

FORET D'ARBRES ALEATOIRES ET CLASSIFICATION D'IMAGES SATELLITES : RELATION ENTRE LA PRECISION DU MODELE D'ENTRAINEMENT ET LA PRECISION GLOBALE DE LA CLASSIFICATION

Aurélien N.G. Matsaguim¹, Emmanuel D. Tiomo²

Département de Géographie- Faculté des Arts, Lettres et Sciences Humaines, Université de Yaoundé I, Cameroun,
Tel : (+237) 694 83 43 37, nquimdoaure@yahoo.fr

(1) Département de Géographie, Faculté des Lettres et Sciences Humaines, Université de Dschang – Cameroun
Tel : (+237) 696 51 62 98, tiomoemmanuel2008@yahoo.fr

Résumé

En télédétection, il existe un grand nombre d'algorithmes permettant de classifier une image satellite. Parmi ces algorithmes de classification, la Forêt d'Arbres Aléatoires apparaît comme particulièrement performante. Cette étude a pour objectif d'évaluer (1) l'importance de la sélection des images pour le niveau de précision du modèle d'entraînement et (2) la nature de la relation qui existe entre le niveau de précision du modèle et celui de la précision globale de la carte thématique résultant de la classification de l'image satellite avec cet algorithme de classification. A partir d'une image Landsat 8 OLI prise au-dessus d'une zone de montagne tropicale : la région de l'Ouest Cameroun, 35 modèles ont été construits et testés. Les résultats montrent que le niveau de la précision globale des résultats de la Forêts d'Arbres Aléatoires est étroitement dépendant d'une part de la précision du modèle d'entraînement utilisé pour classifier l'image satellite, et d'autre part du choix des images utilisées pour entraîner ce modèle. De plus, la sélection de ces images est elle-même dépendante de la qualité des zones d'entraînement qui serviront à la construction du modèle. Il est donc important de mettre un accent particulier sur la qualité des données d'entrée afin de garantir des résultats satisfaisants avec cet algorithme.

Mots clés : Forêt d'Arbres Aléatoires, précision, modèle d'entraînement, télédétection, Cameroun

Abstract

In remote sensing, there is a large number of algorithms to classify a satellite image. Among these classification algorithms, the Random Forest appears to be particularly powerful. The objectives of this study are to evaluate (1) the importance of image selection for the level of accuracy of the training model and (2) the nature of the relationship between the level of accuracy of the model and the overall accuracy of the thematic map resulting from the classification of the satellite image with this algorithm. Based on a Landsat 8 OLI image taken over a tropical mountain area : the West Cameroon region, 35 models were built and tested. Analysis shows that the level of overall accuracy of the Random Forest results are closely dependent on the accuracy of the training model used to classify the satellite image, and on the choice of the images used to train this model. Moreover, the selection of these images is itself dependent on the quality of the Regions of interest (ROI) that will be used to build the model. It is therefore important to place particular emphasis on the quality of the input data in order to guarantee satisfactory results with this algorithm. Otherwise, the performance of this algorithm could be inferior to that of other classification algorithms.

Keywords: Random Forests, Accuracy, training model, remote sensing, Cameroon

Introduction

En télédétection, la Forêt d'Arbres Aléatoires (RF) ou « *Random Forests* » en anglais apparaît comme un algorithme de classification particulièrement performant [Gislason et al., 2006 ; Breiman, 2001]. Il permet d'obtenir des niveaux de précision supérieurs à ceux obtenus avec différents autres algorithmes de classification [Meddens et al., 2016 ; Pelletier et al., 2016 ; Colditz, 2015 ; Jhonnerie et al., 2015 ; Ozdarici et al., 2012 ; Rodriguez-Galiano et al., 2012]. Toutefois, cette performance est conditionnée par le niveau d'hétérogénéité ou « *noise* » des classes d'occupation du sol utilisées pour l'entraînement du modèle ainsi que la taille globale de l'échantillon d'entraînement [Kulkarni et Lowe, 2016 ; Shi et Yang, 2016 ; Millard et Richardson, 2015]. Cette hétérogénéité, très importante dans un paysage fortement fragmenté [Lippitt et al.,

2008], se traduit très souvent par une variabilité spectrale intra-classe forte et inter-classe relativement faible qui est source de confusion au moment de la classification des images satellites [DeFries et Chan, 2000]. En effet, dans le cas où le niveau d'hétérogénéité au sein d'une classe d'entraînement dépasserait le seuil de 20 - 25% [Rodriguez-Galiano et al., 2012], elle sera mal classifiée ; ce qui affectera négativement les performances de l'algorithme contrairement à d'autres. Il s'agit par exemple de l'algorithme des Réseaux de Neurones Artificielles ou de celui des Machines à Vecteur de Support pour ce qui concerne les algorithmes de classification issus du domaine de l'intelligence artificielle, mais également d'algorithmes plus classiques comme le maximum de vraisemblance, [Xie et al., 2019 ; Raczko et Zagajewski, 2017 ; Kulkarni et Lowe, 2016]. La taille et la qualité de l'échantillon d'entraînement sont donc deux critères importants en ce qui concerne les performances de la RF en télédétection [Millard et Richardson, 2015].

Deux autres critères affectent également les performances de l'algorithme RF. Il s'agit du nombre d'arbres de décision qui devra être généré lors de la phase d'entraînement du modèle et du nombre de variables qui seront utilisées pour la construction de chacun des arbres générés. Tous ces critères sont des paramètres de l'algorithme qui doivent être définis par l'utilisateur. C'est une étape cruciale car les résultats de la classification en dépendent étroitement [Shi et Yang, 2016]. Bien que des tentatives de solutions, basées sur la réduction de l'erreur de généralisation (Out-of-bag Error en anglais – OOB) ont été proposées pour permettre de définir ces paramètres [Nguyen et al., 2018 ; Ming et al., 2016], l'idée selon laquelle plus grand est le nombre d'arbres générés par l'algorithme, meilleures seront ses performances reste encore admise. Cependant, il a été démontré que compte tenu des images utilisées et de la qualité des échantillons d'entraînement, cette idée n'était pas valide [Shi et Yang, 2016].

Une autre idée généralement admise est que la RF est peu sensible au nombre de variables explicatives. Il peut par conséquent prendre en compte un très grand nombre de variables de nature différente en entrée [Gislason et al., 2006]. Dans les régions présentant une forte hétérogénéité spatiale comme les montagnes, due entre autres à la topographie, l'altitude, l'exposition et la pente qui ont un impact sur la réflectance des objets à la surface [Dedieu, 1989], ainsi qu'à une grande diversité des paysages [Rodriguez-Galiano et al., 2012], c'est une démarche généralement recommandée [Rodriguez-Galiano et al., 2012] et utilisée [Xie et al., 2019; Ayala-Lzurieta et al., 2017 ; Meddens et al., 2016 ; Mellor et al., 2013]. Cependant, le choix des zones d'entraînement ainsi que celui du nombre de bandes spectrales, ou « *channels* » en anglais, d'une image satellite qui sera utilisée a un impact important sur les résultats issus de l'opération de classification avec cet algorithme. Certains auteurs recommandent à ce propos de n'inclure que les bandes les plus importantes et le moins corrélées possibles [Millard et Richardson, 2015]. Le choix du nombre, de même que la sélection des bandes spectrales d'une image satellite, qui seront utilisées pour la construction du modèle, est crucial [Raczko et Zagajewski, 2017] et dépend de la qualité des zones d'entraînement [Shi et Yang, 2016]. En effet, son niveau de précision en dépend.

