

Rapport de stage de Master 2 Aménagement des Territoires et Télédétection
de l'Université de Toulouse III – Paul Sabatier

APPLICATION D'ALGORITHME D'OCCUPATION DU SOL EN HAÏTI - AIDE À LA RECONSTRUCTION POST OURAGAN

Edna Lized MELO RAMIREZ

Encadrants :

Agwilh COLLET
Responsable technique du
Recovery Observatory
CNES - DSO/SI/2A

Julien MICHEL
Ingénieur en traitement d'images
CNES - DSO/SI/2A

Stage réalisé au CNES du 19 mars 2018 au 14 septembre 2018 soutenu le
17 septembre 2018

Toulouse, France, 2018

Remerciements

Si ce stage a été aussi enrichissant académiquement et professionnellement c'est principalement grâce à l'accompagnement que j'ai eu de la part de mes encadrants. Merci donc Agwilh et Julien de la confiance portée sur moi pour la réalisation de ce travail mais surtout de votre investissement et votre disposition au cours de ces six mois.

Je tiens à remercier vivement Arthur Vincent et Vincent Thierion du CESBIO qui se sont montrés très collaborateurs et prêts à m'aider à surmonter les difficultés techniques auxquelles j'ai été souvent confrontées lors de ce stage.

Merci également à Santiago Peña et à Manuel Grizzonet pour la bonne ambiance dans le bureau et principalement pour leur aide en continu, leurs conseils techniques et leurs commentaires sur ce rapport.

Un grand merci à Robin Faivre du SERTIT qui avec ses commentaires et suggestions m'a beaucoup orienté dans la discussion des résultats. Je le remercie pour l'enthousiasme qu'il a montré pour mon travail et sa volonté de faire la relecture de ce rapport.

De manière générale, merci à tous les membres du service 2A pour leur accueil sympathique et leur bonne énergie.

Et comme dans la vie tout n'est pas que du travail, heureusement, merci au *Cercle restreint* pour les repas prolongés de tous les mardis, et surtout à Celas pour les pauses chargées de rires et de complicité.

Enfin, merci Stéph pour ton regard toujours critique sur mon travail.

Table des matières

Résumé	4
Introduction	5
1. Contexte du stage.....	8
1.1. Centre National d'Études Spatiales	8
1.2. Cahier des charges du stage.....	8
2. Application d'algorithme d'occupation du sol en Haïti.....	9
2.1. Site d'étude.....	9
2.2. Les données	9
2.2.1. Images Sentinel-2	10
2.2.2. Données de référence.....	13
2.3. La chaîne <i>iota</i> ²	14
2.3.1. Fonctionnement	14
2.4. Préparation des données de référence.....	17
2.4.1. Données de type polygone.....	17
2.4.2. Données de type polyligne	23
2.4.3. Masque d'eau <i>Global Surface Water</i>	23
2.5. Lancement de <i>iota</i> ²	24
3. Résultats et discussion	25
4. Conclusions	31
RÉFÉRENCES	32
ANNEXES.....	36

Résumé

Le *Recovery Observatory*, mis en place par le Comité sur les satellites d'observation de la Terre (*Committee on Earth Observation Satellites* ; CEOS) et co-dirigé par le Centre National d'Études Spatiales (CNES), accompagne depuis début 2017 la récupération des infrastructures et des activités agricoles et la réhabilitation environnementale d'Haïti postérieur au passage dévastateur de l'ouragan Matthew en octobre 2016. Parmi les objectifs du *Recovery Observatory* l'on compte la production de la carte d'occupation du sol d'Haïti actualisée en exploitant de données satellitaires optiques ainsi que des outils de traitement et des données en libre accès comme la chaîne de traitement *iota*² développée par le Centre d'Étude Spatiales de la Biosphère (CESBIO) et les données OpenStreetMap. Dans ce contexte, le but de ce stage a été d'identifier le potentiel de *iota*² avec des données de référence limitées. *iota*², conçue pour le territoire de la France métropolitaine, n'avait jamais été testée ailleurs sur l'ensemble d'un pays.

Dans ce travail, la chaîne de classification du sol *iota*² a été exécutée avec la série temporelle 2017 des images Sentinel-2 sur le territoire haïtien et des données de référence principalement issues d'OpenStreetMap. Le pré-traitement des données de référence a été l'étape majeure dans ce stage compte tenu du gros volume de bases de données et de leur qualité en termes de propreté du contenu. Les principaux résultats ont montré que les performances de la chaîne sont d'autant plus satisfaisantes que des données de référence pertinentes sont disponibles. *iota*² s'est avérée opérationnelle sur Haïti avec des productions en accord avec les données de référence. A ce propos, le potentiel des données OpenStreetMap dans la production des cartes d'occupation du sol par classification à large échelle reste discutable vis-à-vis de l'hétérogénéité de leur disponibilité dans le monde. La Précision globale de la carte produite a été de 96,4% et l'indice Kappa de 88,7%. Néanmoins, l'évaluation visuelle du résultat n'est pas aussi satisfaisante. Les confusions entre les classes et la sur-classification, notamment de la classe Végétation boisée, ont mis en évidence le besoin aigu de données de référence pertinentes des classes Sol nu et Prairie.

Au cours de récents échanges avec des partenaires du *Recovery Observatory* du Centre National de l'Information Géo-Spatiale de Haïti et des collaborateurs du Service Régional de Traitement d'Image et de Télédétection (SERTIT), les résultats de ce travail étant jugés encourageants, les partenaires haïtiens ont de nouveau exprimé leur intérêt et leur motivation pour l'obtention des données de référence manquantes.

Mots clés : Classification d'images, télédétection, Sentinel-2, *iota*², occupation du sol.

Introduction

Peu de choses dans la nature peuvent se comparer à la force destructrice d'un ouragan. Connue comme la plus grande tempête sur Terre, un ouragan peut anéantir des zones côtières avec des vents soutenus de 250 km/h ou plus, des zones de précipitations intenses et une onde de tempête. Les ouragans sont classés parmi les risques naturels les plus courants et font chaque année des centaines, quelquefois des milliers, de victimes (Graham et Riebeek, 2006). « Ouragan » est le nom générique attribué à un cyclone tropical dans l'Atlantique nord, le golfe du Mexique, la mer des Caraïbes (dont l'archipel des Antilles) et le Pacifique nord et les zones côtières voisines. Il est appelé Typhon dans le Pacifique nord-ouest et cyclone tropical dans le Pacifique sud et les zones côtières voisines, dont la Nouvelle-Calédonie et la Polynésie française est dans le sud-est de l'océan Indien, entre l'Indonésie et l'Australie (Météo France, 2018).

