait

référence à l'idée que les modèles d'apprentissage machine sont parfois trop adaptés aux données d'apprentissage utilisées, alors qu'ils ne donnent pas de bons résultats avec les données de test (perte de capacité de généralisation). En général, dans le cas de sur-apprentissage, l'erreur sur les données d'apprentissage est minime en comparaison avec l'erreur sur les données de test.

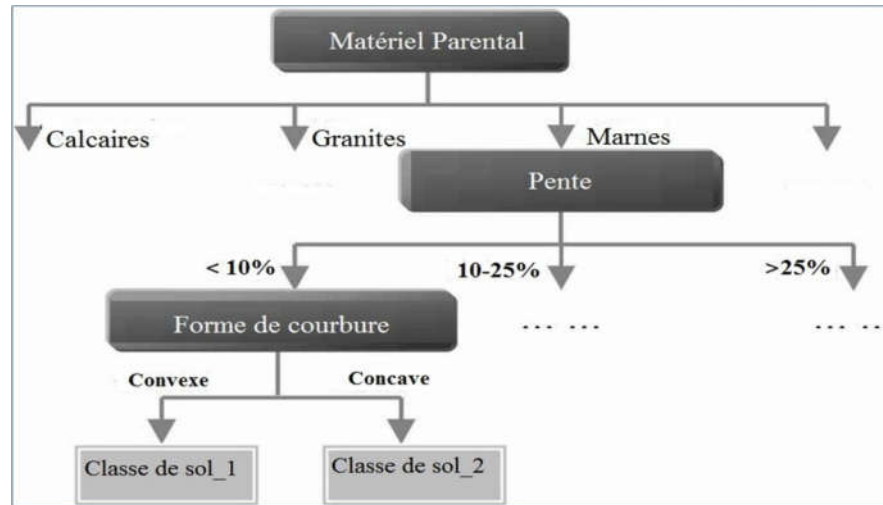


Figure 14. Mode de fonctionnement de l'arbre de décision.

Dans cette étude nous avons utilisé deux types d'algorithmes d'arbres de décision :

#### 2.4.2.1. Forêt aléatoire (RF)

La forêt aléatoire est un modèle ensembliste basé sur l'approche de bagging (Vincenzi et al., 2011). Le modèle consiste à construire une multitude d'arbres de décision indépendants, dont chacun donne une estimation partielle du problème (Figure 15). Les résultats des différents modèles sont combinés pour obtenir un modèle plus performant que les modèles individuels. Le résultat final de RF est un modèle unique déduit à partir de l'ensemble des prédictions par agrégation sur tous les arbres, basé sur le vote à la majorité dans le cas de la classification, et sur la moyenne de toutes les estimations dans le cas de la régression.

Pour la forêt aléatoire, chaque arbre est entraîné sur des sous-ensembles sélectionnés de manière aléatoire et utilisés pour les règles de fractionnement de nœud. Elle offre la possibilité d'estimer l'importance des variables en utilisant une diminution moyenne de la précision ou une diminution moyenne de l'indice de Gini, qui mesure la réduction des impuretés de classes résultantes de la partition de l'ensemble de données (Barthold et al., 2013).

RF a été utilisé dans la cartographie numérique du sol, notamment la cartographie du carbone organique du sol (Grimm et al., 2008; Wiesmeier et al., 2011), la texture du sol (Ließ et al., 2012) ainsi que la classification du matériau parental (Heung et al., 2014) et la classification des sols (Häring et al., 2012; Stum et al., 2010).

Le modèle de forêt aléatoire est puissant et précis sur de nombreux problèmes, y compris le cas de modélisation non linéaires. Il peut être utilisé à la fois pour des problèmes de régression et de classification. Une autre grande qualité de la forêt aléatoire réside dans la facilité de mesurer l'importance relative de chaque variable explicative sur la modélisation. RF évite le problème de sur-apprentissage qui affecte les modèles d'apprentissage machine, cela peut être expliqué par le nombre important d'arbres utilisés dans le modèle, et s'adapte aux données de plus grande dimensionnalité. Par contre, le modèle est considéré comme une boîte noire, en effet les résultats du modèle sont difficiles à interpréter, en particulier quand le nombre de divisions est important.

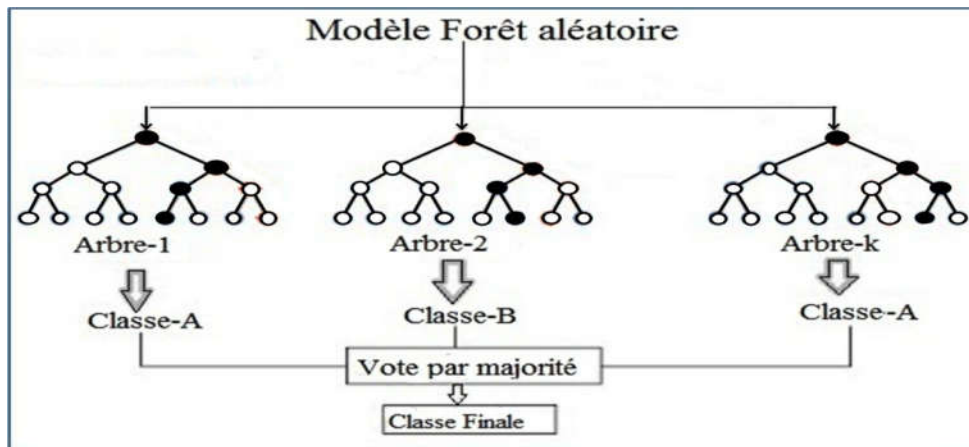


Figure 15. Architecture de l'algorithme de forêt aléatoire.

#### 2.4.2.2. Bagged classification tree (BCT)

Le mot bagging est une combinaison de bootstrapping et d'agrégation (Kuhn et Johnson, 2013; James et al., 2013). Le bootstrapping est une méthode de ré-échantillonnage qui crée des données d'entraînement par échantillonnage aléatoire avec remplacement à partir des données originales. Les données de formation sont utilisées pour construire des arbres de décision et les résultats de chaque arbre sont ensuite agrégés. L'apprenant obtenu est calculé à partir de la moyenne dans le cas de la régression ou à la majorité dans le cas de la classification.

La différence entre BCT et RF réside dans le fait que BCT n'a qu'un seul paramètre qui est le nombre d'arbres (*ntree*), alors que RF en plus du « *ntree* » utilise un paramètre « *mtry* » qui permet de sélectionner le nombre de prédicteurs testés à chaque division d'un nœud. Par ailleurs, dans le cas de modélisation avec BCT, toutes les covariables sont utilisées pour diviser un nœud. Par contre, RF utilise un sous-ensemble de covariables sélectionnées de manière aléatoire permettant de créer une contrainte d'apprentissage pour chaque arbre différente des autres, ce qui réduit la corrélation entre les différents arbres.

### *2.4.3. Machines à vecteur de support (SVM)*

Les machines à vecteur de support (Support Vector Machines) font partie des méthodes d'apprentissage supervisé et sont utilisées pour la classification et la régression (Figure 16). Ces algorithmes sont basés sur l'utilisation d'un hyperplan linéaire optimal séparant les données et dont la séparation (marge) est la plus grande possible entre deux classes (James et al., 2013). À l'origine, les SVM sont des classificateurs binaires, néanmoins, on peut les utiliser dans le cas de classes multiples. Ainsi, la méthode utilisée est basée sur la combinaison de plusieurs classificateurs binaires. Lorsque les données ne sont pas linéairement séparables, on utilise la fonction noyau (kernel) qui permet de projeter les données non linéaires dans un espace de grande dimensionnalité avec une forte probabilité de trouver un hyperplan, ce qui assure une séparation linéaire (Hahn et Gloaguen, 2008; Lorenzetti et al., 2015). Dans cette étude, les machines à vecteur de support utilisées sont :

- Les machines à vecteur de support linéaire (SVML) : qui utilisent un vecteur linéaire pour séparer les données selon un hyperplan.
- Les machines à vecteur de support à base radiale (SVMR) : qui sont généralement utilisées lorsque l'ensemble de données n'est pas linéairement séparable.

Les machines à vecteur de support ont été appliquées avec succès dans la cartographie numérique du sol. Par exemple Kovačević et al. (2010) ont prédit les valeurs des propriétés des sols et la classification des types de sol, en utilisant les machines à vecteur de support. De même, Ballabio (2009) a utilisé les machines à vecteur de support pour modéliser les propriétés des sols dans les zones montagneuses.

Les SVM sont des classificateurs particulièrement puissants, elles ne sont pas trop influencées par les données bruyantes, et sont peu sensibles au sur-apprentissage. Cependant, elles

peuvent être lentes à entraîner sur de grands ensembles de données, et il faut tester diverses combinaisons de noyaux et de paramètres de modèle.

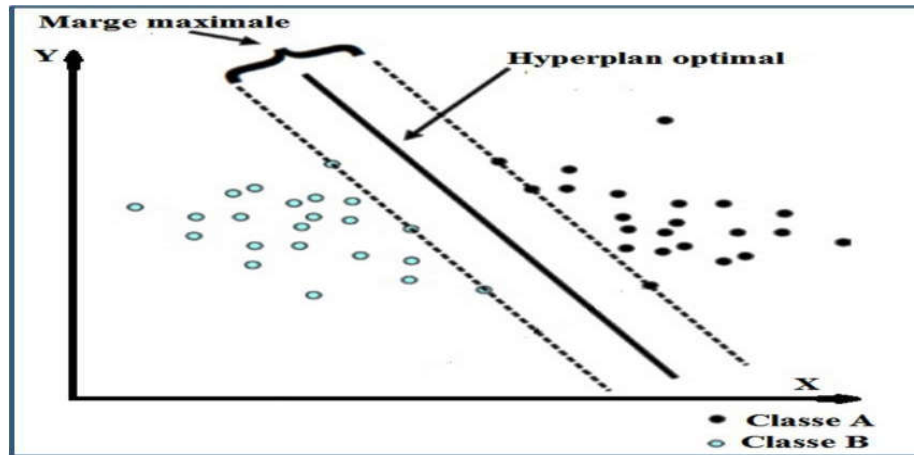


Figure 16. Principe de l'algorithme machines à vecteur de support.

## 2.5. Évaluation métrique des modèles

La statistique de Kappa ( $k$ ) a été utilisée pour évaluer la performance des modèles. Elle permet de mesurer l'accord entre un résultat associé à une classification observée et un résultat associé à une classification prédite (Congalton, 1991). La statistique de Kappa est généralement considérée comme une mesure solide pour traiter les problèmes de multi-classes et des classes avec des données déséquilibrées (Brungard et al., 2015).

Les valeurs de Kappa inférieures à 0,2 sont considérées comme faibles, celles comprises entre 0,2 et 0,4 comme justes, entre 0,4 et 0,6 comme modérées, celles entre 0,6 et 0,8 sont substantielles et les valeurs supérieures à 0,8 sont en accord quasi parfait (Landis et Koch, 1977; Levi, 2017). Le modèle le plus efficace a été déterminé par la valeur de kappa la plus élevée.

Kappa est défini comme suit :

$$k = \frac{p_o - p_e}{1 - p_e} \quad (9)$$

Où,  $p_o$  est l'accord observé, et  $p_e$  est l'accord prédit.

## **2.6. Signification statistique des différences observées entre les modèles**

Dans le cas des comparaisons multiples entre les moyennes effectuées simultanément sur un ensemble de données unique, le test de Student, habituellement utilisé dans la comparaison entre deux échantillons, n'est pas approprié, car l'augmentation du nombre de tests statistiques accroît le risque de trouver à tort une différence statistique. Le risque alpha fixé par défaut à 5 % augmente avec le nombre de tests statistiques, cela correspond à l'inflation du risque alpha. Pour contrôler l'inflation du risque alpha, l'ajustement de Bonferroni peut être utilisé (Bland et Altman, 1995). Il est fondé sur la correction de  $t$  de Student, en diminuant le risque alpha initial. Pour cela, on divise la valeur  $\alpha$  initial par le nombre de comparaisons effectuées. Cette méthode consiste à fixer un taux d'erreur beaucoup plus faible que  $\alpha = 0,05$  de telle sorte que le taux d'erreur de l'ensemble reste raisonnable. Pour  $k$  comparaisons, la nouvelle valeur critique serait  $\alpha' = \alpha/k$ . Dans cette étude nous avons utilisé l'ajustement de Bonferroni pour déterminer si la différence des valeurs de kappa entre les modèles a été statistiquement significative. Pour ce faire, nous avons réalisé de multiples tests de corrélations individuelles à partir de la même matrice de corrélation. Le niveau de signification ajusté de Bonferroni pour chacune des corrélations est de  $0,05 / 15 = 0,003$ . Chaque test dont les résultats ont une valeur  $p$  inférieure à 0,003 est considéré comme statistiquement significatif, alors que les corrélations dont la valeur  $p$  est supérieure à 0,003 sont considérées non significatives.

## **2.7. Évaluation des cartes modélisées**

Nous avons visualisé les cartes afin de vérifier le modèle qui a expliqué le mieux la distribution spatiale des sols de la région d'étude. De plus, les 135 données de validation externe ont été utilisées pour évaluer la précision globale, la précision de l'utilisateur et la précision du producteur à partir de la matrice de confusion. La précision globale (Overall Accuracy ; OA) fournit des informations sur la proportion des sites correctement classés par rapport à toutes les classes. La précision du producteur (Producer's Accuracy ; PA) est la proportion des observations de la classe de référence attribuée à la même classe par la classification. Alors que la précision de l'utilisateur (User's Accuracy ; UA) est la proportion de classe de sol résultant de la classification, qui correspond à la même classe dans les données de référence.

## **2.8. Estimation de l'incertitude**

L'indice de confusion (CI) a été utilisé pour évaluer l'incertitude de la modélisation entre les deux principales classes de sol prédites (Burrough et al., 1997) ¶ :

$$CI = [1 - (\mu_{\max j} - \mu_{(\max-1)j})] \quad (10)$$

Où  $\mu_{\max j}$  est la valeur de probabilité de la classe avec une probabilité maximale au pixel  $j$  et  $\mu_{(\max-1)j}$  est la deuxième plus grande valeur de probabilité à la même pixel. Les valeurs de l'indice de confusion varient entre 0 et 1, quand les valeurs de CI augmentent, l'incertitude devient plus élevée.