Afin d'évaluer ce niveau de précision du modèle, l'un des avantages de la RF réside dans le fait qu'elle évalue l'OOB du modèle à la fin de l'étape d'entraînement. Cette erreur donne une indication du niveau de précision du modèle, c'est-à-dire sa capacité à correctement classifier les différents pixels de l'image satellite. Plus cette erreur est proche de 0, meilleur sera le modèle [Breiman, 2001]. Par conséquent, l'objectif premier de la phase d'entraînement est d'obtenir le niveau d'erreur de généralisation le plus bas possible selon l'hypothèse plus ou moins explicitement admise qu'il existerait une relation linéaire entre le niveau de cette erreur, soit la précision du modèle, et le niveau de la précision globale obtenu à l'issue de l'étape de vérification de l'image classifiée [Breiman, 2001]. Or, cette hypothèse serait invalide, car la méthode de validation de la carte thématique obtenue à l'issue de l'étape de vérification de l'image classifiée diffère de celle utilisée par la RF pour valider le niveau de précision du modèle utilisé pour

classifier l'image satellite [Shi et Yang, 2016 ; Stehman et Foody, 2009].

Cette étude a donc pour objectif d'étudier (1) l'importance de la sélection des images pour le niveau de précision (OOB) du modèle d'entraînement et (2) la nature de la relation qui existe entre le niveau de précision du modèle et celui de la précision globale de la carte thématique résultant de la classification de l'image satellite avec l'algorithme de classification RF. Pour y parvenir, nous procéderons dans un premier temps à un bref rappel des principes théoriques sur lesquels repose la RF à la section 1. La méthodologie suivie tout au long de cette étude sera présentée à la section 2 ; tandis que les principaux résultats et leur discussion seront présentés aux sections 3 et 4 respectivement.

1. Arbre de décision et Forêt d'Arbres Aléatoires : principes théoriques

La Forêt d'Arbres Aléatoires appartient à la famille des algorithmes de classification issus du champ de l'apprentissage machine ou « Machine Learning » au même titre que les Réseaux de Neurones Artificiels et les Machines à Vecteur de Support [Jensen, 2015]. Elle fait partie de la catégorie des algorithmes d'apprentissage supervisés et permet d'expliquer ou de prédire une variable qualitative (classification) ou quantitative (régression). C'est un algorithme de classification de type arbre de décision [Gislason et al., 2006].

1.1. Principe théorique de l'arbre de décision

D'une manière générale, l'arbre de décision fonctionne selon un principe de partitionnement récursif. L'objectif du partitionnement est de créer des groupes d'individus les plus homogènes possible par rapport à la variable que l'arbre cherche à classifier. Pour y parvenir, l'algorithme à l'origine de la construction ou « *pousse* » de l'arbre pose une succession de questions binaires (oui/non) relatives aux variables qui auront été définies comme des variables explicatives. Ce sont les réponses à ces questions qui constituent les branches de l'arbre.

Pris individuellement, Les arbres de décision ne sont pas toujours très efficaces pour prédire ou classifier les variables. En effet, il demeure le risque de surajustement du modèle, c'est-à-dire qu'il prédise ou classifie parfaitement la variable d'entraînement qui aura servi à sa construction. Autrement dit, le risque que le modèle ne puisse être généralisable.

1.2. Principe théorique de la Forêt d'Arbres Aléatoires

La RF a été inventée pour pallier ce problème de surajustement et possède au moins deux avantages importants :

1. le risque de surajustement du modèle final est réduit du fait de la construction de manière totalement aléatoire d'un grand nombre d'arbres de décision conformément à la loi forte des grands nombres ;
2. les prédictions sont plus précises car l'accroissement du nombre d'arbres donne lieu à

une faible erreur de généralisation du modèle [Breiman, 2001].

La RF s'exécute en deux étapes. Premièrement, une « forêt » d'arbres de décision, c'est à dire un grand nombre d'arbres de décision selon le nombre désiré, va être construite de manière totalement aléatoire. Deuxièmement, une fois cette « forêt » construite, un vote sera effectué pour ne retenir que la classe ou la catégorie la plus fréquente [Breiman, 2001].

Il s'agit d'un « *Ensemble Method* » en anglais, car l'algorithme va créer une multitude de modèles faibles et simples sous la forme d'arbres de décision individuels qui, lorsqu'ils seront combinés, vont former un modèle robuste peu sujet au surajustement [Breiman, 2001]. L'algorithme va ainsi agréger les différents modèles faibles en procédant au vote.

Le caractère aléatoire concerne la construction de chaque arbre de la « forêt ». En effet, seule une fraction des individus tests et des variables explicatives sera aléatoirement sélectionnée pour construire les différents arbres. Autrement dit, pour chaque arbre conçu dans le modèle, un échantillon aléatoire d'individus est sélectionné et la construction d'un nœud se fait sur un sous-ensemble de variables lui aussi sélectionné aléatoirement. Une fois le modèle élaboré, les données sont insérées dans tous les arbres, qui votent pour la classe de sortie : pour une régression en faisant une moyenne, pour une classification en votant à la majorité c'est à dire la catégorie la plus fréquente.

Breiman (2001) montre que le fait que la construction de chaque arbre s'effectue de manière totalement aléatoire permet d'améliorer un certain nombre des caractéristiques de la RF, à savoir :

- sa précision ;
- sa robustesse vis-à-vis des valeurs aberrantes et/ou manquantes ;
- sa vitesse d'exécution ;
- le calcul d'un certains nombres d'indicateurs de performances (estimation de l'erreur OOB, niveau de précision de chaque arbre, corrélation entre les arbres, importance de chaque variable) ;
- etc.

En effet, la précision du modèle construit grâce à la RF dépend du niveau de précision de chaque arbre généré au sein du modèle, et du niveau de corrélation entre ces derniers. Il est important que les niveaux de précision des arbres soient élevés et qu'ils soient indépendants les uns des autres ; d'où la nécessité d'introduire l'aléatoire qui permet d'obtenir ce résultat [Breiman, 2001].

Cette précision du modèle est mesurée par son erreur OOB qui doit être le plus faible possible. Plus cette erreur est faible, plus le modèle est précis dans ses prédictions et plus il sera performant en matière de classification des images satellites [Kulkarni et Lowe, 2016 ; Breiman, 2001].

2. Matériels et Méthodes

2.1. Zone d'étude

Notre zone d'étude se situe dans la partie occidentale du Cameroun à savoir la région administrative de l'Ouest qui s'étend entre 4,8° et 6,3° de latitude Nord et entre 9,4 et 11,7° de longitude Est (figure 1). Elle fait partie de la zone dite des Hautes Terres de l'Ouest Cameroun. Son altitude par rapport au niveau de la mer varie entre 242 m et 2 737 m. La partie occidentale de cette région est celle où l'altitude est la plus élevée (>1 000 m). Les pentes y sont généralement assez fortes (>25%). La partie orientale quant à elle se caractérise par une topographie généralement plane (<=12%) en dehors de quelques zones où la topographie est relativement accidentée donnant lieu à des pentes supérieures à 25% (figure 1). La Région de l'Ouest Cameroun est une région anciennement et densément peuplée où l'agriculture représente la principale activité économique, notamment dans sa partie occidentale [Dongmo, 1981]. Par conséquent, le couvert végétal y est dominé principalement par les cultures vivrières et maraichères [Lounang et al., 2014].