Les registres reportant l'occurrence des cyclones dans le monde entier à partir de 1981 permettent d'identifier, de manière générale, la fréquence maximale de cyclones tropicaux vers la fin de l'été, lorsque l'eau est la plus chaude. Chaque bassin océanique a toutefois ses propres caractéristiques saisonnières et le nombre de cyclones tropicaux formés varie grandement d'une année à l'autre. Dans l'Atlantique nord, une saison des ouragans bien démarquée commence au début juin et se termine fin novembre, avec une forte poussée au début de septembre ; le nord-est du Pacifique a une période d'activité plus large mais similaire à celle de l'Atlantique ; le nord-ouest du Pacifique produit des typhons toute l'année, avec un minimum en février et une pointe au début de septembre ; dans le bassin du nord de l'océan Indien, les cyclones tropicaux sont plus fréquents d'avril à décembre, avec des pointes en mai et en novembre ; dans l'hémisphère sud, la formation de cyclones commence à la fin octobre et se termine en mai. Les pointes surviennent mi-février et début mars (*National Oceanic & Atmospheric Administration ; NOAA ; 2017*). Sur Haïti particulièrement, la saison des ouragans atteint son maximum entre août et la fin du mois de novembre.

Les enregistrements montrent que Haïti est régulièrement touché par des forts ouragans créant des dégâts majeurs et des nombreuses pertes humaines. Parmi les ouragans les plus dévastateurs ayant frappé le pays on compte l'ouragan Flora en octobre 1963 faisant près de 5000 morts ; l'ouragan Gordon en octobre 1994 provoquant environ 2000 morts et disparus ; l'ouragan Jeanne en septembre de 2004 avec un bilan mortel de plus de 3000 victimes ; et plus récemment, l'ouragan Matthew qui a occasionné la mort à près de 500 personnes, détruit des routes, de liens de communication et rasé 80% des bâtiments dans la ville de Jérémie ainsi que 90% des maisons de la côte sud de la Péninsule de Tiburon. La perte des récoltes a menacé le pays de famine. Matthew a été le premier ouragan majeur, et aussi le premier ouragan de catégorie 4 à toucher Haïti depuis Cleo en 1964 (Stewart, 2017)

Face à des catastrophes naturelles de ce type, des initiatives internationales ont été mises en place pour travailler sur la résilience des populations et des territoires impactés. C'est ainsi que le Comité sur les satellites d'observation de la Terre (*Committee on Earth Observation Satellites ; CEOS*), qui depuis 2014 travaille sur les moyens d'accroître la contribution des données satellitaires à la réhabilitation des territoires suite à ces

événements majeurs, a mis en place sur Haïti le *Recovery Observatory* (RO), codirigé par le Centre National d'Études spatiales (CNES). L'objectif du RO de Haïti est de suivre la reconstruction des bâtiments, des réseaux de transport, la reprise des activités agricoles et la réhabilitation environnementale pour une période de trois à quatre ans. De nombreux travaux et produits scientifiques à forte valeur ajoutée sont en cours de réalisation et beaucoup d'entre eux reposent sur une première carte d'occupation du sol d'Haïti pertinente. La dernière datant de 1998, le RO vise à produire la carte d'occupation du sol actualisée en exploitant de données satellitaires optiques.

Les images sont exploitées par une technique de classification numérique qui utilise l'information contenue dans les valeurs d'une ou de plusieurs bandes spectrales pour classifier chaque pixel individuellement en lui assignant une classe particulière (*i.e.* : eau, forêt, maïs, bâti, etc.). Le résultat de la classification est une nouvelle image composée d'une mosaïque de pixels qui appartiennent chacun à une classe particulière. Cette image est essentiellement une représentation thématique de l'image originale. Il existe plusieurs approches de la classification dont les méthodes de classification supervisée et les méthodes de classification non-supervisée. La principale différence est que dans la classification supervisée l'utilisateur spécifie les différentes valeurs de pixels ou les signatures spectrales à associer à chaque classe en sélectionnant des sites d'échantillonnage représentatifs de type de couverture connu. L'algorithme informatique utilise ces zones d'entraînement pour classer la totalité de l'image. En revanche, la classification non-supervisée n'utilise pas des données d'entraînement et c'est l'algorithme qui forme les classes spectrales sur la base des informations numériques contenues dans les données (les valeurs de pixels pour chacune des bandes ou indices ; RNCAN, 2018). Plusieurs travaux montrent que les méthodes de classification supervisées sont plus performantes que celles non supervisées (Khatami *et al.*, 2016 ; Szuster *et al.*, 2011). De nombreux algorithmes sont utilisés dans la classification d'occupation du sol. Les plus connus sont les *Support Vector Machines* (SVM), les Arbres de décision et le *Random Forest* (Breiman, 2001) dont plusieurs études montrent des résultats plus satisfaisants par rapport aux deux premiers classificateurs (Pelletier *et al.*, 2016 ; Inglada *et al.*, 2015 ; Rodriguez-Galiano *et al.*, 2012 ; Gislason *et al.*, 2006). Dans le but d'automatiser la procédure de classification, de traiter de volumes de données conséquents sur des grands territoires français en peu de temps, le Centre d'Études Spatiales de la Biosphère (CESBIO) a développé la chaîne de traitement *iota*² basée sur *Random Forest*. Des cartes d'occupation du sol sur la France métropolitaine entière ont déjà été produites par *iota*². Il reste à tester l'opérabilité de la chaîne sur d'autres grands sites d'étude. D'après Ma *et al.* (2017), la plupart des études de classification d'occupation du sol supervisée (96% des 173 publications scientifiques analysées) sont menées sur des zones de taille inférieure à 300 ha, ce qui montre la nécessité de zones d'étude plus vastes dans de futures recherches afin de vérifier l'applicabilité de la technique dans ces cas.

L'utilisation des images satellites dans la production des cartes d'occupation du sol est de plus en plus fréquente. L'imagerie multi-spectrale et multi-temporelle permet de caractériser les variations phénologiques de l'état des couvertures végétales (Rodriguez-Galiano *et al.*, 2012) et de détecter les différents composants de la surface terrestre. D'après Inglada *et al.* (2017) des images à haute résolution spatiale (métrique ou décimétrique) et temporelles

sont nécessaires pour produire des cartes détaillées d'occupation du sol. Les images Sentinel-2 (Drusch *et al.*, 2012) avec ses caractéristiques uniques (fauchée de 290 km, résolution spatiale de 10 à 60 m, revisite de 5 jours avec 2 satellites et 13 bandes spectrales) s'avèrent un outil puissant pour cartographier et surveiller les grands écosystèmes riches, complexes et sensibles (Yesou *et al.*, 2016). Ma *et al.* (2017) constatent également une corrélation positive entre la taille des zones d'étude et les résolutions spatiales des images utilisées. Les images à très haute résolution spatiale (<2 m) sont les plus exploitées. Néanmoins, dû à la facilité d'accès et la grande disponibilité les images Sentinel-2 sont souvent utilisées.