## **TROISIEME PARTIE : RESULTATS ET DISCUSSION**

### 3.1. Résultats

#### 3.1.1. Classes de sol

Dans la zone d'étude, cinq groupes de sol de référence ont été identifiés (Tableau 3 ; Annexe II) :

- Les Gypsisols : ont un horizon pétrogypsiqique à moins de 100 cm de la surface du sol ou bien les deux caractéristiques suivantes, un horizon gypsiqique dans les premiers 100 cm à partir de la surface du sol et l'absence d'un horizon argique au-dessus de l'horizon gypsiqique (WRB, 2014). Dans la région d'étude, les Gypsisols ont été les plus fréquents avec 280 observations, ils ont une teneur élevée en gypse et une accumulation de gypse secondaire sous forme cimentée (encroûtement gypseux) et/ou sous forme de pseudo-mycélium et cristallisée. En général, la texture de ces sols est moyenne à grossière.
- Les Solonchaks : caractérisés par un horizon salique dans les premiers 50 cm à partir de la surface du sol et par l'absence d'un horizon thionique (sulfurique) commençant dans les premiers 50 cm à partir de la surface du sol (WRB, 2014). Ce groupe de sol a été en deuxième position avec un nombre de 150 observations localisées sur des alluvions de texture limono-argileuse à argileuse et à faible drainage interne, ainsi que sur le glacié au niveau des palmeraies. La salinité de ce groupe est très élevée. Parfois, les Solonchaks sont caractérisés soit par la présence de signes d'hydromorphie dans les zones plates où la nappe phréatique est proche de la surface, soit par la présence d'un horizon gypsiqique.
- Les Fluvisols : ont un matériau fluviqique d'une épaisseur égale ou supérieure à 25 cm et commençant dans les premiers 25 cm à partir de la surface du sol (WRB, 2014). Dans la région, les caractéristiques physico-chimiques de ces sols sont très variées en raison

des conditions différentes de leurs dépôts. Ils sont caractérisés par une faible salinité et formés par une succession d'horizons de couleur, d'épaisseur et de texture variables.

- Les Regosols : ont constitué quelques zones en bas des versants. Les regosols sont formés en surface de recouvrement caillouteux d'origine colluvionaires. La texture de ces sols est grossière renfermant une quantité importante de cailloux régulièrement répartis.
- Les Leptosols : sont caractérisés par une roche continue dans les 25 premiers cm à partir de la surface du sol, par la présence d'une forte charge caillouteuse à partir de la surface du sol et par l'absence d'un horizon calcique, pétrocalcique, pétrogypsiq et gypsiq. Ces sols ont été les moins représentés avec seulement 3 observations et ont été localisés sur les quelques monts qui existent dans la région d'étude. Ce groupe de sols est caractérisé par une forte pente et une faible profondeur à cause de la présence d'une roche dure.

En plus, au niveau des groupes de sols de référence avec qualificatifs principaux, 13 classes de sol ont été identifiées (Tableau 3 ; Annexe II) :

- Haplic Petric Gypsisols (94 échantillons): Gypsisols avec un horizon pétrogypsiq commençant dans les 100 premiers cm à partir de la surface du sol (Solum 1 ; Annexe III).
- Haplic Gypsisols (91 échantillons): cette classe de sol n'a pas de caractérisation plus poussée ou plus significative (Solum 2 ; Annexe III).
- Calcic Petric Gypsisols (60 échantillons): Gypsisols avec un horizon pétrogypsiq et un horizon calcique (Solum 3 ; Annexe III).
- Leptic Petric Gypsisols (17 échantillons): Gypsisols avec un horizon pétrogypsiq et une roche continue (Solum 8 ; Annexe III).

- Skeletic Petric Gypsisols (10 échantillons): Gypsisols ayant un horizon pétrogypique et un volume d'éléments grossiers en moyenne égale ou supérieure à 40%, jusqu'à une profondeur de 100 cm (Solum 6 ; Annexe III).
- Haplic Skeletic Gypsisols (8 échantillons): Gypsisols ayant un volume d'éléments grossiers en moyenne égale ou supérieure à 40%, jusqu'à une profondeur de 100 cm (Solum 7 ; Annexe III).
- Gypsic Solonchaks (22 échantillons): Solonchaks avec un horizon gypique (Solum 10 ; Annexe III).
- Gypsic Sodique Solonchaks (36 échantillons): En plus d'un horizon salique et gypique, ces sols ont été caractérisés par une sodisation de l'horizon d'une épaisseur égale ou supérieure à 20 cm commençant dans les 100 premiers cm à partir de la surface du sol (Solum 5 ; Annexe III).
- Gleyic Solonchaks (92 échantillons): Solonchaks avec une couche d'une épaisseur égale ou supérieure à 25 cm commençant dans les premiers 75 cm à partir de la surface du sol, qui a des propriétés gleyiques partout et des conditions réductrices dans certaines parties de chaque sous-couche (Solum 4 ; Annexe III).
- Gypsic Fluvisols (11 échantillons): Fluvisols ayant du matériel gypique partout entre 20 et 100 cm de la surface du sol et n'ont pas d'horizon gypique ou pétrogypique commençant dans les 100 premiers cm à partir de la surface du sol (Solum 11 ; Annexe III).
- Calcaric Fluvisols (18 échantillons): Fluvisols ayant du matériel calcique partout entre 20 et 100 cm de la surface du sol et n'ont pas d'horizon calcique ou pétrocalcique commençant dans les 100 premiers cm à partir de la surface du sol (Solum 9 ; Annexe III).

- Colluvic Regosols (4 échantillons): Regosols ayant du matériau colluvique d'une épaisseur égale ou supérieure à 20 cm commençant à la surface du sol.
- Hyperskeletal Lithic Leptosols (3 échantillons): Leptosols ayant une roche continue commençant dans les 10 premiers cm à partir de la surface du sol et une teneur moyenne en terre fine inférieure à 20%.

Tableau 3. Nombre et fréquence des données de sol pour les deux niveaux taxonomiques de la base de référence mondiale (WRB).

| Groupe de sols de référence | Nombre d'observations | Fréquence (%) | Groupe de sols de référence avec qualificatifs principaux | Nombre d'observations | Fréquence (%) |
|-----------------------------|-----------------------|---------------|-----------------------------------------------------------|-----------------------|---------------|
| Gypsisols                   | 280                   | 60.09         | Haplic Petric Gypsisols                                   | 94                    | 20.17         |
| Solonchaks                  | 150                   | 32.19         | Gleyic Solonchaks                                         | 92                    | 19.74         |
| Fluvisols                   | 29                    | 6.22          | Haplic Gypsisols                                          | 91                    | 19.53         |
| Regosols                    | 4                     | 0.86          | Calcic Petric Gypsisols                                   | 60                    | 12.88         |
| Leptosols                   | 3                     | 0.64          | Gypsic Sodic Solonchaks                                   | 36                    | 7.73          |
|                             |                       |               | Gypsic Solonchaks                                         | 22                    | 4.72          |
|                             |                       |               | Calcic Fluvisols                                          | 18                    | 3.86          |
|                             |                       |               | Leptic Petric Gypsisols                                   | 17                    | 3.65          |
|                             |                       |               | Gypsic Fluvisols                                          | 11                    | 2.36          |
|                             |                       |               | Skeletal Petric Gypsisols                                 | 10                    | 2.15          |
|                             |                       |               | Haplic Skeletal Gypsisols                                 | 8                     | 1.72          |
|                             |                       |               | Colluvic Regosols                                         | 4                     | 0.86          |
|                             |                       |               | Hyperskeletal Lithic Leptosols                            | 3                     | 0.64          |
| Total                       | 466                   |               |                                                           | 466                   |               |

### 3.1.2. Analyse multi-échelle des facteurs environnementaux

Le tableau 4 résume les résultats de l'exploration des indices topographiques utilisés, qui indique d'une part que certains attributs du terrain nécessitent une transformation logarithmique du fait que la distribution des données n'est pas normale, et d'autre par une élimination de la tendance spatiale en utilisant un polynôme d'ordre 1 ou 2. Par exemple,

l'utilisation de l'outil d'analyse de la tendance (Figure 17) a montré que les données d'élévation ont indiqué une forte diminution du nord au sud (la courbe verte), alors qu'elles ont présenté une faible diminution de l'ouest à l'est (courbe bleue). Par conséquent, il existe une tendance spatiale du nord au sud qui peut être ajustée à un polynôme d'ordre 2 de forme U moins accentué.

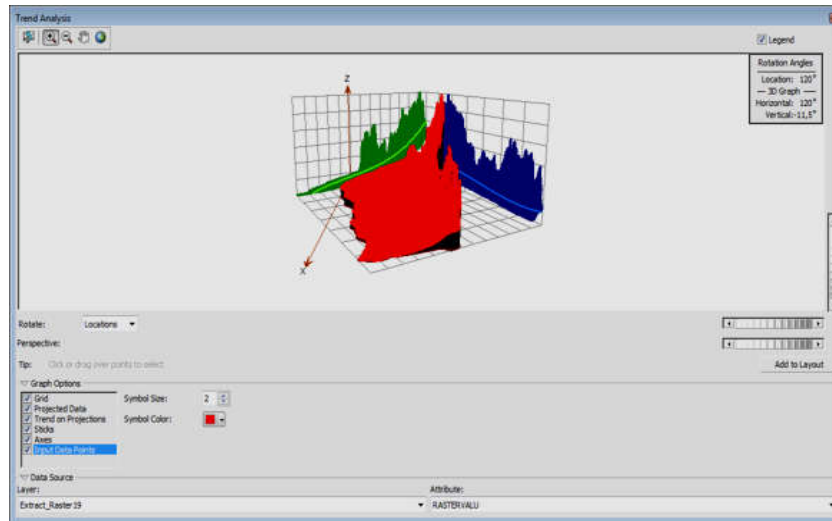


Figure 17. Représentation spatiale de la tendance de l'élévation.

Tableau 4. Tendance globale et transformation des attributs du terrain.

| Attributs du terrain    | Ordre de la tendance globale | Transformation |
|-------------------------|------------------------------|----------------|
| Élévation               | Polynômiale d'ordre 2        | Log-Normale    |
| TWI                     | Absence de tendance          | Log-Normale    |
| Pente (%)               | Polynômiale d'ordre 1        | Log-Normale    |
| TCILow                  | Absence de tendance          | Log-Normale    |
| TRI                     | Polynômiale d'ordre 1        | Log-Normale    |
| Pente du bassin versant | Absence de tendance          | /              |
| Mrvbf                   | Polynômiale d'ordre 2        | /              |
| Profil de courbure      | Polynômiale d'ordre 1        | /              |
| Plan de courbure        | Polynômiale d'ordre 1        | /              |

Le tableau 5 présente les résultats relatifs au type d'ajustement par un modèle du semi-variogramme et l'analyse de la structure spatiale des indices topographiques utilisés. Dans le cas de l'élévation (Figure 18 a), le type du semi-variogramme d'ajustement utilisé est gaussien et nous avons obtenu une portée de 300 mètre équivalente à 10 pixels de résolution spatiale de 30 mètres, ce qui correspond au rayon du filtre appliqué sur la variable (Figure 18 a.1). En raison de la présence d'anisotropie, les variables profil de courbure et plan de courbure ont montré l'existence d'une portée majeure et mineure. En général, les rayons des cercles calculés pour les indices topographiques sélectionnés ont été de 143, 150, 206, 211, 213 et 300 mètres (Tableau 5).

Tableau 5. Propriétés des semi-variogrammes utilisés.

| Covariables             | Modèle du semi-variogramme d'ajustement | Présence d'anisotropie | La portée majeure (mètre) | La portée mineure (mètre) |
|-------------------------|-----------------------------------------|------------------------|---------------------------|---------------------------|
| Élévation               | Gaussien                                | Non                    | 300                       | /                         |
| Pente                   | Gaussien                                | Non                    | 300                       | /                         |
| Mrvbf                   | K-Bessel                                | Non                    | 300                       | /                         |
| TRI                     | Rational Quadratic                      | Non                    | 300                       | /                         |
| TWI                     | K-Bessel                                | Non                    | 211                       | /                         |
| TCILow                  | K-Bessel                                | Non                    | 213                       | /                         |
| Pente du bassin versant | J-Bessel                                | Non                    | 206                       | /                         |
| Profil de courbure      | K-Bessel                                | Oui                    | 211                       | 143                       |
| Plan de courbure        | Rational Quadratic                      | Oui                    | 206                       | 150                       |

À partir des valeurs de la portée des semi-variogrammes, onze covariables ont été calculés incluant : l'élévation avec un rayon de 300 mètres, la pente avec un rayon de 300 mètres, l'indice topographique d'humidité avec un rayon de 211 mètres, l'indice de rugosité de terrain

avec un rayon de 300 mètres, l'indice de classification du terrain pour les basses terres avec un rayon de 213 mètres, le profil de courbure avec un rayon de 211 mètres, le profil de courbure avec un rayon de 143 mètres, le plan de courbure avec un rayon de 206 mètres, le plan de courbure avec un rayon de 150 mètres, l'indice de planéité du fond de vallée à résolutions multiples avec un rayon de 300 mètres et la pente du bassin versant avec un rayon de 206 mètres.

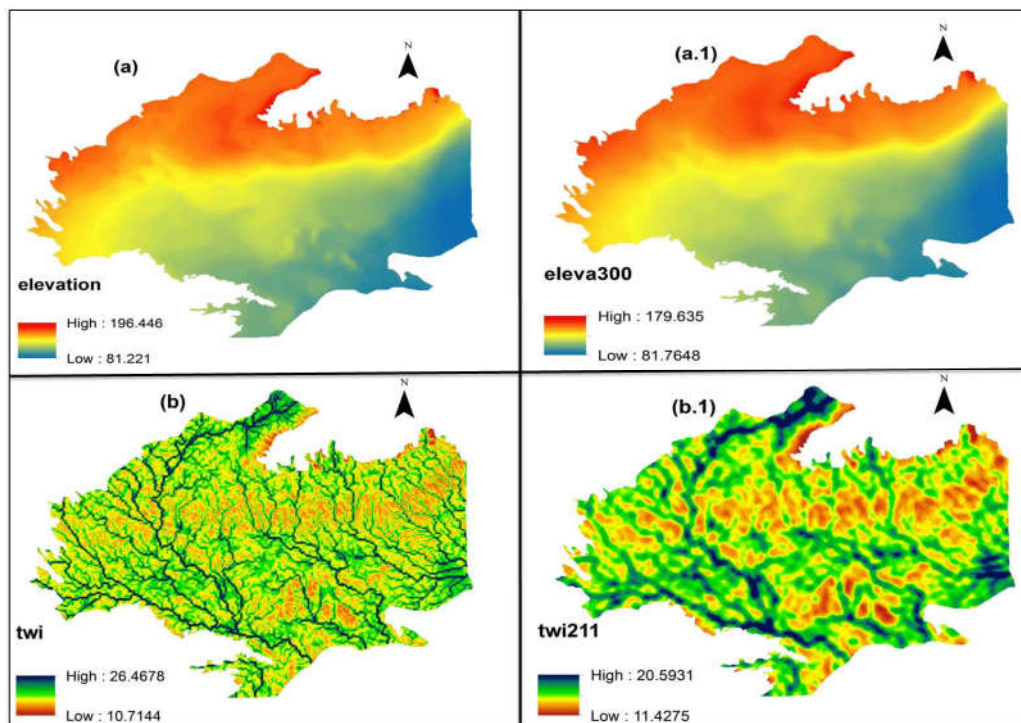


Figure 18. Exemples de dérivés MNA avant (a et b) et après l'application du filtre spatial (a.1 et b.1).