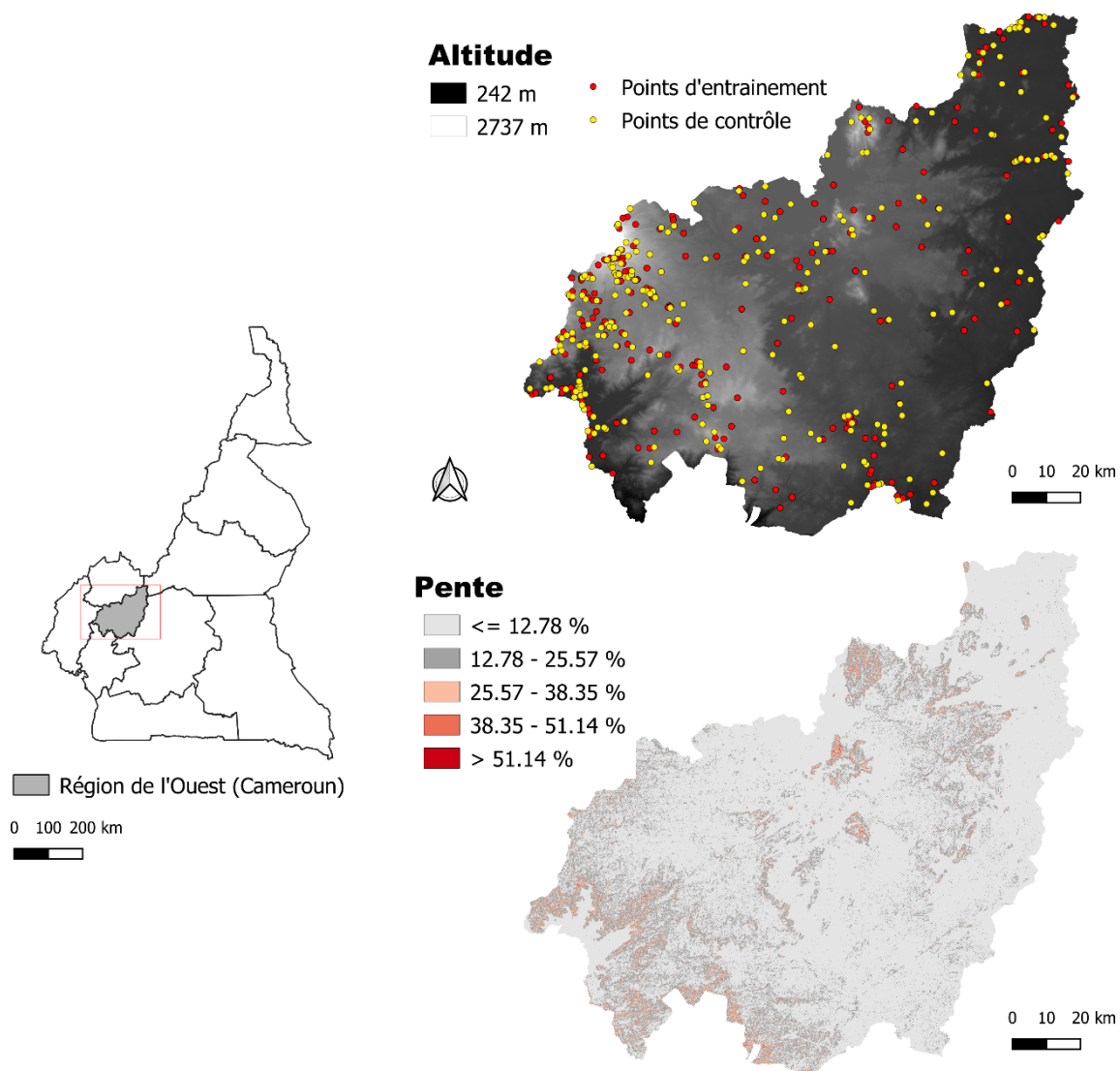


Figure 1 : Localisation de la zone d'étude et des régions d'intérêt utilisées pour la classification

2.2. Le choix des classes et des régions d'intérêt (ROI- Region Of Interest)

Pour tester les performances de la RF pour classier les différents types d'occupation du sol dans notre zone d'étude, nous avons retenu quinze classes d'occupation du sol (tableau 1) conformément aux travaux de

Matsaguim et al. (2019). Par ailleurs, un total de 512 ROI principalement localisées sur les secteurs à fortes pentes de la zone d'étude a été défini (figure 1). La définition de ces ROI s'est faite selon le principe de proportionnalité : le nombre de ROI par classe étant fonction de sa représentativité sur l'image.

| N° | Classes d'occupation du sol | N° | Classes occupation du sol |
|----|---|----|---|
| 1 | Villes, Zone d'habitation, Infrastructures | 9 | Dépôt de sable |
| 2 | Cultures (vivrier, maraicher) | 10 | Zones humides mises en valeur (Agriculture) |
| 3 | Lacs / Retenues d'eau | 11 | Prairies / Pâturages |
| 4 | Cultures industrielles (Thé) | 12 | Raphiale / Végétation hygrophyte |
| 5 | Végétation ligneuse peu dense | 13 | Zones humides non mises en valeur |
| 6 | Végétation ligneuse dominante et dense | 14 | Cours d'eau |
| 7 | Végétation mixte (arbres + herbes + sols nus) | 15 | Surfaces brûlées |
| 8 | Sols nus / Rochers / Carrières | | |

Tableau 1 : Les différentes classes d'occupation du sol retenues

2.3. Les données utilisées

L'image satellite utilisée provient du capteur Landsat Operational Land Imager (OLI)-8 acquise le 27 janvier 2018 à 09:32:22. Cette période correspond à la saison sèche dans notre zone d'étude. Durant cette saison qui s'étend de décembre à la mi-mars, l'humidité de l'air et la nébulosité sont les plus faibles de l'année, tandis que l'éclairement solaire est élevé. Il s'agit de la scène 186/056 (PATH/ROW) d'une résolution spatiale de 30 x 30 m. Elle recouvre à plus de 95% notre zone d'étude. Elle est sans couverture nuageuse et a été prise dans des conditions d'illumination moyennes (élévation du soleil = 52,34° et azimut = 130,63°). Lors de la phase

de prétraitement, cette image a été corrigée des perturbations atmosphériques grâce au modèle Landsat Ecosystem Disturbance Adaptive Processing System – LEDAPS [Schmidt et al., 2013]. Ce modèle permet de réduire la variabilité spectrale intra-classe comme le montre la figure 2 comparativement à ce qui s'observe sur l'image non corrigée. L'intervalle de variation a été obtenu en calculant la moyenne et l'écart-type de l'ensemble des ROI définies pour chaque classe. A titre d'illustration, seules les valeurs obtenues à partir de la bande 1 ont été présentées (figure 2). La correction des effets de la topographie sur les valeurs spectrales des pixels n'a pas été effectuée au regard des résultats obtenus par Tiomo et Matsaguim (2020).