Dans cette optique, un des objectifs ponctuels du RO est de produire à partir des images Sentinel-2 la carte d'occupation du sol d'Haïti post-ouragan Matthew en mettant en œuvre la chaîne de classification *iota*². Dans ce contexte, l'apport particulier du présent stage sera d'identifier le potentiel de *iota*² avec des données de référence limitées. L'obtention de performances correctes avec des données en accès libre permettrait la production des cartes d'occupation du sol à grande échelle en peu de temps grâce à la possibilité d'automatisation de la procédure. Dans le contexte actuel d'Haïti, cela signifierait une avancée en termes de connaissance de l'état de leur territoire qui pourrait être exploitée dans les nombreux produits en cours de réalisation nécessitant des cartes d'occupation du sol pertinentes.

Ce document est divisé en quatre chapitres contenant : i) la présentation de la structure d'accueil et le cahier des charges du stage ; ii) la description détaillée de la méthodologie appliquée ; iii) la présentation et discussion des résultats ; et iv) une brève conclusion et les perspectives du travail.

1. Contexte du stage

1.1. Centre National d'Études Spatiales

Le Centre National d'Études Spatiales (CNES) est un établissement public à caractère industriel et commercial (EPIC) chargé d'élaborer et de proposer au gouvernement français le programme spatial du pays et de le mettre en œuvre. Il a été fondé en 1961 en tant que structure chargée de coordonner et animer les activités spatiales françaises centrées à l'époque sur le développement du lanceur-fusée *Diamant*. Depuis, le CNES travaille sur un ensemble d'activités principales réparties en 5 domaines d'intervention : Accès à l'espace (lanceurs), Sciences et innovation (astronomie et physique fondamentale), Observation de la Terre (Terre, environnement et climat), Télécommunications, et Sécurité et Défense.

Le CNES est réparti en quatre centres stratégiques : le siège social à Paris, la direction des lanceurs à Paris Daumesnil, le centre spatial guyanais, et le centre spatial de Toulouse où s'est déroulé ce stage.

Le Centre Spatial de Toulouse (CST) a pour mission, à l'exception des lanceurs et de leurs lancements, d'effectuer l'essentiel des tâches techniques et d'assistance aux scientifiques qui sont de la responsabilité du CNES. On distingue dans ses activités : la gestion des projets, les études de recherche et technologie, les centres d'opération pour les mises à poste et la gestion en orbite, les moyens informatiques et d'études mathématiques, et les supports : Administration, logistique et Communication. Le CST héberge entre autres le Centre d'aide au développement des activités en micropesanteur et des opérations spatiales (CADMOS).

Ce stage a été réalisé au sein du service Analyse & Algorithme (2A) qui fait partie de la sous-direction des Systèmes Instrumentaux (SI) appartenant à la Direction des Systèmes Orbitaux (DSO) du CNES dans le CST. Le rôle principal du service 2A est le développement d'outils, de méthodes et algorithmes pour l'accompagnement des utilisateurs dans le domaine de l'imagerie spatiale. Parmi les produits du service on compte la librairie de traitement d'images par télédétection Orfeo Toolbox, la plateforme de données satellite de référence Kalideos et la mise en place des programmes d'aide à la réhabilitation suite à des catastrophes naturelles tel que le *Recovery Observatory*.

1.2. Cahier des charges du stage

Le travail pendant les six mois du stage a été consacré à trois principales activités : la préparation des données de référence pour les classifications, les lancements de la chaîne *iota*² et la rédaction du rapport et préparation de la soutenance. Ces activités ont été réalisées exclusivement par moi avec l'accompagnement, bien évidemment, et le support technique indispensable des encadrants, des membres du service et des experts du CESBIO. Un suivi hebdomadaire du travail a été réalisé par les encadrants. Les principales tâches accomplies, listées par activités ci-dessous, ont été :

- I. Préparation des données de référence.
 - Récupération et organisation des données de référence disponibles pour les classifications.
 - Prétraitement des données sous QGIS ou sous Linux avec des libraires de traitement vecteur en langage Python. Cela m'a demandé de me familiariser avec l'environnement Linux, le langage Python et le fonctionnement du Cluster du CNES.
- II. Lancements de *iota*²
 - La préparation des fichiers de configuration de la chaîne à partir des fichiers modèles disponibles.
 - Les multiples lancements de *iota*². Initialement l'accompagnement des tuteurs ainsi que des développeurs du CESBIO a été indispensable dans cette étape.
- III. Rédaction du rapport et préparation de la soutenance
 - La rédaction du rapport a été faite par étapes toujours avec relecture de la part des tuteurs. Une présentation finale du stage devant les membres du service aura lieu avant la soutenance à l'Université Paul Sabatier.

2. Application d'algorithme d'occupation du sol en Haïti

2.1. Site d'étude

La République d'Haïti est un pays des Grandes Antilles occupant le tiers occidental de l'île Hispaniola et s'étendant sur une superficie d'environ 27560 km². Il est bordé au nord par l'océan Atlantique, à l'est par la République dominicaine, au sud par la mer des Caraïbes et à l'ouest par le canal du Vent, qui le sépare de l'île de Cuba (Figure 1). Haïti est formé par deux péninsules séparées par le golfe de la Gonâve. Le pays présente un relief majoritairement montagneux, avec des massifs orientés est-ouest, séparés par des vallées étroites.



Figure 1. Localisation d'Haïti.

Par sa position en latitude, Haïti présente un climat tropical caractérisé par l'alternance entre une saison humide, de mai à novembre, et une saison sèche, de novembre à mai. Le pays est fréquemment frappé par des tempêtes et cyclones tropicaux notamment entre août et novembre, dont le cyclone Jeanne de catégorie 3 qui a provoqué la mort de plus de 3000 personnes dans la ville de Gonaïves en septembre de 2004, et l'ouragan Matthew de catégorie 5 qui a fait près de 500 morts en octobre de 2016.

2.2. Les données

Pour réaliser la carte d'occupation des sols par apprentissage supervisé, on dispose d'une part d'imagerie satellite, et d'autre part de données de référence, souvent au format vectoriel

(points, polygones, polygones ; Humboldt S.U., 2015). Dans le cadre de ce stage, concernant les données optiques le choix s'est porté vers les images de Sentinel-2 en raison de leur haute résolution spatiale et temporelle très avantageuse pour le suivi de l'occupation du sol ainsi que de leur disponibilité. À défaut de données de terrain, OpenStreetMap (osm ; © OpenStreetMap, 2018) a été la principale source de données de référence. L'ensemble de ces données géographiques sont disponibles sous licence libre *Open Data Commons Open Database License* (ODbL ; <https://www.openstreetmap.org/copyright>) sur le portail <https://www.openstreetmap.org>.

2.2.1. Images Sentinel-2

Les images Sentinel-2 sont issues d'une mission opérationnelle de l'Agence Spatiale Européenne (ESA) dédiée à l'observation des surfaces continentales dans le cadre du programme Copernicus. Cette mission, constituée de 2 satellites à orbite polaire (786 km d'altitude), Sentinel-2A (depuis juin 2015) et Sentinel-2B (depuis mars 2017), permet une couverture totale de la Terre avec une répétitivité de 5 jours et fournit des images d'une fauchée de 290 km et une résolution de 10 à 60 m selon les bandes spectrales allant du visible au moyen infrarouge (Figure 2). Au total 13 bandes spectrales dont 3 dans l'infrarouge court (*Short-Wave Infrared* ; SWIR) constituent les produits Sentinel-2 contenant des données de réflectance de surface.