### 3.1.3. Comparaison des modèles

La forêt aléatoire a été le modèle le plus précis pour les deux niveaux taxonomiques (Figure 19). En effet, pour les groupes de sol de référence (RSGs), la valeur de kappa la plus élevée a été RF ( $k=0,45\pm0,07$ ) suivie par BCT ( $k=0,43\pm0,08$ ). De même, pour les RSGs avec qualificatifs principaux, le modèle RF a été le plus efficace, avec  $k=0,47\pm0,05$ , suivie par

BCT avec  $k = 0.44 \pm 0.05$ . Dans les deux cas, la valeur de kappa de RF a montré une concordance modérée entre les classes de sol observées et prédites. La figure 19 montre que la valeur de kappa pour la majorité des modèles utilisés (RF, BCT, MLP, NNET et SVMR), diminue lorsque les RSGs avec qualificatifs principaux basés sur 13 classes sont agrégés à un niveau de groupes de sols de référence moins détaillés avec 5 classes de sol.

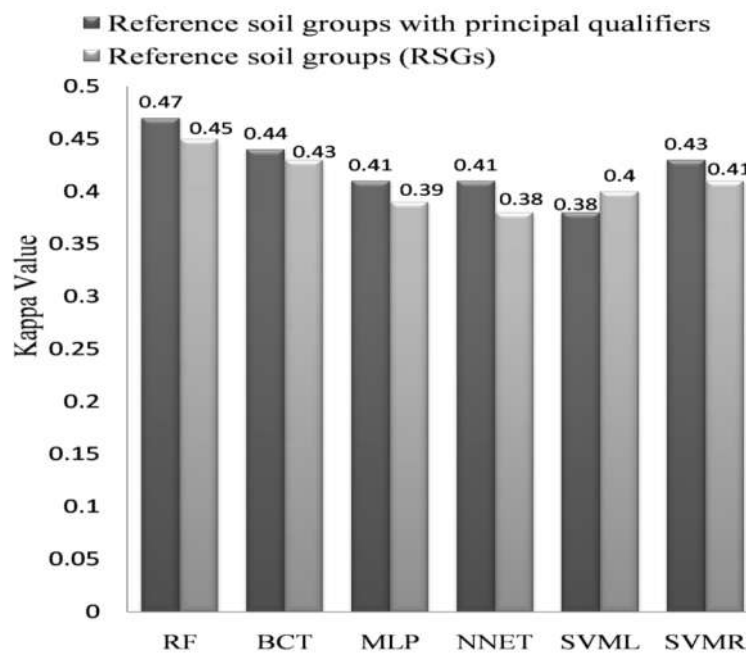


Figure 19. Les valeurs moyennes de kappa pour les deux niveaux taxonomiques de la WRB.

Dans le cas des RSGs avec qualificatifs principaux, l'ajustement de Bonferroni a montré une différence significative entre le modèle forêt aléatoire et les autres modèles. Ainsi, les valeurs de  $p$  entre RF et les autres modèles ont été inférieure à 0.003 (Tableau 6). Toutefois, pour les RSGs, la différence entre RF et BCT a été non significative, car la valeur de  $p$  a été supérieure à 0.003 (Tableau 7).

Tableau 6. Ajustement de Bonferonni au niveau des groupes de sol de référence avec qualificatifs principaux.

| p-value adjustment: bonferroni                 |           |           |           |           |           |            |
|------------------------------------------------|-----------|-----------|-----------|-----------|-----------|------------|
| Upper diagonal: estimates of the difference    |           |           |           |           |           |            |
| Lower diagonal: p-value for H0: difference = 0 |           |           |           |           |           |            |
| Kappa                                          | RF        | BCT       | MLP       | NNET      | SVML      | SVMR       |
| RF                                             |           | 2.702e-02 | 5.686e-02 | 5.688e-02 | 8.636e-02 | 4.020e-02  |
| BCT                                            | 1.826e-06 |           | 2.984e-02 | 2.986e-02 | 5.934e-02 | 1.318e-02  |
| MLP                                            | < 2.2e-16 | 2.274e-06 |           | 2.122e-05 | 2.950e-02 | -1.666e-02 |
| NNET                                           | < 2.2e-16 | 1.524e-05 | 1.00000   |           | 2.948e-02 | -1.668e-02 |
| SVML                                           | < 2.2e-16 | 8.582e-16 | 3.314e-07 | 1.387e-06 |           | -4.616e-02 |
| SVMR                                           | 2.971e-11 | 0.33434   | 0.02773   | 0.05111   | 3.396e-11 |            |

Tableau 7. Ajustement de Bonferonni au niveau des groupes de sol de référence (RSGs).

| p-value adjustment: bonferroni                 |           |           |          |          |           |           |
|------------------------------------------------|-----------|-----------|----------|----------|-----------|-----------|
| Upper diagonal: estimates of the difference    |           |           |          |          |           |           |
| Lower diagonal: p-value for H0: difference = 0 |           |           |          |          |           |           |
| Kappa                                          | RF        | BCT       | MLP      | NNET     | SVML      | SVMR      |
| RF                                             |           | 0.022257  | 0.060042 | 0.072002 | 0.053804  | 0.039920  |
| BCT                                            | 0.012077  |           | 0.037785 | 0.049745 | 0.031547  | 0.017663  |
| MLP                                            | 3.636e-10 | 0.001383  |          | 0.011960 | -0.006237 | -0.020121 |
| NNET                                           | 2.607e-12 | 4.398e-06 | 1.000000 |          | -0.018197 | -0.032082 |
| SVML                                           | 6.919e-11 | 0.008651  | 1.000000 | 0.369283 |           | -0.013884 |
| SVMR                                           | 2.660e-08 | 0.322552  | 0.285917 | 0.001453 | 0.788206  |           |

### 3.1.4. Sélection des principaux facteurs environnementaux

Le tableau 8 indique que le groupe de covariables produit par RFE (12 covariables) a été le plus précis, déterminé par la valeur de kappa la plus élevée avec  $k = 0.54$ , par rapport à l'utilisation de la totalité des covariables qui donnent un indice de kappa de 0.49. En plus, L'algorithme RFE a fourni les types de covariables qui ont déterminé la distribution des classes de sol dans la zone d'étude (Tableau 9; Annexe I)

Tableau 8. Comparaison de la performance entre l'ensemble des covariables et celles sélectionnée par RFE.

|                                   | Nombre des covariables | Précision | Kappa |
|-----------------------------------|------------------------|-----------|-------|
| Ensemble des covariables          | 24                     | 0.58      | 0.49  |
| Covariables sélectionnées par RFE | 12                     | 0.61      | 0.54  |

Dans la zone d'étude, les RSGs avec qualificatifs principaux ont correspondu à notre connaissance de la relation entre les facteurs environnementaux et la distribution des sols. À ce niveau taxonomique, le modèle RF a montré que l'élévation (elevation), l'élévation avec un rayon de 300 mètres (eleva300), qui reflète le contexte spatial, l'indice de taille de grain (gsi\_indx) et mrvbf ont été les facteurs les plus importants dans la différenciation des classes (Figure 20). Par ailleurs, la pente et les indices spectraux, c'est-à-dire gyps\_indx et sali\_indx n'ont pas eu une influence significative sur la séparation des sols.

L'élévation a permis de distinguer les Gleyic Solonchaks, les Gypsic Solonchaks et les Gypsic Fluvisols à des altitudes basses, comparativement aux Colluvic Regosols et aux Hyperskeletal Lithic Leptosols qui ont dominé les parties élevées du paysage (Figure 21.a). Pour l'indice de taille de grain, les sols proches de la valeur zéro ont été associés à des zones de palmeraies comme les Haplic Gypsisols et les Gypsic Sodique Solonchaks. Mais à mesure

que les valeurs ont augmenté, les sols ont été associés à des zones sans couverture végétale (Figure 21.b). En ce qui concerne le mrvbf, nous avons observé que les sols proches de la valeur zéro ont été fortement associés à des zones avec un relief élevé comme les Hyperskeletal Lithic Leptosols. Par contre, quand les valeurs de mrvbf ont augmenté, les sols ont été identifiés dans des zones graduellement plus plates (Figure 21.c).

Tableau 9. Covariables sélectionnées par la fonction d'élimination de variables par récursion (RFE).

| Source des données                                                                                              | Facteurs de<br>pédogenèse | Co-variables sélectionnées                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------------------------------------------------------------------------------------------------|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Modèle élévation numérique<br>( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> ) | Relief                    | élévation (elevation), élévation avec un rayon de 300 mètres (eleva300), pente avec un rayon de 300 mètres (slope300), indice de rugosité de terrain avec un rayon de 300 mètres (tri300), indice de classification du terrain pour les basses terres avec un rayon de 213 mètres (tcilow213), indice de planéité du fond de vallée à résolution multiple (mrvbf), indice de planéité du fond de vallée à résolution multiple avec un rayon de 300 mètres (mrvbf300), pente du bassin versant avec un rayon de 206 mètres (catchslpe2). |
| Landsat 5 TM<br>( <a href="https://earthexplorer.usgs.gov/">https://earthexplorer.usgs.gov/</a> )               | Matériau<br>parental      | bande 1/bande 7 (rat_7_1), indice de taille de grain (gsi_indx), indice gypsique (gyps_indx), indice de salinité 3 (sali_indx).                                                                                                                                                                                                                                                                                                                                                                                                         |

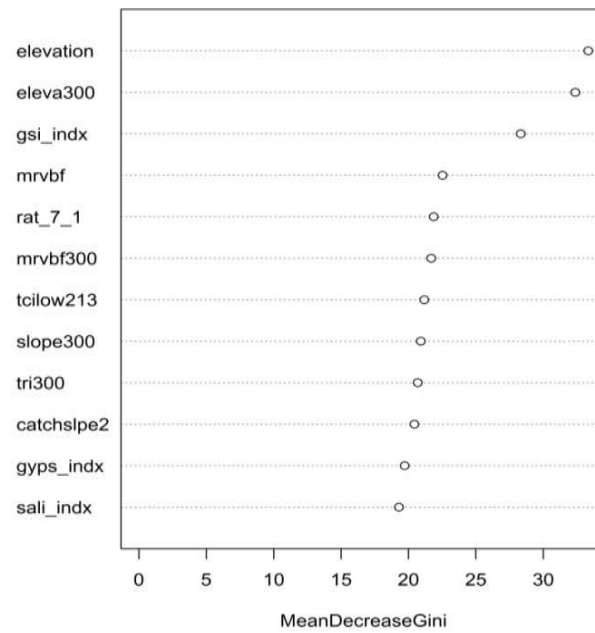


Figure 20. Importance des covariables utilisées par la forêt aléatoire au niveau des groupes de sols de référence avec qualificatifs principaux. Les abréviations des covariables sont expliquées dans le tableau 9.

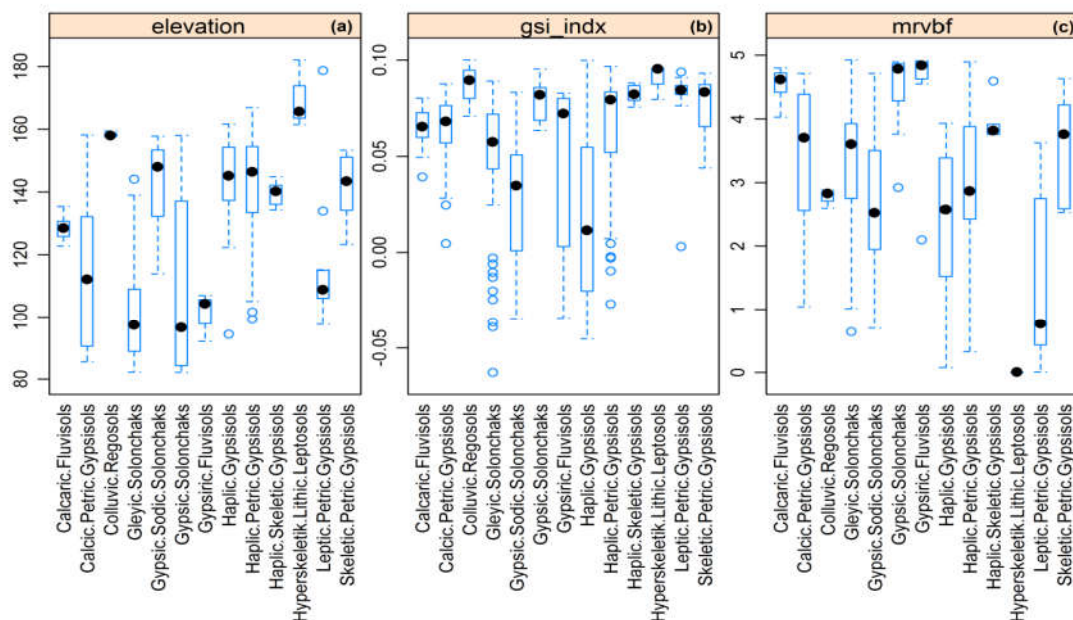


Figure 21. Distribution des groupes de sols de référence avec qualificatifs principaux en fonction de l'élévation (a), de l'indice de taille de grain (b) et de mrvbf (c).

### 3.1.5. Qualité de la carte modélisée

La carte des sols des RSGs avec qualificatifs principaux modélisée par RF a été la plus représentative de l'organisation spatiale des sols dans la zone d'étude (Figure 22). Les classes les plus cartographiées ont été les Haplic Petric Gypsisols, qui ont couvert une superficie d'environ 6377.04 hectares (29,21%), suivi des Gleyic Solonchaks avec 5141.07 ha (23,55%) et des Haplic Gypsisols avec 3209.40 ha (14,70%). Ces trois classes ont occupé 67,46% de la superficie totale de la région d'étude. Tandis que, les trois dernières classes (Colluvic Regosols, Skeletic Petric Gypsisols et Hyperskeletic Lithic Leptosols) n'ont pas dépassé les 1% de la superficie prédite (Tableau 10).

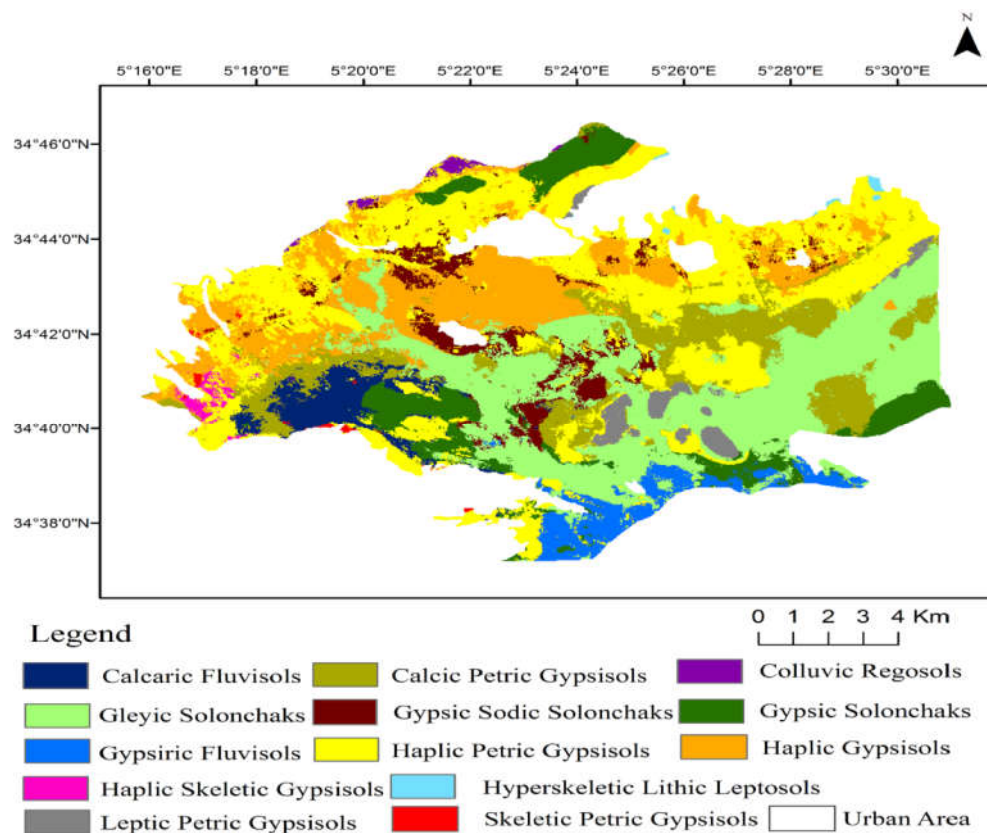


Figure 22. Carte des classes de sol de Zeb El Gherbi modélisée par la forêt aléatoire au niveau des groupes de référence du sol avec qualificatifs principaux.