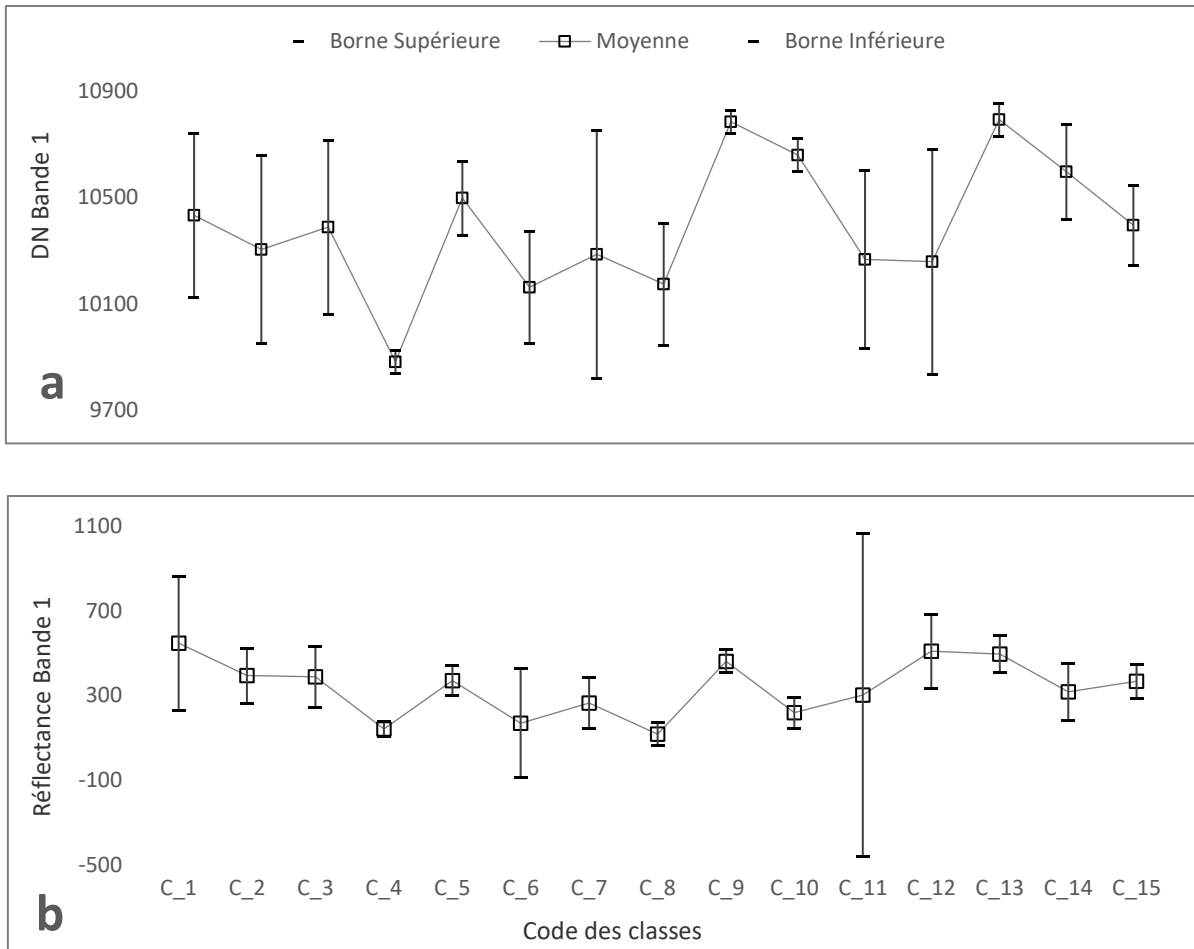


Figure 2 : Comparaison des intervalles de variation des valeurs spectrales des différentes classes d'occupation du sol pour l'image (a) non corrigée et (b) celles corrigées des influences de l'atmosphère avec le modèle LEDAPS

Afin de tenir compte de l'impact de la nature des images satellites utilisées ainsi que de leur nombre sur la précision du modèle [Raczko et Zagajewski, 2017], nous avons utilisé quatre groupes d'images. Le premier regroupe les bandes multispectrales de l'image satellite ; tandis que les trois autres groupes sont constitués d'images dérivées de l'image satellite : Transformation ACP (Analyse en Composantes Principales) ; Transformation Tasseled Cap ; Indices spectraux (tableau 2). En ce qui concerne le groupe des indices

spectraux, compte tenu du très grand nombre d'indices qui existent dans la littérature, nous avons retenu ceux qui permettent de mettre en évidence l'activité végétale (NDVI & MSAVI), l'humidité des plantes et du sol (MDMI), les surfaces brûlées (NBR & NBR 2), la brillance des sols (BI & MBI), et les étendues d'eau libre telles les cours d'eau, les lacs et retenues d'eau (FWI) (tableau 2). Ce choix a tenu compte des différentes classes d'occupation du sol retenues (tableau 1).

| Groupes | Nomenclature | Description |
|---------------------------------|---|--|
| (1) Multispectrale | Bleu ; Vert ; Rouge, Proche Infrarouge (PIR), Moyens infrarouges (MIR 1 & 2) | Il s'agit des différentes bandes ou « channels » issues de l'image satellite Landsat 8 OLI. |
| (2) Transformation ACP | Composantes 1 à 6 | Il s'agit des images issues de l'analyse en composantes principales effectuée sur les images du groupe 1 |
| (3) Transformation Tasseled Cap | Brightness ; Greenness ; Wetness | Il s'agit des images issues de méthode de transformation Tasseled Cap appliquée aux images du groupe 1. L'algorithme utilisé est celui développé par Baig et al. (2014) pour les images issues du capteur Landsat 8 OLI. |
| (4) Indices spectraux | Normalized Difference Vegetation Index (NDVI) | $\frac{PIR - Rouge}{PIR + Rouge}$ |
| | Modified Soil Adjusted Vegetation Index (MSAVI) | $\frac{(2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - R)})}{2}$ |
| | Normalized Difference Moisture Index (NDMI) | $\frac{PIR - MIR 1}{PIR + MIR 1}$ |
| | Normalized Burn Ratio (NBR) | $\frac{PIR - MIR 2}{PIR + MIR 2}$ |
| | Normalized Burn Ratio 2 (NBR2) | $\frac{MIR 1 - MIR 2}{MIR 1 + MIR 2}$ |
| | Brightness Index (BI) | $\sqrt{Rouge^2 + PIR 2^2}$ |
| | Modified Brightness Index (MBI)* | $\sqrt{Bleu^2 + MIR 2^2}$ |
| Free water Index (FWI)* | $\frac{\sqrt{Bleu^2 + Rouge^2} - \sqrt{Bleu^2 + MIR 2^2}}{\sqrt{Bleu^2 + Rouge^2} + \sqrt{Bleu^2 + MIR 2^2}}$ | |

(*): il s'agit d'indices spectraux définis par les auteurs dans le cadre de cette étude.

Tableau 2 : les différents groupes de données images utilisées

2.4. Démarche de l'opération de classification des différentes images

2.4.1. Paramétrage de l'algorithme de classification : Forêt d'Arbres Aléatoires

Afin de réaliser la classification de nos différents groupes d'images, nous avons utilisé le logiciel ArcMap 10.6 qui propose une version adaptée de l'algorithme RF proposée par Breiman (2001). En plus des variables d'entrées, l'utilisateur doit définir trois autres critères : le nombre d'arbres à générer, le nombre maximal de règles ou « branches » par arbre, le nombre maximal de pixels échantillonnés qui sera utilisé pour entraîner chaque classe d'occupation du sol qui a été retenue. La définition de ces différents paramètres s'est appuyée sur les travaux de [Shi et Yang, 2016 ; Rodriguez-Galiano et al., 2012]. Des tests préliminaires sur différents groupes d'images nous ont conduit à retenir empiriquement un nombre maximum de 100 arbres générés et un nombre maximum de règles (branches) de 50 par arbre. En effet, nous avons pu constater qu'au-delà de ces valeurs, le temps de calcul de la RF augmentait significativement pour une amélioration statistiquement non significative au seuil de 95% de la précision du modèle obtenu en utilisant un test de comparaison de la moyenne. De plus, nous avons choisi d'inclure tous les pixels de l'échantillon d'entraînement qui ont été définis pour chaque classe pour la construction des modèles.