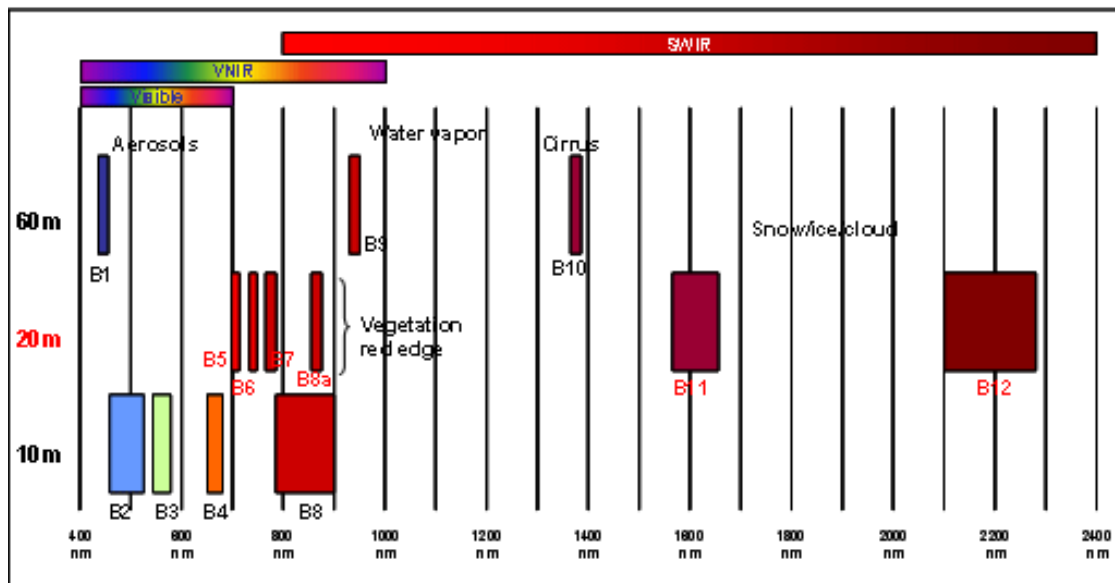


Figure 2. Bandes spectrales de Sentinel-2. Source : CESBIO

La réflectance (ρ), définie comme la capacité d'une surface à réfléchir la lumière incidente est exprimée par le rapport entre le flux lumineux réfléchi (ϕ_r) et incident (ϕ_0) : $\rho = \phi_r / \phi_0$. Elle varie en fonction de la position du soleil par rapport à la surface observée et de la direction suivant laquelle le capteur vise la surface. Elle dépend aussi des caractéristiques intrinsèques des surfaces et de la longueur d'onde de la lumière incidente, créant ainsi une signature spectrale propre à chaque surface en fonction de la longueur d'onde. Dans cette optique, la réflectance, à travers la signature spectrale permet de caractériser les surfaces naturelles (sol nu, prairie, eau, forêt...) dans un certain état (sol sec ou mouillé, prairie verte ou sèche... ; GDTA, 1991). La réflectance mesurée au niveau des satellites diffère de la

réflectance effective au niveau de la surface de la Terre à cause des déformations provoquées par l'atmosphère et les nuages (Girard et Girard, 2010). Les images brutes sont ortho-rectifiées et traitées afin de corriger ces perturbations et de fournir les valeurs de réflectances au niveau de la surface terrestre (Hagolle *et al.*, 2015^a ; Hagolle *et al.*, 2015^b ; Hagolle *et al.*, 2010 ; Hagolle *et al.*, 2008).

En fonction des traitements effectués sur les images Sentinel-2, les produits sont classés par niveaux. Le niveau 1C correspond à des images ortho-rectifiées sans correction atmosphérique, distribuées pas l'ESA. Et le niveau 2A, correspond à des images ortho-rectifiées en réflectance de surface après corrections atmosphériques, auxquelles sont associés un masque des nuages et de leurs ombres, ainsi qu'un masque des surfaces d'eau et de neige.

Les images Sentinel-2 sont fournies selon un découpage précis et fixe créant ainsi des tuiles de 110 km par 110 km dans la projection UTM / WGS84 (*Universal Transverse Mercator / World Geodetic System 1984*), avec un recouvrement de 10 km des tuiles adjacentes (Figure 3). Les images Sentinel-2 niveau 2A sont traitées différemment par l'ESA et par le centre de traitement MUSCATE du CNES. Ces produits sont disponibles accès libre dans des plateformes différentes. L'ESA effectue des corrections atmosphériques avec Sen2Cor tandis que MUSCATE utilise la chaîne de traitement MAJA (*MACCS ATCOR® Joint Algorithm ; Multi-sensor Atmospheric Correction and Cloud Screening ; Atmospheric and Topographic Correction*). Les produits niveau 2A du MUSCATE ici utilisés sont téléchargeables depuis la plateforme du pôle thématique THEIA (<http://www.theia-land.fr/fr/produits/réflectance-sentinelle-2>). Cette facilité d'obtention rend ces images très intéressantes dans le contexte haïtien. La possibilité de reprendre sur place la suite de ce travail avec un moindre investissement économique est un aspect fortement apprécié.