La précision globale a montré que 64,44% des classes de sols ont été correctement classées. Les valeurs les plus élevées de la précision de l'utilisateur et du producteur ont été attribuées aux Colluvic Regosols et aux Gypsic Fluvisols avec 100% (Tableau 11). Cependant, ces chiffres doivent être interprétés avec prudence, car le nombre de sites de validation est faible. De même, les Leptic Petric Gypsisols ont été correctement prédits avec une précision de l'utilisateur de 100%, alors que la précision du producteur a été de 40%. Le plus mauvais score de précision de l'utilisateur a été attribué aux Calcic Petric Gypsisols, où 42,86% des classes prédites ont correspondu à ce type de sol. De plus, la précision du producteur dans cette classe a été très faible (33,33%). Le tableau 11 montre que les Haplic Petric Gypsisols (UA = 61,29%), les Gleyic Solonchaks (UA = 72,73%) et les Haplic Gypsisols (UA = 62,50%) ont une précision de l'utilisateur modérée, en dépit du grand nombre d'échantillons et de la superficie importante qu'ils occupent dans la région (Tableau 3; Tableau 10).

La carte de l'indice de confusion a montré qu'une grande partie de la zone d'étude a un faible indice de confusion. On peut toutefois visualiser quelques secteurs avec un indice relativement élevé (Figure 23).

Tableau 10. Superficie couverte par les 13 classes de sol prédite par le modèle forêt aléatoire.

| Groupes de sols de référence avec qualificatifs principaux | Superficie prédite par hectare (ha) | Superficie prédite (%) |
|------------------------------------------------------------|-------------------------------------|------------------------|
| Haplic Petric Gypsisols                                    | 6377.04                             | 29.21                  |
| Gleyic Solonchaks                                          | 5141.07                             | 23.55                  |
| Haplic Gypsisols                                           | 3209.40                             | 14.70                  |
| Calcic Petric Gypsisols                                    | 2310.93                             | 10.59                  |
| Gypsic Solonchaks                                          | 1612.98                             | 7.39                   |
| Gypsic Sodic Solonchaks                                    | 960.75                              | 4.40                   |
| Gypsic Fluvisols                                           | 802.98                              | 3.68                   |
| Calcaric Fluvisols                                         | 734.13                              | 3.36                   |
| Leptic Petric Gypsisols                                    | 407.7                               | 1.87                   |
| Haplic Skeletic Gypsisols                                  | 103.14                              | 0.47                   |
| Colluvic Regosol                                           | 83.79                               | 0.38                   |
| Skeletic Petric Gypsisols                                  | 48.24                               | 0.22                   |
| Hyperskeletic Lithic Leptosols                             | 39.51                               | 0.18                   |
| Total                                                      | 21831.66                            | 100                    |

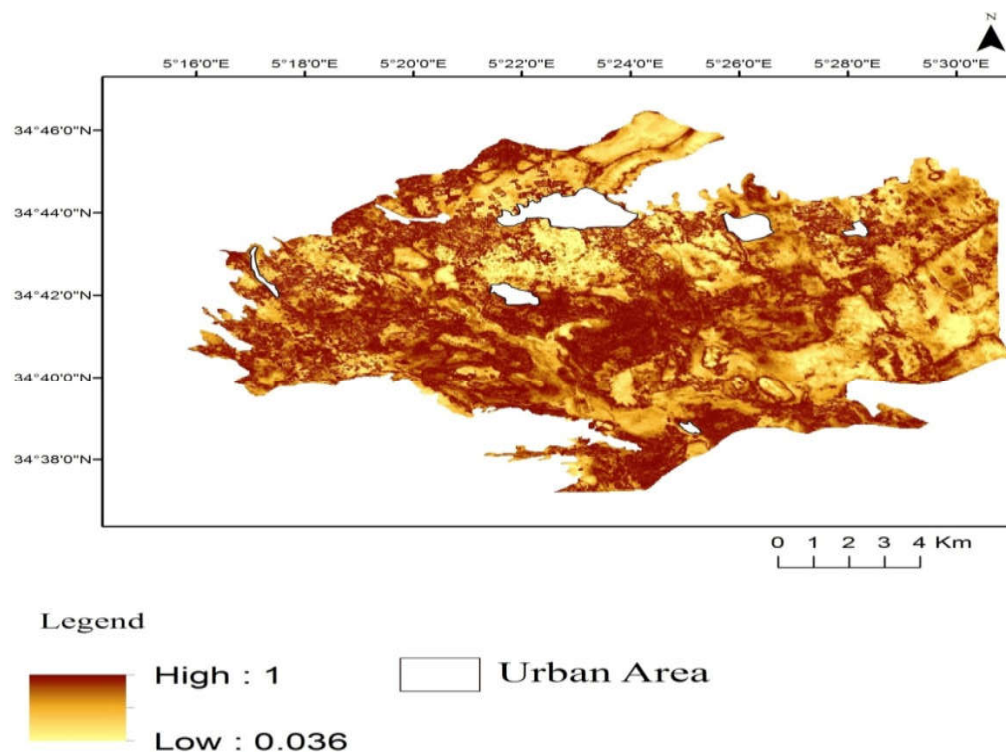


Figure 23. Carte de l'indice de confusion de Zeb El Gherbi au niveau des groupes de référence du sol avec qualificatifs principaux.

Tableau 11. Matrice de confusion des 13 classes de sol utilisant des données de validation externe (N = 135).

|                                   | Référence |          |          |           |          |          |          |           |           |          |          |          |          |                    |
|-----------------------------------|-----------|----------|----------|-----------|----------|----------|----------|-----------|-----------|----------|----------|----------|----------|--------------------|
| Prédiction                        | 1         | 2        | 3        | 4         | 5        | 6        | 7        | 8         | 9         | 10       | 11       | 12       | 13       | UA (%)             |
| Calcaric Fluvisols-1              | <b>4</b>  | 3        | 0        | 0         | 0        | 0        | 0        | 0         | 0         | 0        | 0        | 0        | 0        | 57.14              |
| Calcic Petric Gypsisols-2         | 0         | <b>6</b> | 0        | 2         | 0        | 1        | 0        | 0         | 4         | 0        | 0        | 1        | 0        | 42.86              |
| Colluvic Regosols -3              | 0         | 0        | <b>1</b> | 0         | 0        | 0        | 0        | 0         | 0         | 0        | 0        | 0        | 0        | 100                |
| Gleyic Solonchaks-4               | 0         | 5        | 0        | <b>24</b> | 2        | 0        | 0        | 1         | 0         | 0        | 0        | 1        | 0        | 72.73              |
| Gypsic Sodic Solonchaks -5        | 0         | 0        | 0        | 1         | <b>2</b> | 0        | 0        | 1         | 0         | 0        | 0        | 0        | 0        | 50                 |
| Gypsic Solonchaks -6              | 1         | 0        | 0        | 0         | 0        | <b>5</b> | 0        | 0         | 0         | 0        | 0        | 0        | 0        | 83.33              |
| Gypsic Fluvisols -7               | 0         | 0        | 0        | 0         | 0        | 0        | <b>3</b> | 0         | 0         | 0        | 0        | 0        | 0        | 100                |
| Haplic Gypsisols -8               | 0         | 1        | 0        | 0         | 5        | 0        | 0        | <b>20</b> | 5         | 0        | 0        | 1        | 0        | 62.50              |
| Haplic Petric Gypsisols -9        | 0         | 3        | 0        | 0         | 1        | 0        | 0        | 4         | <b>19</b> | 1        | 0        | 0        | 3        | 61.29              |
| Haplic Skeletic Gypsisols -10     | 0         | 0        | 0        | 0         | 0        | 0        | 0        | 1         | 0         | <b>1</b> | 0        | 0        | 0        | 50                 |
| Hyperskeletic Lithic Leptosols-11 | 0         | 0        | 0        | 0         | 0        | 0        | 0        | 0         | 0         | 0        | <b>0</b> | 0        | 0        | —                  |
| Leptic Petric Gypsisols -12       | 0         | 0        | 0        | 0         | 0        | 0        | 0        | 0         | 0         | 0        | 0        | <b>2</b> | 0        | 100                |
| Skeletic Petric Gypsisols -13     | 0         | 0        | 0        | 0         | 0        | 0        | 0        | 0         | 0         | 0        | 0        | 0        | <b>0</b> | —                  |
| <b>PA (%)</b>                     | 80        | 33.33    | 100      | 88.89     | 20       | 83.33    | 100      | 74.07     | 67.86     | 50       | —        | 40       | 0        | <b>OA= 64.44 %</b> |

— = Non estimé; PA =Précision du producteur; UA= Précision de l'utilisateur ; OA=Précision globale; N= Nombres des données de sol

### 3.2. Discussion

#### 3.2.1. Modèle optimal

La supériorité de RF (Figure 19) a été confirmée par d'autres chercheurs qui utilisent des données ponctuelles pour la modélisation (Heung et al., 2017; Levi, 2017). Brungard et al. (2015) attribuent cette performance à l'utilisation de RF comme fonction de contrôle dans l'algorithme d'élimination des caractéristiques récursives (RFE). La précision de RF a été améliorée avec l'augmentation du nombre de classes (Figure 19), ce qui corrobore les résultats d'autres auteurs qui concluent que la précision du modèle augmente avec des classes taxonomiques plus détaillées (Mirakzahi et al., 2018 ; Zeraatpisheh et al., 2017). Pahlavan-Rad et al. (2016) expliquent que RF est plus appropriée pour l'identification de la relation complexe entre les sols et l'environnement à des niveaux plus détaillés, en particulier lorsque la relation n'est pas linéaire.

De manière générale, le concept sol-paysage (pédopaysage) semble le plus approprié pour comprendre la relation entre la répartition des sols et les facteurs environnementaux dans la région. Ce concept est formalisé par le modèle RF, qui montre qu'à l'échelle du paysage, les sols varient selon l'approche déterministe de type SCORPAN.

#### 3.2.2. Variables environnementales contrôlant la distribution des classes de sol

Les résultats du tableau 8 ont montré la performance de RFE dans la sélection d'un groupe optimal de prédicteurs déterminés en mesurant les indicateurs de performance, par rapport à l'ensemble des covariables sélectionnées en fonction de notre connaissance de la relation sol-paysage. Ceci est cohérent avec les résultats de Brungard et al. (2015) qui montrent que les covariables sélectionnées par le pédologue sur le terrain sont moins précises que les covariables choisis automatiquement par l'algorithme de RFE. Brungard et al (2015) expliquent que le choix des covariables par les pédologues n'a pas fourni les modèles les plus

précis en raison de l'incapacité à trouver les covariables les plus favorables pour la modélisation des sols. Ces derniers sont formés à partir de facteurs de pédogénèse agissant sur de longues périodes et à différentes échelles. À cela s'ajoute la subjectivité du pédologue dans le choix des covariables face à un algorithme capable d'éliminer toute les informations non pertinentes.

Dans la région d'étude, le rôle joué par le relief (topographie) et le matériau parental dans la détermination de la répartition des sols a été prédominant. La variabilité des sols à l'échelle du paysage est liée à la nature des processus mise en jeu par l'action de la topographie, en particulier l'élévation. Cet ensemble de processus peut être classé comme suit :

1) les processus d'érosion et de dépôt qui sont des phénomènes générés par l'eau et/ou le vent qui enlèvent et transportent les matériaux solides de la surface de certains sols et les déposent à la surface d'autres sols. En effet, dans la partie supérieure du paysage où le processus d'érosion est dominant, les Colluvic Regosols et Hyperskeletal Lithic Leptosols ont été identifiés. Tandis que dans la partie inférieure du paysage, qui a tendance à recevoir plus de sédiments on trouve les Gypsic Fluviosols, Gypsic Solonchaks et Gleyic Solonchaks (Figure 21.a) ;

2) les processus de transport des éléments géochimiques sous forme soluble, notamment le gypse et les sels les plus solubles qui précipitent selon leur degré de dissolution dans le paysage, cela conduit à la formation des Gypsisols et des Solonchaks.

Nos résultats concordent avec ceux de Chagas et al. (2017) et Odgers et al. (2014), qui confirment l'influence de l'élévation sur la distribution spatiale des sols. De plus, Chaney et al. (2016) notent le pouvoir prédictif de l'élévation même lorsque le relief n'est pas aussi prononcé.

L'importance de l'indice de taille de grain (GSI) peut être liée au fait que cet indice spectral permet de distinguer entre les sols sous couverture végétale et les sols nus (Figure 21.b). En

plus, il permet d'identifier la composition granulométrique du matériau parental. Dans la région d'étude, l'indice de taille de grain représente la covariable liée au matériau parental qui contient une quantité importante de sables fins dans l'horizon superficiel (Xiao et al., 2006).

Le mrvbf qui identifie dans le paysage des zones de dépôt sédimentaire relativement plates à différents niveaux (Mckenzie et Gallant 2007) confirme l'influence des processus d'érosion et de dépôt dans la formation des sols de la région. La covariable mrvbf a séparé les sols situés sur des terrains élevés, issus de processus d'érosion, de ceux formés sur des alluvions et localisés dans des zones plates, tels que les Calcaric Fluviosols, Gypsic Fluvisols et Gypsic Solonchaks (Figure 21c). L'importance de mrvbf en tant que facteur explicatif de la distribution des classes de sols est citée dans des études menées dans des régions arides (Roecker et al., 2010; Taghizadeh-Mehrjardi et al., 2014; Jafari et al., 2012; Afshar et al., 2018).

Les résultats obtenus ont montré que l'indice gypsiq et l'indice de salinité ont une faible influence sur la modélisation des sols, en raison de l'absence de différence significative dans la réflectance spectrale entre la surface des sols gypseux et salins (Figure 20). Ces sols sont souvent situés dans des positions topographiques similaires. En outre, la faible influence de la pente dans la différenciation des classes de sol s'explique par l'existence d'un relief peu accentué dans la région (Figure 20).