2.4.2. Déroulement de la classification

Afin de tenir compte de l'impact du nombre et de la corrélation entre les bandes spectrales utilisées pour entraîner le modèle et classifier l'image, nous avons réalisé la classification sur les différents groupes d'images séparément dans une première phase. Lors de cette phase, pour chaque groupe d'images, la classification a été réalisée de manière répétitive sur les mêmes bandes spectrales en prenant soin d'extraire à chaque nouvelle classification l'image qui a été considérée comme la moins importante pour la construction du modèle lors de la classification précédente. En effet, après chaque phase d'entraînement du modèle, l'importance de chacune des images, qui est calculée par la RF, était notée. L'image la moins importante était extraite et un nouveau modèle était construit à partir des images restantes. Le processus se poursuivait ainsi jusqu'à atteindre un nombre minimum d'images au sein du groupe. Vu les résultats obtenus par Shi et Yang (2016) concernant la relation entre le nombre de bandes spectrales utilisé en entrée et le niveau de précision du modèle, nous avons retenu un nombre minimal de trois bandes spectrales compte tenu des classes d'occupation du sol retenues et du niveau de précision du modèle obtenu. Dans une seconde phase, l'ensemble des images des différents groupes a été regroupé pour servir à la création d'une série de nouveaux modèles conformément à la procédure décrite précédemment.

Au final, 35 modèles ont été construits. La figure 3 représente schématiquement la démarche décrite dans ce paragraphe.

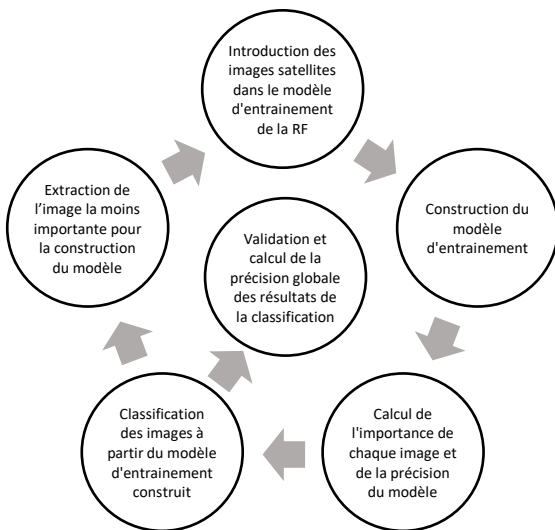


Figure 3 : Démarche de construction des modèles d'entraînement de classification des images satellites

2.5. L'analyse de la relation entre la précision du modèle d'entraînement et la précision globale de la classification

Afin de tester l'hypothèse de l'existence d'une relation linéaire positive entre le niveau de précision du modèle d'entraînement et celui de la classification de l'image satellite, nous avons utilisé le test de corrélation de Pearson avec seuil de significativité de 95%, soit un p-value < 0,05. Ce test a également été utilisé pour évaluer l'indépendance entre les différentes bandes spectrales d'une image notamment en ce qui concerne le groupe des indices spectraux.

3. Résultats

3.1. Influence du nombre et de la qualité des images sélectionnées sur le niveau de précision du modèle

Nos premiers résultats montrent que d'une manière générale, le nombre d'images utilisées lors de la phase d'entraînement du modèle ne détermine pas le niveau de précision du modèle obtenu à la fin de cette phase. La figure 4 permet de se rendre compte de cela. Elle montre bien dans notre cas qu'au-delà d'un certain nombre d'images, l'ajout de nouvelles images n'améliore pas la précision du modèle. Dans un premier temps, nous constatons que l'accroissement du nombre d'images passant de trois jusqu'à six permet d'améliorer la précision du modèle de 80 à 95% (figure 4). Dans un second temps, au-delà de six images, le niveau de précision du modèle obtenu n'augmente plus et tend à plafonner autour de 95%. Il n'existe donc pas une relation linéaire entre le nombre de variables explicatives utilisées pour entraîner le modèle et son niveau de précision. La stratégie qui consisterait à introduire un grand nombre de variables explicatives de nature différente pour tenter de classer une image satellite avec la RF n'est pas nécessairement la bonne, y compris

dans les zones montagneuses. Bien entendu, ce résultat est en partie lié à la qualité des ROI qui a été utilisé comme l'ont indiqué Shi et Yang (2016). La figure 4 tend à démontrer que bien plus que le nombre d'images introduites, c'est davantage leur qualité qui importe.

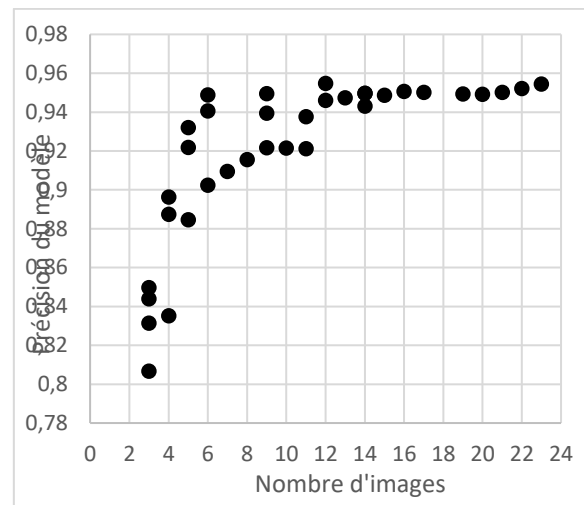


Figure 4 : Relation entre le nombre de bandes spectrales et la précision du modèle d'entraînement

Dès lors s'agissant de la qualité des images à retenir comme variables explicatives permettant de correctement classer les différents pixels de l'image, nous avons pu constater que deux critères sont avancés pour effectuer ce choix [Raczko et Zagajewski, 2017 ; Millard et Richardson, 2015]. Il s'agit de l'indépendance des images les unes par rapport aux autres et de leur importance dans le modèle. La figure 5 présente dans l'ordre décroissant le niveau de précision des modèles obtenus en utilisant les différents groupes d'images séparément. Il en ressort que le groupe 2, qui est composé uniquement des images issues de l'ACP réalisée sur les images multispectrales, est celui qui permet d'obtenir la précision la plus élevée, soit 0,95. Cette valeur est assez proche de celle obtenue à partir des images multispectrales du groupe 1 (0,94) avec lequel il compte le même nombre d'images, soit six chacun. Par conséquent, cette différence ne peut être expliquée que par le faible niveau de corrélation entre les différentes images du groupe 2 contrairement à ceux du groupe 1. En effet, l'une des propriétés de l'ACP en télédétection est de produire des composantes sous la forme d'images qui sont plus ou moins totalement indépendantes les unes des autres. Les images du groupe 3 (Transformation Tasseled Cap) sont celles qui produisent le niveau de précision le plus bas parmi les quatre groupes (figure 5). Tout comme l'ACP, la méthode transformation Tasseled Cap permet d'obtenir des images indépendantes les unes des autres. Cependant, nous n'avons retenu dans ce groupe que les trois principales images sur les six que produit cette méthode de transformation. Ce choix semble avoir conduit à une perte importante d'informations compte tenu de la qualité de nos ROI. Le groupe 3 composé des indices spectraux permet d'obtenir un niveau de précision de près de 0,92. Ces indices ont été sélectionnés en tenant compte des différentes classes d'occupation du sol qui ont été retenues pour notre étude, mais également du niveau d'indépendance entre chaque indice. Cette indépendance a été déterminée en

effectuant un test de corrélation de Pearson entre les images des différents indices pris deux à deux.