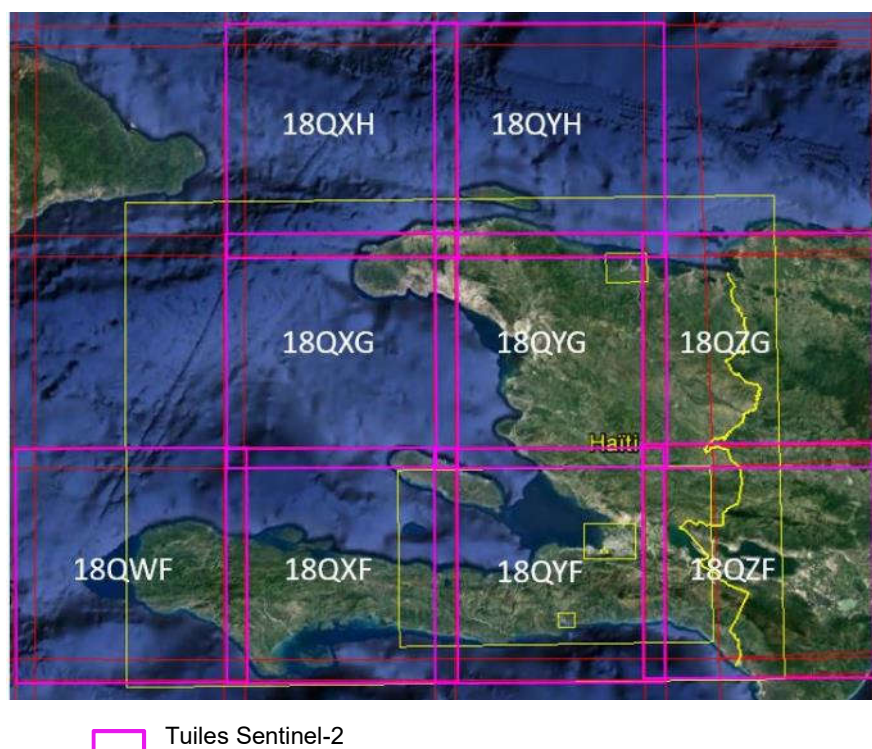


Figure 3. Tuiles Sentinel-2 sur Haïti. Source : Kalideos

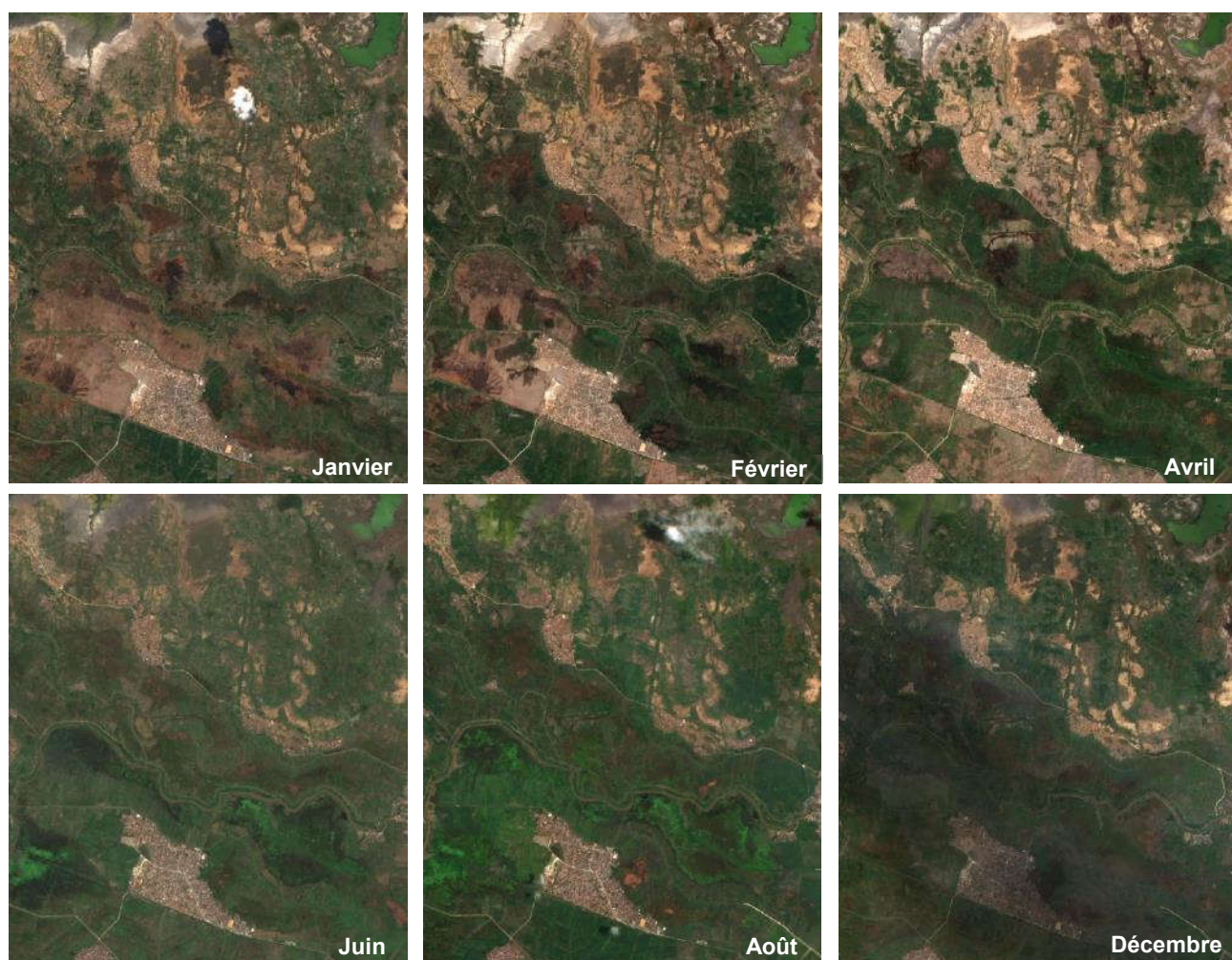
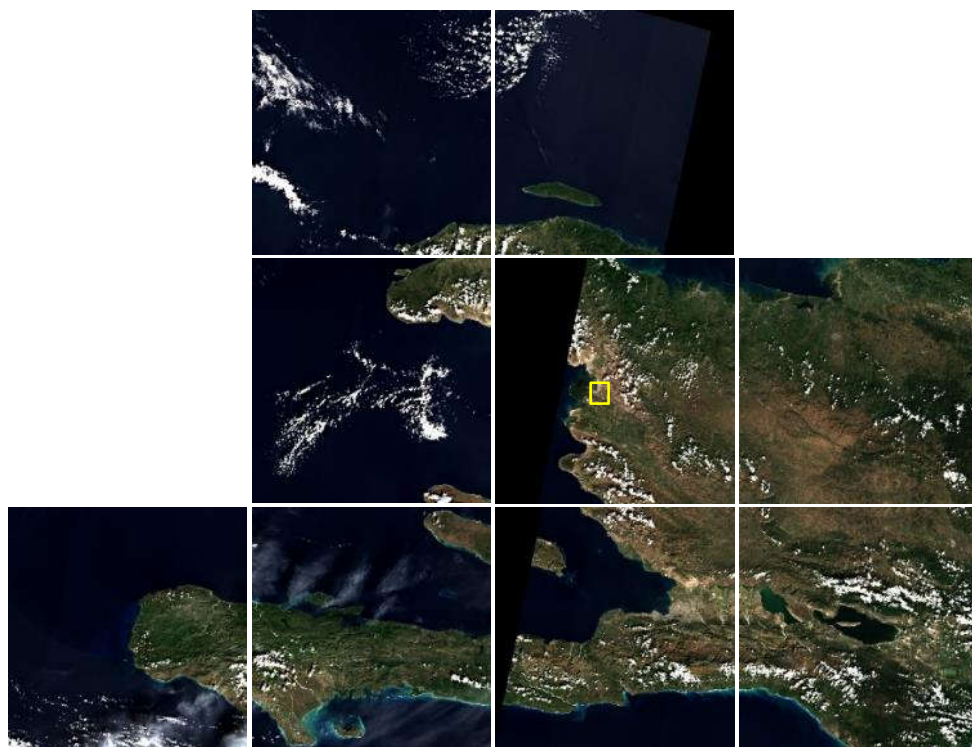


Figure 4. Mosaïque d'images Sentinel-2 sur Haïti et extrait en pleine résolution de série temporelle 2017. Source : Theia.