### 3.2.3. Estimation des classes de sol

Dans la région d'étude parmi les groupes de référence du sol (RSGs), les Gypsisols ont été les plus dominants avec 60.09% des données ponctuelles du sol suivis des Solonchaks avec 32.19%. Tandis que, pour les RSGs avec qualificatifs principaux, les Haplic Petric Gypsisols, les Gleyic Solonchaks et les Haplic Gypsisols ont été les plus répandus avec un total de 59.44% des observations du sol (Tableau 3). La dominance des Gypsisols et des Solonchaks s'explique, par l'aridité du climat qui favorise la présence de vaste ensemble de sols contenant

des sels, par un relief faiblement accidenté (pente faible) où sont généralement localisés des encroûtements gypseux (Vieillefon, 1976) et qui aussi favorise l'accumulation des sels. De plus, la nature du matériau parental riche en gypse et en sels est également une source de formation de ces sols dans la région.

L'amélioration de la qualité de l'information sur les sols passe par un modèle de RF plus performant, la figure 19 a montré une valeur de kappa modérée qui est due d'une part à l'incapacité du modèle à distinguer entre les classes de sol, en raison d'une confusion entre les types de sol (Jafari et al., 2013), et d'autre part aux déséquilibres dans la répartition du nombre d'échantillons entre les classes de sol, limitant ainsi la précision du modèle (Tableau 3). En effet, la performance des algorithmes d'apprentissage machine semble être affectée par le problème de déséquilibre des classes de sol, qui sont plus efficaces avec des données d'apprentissage également réparties (Sun et al., 2009; Ali et al., 2015). Dans cette étude, le déséquilibre des données entre les classes de sols est lié à l'importance en termes de superficie des Gypsisols et des Solonchaks par rapport aux Fluvisols, Regosols et Leptosols. Pour réduire le déséquilibre des données entre les classes de sols, Brungard et al. (2015) et Levi (2017) proposent une combinaison de méthodes qui consiste à augmenter le nombre d'échantillons dans les classes sous représentées (sur-échantillonnage), à réduire le nombre d'échantillons des classes majoritaires et/ou à regrouper les classes proche des classes majoritaires (sous échantillonnage).

Le tableau 11 a montré une précision globale de 64,44%, ce qui signifie que pour 64% de la surface cartographiée, les classes de sol ont été correctement prédites. En effet, une précision globale comprise entre 50 et 60% peut être considérée comme un résultat adéquat pour une carte construite par la cartographie numérique du sol (Kempen et al., 2009 ; Kempen et al., 2012). De plus, la valeur de la précision globale a été supérieure à la précision globale rapportée par Dornik et al. (2017) qui est de 48% avec 50 profils et 10 classes de sol prédites

par le modèle RF. Deux facteurs importants contribuent à la supériorité de notre résultat, premièrement la taille de l'échantillon par rapport à l'ensemble de la zone d'étude (i.e., 2.13/km<sup>2</sup> pour notre étude versus 0.54/km<sup>2</sup> pour l'étude de Dornik). En effet, une faible densité d'échantillonnage réduit la précision des modèles (Taghizadeh-Mehrjardi et al., 2014). Deuxièmement, l'utilisation des covariables multi-échelles pour modéliser le sol, ce qui augmente la performance du modèle (Behrens et al., 2010 ; Levi, 2017 ; Miller et al., 2015). En revanche, notre résultat a été inférieur à celui de Barthold et al. (2013) qui ont obtenu une précision globale de 71,1% avec 52 profils de sols et 9 classes de sols. Cela peut s'expliquer d'une part, par le fait que Barthold et al. (2013) utilisent des covariables catégorielles (la carte géologique et la carte de l'utilisation des terres) pour calibrer le modèle, ce qui augmente la performance du modèle (Pahlavan-Rad et al., 2016 ; Zeraatpisheh et al., 2017), et d'autre part la méthode d'échantillonnage aléatoire stratifié utilisée par Barthold et al. (2013) pour échantillonner les sols est considérée comme robuste et efficace (Biswas et Zhang, 2018; Brus et al., 2011 ; Hengl et al., 2007).

En ce qui concerne la précision des classes de sol, les Leptic Petric Gypsisols (UA = 100%) ont été correctement prédits, probablement liés à leur localisation géographique et topographique identifiable dans le paysage. En fait, les Leptic Petric Gypsisols sont situés dans la partie centrale de la région sur les collines et vers le nord sur le glacis (Figure 22). Cependant, la précision du producteur a été de 40%, ce qui signifie qu'une grande partie des zones où ces sols sont observés ne sont pas cartographiées comme des Leptic Petric Gypsisols. La précision de l'utilisateur et la précision du producteur des Calcic Petric Gypsisols ont été faibles en raison de leurs emplacements avec d'autres classes sur des paysages identiques (Gleyic Solonchaks, Gypsic Solonchaks, Leptic Petric Gypsisols et Haplic Petric Gypsisols). Dans la zone d'étude, les Haplic Petric Gypsisols, les Gleyic Solonchaks et les Haplic Gypsisols ont été les plus importants en termes de superficie

(Tableau 10, Figure 22), ceci est liée à la grande taille de l'échantillon du sol (Pahlavan Rad et al., 2016; Pahlavan Rad et al., 2014). Toutefois, nos résultats ont montré que la taille de l'échantillon n'a pas été suffisante pour assurer une grande précision des Haplic Petric Gypsisols, des Gleyic Solonchaks et des Haplic Gypsisols, alors que d'autres études indiquent le contraire (Barthold et al., 2013; Taghizadeh-Mehrjardi et al., 2014). Cela peut être attribué au fait que les Haplic Petric Gypsisols et les Haplic Gypsisols sont associés à d'autres classes de sol sur différentes formes de relief, comme les glacis et les terrasses. Ainsi, Lorenzetti et al. (2015) indiquent que les classes de sol les plus fréquentes et que l'on trouve dans des environnements contrastés peuvent avoir une faible précision. De plus, les Haplic Petric Gypsisols, les Gleyic Solonchaks et les Haplic Gypsisols présentent des similarités taxonomiques avec les sols du même groupe, ce qui peut réduire la précision de ces classes (Minasny et McBratney, 2007).

#### 3.2.4. Analyse spatiale des sols

L'évaluation visuelle de la carte élaborée par le modèle RF montre que l'approche déterministe SCORPAN permet de modéliser et d'interpréter l'organisation spatiale des sols dans la région. La carte indique que les sols gypseux de type Skeletic Petric Gypsisols et Haplic Skeletic Gypsisols caractérisés par une quantité importante de cailloux et de galets arrondis sont déterminés à l'ouest de la région sur des traces d'anciens talwegs et des lits d'oueds (Figure 22). La présence d'éléments grossiers dans ces sols indiquent que le gypse est transporté par des eaux superficielles à la suite d'un mouvement de masse et/ou d'un ruissellement avec dissolution et précipitations ultérieures ou bien lessivage (Casby-Horton et al., 2015). Les Haplic Gypsisols sont le plus souvent cartographiés en association avec les Haplic Petric Gypsisols localisés à mi-altitude sur des glacis subhorizontaux et sur des anciennes terrasses alluviales stables. Au fur et à mesure que l'élévation diminue, on observe le long d'un axe de ruissellement nord-sud un passage assez net des Gypsisols, c'est-à-dire les

Haplic Gypsisols et les Haplic Pertic Gypsisols, vers les Gleyic Solonchaks situés à une basse altitude sur des terrasses alluviales où des tâches d'hydromorphie et des quantités importantes de sels existent. Cela peut s'expliquer par l'accumulation des eaux de ruissellement et/ou par la présence d'une nappe phréatique saline. Les Gypsic Solonchaks sont formés à des altitudes basses sur des zones relativement plates avec un mauvais drainage, probablement en contact avec une nappe phréatique saline ou bien à cause de l'accumulation des eaux superficielles chargé en sels. La carte des classes de sol a montré la présence des Leptic Petric Gypsisols au centre de la région sur les anciennes collines encroûtées, et au nord sur les glacis. La présence d'un encroûtement gypseux dans ces sols, caractérisés par la présence de fossiles est liée à la genèse des glacis et se différencie par son ancienneté des autres encroûtements gypseux localisés dans la région (Figure 22). Les Calcaric Fluvisols et les Gypsic Fluvisols qui ont comme substrat des alluvions sont déterminés sur des terrasses et dans les zones d'épandage des grands oueds tel que l'oued Djedi. Les Hperskeletalic Lithic Leptosols qui sont en contact avec une roche continue sont cartographiés sur une altitude élevée avec une pente forte et abrupte. En bas de versants et à une altitude élevée sont prédit les Colluvic Regosols qui sont formés sur du matériau colluvial issu du transport par ruissèlement sur une pente douce. Le plus souvent, les Gypsic Sodique Solonchaks sont cartographiés à moyenne altitude dans les palmeraies en association avec les Haplic Gypsisols et les Haplic Pertic Gypsisols, l'origine de la salinité de ces sols s'explique par l'irrigation des palmiers avec des eaux salées. Les Calcic Petric Gypsisols sont déterminés plus fréquemment à basse altitude et dans des zones plates en association avec les Gleyic Solonchaks et les Calcaric Fluvisols. La formation de l'encroûtement des Calcic Petric Gypsisols peut être expliquée par la présence d'une nappe phréatique qui joue un rôle important dans la pédogénèse des sols de la région.

### *3.2.5. Importance de l'approche multi-échelle dans la modélisation des sols*

La perception de la dimension spatiale des variables explicatives et la prise en compte de l'existence dans le paysage de plusieurs niveaux d'organisation partiellement imbriqués sont des enjeux importants pour comprendre l'organisation spatiale des sols (Behrens et al., 2010). Selon Grinand et al. (2008), le filtrage spatial est un moyen efficace d'intégrer diverses échelles de variables explicatives. Dans cette étude la comparaison visuelle entre les attributs du terrain bruts et les attributs du terrain dérivés des filtres spatiaux a montré que la méthode appliquée pour produire des attributs du terrain à différentes échelles et qui est basée sur le calcul de la portée du semi-variogramme (Figure 18b.1) est plus robuste que les attributs du terrain bruts (Figure 18b). En effet, la taille de la fenêtre des attributs du terrain bruts dérivés à partir du MNA est donnée par défaut et ne repose sur aucune connaissance solide sur l'échelle de la variabilité spatiale des attributs du terrain dans le paysage.

L'approche multi-échelle, qui consiste à intégrer le contexte spatial en utilisant une combinaison des attributs du terrain à des échelles d'analyse différentes produit un modèle performant. Ainsi, Chaplot et Walter (2007) expliquent que l'approche multi-échelle est nécessaire non seulement pour détecter le schéma de répartition des sols mais aussi pour développer des modèles de prédiction à haute résolution dans le but de caractériser des grandes superficies. Dans la zone d'étude, l'importance de l'approche multi-échelle réside dans la complexité de l'association entre le sol et le paysage. Il est vrai que les facteurs de pédogénèse influencent la formation des sols à différentes échelles (Behrens et al., 2010). À titre d'exemple, grâce à la dimension multi-échelle de l'élévation, les processus d'érosion et de dépôt dans la région déterminent les types de sol à différents niveaux, depuis la toposéquence jusqu'au paysage. Ces processus sont responsables de la circulation des matériaux et des éléments chimiques contribuant à la différenciation des sols. Par conséquent, la combinaison

de covariables à différentes échelles peut aider à améliorer les modèles quantitatifs de pédogenèse des sols (Levi, 2017).

### *3.2.6. Prédiction de l'incertitude*

Certaines parties de la zone d'étude ont été représentées par un indice de confusion relativement élevé (Figure 23), cela concorde avec les résultats d'autres études de modélisation des classes de sol (Brungard et al., 2015 ; Levi, 2017, Odgers et al., 2014). Dans notre cas, ces valeurs élevées peuvent être expliquées par ;

- 1) La taille de l'échantillon d'apprentissage, qui avec un total de 331 données n'est pas suffisante pour réduire l'incertitude dans certaines parties de la région caractérisée en général par un faible relief. Selon Somarathna et al. (2017), l'incertitude du modèle diminue avec l'augmentation de la taille des échantillons d'apprentissage ;
- 2) La précision des covariables, en effet le modèle numérique d'altitude brut présente une incertitude verticale intrinsèque au type de MNA utilisé dans notre cas, le SRTM. Cette incertitude verticale se propage aux différents attributs du terrain ;
- 3) La relation entre les classes de sol, qui portent des noms différents mais peuvent être proches du point de vue taxonomique (Chaney et al., 2016). En effet, la WRB est basée sur le principe que les sols dans le paysage ne sont pas des entités distinctes suffisamment définies comme les espèces animales ou végétales, ils forment un ensemble continu dans lequel on trouve toutes les transitions. Contrairement aux classifications hiérarchiques telle que la soil taxonomy, la WRB est un système référentiel qui définit plusieurs intergardes entre les groupes de sol. De plus, la classification des sols par la WRB est basée sur des caractéristiques de diagnostic mesurables et observables sur le terrain (horizons, propriétés et matériaux). Ces caractéristiques de diagnostic peuvent être associées entre les classes de sol,

notamment au niveau des qualificatifs principaux. À cet égard, 6 des 13 classes de sol appartiennent aux Gypsisols et 3 des 13 classes sont en relation avec les Solonchaks (Tableau 11). L'incertitude liée à l'existence d'une similarité taxonomique entre les sols peut être réduite en introduisant dans le modèle la distance taxonomique, qui est basée sur le calcul de la distance euclidienne entre les classes de sol (Minasny et McBratney, 2007; Hengl et al., 2007) ;

4) Les problèmes liés au déséquilibre des données de sol entre les classes. Certaines classes de sol ont été représentées en petit nombre dans les données de formation utilisées, ce qui réduit la capacité des modèles à identifier correctement les classes de sol (Levi, 2017).

## **CONCLUSION GENERALE**

Ce travail de recherche a permis d'aborder la variabilité spatiale des pédopaysages arides dans la région de Zeb El Gherbi. Ainsi, l'étude a montré que la variabilité spatiale des sols peut être formalisée par une approche déterministe de type SCORPAN. De plus, l'étude nous a permis de décrire et de cartographier l'organisation spatiale des classes de sols en comparant six modèles d'apprentissage machine. Les principales conclusions tirées sont les suivantes :

- Le modèle forêt aléatoire a été le plus performant pour les deux niveaux taxonomiques des sols de la WRB.
- La précision du modèle forêt aléatoire a augmenté avec le niveau taxonomique le plus détaillé, c'est-à-dire. les groupes de sol de référence avec les principaux qualificatifs. Ce modèle a la capacité d'identifier la relation complexe et non linéaire entre les facteurs environnementaux et les sols.
- Les covariables optimales retenues par la fonction d'élimination de variables par récursion (RFE) ont été les plus précises par rapport à l'ensemble des covariables que nous avons sélectionnées.
- Le relief à travers l'élévation et le matériau parental représenté par la covariable indice de taille de grain ont été les principaux facteurs environnementaux contrôlant la distribution spatiale des sols dans la région.
- L'élévation est la force motrice de différents processus tels que les processus d'érosion et de dépôt de matériaux solides et le transport d'éléments géochimiques solubles, qui à leur tour contrôlent l'organisation des sols à l'échelle du paysage.
- L'intégration du contexte spatial par l'utilisation de filtres spatiaux adaptés a permis de produire des attributs du terrain à différents échelles. Ces derniers ont permis d'intégrer les relations spatiales dans le modèle. À eux seuls, les attributs du terrain

bruts ne sont pas en mesure de saisir les différents processus qui contrôlent la formation des sols à différents niveaux.