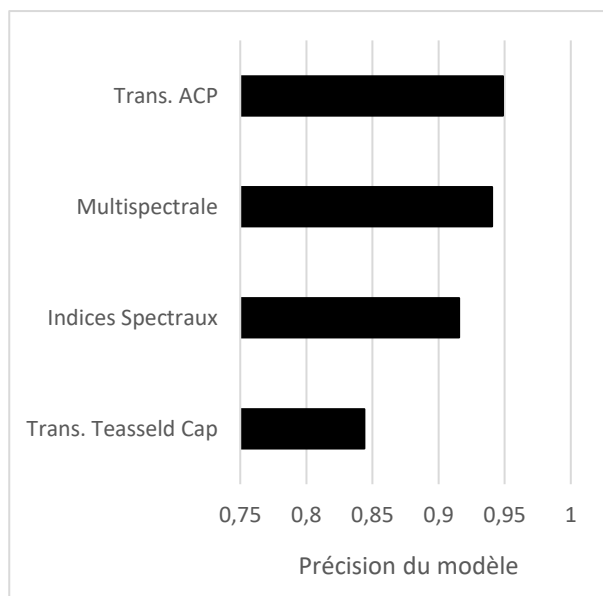


Figure 5 : Précisions du modèle obtenues pour chaque groupe d'images séparément

Enfin, concernant l'importance de chaque variable explicative dans la construction du modèle comme critère de sélection des images, la figure 6 montre que dans certains cas de figure, la stratégie qui consiste à retirer des modèles, les variables les moins importantes pour ne retenir que celles qui participent le plus à la classification peut se révéler très préjudiciable au vu de la qualité des données d'entraînement. Nous nous rendons compte qu'en retirant successivement les bandes les moins importantes pour ne retenir que celles les plus importantes au sein des différents groupes d'images, cela a un impact négatif sur la précision des différents modèles. Sur la figure 6, l'axe des abscisses a été volontairement inversé pour les besoins de représentation. Elle montre qu'il existe une relation non linéaire positive. Autrement dit, l'augmentation du nombre d'images se traduit par une amélioration de la précision du modèle. Cette amélioration de la précision est assez importante dans un premier temps, puis elle diminue à partir d'un certain nombre.

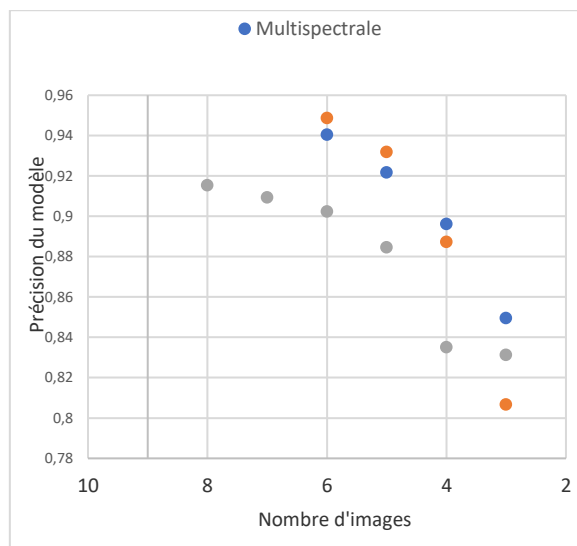


Figure 6 : Relation entre le nombre d'images et la précision du modèle pour les différents groupes

3.2. Relation entre la précision du modèle et celle du résultat de la classification

L'hypothèse qu'il existe une relation linéaire entre le niveau de précision (OOB) du modèle issu de l'étape d'entraînement et celui de la précision globale de la carte thématique résultant de la classification de l'image satellite avec l'algorithme de classification RF se vérifie. En effet, la figure 7 (a et b) met en évidence l'existence d'une relation linéaire entre ces deux variables. Un niveau de précision du modèle d'entraînement élevé permet d'obtenir une carte thématique plus précise. Cependant, dans notre cas et en tenant compte de la qualité de nos données d'entraînement, il apparaît que cette relation n'est plus valable au-delà d'une certaine limite. En observant la figure 7, nous constatons qu'il semble exister un seuil au-delà duquel l'accroissement de la précision du modèle obtenu après entraînement ne conduit plus à une amélioration des résultats de la classification, mais bien au contraire à leur détérioration. Il est possible de se rendre compte qu'au-delà d'un niveau de précision du modèle d'entraînement égal à 0,95, il s'observe systématiquement une détérioration de la qualité des cartes thématiques comparativement à celle obtenue à partir des modèles d'une précision supérieure à 0,945. Bien que cette détérioration soit assez faible, c'est surtout son caractère systématique pour tous les modèles ayant obtenu une précision supérieure à 0,95 qui est surprenant. Cela est valable pour les deux indicateurs de précision que sont la précision globale (figure 7a) et le coefficient de Kappa (figure 7b). Une analyse des résultats permet de se rendre compte que les classifications les plus précises sont obtenues à partir des modèles d'entraînement dont la précision est comprise entre [0,945 – 0,95].

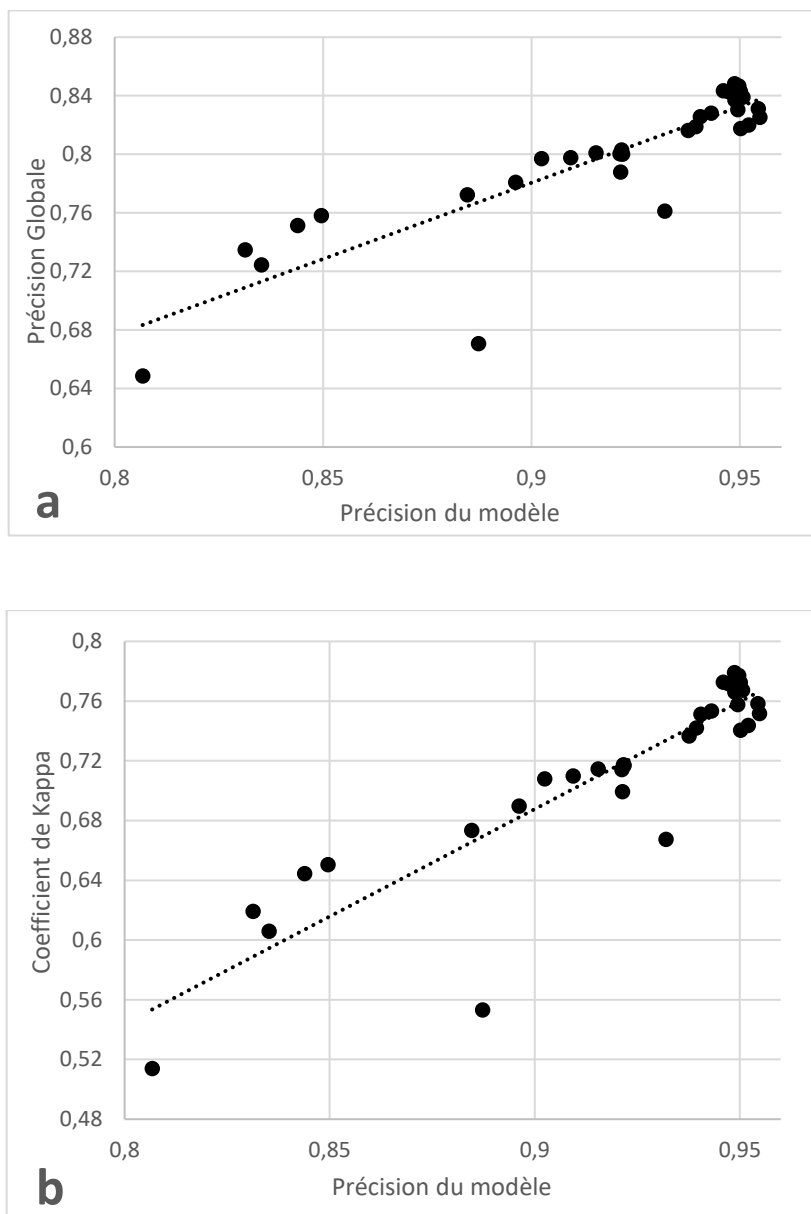


Figure 7 : Relation entre le niveau de précision du modèle et les différents indicateurs de mesure de la précision globale des résultats de la classification