*iota*² fonctionnant par des cycles annuels de couverture du sol (Section 2.3), des séries temporelles, de janvier à décembre 2017, des images niveau 2A des 9 tuiles couvrant le territoire haïtien ont été utilisées dans ce travail.

La mosaïque de la Figure 4 créée à partir des images Sentinel-2 permet de voir le recouvrement entre des tuiles voisines. Haïti se situe sur une limite d'orbite (orbites relatives R025 et R068), ce qui explique que certaines tuiles soient partiellement imagées, tel que l'on peut l'observer sur la mosaïque de la Figure 4. Un extrait à pleine résolution de la série temporelle de 2017 montre l'évolution du couvert végétal au cours de l'année ainsi que la présence de nuages notamment de cirrus en décembre (ANNEXE 1). Environ 30% des images téléchargées dans ce travail présentent une couverture de nuages supérieure à 30%. Cependant, aucun tri sous ce critère n'ayant été réalisé, toutes les images ont été utilisées.

2.2.2. Données de référence

Des données vectorielles des différentes classes d'occupation du sol ont été utilisées dans la phase d'apprentissage de *iota*² (Section 2.4). Aucune donnée de terrain n'a été recueillie spécifiquement pour ce travail. Des données provenant de sources externes ont donc été utilisées.

Les deux premières semaines du stage ont été consacrées à la récupération de données à partir de deux principales sources : i) l'archive des données du service dans lequel le stage se déroule (DD\CNES-Haïti). Il s'agit d'une compilation des données d'origines variées (gouvernementale, privée, UNESCO, etc.) concernant le territoire haïtien et la République Dominicaine ; et ii) OpenStreetMap dans sa version la plus récente (2018). Au total, 40 couches en format *shapefile* (.shp) de différents Systèmes de Coordonnées de Référence (SCR) ont été récupérées. Dans une couche *shapefile* les échantillons ou objets réels sont représentés par des entités vectorielles. Une entité est un objet numérique possédant une géométrie vectorielle qui stocke la représentation géographique (souvent des points, des polygones ou des polygones) des objets réels sous une forme en une partie ou multi-parties. Chaque entité, représentée par une ligne dans le tableau d'attributs de la couche, peut être accompagnée des informations/attributs décrivant l'objet qu'elle représente.

Les couches récupérées contiennent des données de deux types de géométrie, polygone ou polyligne, des classes comme le Bâti, l'Hydrographie, le Réseau routier, l'Utilisation du sol, ainsi que des couches des limites administratives (ANNEXE 2). Ces couches, contenant jusqu'à 904250 entités, peuvent regrouper des données de plusieurs classes d'occupation du sol en un seul jeu de données. Dans certains cas, il y a de la redondance entre les couches, des décalages ou des incompatibilités dans la donnée. Parfois, la source et la date de création sont inconnues et des informations importantes relatives aux données ne sont pas renseignées (ex. : routes revêtues ou chemin de terre). Malgré le volume de données obtenu, certaines classes restent mal voire pas du tout représentées. C'est le cas de la classe Culture dans laquelle le type de culture n'est pas identifié et la classe Sol nu pour laquelle aucune donnée n'était disponible. Bien que les données récupérées ne réunissent

pas tous les critères nécessaires pour garantir des performances correctes de *iota*², ce sont les seules données dont on dispose à l'heure actuelle pour réaliser la carte d'occupation du sol d'Haïti. Cet exercice permettra, entre autres, de tester la performance de données en accès libre à ce propos ouvrant la possibilité de les utiliser de manière automatique sur d'autres sites et à des échelles différentes.

2.3. La chaîne *iota*²

L'Infrastructure pour l'Occupation des sols par Traitement Automatique Incorporant les Orfeo *Toolbox* (OTB) *Applications* (*iota*²) est une chaîne de traitement développée par le CESBIO pour produire des classifications d'occupation du sol à partir des images satellites d'un ou plusieurs capteurs et des données de référence. *iota*² permet la gestion des gros volumes de données, provenant de différentes tuiles, de différentes dates d'acquisitions et sur des grandes étendues comme la France métropolitaine. Disponible comme logiciel libre, la chaîne se base sur l'algorithme *Random Forest* (Forêts aléatoires) avec stratification éco-climatique (Inglada et al., 2017). Elle effectue du *gapfilling* (remplissage de vides) temporel et nécessite des bibliothèques de traitement vecteur (Arias et Morin, 2016). *iota*² a été créée pour être exécutée indifféremment sur les systèmes d'exploitation POSIX (*Portable Operating System Interface* – UNIX). Il est possible de paralléliser les traitements et de l'utiliser à la fois sur des machines multi-cœurs à mémoire partagée et sur des clusters de calcul à haute performance avec des centaines de nœuds (Inglada, 2016). Le principal produit de *iota*² jusqu'à présent est la carte d'Occupation du SOL de la France métropolitaine « OSO » à partir des images Landsat-8 et Sentinel-2 à, respectivement, 10 et 20 m de résolution (<http://osr-cesbio.ups-tlse.fr/~oso/>). Des applications sur d'autres territoires comme l'île de La Réunion sont actuellement menées par le Centre de Coopération Internationale et Recherche Agronomique pour le Développement (CIRAD).

2.3.1. Fonctionnement

La méthodologie générale de classification appliquée par *iota*² est basée sur une procédure classique de classification supervisée (Figure 5), avec l'avantage de pouvoir traiter de très grands territoires et de gros volumes de données de manière entièrement automatique en peu de temps. Du fait de cette approche automatique, *iota*² a été conçue pour être applicable indépendamment de la nomenclature des classes du paysage et par conséquent aucune sélection de date en termes de saison ou de phénologie de la végétation n'est appliquée. Il est donc conseillé d'utiliser autant d'images que possible afin de mieux caractériser le cycle annuel de couverture du sol (Inglada et al., 2017).

*iota*² est un logiciel récent et en développement actif qui est relativement facile à utiliser par des non-experts car son lancement repose sur un fichier de configuration en langage Python qui ne nécessite pas de connaissances détaillées sur les méthodes mises en œuvre. Néanmoins, la chaîne reste difficile à installer et n'a pas de documentation utilisateur suffisamment détaillée pour une prise en main autonome. L'installation et les mises à jour de *iota*² sur le cluster du CNES ont été accompagnées par les experts du CESBIO.

Les entrées du traitement montrées dans la Figure 5 sont : i) Les données de référence. Il s'agit des échantillons géoréférencés et étiquetés avec une classe d'occupation du sol connue ; ii) les masques de validité des séries temporelles des images. Chaque image correspondant à une date est accompagnée d'un masque indiquant les pixels valides (réflexion de surface) et invalides (détection de nuages, ombre de nuages, saturation) ; iii) les séries temporelles des images satellite niveau 2A ; iv) entrée facultative, un masque de région d'intérêt (*Region Of Interest* ; ROI) permettant d'exclure des zones lors de la classification.