- La présence des Gypsisols en association avec les Solonchaks est représentative de la région d'étude. Ainsi, les Haplic Petric Gypsisols, Gleyic Solonchaks et Haplic Gypsisols ont été identifiés comme les sols les plus répandus. La dominance des Gypsisols et des Solonchaks s'expliquent, par l'aridité du climat qui favorise la présence de vaste ensemble de sols contenant des sels, par un relief faiblement accidenté ainsi que par la nature des eaux et du matériau parental riches en gypse et en sels.
- La localisation topographique contrastée et la similarité taxonomique des Haplic Petric Gypsisols, Gleyic Solonchaks et Haplic Gypsisols avec d'autres classes de sol ont réduit la précision de ces classes, en dépit du nombre élevé des échantillons du sol.
- L'existence d'une incertitude de prédiction élevée dans quelques parties de la zone d'étude est probablement liée à la taille de l'échantillon d'apprentissage, à l'incertitude verticale intrinsèque aux covariables utilisées, à l'existence d'une relation taxonomique entre les classes de sol et au déséquilibre des données entre les classes de sol.

L'information spatiale sur les sols est de plus en plus recherchée pour diverses activités agricoles et environnementales, cependant elle est souvent rare dans les régions désertiques. Ainsi l'utilisation de la cartographie numérique du sol peut fournir, grâce à la gratuité des covariables environnementales via internet (MNA et images satellitaires) et à l'utilisation de différents modèles d'apprentissage machine, les outils nécessaires pour améliorer et accélérer la cartographie des sols dans les zones sahariennes d'Algérie. Les travaux futurs devront être axés sur l'utilisation d'autres techniques modernes d'information spatiale, telles que les modèles d'apprentissage profond (deep learning) et la mise au point de nouvelles variables

explicatives fondées sur des modèles numérique d'altitude et des images satellitaires multispectrales à haute résolution, qui peuvent réduire l'incertitude et améliorer la modélisation des sols dans les régions arides.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Afshar, A.F., Ayoubi, S., Jafari, A. 2018. The Extrapolation of Soil Great Groups Using Multinomial Logistic Regression at Regional Scale in Arid Regions of Iran. *Geoderma* 315 : 36–48. <https://doi.org/10.1016/j.geoderma.2017.11.030>.
- Aidaoui S. 1994. Ressource en eau et aménagement hydro-Agricole dans la région de Biskra "Ziban" (Algérie). Thèse Doctorat, université de Nancy II. 327 p.
- Ali, A., Shamsuddin S.M., Ralescu, Anca L., 2015. Classification with Class Imbalance Problem: A Review. *Int. J. Advance Soft Compu. Appl* 7 (3).
- ANRH., 2006. Etude pédologique par télédétection dans la wilaya de Biskra. Rapport de l'Agence Nationale de Ressources Hydrauliques. Département de pédologie, Alger, Algerie.
- Assami. T, Hamdi-Aissa. B, Haltim. A, Mostefaoui. T, Rouhana. H. 2014. Apport de la télédétection dans l'étude de la relation entre états de surface et pédopaysage en milieu aride : exemple de la région d'El Outaya (Biskra, Algérie). *Algerian Journal of Arid Environment*, 59 vol. 4, n° 1, 53-59.
- Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma* 151, 338–350. <https://doi.org/10.1016/j.geoderma.2009.04.022>.
- Baize D., Girard M.C. 2008. Référentiel pédologique. Eds. Quae, Paris, 405 p.
- Barthold,F.K.,Wiesmeier,M.,Breuer, L., Frede, H.-G., Wu, J., Blank, F.B., 2013. Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. *J. Arid Environ.* 88, 194–205. <https://doi.org/10.1016/j.jaridenv.2012.08.004>.
- Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. *J. Plant Nutr. Soil Sci.* 168, 21–33. <https://doi.org/10.1002/jpln.200421414>.
- Behrens, T., Zhu, A.-X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* 155, 175–185. <https://doi.org/10.1016/j.geoderma.2009.07.010>.
- 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).
- Bland, j. M., Altman, D.G., 1995. Multiple significance tests: The Bonferroni method. *BMJ* 310, 170. <https://doi.org/10.1136/bmj.310.6973.170>.

- Bock, M., Böhner, J., Conrad, O., Köthe, R., Ringeler, A., 2007. Methods for creating functional soil databases and applying digital soil mapping with SAGA GIS, in: Status and Prospect of Soil Information in South-Eastern Europe: Soil Databases, Projects and Applications. EUR 22646 EN Scientific and Technical Research Series, Office for Official Publications of the European Communities. pp. 149–162.
- Breunig, F.M., Galvao, L.S., Foramggio A.R. 2008. Detection of sandy soil surface using ASTER- derived reflectance, emissivity and elevation data: potential for the identification of land degradation. *Int. J. remote sensing*, 29(24-276), 1833-1840. <https://doi.org/10.1080/01431160701851791>.
- Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma* 264, 256–274. <https://doi.org/10.1016/j.geoderma.2015.05.017>.
- Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma* 239–240, 68–83. <https://doi.org/10.1016/j.geoderma.2014.09.019>.
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *Eur. J. Soil Sci.* 62, 394–407. <https://doi.org/10.1111/j.1365-2389.2011.01364.x>.
- Burrough, P.A., Van Gaans, P.F.M., Hootsmans, R., 1997. Continuous classification in soil survey: Spatial correlation, confusion and boundaries, in: *Geoderma*. pp. 115–135. [https://doi.org/10.1016/S0016-7061\(97\)00018-9](https://doi.org/10.1016/S0016-7061(97)00018-9).
- Casby-Horton, S., Herrero, J., Rolong, N.A., 2015. Gypsum Soils—Their Morphology, Classification, Function, and Landscapes. *Adv. Agron.* 130, 231–290. <https://doi.org/10.1016/BS.AGRON.2014.10.002>
- Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., Fealy, R., 2013. Are fine resolution digital elevation models always the best choice in digital soil mapping? *Geoderma*. <https://doi.org/10.1016/j.geoderma.2012.11.020>
- Chagas, C.S., Helena, S.K.P., Waldir C.J., Lúcia, H.C.A., Nilson, R.P., Silvio, B.B., 2017. Data mining methods applied to map soil units on tropical hillslopes in Rio de Janeiro, Brazil. *Geoderma Regional*, Digital soil mapping across the globe, 9: 47-55. <https://doi.org/10.1016/j.geodrs.2017.03.004>.

- 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., Walter, C., Curmi, P., 2000. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma* 97, 405–422. [https://doi.org/10.1016/S0016-7061\(00\)00048-3](https://doi.org/10.1016/S0016-7061(00)00048-3).
- Chaplot, V., Walter, C., 2007. Chapter 38. Improving the spatial prediction of soils at local and regional levels through a better understanding of soil-landscape relationships: soil hydromorphy in the Armorican Massif of Western France. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), *Digital Soil Mapping, an introductory perspective. Developments in soil science*, vol. 31. Elsevier, Amsterdam, pp. 507–522.
- Chávez, P.S.J., 1996. Image-Based Atmospheric Corrections - Revisited and Improved. *Photogramm. Eng. Remote Sensing*. <https://doi.org/0099-1112/96/6209-1025>.
- CNULD. 2011. Convention des Nations unies sur la lutte contre la désertification. Désertification : une synthèse visuelle. Rapport de synthèse. 51 p.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B).
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* 8, 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>.
- CPCS., 1967. Classification des sols. Commission de Pédologie et de Cartographie des Sols, France.
- DEMRH., 1972. Etude agro-pédologique de la zone d'extension possible des palmeraies de la région de Tolga. Rapport de la Direction des Etudes de Milieu et de la Recherche Hydraulique. Sous direction des ressources en sols. Algérie.
- DEMRH., 1973. Etude agro-pédologique des régions El Amri-Tolga et Ourellal. Rapport de la Direction des Etudes de Milieu et de la Recherche Hydraulique. Sous direction des ressources en sols. Algérie.

- Ding, J.L., WU, M.C., Tiyyip, T. 2011. Study of soil salinization information in arid region using remote sensing technique. *Agricultural Science in China*, 10(3), 404-411. [https://doi.org/10.1016/S1671-2927\(11\)60019-9](https://doi.org/10.1016/S1671-2927(11)60019-9).
- Dornik, A., Drăguț, L., Urdea, P., 2017. Classification of soil types using geographic object-based image analysis and Random Forest. *Pedosphere*. [https://doi.org/10.1016/S1002-0160\(17\)60377-1](https://doi.org/10.1016/S1002-0160(17)60377-1).
- Douaoui, A.E.K., Nicolas, H., Walter, C., 2006. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* 134, 217–230. <https://doi.org/10.1016/j.geoderma.2005.10.009>.
- DPAT., 2015. Monographie de la wilaya de Biskra, 209 p
- Duchaufour Ph. 1998. Réflexion sur les classifications des sols. *Etude et Gestion des sols*, 5, 3, 201-205.
- Duda, R.O., P.E. Hart, and D.G. Stork. 2001. *Pattern classification*. John Wiley & Sons, New York, NY.
- Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. *Trends Ecol. Evol.* 17, 177–183. [https://doi.org/10.1016/S0169-5347\(02\)02496-5](https://doi.org/10.1016/S0169-5347(02)02496-5).
- ESRI, 2016. ArcGIS Desktop: release 10.5. Environmental Systems Research Institute, Redlands, CA.
- Evans, D.M., Hartemink, A.E., 2014. Digital soil mapping of a red clay subsoil covered by loess. *Geoderma* 230–231, 296–304. <https://doi.org/10.1016/j.geoderma.2014.03.013>.
- FAO., 2005. Profil Nutritionnel de l'Algérie. Division de l'Alimentation et de la Nutrition. Algerie. Food and Agriculture Organisation.
- Franklin, R.B., Mills, A.L., 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiol. Ecol.* 44, 335–346. [https://doi.org/10.1016/S0168-6496\(03\)00074-6](https://doi.org/10.1016/S0168-6496(03)00074-6)
- Gaddas, F., 2001. Proposition d'une méthode de cartographie des pédopaysages. Application à "la moyenne vallée du Rhône". Thèse de doctorat. Paris, Institut national d'agronomie de Paris Grignon.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* 39. <https://doi.org/10.1029/2002WR001426>.

- Gray, J.M., Bishop, T.F.A., Wilford, J.R., 2016. Lithology and soil relationships for soil modelling and mapping. *CATENA* 147, 429–440. <https://doi.org/10.1016/j.catena.2016.07.045>.
- Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. *Geoderma* 146, 102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>.
- Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma* 143, 180–190. <https://doi.org/10.1016/j.geoderma.2007.11.004>.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2009.06.003>.
- Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* 46, 389–422. <https://doi.org/10.1023/A:1012487302797>.
- Hahn, C., Gloaguen, R., 2008. Estimation of soil types by non linear analysis of remote sensing data. *Nonlinear Process. Geophys.* 15, 115–126. <https://doi.org/10.5194/npg-15-115-2008>.
- Hamdi-Aïssa, B., Girard, M.-C., 2000. Utilisation de la télédétection en régions sahariennes, pour l'analyse et l'extrapolation spatiale des pédopaysages. *Science et changements planétaires / Sécheresse*. V11, numéro 3, John Libbey Eurotext Ltd.
- 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>
- Hartemink, A.E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R.A., Minasny, B., Montanarella, L., de Mendonça Santos, M.L., Sanchez, P., Walsh, M., Zhang, G.-L., 2010. GlobalSoilMap.net – A New Digital Soil Map of the World, in: *Digital Soil Mapping*. Springer Netherlands, Dordrecht, pp. 423–428. [https://doi.org/10.1007/978-90-481-8863-5\\_33](https://doi.org/10.1007/978-90-481-8863-5_33).

- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed, Springer Series in Statistics. Springer-Verlag, New York.
- Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma* 140, 417–427. <https://doi.org/10.1016/j.geoderma.2007.04.022>.
- Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12, e0169748. <https://doi.org/10.1371/journal.pone.0169748>.
- Hengl, T., 2018. Predictive Soil Mapping with R, in: Predictive Soil Mapping with R.
- 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>.
- Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62–77. <https://doi.org/10.1016/j.geoderma.2015.11.014>.
- Heung, B., Hodúl, M., Schmidt, M.G., 2017. Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes. *Geoderma* 290, 51–68. <https://doi.org/10.1016/j.geoderma.2016.12.001>.
- Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. *Geoderma* 100, 269–301.
- Hudson, H.D., 1992. The Soil Survey as Paradigm-based Science. *Soil Sci. Soc. Am. J.* 56, 836–841. <https://doi.org/10.2136/sssaj1992.03615995005600030027x>.
- IUSS Working Group WRB, 2014. International soil classification system for naming soil and creating legends for soil maps. World Soil Resources Reports No 106. FAO, Rome. World Reference Base for Soil Resources.
- Jafari, A., Finke, P.A., Vande Wauw, J., Ayoubi, S., Khademi, H., 2012. Spatial prediction of USDA- great soil groups in the arid Zarand region, Iran: Comparing logistic regression

approaches to predict diagnostic horizons and soil types. *Eur. J. Soil Sci.* 63, 284–298.  
<https://doi.org/10.1111/j.1365-2389.2012.01425.x>.

Jafari, A., S. Ayoubi, H. Khademi, P. A. Finke, and N. Toomanian. 2013. Selection of a Taxonomic Level for Soil Mapping Using Diversity and Map Purity Indices: A Case Study from an Iranian Arid Region. *Geomorphology* 201: 86–97.  
<https://doi.org/10.1016/j.geomorph.2013.06.010>.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An Introduction to Statistical Learning: with Applications in R*, Springer Texts in Statistics. Springer-Verlag, New York.

Jenny, H., 1941. *Factors of Soil Formation, A System of Quantitative Pedology*. McGraw-Hill, New York.

Jeong, G., Oeverdieck, H., Park, S.J., Huwe, B., Ließ, M., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. *Catena* 154, 73–84. <https://doi.org/10.1016/j.catena.2017.02.006>.

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>

Kempen, B., Brus, D.J., Heuvelink, G.B.M., 2012. Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland. *Geoderma* 189–190, 540–553. <https://doi.org/10.1016/j.geoderma.2012.05.028>

Kerry, R., Oliver, M.A., 2011. Soil geomorphology: Identifying relations between the scale of spatial variation and soil processes using the variogram. *Geomorphology* 130, 40–54.  
<https://doi.org/10.1016/j.geomorph.2010.10.002>

Kovačević, M., Bajat, B., Gajić, B., 2010. Soil type classification and estimation of soil properties using support vector machines. *Geoderma* 154, 340–347.  
<https://doi.org/10.1016/j.geoderma.2009.11.005>.