L'analyse de la corrélation entre les différentes images utilisées comme variables explicatives pour la construction de ces différents modèles permet de se rendre compte que tous les modèles d'une précision supérieure à 95% se caractérisent par une dichotomie au sein des images utilisées pour sa construction. Au sein de ces images, on distingue deux groupes du point de vue de leur dépendance les unes par rapport aux autres : d'un côté un groupe d'images très fortement corrélées entre elles et de l'autre un groupe d'images peu ou pas corrélées entre elles. Par contre, les modèles

dont la précision est comprise entre $[0,945 - 0,95[$ qui ont permis d'obtenir les classifications les plus précises sont tous caractérisés par la faible à très faible corrélation entre toutes les images utilisées pour leur construction. Pour ces modèles, il est de ce fait très important que la totalité des images utilisées permettent toutes ensembles de correctement discriminer entre les différentes classes d'occupation du sol qui ont été retenues. Dans notre étude, le modèle qui a obtenu la classification la plus précise associait les images des groupes 2 et 4 (tableau 2).

4. Discussion

La RF est un algorithme de classification qui permet de prendre en compte un nombre important d'images différentes. Cela s'avère très utile surtout lorsque les phénomènes spatiaux considérés présentent une grande hétérogénéité [Rodriguez-Galiano et al., 2012]. Cependant, il apparaît clairement que le choix des différentes classes d'occupation du sol et la qualité des ROI ont un impact important sur la sélection des images à utiliser. Par ailleurs, nos résultats confirment ceux obtenus précédemment [Raczko et Zagajewski, 2017 ; Millard et Richardson, 2015] et mettent en avant l'importance de la sélection des images par rapport à leur nombre : plus d'images n'implique pas nécessairement de meilleurs résultats. Le choix des images à utiliser doit se faire en tenant compte d'un certain nombre de critères. A ce propos, Shi et Yang (2016) précisent que le nombre d'images à utiliser dépend du niveau d'homogénéité spectrale des différentes classes d'occupation du sol retenues pour la classification. Nos résultats ont montré que, compte tenu des classes d'occupation du sol retenues (tableau 1) et de la variabilité spectrale des ROI (figure 2), dans le cas où les ROI sont tous homogènes, l'utilisation des images situées dans le spectre du Visible (Bleu, Vert, Rouge) et du Proche et Moyen infrarouge suffisent à correctement classer une image satellite grâce à la RF. Dans le cas contraire, il sera nécessaire d'augmenter le nombre d'images pour correctement classer les classes qui présentent une forte hétérogénéité spectrale. Cependant, cette augmentation n'est pas sans risque comme nous avons pu le constater. Il faut tenir compte de l'indépendance entre les différentes images. Raczko et Zagajewski (2017) et Millard et Richardson (2015) suggèrent par exemple de retenir un grand nombre d'images plus ou moins corrélées entre elles dans un premier temps, puis dans un second temps de procéder à une réduction de leur dimension au moyen d'une analyse en composantes principales afin d'obtenir un nombre restreint d'images indépendantes les unes des autres. Ce critère d'indépendance est particulièrement important en ce qui concerne la RF car il détermine sa capacité à prendre en compte des images qui apportent peu d'informations pour la classification des différents pixels d'une image satellite [Breiman, 2001].

D'un autre côté, la qualité des ROI n'est pas que déterminante pour la sélection des images à utiliser. Les résultats de la classification en dépendent étroitement indépendamment de la robustesse de la RF vis-à-vis des données manquantes ou aberrantes et du surajustement [Gislason et al., 2006 ; Breiman, 2001]. En effet, en matière de classification d'images satellites, Kulkarni et Lowe (2016) montrent que la performance de la RF dépend de trois critères : [1] la taille de chaque classe d'occupation du sol utilisée pour entraîner l'algorithme de classification ; [2] l'homogénéité spectrale au sein de chacune des classes ; et [3] la taille globale de l'échantillon d'entraînement ou ROI. Par conséquent, cela implique que tout résultat est étroitement dépendant de la qualité des ROI. Ainsi, Millard et Richardson (2015) recommandent que la distribution des valeurs spectrales au sein de chaque classe d'entraînement doit être aléatoire. Pour les classes hétérogènes, Shi et Yang (2016) préconisent de les scinder en deux classes afin d'assurer un niveau de précision satisfaisant des résultats de la classification.

A cet effet, l'hypothèse selon laquelle il existe une relation linéaire positive entre le niveau de précision du modèle d'entraînement et celui de la carte thématique obtenu après classification de l'image semble être valide. C'est ce que montrent du moins nos résultats. En effet, le calcul du coefficient de corrélation de Pearson entre ces deux variables donne une valeur de 0,88. Cependant, il est important de relever que l'hypothèse de normalité des valeurs que suppose le test de corrélation de Pearson n'est pas vérifiée. En effet, nos variables ont toutes deux une distribution asymétrique négative. Néanmoins, le recours au test de corrélation non paramétrique de Kendall's tau qui est plus robuste sur de petits échantillons nous donne une valeur du tau égale à 0,65. Ce résultat valide bien l'hypothèse d'une relation linéaire positive entre le niveau de précision du modèle d'entraînement et celui des résultats de la classification et confirme ainsi le point de vue de Breiman (2001).

Conclusion

En définitif, les résultats de cette étude mettent clairement en avant l'importance du choix des images à utiliser lors de l'étape d'entraînement de l'algorithme de classification : Forêt d'Arbres Aléatoires. Ce choix doit être guidé par au moins trois critères importants. Il s'agit de [1] les différentes classes d'occupation du sol qui ont été retenues ; [2] la qualité des zones d'entraînement ou ROI qui vont servir lors de l'étape d'entraînement ; et [3] l'indépendance entre les différentes images qui seront retenues. La qualité des ROI renvoie d'une part à l'homogénéité intra-classe, et d'autre part à la taille de l'échantillon. C'est cette qualité qui va déterminer le niveau de précision du modèle qui sera obtenu à la fin de l'étape d'entraînement et qui servira par la suite à classer l'image satellite. Ce niveau de précision du modèle est un indicateur important du niveau de la précision globale de la carte thématique qui sera obtenue à la fin de l'étape de classification. En effet, ces deux paramètres sont liés positivement pour ce qui est de notre zone d'étude et de l'image satellite utilisée. Par conséquent, il semblerait que plus un modèle est précis, plus sera précise la carte thématique qui en résultera. Dès lors, malgré la robustesse intrinsèque de la Forêt d'Arbres Aléatoires comme algorithme de classification [Breiman, 2001], ses performances dépendent avant toute chose de la qualité des données d'entrée ; meilleures elles seront, meilleurs seront les résultats. Il est donc important d'y accorder une attention particulière comme l'ont fait remarquer Xie et al. (2019) pour qui, en matière de classification des images satellites, les performances des différents algorithmes de classification dépendent [1] des images satellites utilisées, [2] des différents objets ou classes d'objets que nous cherchons à identifier, et [3] des caractéristiques des données d'entraînement qui seront utilisées pour classer l'image. Cette remarque permet de soulever un certain nombre de questions qu'il serait intéressant d'aborder ultérieurement en ce qui concerne l'évaluation des facteurs qui influencent les performances de la RF pour la classification de types d'occupation du sol en zone de montagne, à savoir de connaître l'influence : (1) du type d'images satellite utilisées et de la période d'acquisition ; (2) du nombre et du type de classes retenues ; (3) de la taille globale des ROI et de leur nombre par classe.