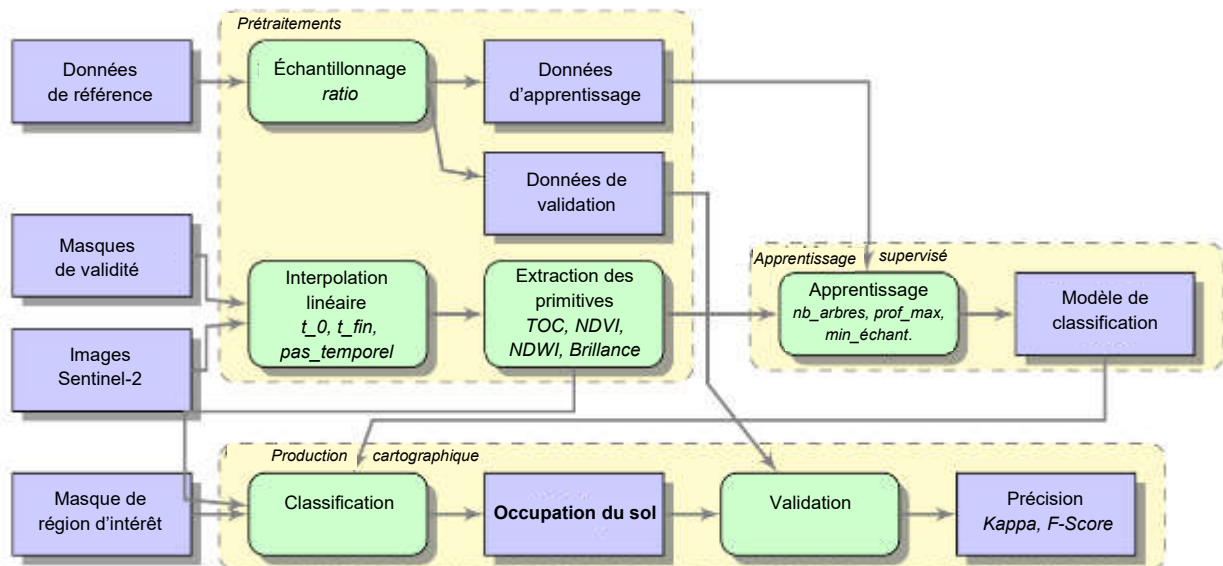


Figure 5. Schéma fonctionnel de la procédure de production des cartes d'occupation du sol.
Source : Traduit de Inglada et al., 2017

Le traitement est divisé en 6 étapes principales (boîtes vertes dans la Figure 5) :

Échantillonnage. *lota*² sépare aléatoirement les données de référence en données d'apprentissage et données de validation suivant le ratio demandé dans le fichier de configuration (0.7 : 0.3 dans ce travail). La séparation s'effectue au niveau du polygone et non au niveau du pixel. Cela afin d'éviter que des pixels provenant d'un même polygone, avec des caractéristiques similaires soient utilisés pour l'apprentissage et la validation, ce qui biaiserait l'évaluation de la qualité des classifications vers des résultats optimistes (Précision, F-Score et Indice Kappa).

Interpolation linéaire. Les séries temporelles des images sont prétraitées avec du *gapfilling* temporel et du ré-échantillonnage afin d'assurer l'homogénéité spatiale et temporelle. L'approche consiste en une interpolation linéaire des pixels invalides en utilisant les valeurs de réflectance de surface des dates précédentes et suivantes aux dates avec des nuages. Concernant le ré-échantillonnage temporel, l'interpolation linéaire est appliquée sur toutes les valeurs de réflectance de surface de toutes les dates (pixels valides et invalides), afin d'avoir des dates communes pour tous les pixels de la zone d'étude.

Extraction des primitives. Les séries temporelles obtenues avec les interpolations sont utilisées pour le calcul des indices spectraux (NDVI : *Normalised Difference Vegetation Index* ; NDWI : *Normalized Difference Water Index* et luminosité) de chaque pixel à chaque date d'acquisition. Ces indices viennent s'ajouter aux données de réflectance de surface de chaque pixel ce qui permet d'améliorer les résultats des classifications notamment lorsque les zones d'études sont très étendues et avec des paysages très variables. Ces indices servent à mettre en évidence des propriétés particulières des surfaces observées comme la présence de végétation avec le NDVI, et de l'eau et des zones humides avec le NDWI (Baghdadi et Zribi, 2017 ; Pelletier *et al.*, 2016).

Apprentissage. Les données d'apprentissage sont utilisées pour apprendre au classifieur à identifier les classes d'occupation du sol. C'est lors de cette étape que le modèle de classification est produit. *Iota*² se base sur l'algorithme de classification *Random Forest* de Breiman (2001) qui a montré des précisions globales supérieures à celles des méthodes traditionnelles comme les Arbres de décision et les SVM. De plus, ce classifieur demande des temps de traitement plus courts avec un paramétrage plus simple (Pelletier *et al.*, 2016 ; Inglada *et al.*, 2015 ; Rodriguez-Galiano *et al.*, 2012 ; Gislason *et al.*, 2006). *Random Forest* est une combinaison d'arbres de décision, de façon que chaque arbre dépend d'un échantillon initial et à chaque étape, la construction d'un nœud de l'arbre se fait sur un sous-ensemble de variables tirées aléatoirement, c'est à dire des tirages avec remise (technique connue sous le nom de *bootstrap*). Après avoir généré un grand nombre d'arbres, la prédiction est le résultat d'un vote majoritaire (*Ensemble learning*). Autrement dit, la classe affectée à chaque pixel est la plus fréquente.

Classification. Cette étape consiste à affecter des classes particulières à tous les pixels de l'image en utilisant les séries temporelles de réflectance de surface et des indices spectraux avec le modèle de classification. Le produit de cette étape est une carte d'occupation du sol comportant les mêmes classes que les données d'apprentissage.

Validation. La qualité des cartes d'occupation du sol produites par *iota*² est évaluée avec un ensemble d'indices dérivés d'une matrice de confusion où les valeurs dans les cellules correspondent au comptage des pixels de validation. Les lignes correspondent à la classe de référence, dite « la vraie classe », et les colonnes à la classe obtenue par la classification. Les indices correspondent à des statistiques globales qui donnent des informations résumées sur la classification, calculés à partir des données de validation utilisées au niveau du pixel : i) la Précision globale, calculée par la somme de la diagonale divisée par la somme de tous les éléments de la matrice de confusion, indique la proportion des pixels qui ont été bien classés, toutes classes confondues ; ii) le Rappel, indiquant la fraction des pixels correctement classifiés par rapport à la vérité de terrain ; iii) la Précision, indiquant la fraction des pixels correctement classifiés par rapport à tous les pixels classifiés dans la classe ; iv) le F-Score, désignant la moyenne harmonique de la Précision et le Rappel ; v) et l'Indice Kappa, qui prend en compte la partie de l'accord entre la sortie du classifieur et les données de référence qui peut être due au hasard. Il exprime donc la différence relative entre l'accord observé et l'accord aléatoire auquel l'on peut s'attendre si le classifieur était aléatoire (Inglada *et al.*, 2017).