Kuhn, M., 2008. *Building Predictive Models in R Using the caret Package*. J. Stat. Softw. 028.

Kuhn, M., Johnson, K., 2013. *Applied Predictive Modeling*. Springer-Verlag, New York.

Lacoste, M. 2012. *Evolution des sols à l'échelle du paysage sous des conditions de changements climatique et de structure du paysage*. Thèse de doctorat Agrocampus Ouest Rennes. INRA-Agrocampus Ouest, UMR Sol Agro et hydrosystème Spatialisation : 241 p.

- Lagacherie P, 2002. Cartographies des sols et de leurs propriétés à un niveau sub-régional, Habilitation à Diriger des recherches, 45 p.
- Lagacherie, P., McBratney, A.B., 2006. Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. *Dev. Soil Sci.* [https://doi.org/10.1016/S0166-2481\(06\)31001-X](https://doi.org/10.1016/S0166-2481(06)31001-X).
- Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. <https://doi.org/10.2307/2529310>.
- Legros J.P., 1996. Cartographie des sols : de l'analyse spatiale à la gestion des territoires. Presses polytechniques et universitaires romandes, Lausanne (Suisse). 321 p.
- Levi, M.R., 2017. Neighborhood Size of Training Data Influences Soil Map Disaggregation. *Soil Sci. Soc. Am. J.* 81, 354–368. <https://doi.org/10.2136/sssaj2016.08.0258>.
- Li, W., Zhang, C., 2007. A Random-Path Markov Chain Algorithm for Simulating Categorical Soil Variables from Random Point Samples. *Soil Sci. Soc. Am. J.* 71, 656–668. <https://doi.org/10.2136/sssaj2006.0173>.
- Lin, H., Wheeler, D., Bell, J., Wilding, L., 2005. Assessment of soil spatial variability at multiple scales. *Ecol. Model., Scaling, fractals and diversity in soils and ecohydrology* 182, 271–290. <https://doi.org/10.1016/j.ecolmodel.2004.04.006>.
- Ließ, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models. *Geoderma* 170, 70–79. <https://doi.org/10.1016/j.geoderma.2011.10.010>.
- Lorenzetti, R., Barbetti, R., Fantappiè, M., L'Abate, G., Costantini, E.A.C., 2015. Comparing data mining and deterministic pedology to assess the frequency of WRB reference soil groups in the legend of small scale maps. *Geoderma* 237, 237–245. <https://doi.org/10.1016/j.geoderma.2014.09.006>.
- Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* 160, 614–626. <https://doi.org/10.1016/j.geoderma.2010.11.013>.
- Malone, B.P., McBratney, A.B., Minasny, B., 2013. Spatial Scaling for Digital Soil Mapping. *Soil Sci. Soc. Am. J.* 77, 890. <https://doi.org/10.2136/sssaj2012.0419>

- Massawe, B.H.J., Subburayalu, S.K., Kaaya, A.K., Winowiecki, L., Slater, B.K., 2018. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. *Geoderma* 311, 143–148. <https://doi.org/10.1016/j.geoderma.2016.11.020>.
- Maynard, J.J., Johnson, M.G., 2014. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2014.03.021>
- McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 293–327. [https://doi.org/10.1016/S0016-7061\(00\)00043-4](https://doi.org/10.1016/S0016-7061(00)00043-4).
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma*. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4).
- Mckenzie, N. J., Gallant, J. C. 2007. Digital Soil Mapping with Improved Environmental Predictors and Models of Pedogenesis. In *Digital Soil Mapping : An Introductory Perspective*, 31, 327–349. [https://doi:10.1016/s0166-2481\(06\)31024-0](https://doi:10.1016/s0166-2481(06)31024-0).
- Messedi, A.G., Delaître, E. 2007. Les états de surface en zone aride à partir d'indices radiométriques et de classifications multitemporelles d'images Landsat TM prises sur la région de Menzel Habib (Tunisie méridionale). *Cahier Sécheresse*, n°4, vol 18, 305-313.
- Miller, B.A., Schaetzl, R.J., 2014. The historical role of base maps in soil geography. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2014.04.020>.
- Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Impact of multi-scale predictor selection for modeling soil properties. *Geoderma*. 239, 97–106. <https://doi.org/10.1016/j.geoderma.2014.09.018>.
- Minasny, B., McBratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* 142, 285–293. <https://doi.org/10.1016/j.geoderma.2007.08.022>.
- Minasny, B., Whelan, B.M., Triantafilis, J., McBratney, A.B., 2013. Pedometrics Research in the Vadose Zone—Review and Perspectives. *Vadose Zo. J.* 12. <https://doi.org/10.2136/vzj2012.0141>.

- Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma*, Soil mapping, classification, and modelling: history and future directions 264, 301–311. <https://doi.org/10.1016/j.geoderma.2015.07.017>.
- Mirakzehi, K., Pahlavan-Rad, M.R., Shahriari, A., Bameri, A., 2018. Digital soil mapping of deltaic soils: A case of study from Hirmand (Helmand) river delta. *Geoderma* 313, 233–240. <https://doi.org/10.1016/j.geoderma.2017.10.048>.
- Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrol. Process.* 5, 3–30. <https://doi.org/10.1002/hyp.3360050103>.
- Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. *Soil Sci. Soc. Am. J.* 57, 443–452. <https://doi.org/10.2136/sssaj1993.03615995005700020026x>.
- Moran, C.J., Bui, E.N., 2002. Spatial data mining for enhanced soil map modelling. *Int. J. Geogr. Inf. Sci.* 16, 533–549. <https://doi.org/10.1080/13658810210138715>.
- Mostephaoui, T., Bensaid, R., Saker, M.L., 2013. Localization and delimitation of the arid soils by remote sensing and in-situ measurements in an arid area: Case of Oued Djedi watershed, Biskra, Algeria. *World Appl. Sci. J.* 24, 370–382. <https://doi.org/10.5829/idosi.wasj.2013.24.03.972>.
- Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping-A review. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2010.12.018>.
- Nield, S.J., Boettinger, J.L., Ramsey, R.D., 2007. Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data. *Soil Sci. Soc. Am. J.* 71, 245–252. <https://doi.org/10.2136/sssaj2006-0049>.
- Odgers, N.P., Sun, W., McBratney, A.B., Minasny, B., Clifford, D., 2014. Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma* 214–215, 91–100. <https://doi.org/10.1016/j.geoderma.2013.09.024>.
- Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., Bogaert, P., 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma* 232–234, 97–106. <https://doi.org/10.1016/j.geoderma.2014.04.036>.

- Pahlavan-Rad, M.R., Khormali, F., Toomanian, N., Brungard, C.W., Kiani, F., Komaki, C.B., Bogaert, P., 2016. Legacy soil maps as a covariate in digital soil mapping: A case study from Northern Iran. *Geoderma* 279, 141–148. <https://doi.org/10.1016/j.geoderma.2016.05.014>.
- Park, E., Elfeki, A.M.M., Song, Y., Kim, K., 2007. Generalized Coupled Markov Chain Model for Characterizing Categorical Variables in Soil Mapping. *Soil Sci. Soc. Am. J.* 71, 909–917. <https://doi.org/10.2136/sssaj2005.0386>.
- Pásztor, László, Annamária, L., Zsófia B., József S., Gábor I.. 2018. Compilation of a national soil-type map for Hungary by sequential classification methods. *Geoderma* 311: 93-108. <https://doi.org/10.1016/j.geoderma.2017.04.018>.
- Pedro G, 1989. L'approche spatiale en pédologie. *Science du Sol. AFES*. Vol 27 N°4, 287-300.
- Peng, J., Biswas, A., Jiang, Q., Zhao, R., Hu, J, Hu, B, Shia, Z. 2019. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* 337, 1309-1319. <https://doi.org/10.1016/j.geoderma.2018.08.006>.
- Planchon, O., Darboux, F., 2002. A fast, simple and versatile algorithm to fill the depressions of digital elevation models, in: *Catena*. pp. 159–176. [https://doi.org/10.1016/S0341-8162\(01\)00164-3](https://doi.org/10.1016/S0341-8162(01)00164-3).
- Pouget, M. 1980. Les relations sol-végétation dans les steppes Sud-Algéroises. Travaux et Documents de l'ORSTOM n° 116. Paris : ORSTOM.
- Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. *Soil Sci. Soc. Am. J.* 82, 186–201. <https://doi.org/10.2136/sssaj2017.04.0122>.
- R Development Core Team, 2016. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. <https://doi.org/10.1007/978-3-540-74686-7>.
- Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. *Intermt. J. Sci.* 5, 23–27. <https://doi.org/citeulike-article-id:8858430>.
- Roecker, S. M., Howell, D. W. C. A., Haydu-Houdeshell., Blinn, C., 2010. A Qualitative Comparison of Conventional Soil Survey and Digital Soil Mapping Approaches. In *Digital Soil Mapping*, 369–84. Progress in Soil Science. Springer, Dordrecht. [https://doi.org/10.1007/978-90-481-8863-5\\_29](https://doi.org/10.1007/978-90-481-8863-5_29).

- Roecker, S., Thompson, J., 2010. Scale effects on terrain attribute calculation and their use as environmental covariates for digital soil mapping. In: Boettinger, J.L. (Ed.), *Digital Soil Mapping: Bridging Research, Production, and Environmental Application*. Springer, pp. 55–66.
- Shi, X., Girod, L., Long, R., DeKett, R., Philippe, J., Burke, T., 2012. A comparison of LiDAR-based DEMs and USGS-sourced DEMs in terrain analysis for knowledge-based digital soil mapping. *Geoderma* 170, 217–226. <https://doi.org/10.1016/j.geoderma.2011.11.020>
- Smith, M.P., Zhu, A.X., Burt, J.E., Stiles, C., 2006. The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma*. <https://doi.org/10.1016/j.geoderma.2006.07.002>.
- Soil Science Division Staff. 2017. *Soil Survey Manual*. USDA Handbook 18. Washington: United States Department of Agriculture.. Government Printing Office, Washington, D.C.
- Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon. *Soil Sci. Soc. Am. J.* 81, 1413. <https://doi.org/10.2136/sssaj2016.11.0376>
- Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random Forests Applied as a Soil Spatial Predictive Model in Arid Utah, in: *Digital Soil Mapping*. pp. 179–190. [https://doi.org/10.1007/978-90-481-8863-5\\_15](https://doi.org/10.1007/978-90-481-8863-5_15).
- Sun, Y., Wong, Andrew K. C., Kamel, Mohamed S., 2009. Classification of Imbalanced Data: A Review. *International Journal of Pattern Recognition and Artificial Intelligence* 23 (04): 687–719. <https://doi.org/10.1142/S0218001409007326>.
- Taghizadeh-Mehrjardi, R., F. Sarmadian, B. Minasny, J. Triantafilis, and M. Omid. 2014. Digital Mapping of Soil Classes Using Decision Tree and Auxiliary Data in the Ardakan Region, Iran. *Arid Land Research and Management* 28 (2): 147–68. <https://doi.org/10.1080/15324982.2013.828801>.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., Triantafilis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma* 253–254, 67–77. <https://doi.org/10.1016/j.geoderma.2015.04.008>.

- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266: 98-110. <https://doi.org/10.1016/j.geoderma.2015.12.003>.
- Thompson, J.A., Bell, J.C., Butler, C.A., 2001. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma*. [https://doi.org/10.1016/S0016-7061\(00\)00081-1](https://doi.org/10.1016/S0016-7061(00)00081-1)
- Vieillefon, J., 1976. Inventaire critiques des sols gypseux en Tunisie-Etude préliminaire. O.R.S.T.O.M., 81p.
- Vincenzi, S., Zucchetta, M., Franzoi, P., Pellizzato, M., Pranovi, F., De Leo, G.A., Torricelli, P., 2011. Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecol. Modell.* 222, 1471–1478. <https://doi.org/10.1016/j.ecolmodel.2011.02.007>.
- Walter C., 2002. Analyse spatiale des sols en vue de leur gestion précise et de leur surveillance. Mémoire d'Habilitation à Diriger des Recherches, ENSAR, Rennes, 96 p.
- Webster, R., 2000. Is soil variation random? *Geoderma* 97, 149–163. [https://doi.org/10.1016/S0016-7061\(00\)00036-7](https://doi.org/10.1016/S0016-7061(00)00036-7).
- Wu., Wei., Ai-Di Li., Xin-Hua He., Ran Ma., Hong-Bin L., Jia-Ke, L., 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. *Computers and Electronics in Agriculture* 144: 86-93. <https://doi.org/10.1016/j.compag.2017.11.037>.
- Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* 340, 7–24. <https://doi.org/10.1007/s11104-010-0425-z>.
- Xiao, J., Shen, Y., Tateishi, R., Bayaer, W., 2006. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. *Int. J. Remote Sens.* 27, 2411–2422. <https://doi.org/10.1080/01431160600554363>.
- 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 Sci. Soc. Am. J.* 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, G. lin., Liu, F., Song, X. dong., 2017. Recent progress and future prospect of digital soil mapping: A review. *J. Integr. Agric.* [https://doi.org/10.1016/S2095-3119\(17\)61762-3](https://doi.org/10.1016/S2095-3119(17)61762-3).

Zhu, A.X., Burt, J.E., Smith, M., Wang, R., Gao, J., 2008. The impact of neighbourhood size on terrain derivatives and digital soil mapping, in: *Lecture Notes in Geoinformation and Cartography*. [https://doi.org/10.1007/978-3-540-77800-4\\_18](https://doi.org/10.1007/978-3-540-77800-4_18)

## **ANNEXES**

**Liste des annexes**

Annexe I. Fonction d'élimination de variables par récursion (RFE).

Annexe II. Définitions des qualificatifs, horizons, propriétés et matériaux de diagnostic de la WRB (2014).

Annexe III. Solums représentatifs.