Références

- Ayala-Izurieta J.E., Márquez C.O., Garcia V.J., Recalde-Moreno C.G., Rodríguez-Llerena M.V., Damián-Carrión D.A., 2017. *Land cover classification in an Ecuadorian mountain geosystem using a random forest classifier, spectral vegetation indices, and ancillary geographic data*. *Geosciences* 7 (34), 21 p. DOI :10.3390/geosciences7020034
- Baig M.H.A., Zhang L.T., Shuai T., Tong Q., 2014. *Derivation of a Tasseled Cap Transformation based on Landsat 8 at satellite reflectance*. *Remote Sensing Letters* 5 (5), pp. 423-431.
- Breiman L., 2001. *Random Forests*. *Machine Learning* 45, pp. 5-32.
- Colditz R.R., 2015. *An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using Decision Tree-Based Algorithms*. *Remote Sensing* 7, pp. 9655- 9681. DOI : 10.3390/rs70809655
- DeFries, R.S., Chan, J.C.-W., 2000. *Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data*. *Remote Sensing of Environment* 74 (3), pp. 503-515.
- Dedieu, 1989. *Téledétection et montagnes. Un outil privilégié pour l'étude des milieux d'altitude ?* *Revue de géographie alpine* 77(1-3), pp. 105-123. DOI :10.3406/rga.1989.2734
- Dongmo J.L., 1981. *Le dynamisme Bamiléké (Cameroun). Volume I : la maîtrise de l'espace agricole*. Yaoundé : CEPER, 424 p.
- Gislason P.O., Benediktsson J.A., Sveinsson J.R., 2006. *Random forests for land cover classification*. *Pattern Recognition Letters* 27, pp : 294-300.
- Jensen J.R., 2015. *Introductory digital image processing: a remote sensing perspective*. USA : Pearson.
- Jhonnerie R., Siregar V.P., Nababan B., Prasetyo L.B., Wouthuyzen S., 2015. *Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries*. *Procedia Environmental Sciences* 24, pp. 215-221. DOI: 10.1016/j.proenv.2015.03.028
- Kolden C.A., Lutz J.A., 2016. *Detecting unburned areas within wildfire perimeters using Landsat and ancillary data across the northwestern United States*. *Remote Sensing of Environment* 186, pp. 275-285. <http://dx.doi.org/10.1016/j.rse.2016.08.023>
- Kulkarni A.D., Lowe B., 2016. *Random forest algorithm for land cover classification*. *International Journal on Recent and Innovation Trends in Computing and Communication* 4 (3), pp. 58 - 63.
- Lippitt, C.D., Li, Z., Eastman, J.R., Jones, T.G., 2008. *Mapping selective logging in mixed deciduous forest: a comparison of machine learning algorithms*. *Photogrammetric Engineering and Remote Sensing* 74 (10), pp. 1201-1211.
- Loungang T.F.C., Djamfa R.C., Youta H.J., Tchawa P., Grozavu A., 2014. *Expansion des cultures de contre-saison, changements d'utilisation du sol et les implications environnementales dans les paysannes de l'ouest-Cameroun*. *Scientific Annals of "Al I. CUZA" University of IASI LX* (2), pp.41-57. DOI : 10.15551/scigeo.v60i2.316
- Matsaguim N.C.A., Tiomo D.E., Ngoufo R., 2019. *Méthode simplifiée pour la cartographie du sol en zone de montagne : cas du bassin versant de la Menoua (Région de l'ouest-Cameroun)*. *Revue Scientifique et Technique Forêt et Environnement du Bassin du Congo* 13, pp.21-33. Meddens A.J.H.,
- Mellor A., Haywood A., Stone C., Jones S., 2013. *The Performance of Random Forests in an Operational Setting for Large Area Sclerophyll Forest Classification*. *Remote Sensing* 5, pp. 2838-2856. DOI : 10.3390/rs5062838
- Millard K., Richardson M., 2015. *On the importance of training data sample selection in Random Forest image classification : A case study in Peatland Ecosystem Mapping*. *Remote Sensing* 7, pp. 8489-8515. DOI : 10.3390/rs70708489
- Ming D., Zhou T., Wang M., Tan T., 2016. *Land cover classification using random forest with genetic algorithm-based parameter optimization*. *Journal of Applied Remote Sensing* 10 (3), 16 p.
- Nguyen H.T.T., Doan T.M., Radeloff V., 2018. *Applying random forest classification to map land use/land cover using Landsat 8 OLI*. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3/W4*, pp. 363 - 367.
- Ozdarici Ok A., Akar O., Gungor O., 2012. *Evaluation of random forest method for agricultural crop classification*. *European Journal of Remote Sensing* 45 (1), pp. 421-432. DOI: 10.5721/EuJRS20124535
- Pelletier C., Valero S., Inglada J., Champion N., Dedieu G., 2016. *Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas*. *Remote Sensing of Environment* 187, pp. 156-168. <http://dx.doi.org/10.1016/j.rse.2016.10.010>
- Raczko E., Zagajewski B., 2017. *Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images*. *European Journal of Remote Sensing* 50 (1), pp.: 144-154. DOI:10.1080/22797254.2017.1299557
- Rodriguez-Galiano V.F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J.P., 2012. *An assessment of the effectiveness of a random forest classifier for land-cover classification*. *ISPRS Journal of Photogrammetry and Remote Sensing* 67, pp. 93-104.
- Schmidt G.L., Jenkerson C.B., Masek J., Vermote E., Gao F., 2013. *Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description*. U.S. Geological Survey Open-File Report 2013-1057, 17 p.
- Shi D., Yang X., 2016. *An assessment of algorithmic parameters affecting image classification accuracy by Random Forests*. *Photogrammetric Engineering & Remote Sensing* 82 (6), pp. 407-417. DOI: 10.14358/PERS.82.6.407
- Stehman S.V., Foody G.M., 2009. *Accuracy Assessment*. in Warner T.A., Nellis M.D., Foody G.M.,

(ed), *The SAGE Handbook of Remote Sensing*. SAGE Publications Ltd, London, pp. 297–309.

Tiomo D.E., Matsaguim N.C.A., 2020, *Evaluation de l'impact de la correction topographique sur la précision globale de la cartographie de l'occupation du sol en zone de montagne tropicale : cas de la 'Région de l'ouest-*

Cameroun. Revue Scientifique et Technique Forêt et Environnement du Bassin du Congo 14, pp.22-31.

Xie Z., Chen Y., Lu D., Li G., Chen E., 2019. *Classification of land cover, forest, and tree species classes with ziyuan-3 multispectral and stereo data*. Remote Sensing 11 (164), 21 p. DOI : 10.3390/rs11020164