**Annexe I.** Fonction d'élimination de variables par récursion (RFE).

Outer resampling method: Cross-Validated (10 fold, repeated 5 times)

Resampling performance over subset size:

| Variables | Accuracy | Kappa  | AccuracySD | KappaSD | Selected |
|-----------|----------|--------|------------|---------|----------|
| 3         | 0.4379   | 0.3345 | 0.07252    | 0.08538 |          |
| 4         | 0.5519   | 0.4703 | 0.07029    | 0.08175 |          |
| 5         | 0.5698   | 0.4908 | 0.07474    | 0.08698 |          |
| 6         | 0.5742   | 0.4949 | 0.06343    | 0.07415 |          |
| 7         | 0.5779   | 0.4988 | 0.06437    | 0.07482 |          |
| 8         | 0.5920   | 0.5140 | 0.06258    | 0.07391 |          |
| 9         | 0.5950   | 0.5178 | 0.05897    | 0.06903 |          |
| 10        | 0.6019   | 0.5268 | 0.05927    | 0.06879 |          |
| 11        | 0.6080   | 0.5338 | 0.05661    | 0.06587 |          |
| 12        | 0.6111   | 0.5371 | 0.05435    | 0.06444 | *        |
| 13        | 0.6092   | 0.5343 | 0.05240    | 0.06218 |          |
| 14        | 0.6098   | 0.5351 | 0.06278    | 0.07468 |          |
| 15        | 0.6084   | 0.5326 | 0.05794    | 0.06979 |          |
| 16        | 0.6025   | 0.5258 | 0.06264    | 0.07447 |          |
| 17        | 0.6034   | 0.5264 | 0.06490    | 0.07813 |          |
| 18        | 0.6055   | 0.5288 | 0.06902    | 0.08364 |          |
| 19        | 0.6009   | 0.5231 | 0.06386    | 0.07671 |          |
| 20        | 0.5963   | 0.5169 | 0.07308    | 0.08790 |          |
| 21        | 0.5982   | 0.5191 | 0.07068    | 0.08512 |          |
| 22        | 0.5965   | 0.5171 | 0.06503    | 0.07829 |          |
| 23        | 0.5769   | 0.4927 | 0.07211    | 0.08660 |          |
| 24        | 0.5784   | 0.4944 | 0.06422    | 0.07697 |          |

The top 5 variables (out of 12):

elevation\_, eleva300, gsi\_indx, mrvbf, rat\_7\_1

predictors(results)

[1] "elevation" "eleva300" "gsi\_indx" "mrvbf" "rat\_7\_1" "mrvbf300"

[7] "tcilow213" "slope300" "tri300" "catchslpe2" "gyps\_indx" "sali\_indx"

**Annexe II.** Définitions des qualificatifs, horizons, propriétés et matériaux de diagnostic de la WRB (2014).

Haplic (du grec haplous, simple): présentant certaines caractéristiques de manière typique (dans le sens qu'il n'y a pas de caractérisation plus poussée ou plus significative) et utilisé seulement si aucun des qualificatifs listés auparavant ne s'applique.

Hyperskeletal (du grec hyper, au-dessus, et skeletos, asséché) : ayant une teneur en terre fine  $< 20 \%$  (en volume) en moyenne pondérée sur une profondeur de 75 cm à partir de la surface du sol ou jusqu'à une roche continue, un matériau technique dur ou une couche cimentée ou indurée commençant à  $> 25$  cm de la surface du sol, si la profondeur est moindre.

Leptic (du grec *leptos*, mince): ayant une roche continue ou un matériau technique dur commençant à  $\leq 100$  cm de la surface du sol.

Lithic (du grec *lithos*, pierre): ayant une roche continue ou un matériau technique dur commençant à  $\leq 10$  cm de la surface du sol (uniquement pour les Leptosols).

Petric (du grec petros, roche): ayant une couche cimentée ou indurée commençant à  $\leq 100$  cm de la surface du sol (se rapporte à un horizon diagnostique du RSG correspondant).

Skeletal (du grec skeletos, asséché): ayant  $\geq 40 \%$  (en volume) d'éléments grossiers, en moyenne jusqu'à une profondeur de 100 cm ou jusqu'à une roche continue, un matériau technique dur ou une couche cimentée ou indurée, si cette profondeur est plus mince.

Sodic (de l'espagnol soda, eau gazeuse): ayant une couche d'une épaisseur  $\geq 20$  cm, commençant à  $\leq 100$  cm de la surface du sol, ayant  $\geq 15 \%$  Na plus Mg et  $\geq 6 \%$  Na sur le complexe d'échange, et n'ayant pas un horizon natrique commençant à  $\leq 100$  cm de la surface du sol.

Horizon calcique (du latin calx, chaux): est un horizon dans lequel du carbonate de calcium secondaire ( $\text{CaCO}_3$ ) s'est accumulé sous forme diffuse (le carbonate de calcium imprègne le fond matriciel ou forme de fines particules de calcite de  $< 1$  mm qui y sont dispersées) ou sous forme de concentrations discontinues (veines, pseudo-mycélium, revêtements, nodules tendres et/ou durs).

Horizon gypsique (du grec gypsos, gypse): est un horizon non cimenté contenant des accumulations de gypse secondaire ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ) sous différentes formes.

Horizon salique (du latin sal, sel): est un horizon de surface ou subsuperficiel à faible profondeur qui contient des quantités importantes de sels facilement solubles, c'est-à-dire des sels plus solubles que le gypse ( $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$ ;  $\log K_s = -4,85$  à  $25^\circ\text{C}$ ).

Gleyiques : les matériaux de sol développent des propriétés gleyiques (du russe gley, masse de sol fangeuse) quand ils sont saturés par la nappe (ou, s'ils sont drainés, qu'ils aient été saturés dans le passé) pendant une période qui permet à des conditions réductrices de s'installer (cela peut aller de quelques jours sous les tropiques à quelques semaines dans les autres régions).

Matériau calcaire (du latin calcarius, contenant de la chaux) : est un matériau qui contient  $\geq 2\%$  d'équivalent carbonate de calcium. Ces carbonates proviennent du matériau parental.

Matériau colluvique (du latin colluvio, mélange): est un mélange hétérogène de matériaux qui, sous la force gravitaire, se retrouve au bas d'un versant. Il a été transporté par érosion en nappe ou par reptation, et le transport peut avoir été accéléré par l'aménagement du territoire (déforestation, labour dans le sens de la pente, dégradation de la structure). Il s'est formé dans des temps géologiques récents. Il s'accumule normalement dans des pentes, des dépressions ou en amont d'un obstacle naturel ou anthropique dans une pente faible.

Matériau fluviatile (du latin fluvius, cours d'eau): se rapporte à des sédiments fluviaux, marins ou lacustres qui reçoivent du matériau frais, ou qui en ont reçu dans le passé, et qui présentent des stratifications.

Matériau gypsifère (du grec gypsos, plâtre): est un matériau minéral dont la teneur en gypse est  $\geq 5\%$  (en volume) dans les parties de terre fine qui ne contiennent pas de gypse secondaire.

### **Annexe III. Solums représentatifs.**

#### ***Solum 1***

##### Description de l'environnement

- Localisation : Tolga.
- Coordonnées :  
Longitude : 5°21'14.16"E  
Latitude : 34°44'23.72"N
- Géomorphologie : Glacis
- Topographie : plane.
- Végétation : *Salsola vermiculata*,  
*Phoenix dactylifera*.

##### Classification :

WRB (2014): Haplic Petric Gypsisols

CPCS (1967):

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : à encroûtement gypseux



##### Description morphologique

0-45 cm : sec, brun jaunâtre (10 YR 5/6), sablo-limoneuse, structure polyédrique, poreux, friable, tâche blanchâtre de gypse, transition, faible effervescence à l'HCl.

40-100 cm : encroûtement gypseux.

## ***Solum 2***

### Description de l'environnement

- Localisation : Tolga.
- Coordonnées :  
Longitude : 5°24'35.44"E  
Latitude : 34°43'44.75"N
- Géomorphologie : Glacis
- Topographie : plane
- Végétation : Phoenix dactylifera.

Classification :

WRB (2014) : Haplic Gypsisols

CPCS (1967) :

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : Modal

### Description morphologique

0-20 cm : sec, brun, beaucoup de racines, limono-sableuse, friable, faible effervescence à l'HCl.

20-43 cm : humide, brun, sablo-limoneuse, présence de tâches blanchâtre, structure fine, friable, présence de cristaux de gypse.

43-81 cm : humide, brun, sableuse avec intercalation de gypse.

### ***Solum 3***

#### Description de l'environnement

- Localisation : Tolga
- Coordonnées :  
Longitude : 5°23'46.67"E  
Latitude : 34°42'45.97"N
- Topographie : pente
- Végétation : *Salsola vermiculata*,  
*Phoenix dactylifera*

#### Classification :

WRB (2014) : Calcic Petric Gypsisols

CPCS (1967) :

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : à encroûtement gypseux



#### Description morphologique

0-20 cm : sec, gris-rosâtre, sablo-limoneuse, traces de pseudo-mycélium blanchâtres, moyenne effervescence à l'HCl.

20-60 cm : sec, blanc, tâches brunes, sableuse, faible effervescence à l'HCl, présence de cristaux de gypse.

60-100 cm : encroûtement gypseux.

#### ***Solum 4***

##### Description de l'environnement

- Localisation : Ouarlal.
- Coordonnées :  
Longitude : 5°30'07.94"E  
Latitude : 34°40'55.30"N
- Géomorphologie : Terrasse
- Topographie : Plane
- Végétation : *Atriplex halimus*, *Salsola vermiculata*, *Tamarix articulata*

##### Classification :

WRB (2014) : Gleyic Solonchaks

CPCS (1967) :

Classe : Sodiques

S/classe : structure non dégradée

Groupe : salins

S/groupe : à efflorescences salines

##### Description morphologique

0-22 cm : sec, gris à gris foncé, limono-argileuse, faible effervescence à l'HCl, horizon salé.

22-58 cm : humidité moyenne, gris, limoneuse, faible effervescence à l'HCl, tâches salines, tâches verdâtres, jaunâtres et grisâtre.

58-77 cm : humidité moyenne, limono-sableuse, faible effervescence à l'HCl, tâches verdâtres, jaunâtres et grises, petits nodule carbonatés et intercalation de gypse, compact.

77-93 cm : humidité moyenne, limoneuse, faible effervescence à l'HCl, tâches ocres et verdâtres, compact, inclusions salines et intercalations de gypse.

### ***Solum 5***

#### Description de l'environnement

- Localisation : Bordj Ben Azzouz
- Coordonnées :  
Longitude : 5°21'9.89"E  
Latitude : 34°41'51.15"N
- Géomorphologie : Glacis
- Topographie : Plane
- Végétation : *Salsola vermiculata*,  
*Tamarix articulata*, *Phoenix dactylifera*.

#### Classification :

WRB (2014) : Gypsic Sodic Solonchaks

CPCS (1967) :

Classe : Sodiques

S/classe : structure non dégradée

Groupe : salins

S/groupe : à efflorescences saline



#### Description morphologique

0-43 cm : sec, brun jaunâtre (10YR5/6), limono-sableuse, structure particulière, sous forme de pseudo-sable, quelques racines, friable, présence de tâches blanchâtre, goût salé, faible effervescence à l'HCL, transition graduelle.

43-110 cm : peu humide, brun jaunâtre (10YR5/6), limoneuse, structure polyédrique, faible effervescence à l'HCL, peu friable.

### ***Solum 6***

#### Description de l'environnement

- Localisation : El Amri
- Coordonnées :  
Longitude : 5°19'1.70"E  
Latitude : 34°39'57.39"N
- Géomorphologie : crête
- Topographie : pente douce
- Végétation : *Salsola vermiculata*.

#### Classification :

WRB (2014): Skeletic Petric Gypsisols

CPCS (1967):

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : à encroûtement gypseux

#### Description morphologique

0-40 cm : sec, blanc, sableuse, croûte gypseuse avec intercalations de galets, moyenne effervescence à l'HCL.

40-90 cm : sec, brun foncé, sableuse, quantité importante de galets enduré avec du gypse, moyenne effervescence à l'HCL.

90-105 cm : encroûtement gypseux.

### ***Solum 7***

#### Description de l'environnement

- Localisation : El Amri
- Coordonnées :  
Longitude : 5°17'13.06"E  
Latitude : 34°40'30.30"N
- Géomorphologie : crête
- Topographie : pente douce
- Végétation : *Salsola vermiculata*, *Atriplex*  
*Halimus*, *Limoniastrum guyonianum*.



#### Classification :

WRB (2014) : Haplic Skeletic Gypsisols

CPCS (1967) :

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : Modal

#### Description morphologique

0-40 cm : sec, gris, sablo-limoneuse, nombreux galets, moyenne effervescence à l'HCL, cristaux de gypse.

40-120 cm : sec, gris, compact, sablo-limoneuse, moyenne effervescence à l'HCL, quantité importante de galets enduré avec du gypse.

### ***Solum 8***

#### Description de l'environnement

- Localisation : Sahira
- Coordonnées :  
Longitude : 5°26'48.90"E  
Latitude : 34°39'12.80"N
- Géomorphologie : Colline résiduelle
- Topographie : pente douce
- Végétation : *Zygophyllum album*.

#### Classification :

WRB (2014) : Leptic Petric Gypsisols

CPCS (1967) :

Classe : Calcimagnésiques

S/classe : Gypseux

Groupe : Bruns gypseux

S/groupe : à encroûtement gypseux

#### Description morphologique

0-15 cm : sec, gris à brun, limono-sableuse, faible effervescence à l'HCL, nodules de gypse.

15-50 cm : sec, gris à blanc, sableuse, compact, faible effervescence à l'HCL, croûte pierreuse-gypseuse, gros bloc du gypse filamenteux.

70-136 cm : encroûtement gypseux.



### ***Solum 9***

#### Description de l'environnement

- Localisation : El Amri
- Coordonnées :  
Longitude : 5°18'50.29"E  
Latitude : 34°40'11.11"N
- Topographie : plane
- Géomorphologie : Terrasse
- Végétation : Tamarix articulata, Atriplex halimus, Phoenix dactylifera.

#### Classification :

WRB (2014) : Calcaric Fluvisols

CPCS (1967) :

Classe : Peu évolués

S/classe : Non climatique

Groupe : d'apport alluvial

S/groupe : Modal

#### Description morphologique

0-30 cm : sec, brun jaunâtre, friable, sableuse, quelques racines, moyenne effervescence à l'HCL, transition nette.

30-50 cm : sec, brun, motteux, limono-sableuse, quelques racines, moyenne effervescence à l'HCL.

### ***Solum 10***

#### Description de l'environnement

- Localisation : Sahira
- Coordonnées :  
Longitude : 5°27'31.40"E  
Latitude : 34°39'6.22"N
- Topographie : plane
- Géomorphologie : Terrasse
- Végétation : *Salsola vermiculata*,  
*Atriplex halimus*.

#### Classification :

WRB (2014) : Gypsic Solonchaks

CPCS (1967) :

Classe : Sodiques

S/classe : structure non dégradée

Groupe : salins

S/groupe : superficiel friable



#### Description morphologique

0-28 cm : sec, 10YR6/8, sableuse, structure particulière, friable, quelques racines, moyenne effervescence à l'HCL, transition graduelle.

28-90 cm : sec, 10YR5/6, limono-sableuse, structure polyédrique, présence de tâches blanchâtres, moyenne effervescence à l'HCL, cristaux de gypse.

### ***Solum 11***

#### Description de l'environnement

- Localisation : Lioua
- Coordonnées :  
Longitude : 5°25'2.70"E  
Latitude : 34°38'6.53"N
- Géomorphologie : Terrasse d'Oued Djedi
- Topographie : plane
- Végétation : *Limoniastrum guyonianum*, *Phoenix dactylifera*.

#### Classification :

WRB (2014): Gypsic Fluvisols

CPCS (1967):

Classe : Peu évolués

S/classe : Non climatique

Groupe : d'apport alluvial

S/groupe : Modal

#### Description morphologique

0-27 cm : sec, 10YR6/8, sableuse, structure particulière, poreux, friable, quelques graviers et cailloux, faible effervescence à l'HCL, transition graduelle.

27-50 cm : sec, 10YR5/6, limono-sableuse, structure polyédrique, poreux, faible effervescence à l'HCL.