2.4. Préparation des données de référence

Cette étape a été celle qui a demandé le plus d'investissement en temps. Le gros volume de données à traiter et leur qualité, la performance des logiciels utilisés pour le prétraitement, le manque et la difficulté d'installation des modules nécessaires pour l'exécution des scripts de traitement ont fait de ce travail de préparation de données la principale activité lors de quatre premiers mois du stage.

Les données vectorielles récupérées ont fait l'objet d'une préparation minutieuse afin de les rendre exploitables par *iota*². Les traitements ont été réalisés sous QGIS 2.14.0 (www.qgis.org) et avec la librairie de traitement vecteur *Vector-tools* développée par le CESBIO. Il s'agit d'un ensemble d'environ 40 scripts en langage Python disponibles en accès libre dans le dépôt Git de *iota*² à l'adresse suivante : <https://framagit.org/inglada/iota2/tree/develop/scripts>.

La préparation des données de référence a consisté en une suite des traitements en fonction des couches. C'est-à-dire que le type de traitements ou leur ordre était dépendant de l'état initial de chaque jeu de données, d'où la difficulté à établir une méthodologie commune. Toutes les couches récupérées n'ont finalement pas été intégrées dans les données de référence par manque de pertinence, de précision ou de redondance avec d'autres couches (par exemple : Macaya, Tsunami UNESCO, Zonages, Lieux d'intérêt, Réseau routier, Hydro en polyligne). Initialement les données en polyligne, le réseau routier et les cours d'eau, avaient été traitées et intégrées dans les premiers tests de classification. Cependant, ces échantillons, souvent très décalés par rapport aux éléments sur les images de référence, induisaient des confusions entre les classes. Ils ont donc été abandonnés ultérieurement. En ce qui concerne les cours d'eau, le masque d'eau *Global Surface Water* (Pekel *et al.*, 2016) a été utilisé comme solution alternative. Elle sera développée dans la Section 2.4.3. En ce qui concerne la classe Route aucune alternative n'a été trouvée.

2.4.1. Données de type polygone

Chaque couche a été traitée séparément avant d'être fusionnée avec les autres. Les premières données traitées ont été celles de la classe Bâti. Bien que toutes les couches aient chacune leurs particularités, de manière générale, pour les jeux de données en polygone la procédure de traitement suivie a été la suivante :

- 1) Enregistrement des couches d'origine sous des nouvelles couches renommées en fonction des classes (Tableau 1). Cela permet d'avoir une sauvegarde de la donnée brute à laquelle s'adresser en cas de besoin.

Tableau 1. Exemple de renommage des couches exportées.

Couche initiale	Couche de traitement	Nombre d'entités
buildings	Bâti1	37346
gis.osm_buildings_a_free_1	Bâti_osm	904250
Bâti_Jérémie_2014_2016_207	Bâti_RO	12857
roads	Route1	68076
gis.osm_roads_a_free_1	Route_osm	186826
Haïti_lakes	Eau1	91
Lac	Eau2	138
gis.osm_water_free_1	Eau_osm	2053
gis.osm_landuse_free_1	Culture_osm	1513
gis.osm_landuse_free_1	VégétationBoisée_osm	2089
gis.osm_nature_free_1	SMinerale_osm	300
...

- 2) Changement du SCR de la couche pour un SCR commun et compatible avec *iota*². Dans ce travail, l'EPSG (*European Petroleum Survey Group*) : 32618, WGS 84/ UTM zone 18N a été utilisé.
- 3) Découpage à l'emprise du pays (Figure 6). Pour ce faire, il est impératif que la couche à découper et celle servant au découpage (la couche « Limite_pays » en l'occurrence) soient dans le même SCR, d'où l'importance de les avoir modifiés si ce n'était déjà pas le cas.

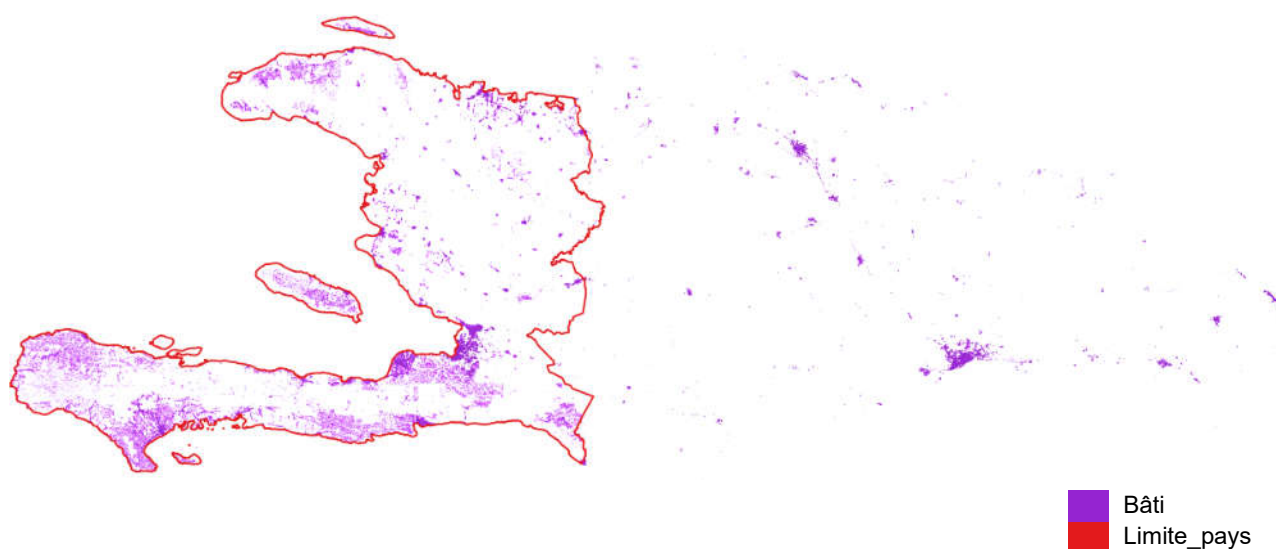


Figure 6. Couche de Bâti à découper par l'emprise d'Haïti.

- 4) Sélection par attributs (Figure 7). Souvent, les couches d'origine contiennent des échantillons de plusieurs classes, sous-classes ou autre qui ne sont pas toujours pertinentes dans le but d'une classification d'occupation du sol à l'échelle du pays. Dans ces cas, des sélections en fonction de la classe, la taille, la longueur, le type d'écoulement (permanent ou temporaire pour les surfaces en eau) doivent être faites.

Il est possible, soit de sélectionner en fonction de ce qui ne doit pas être conservé puis le supprimer, soit de sélectionner ce qui va être utilisé puis l'exporter sous une nouvelle couche. Pour ce travail, les résultats de ces sélections ont été exportés sous des nouvelles couches afin de garder une trace de différents traitements effectués sur les données.

	NOM	CLASSEMENT	ECOULEMENT
148	Rivière Ti Bras	Principale	Permanent
87	Rivière Thomonde	Principale	Permanent
144	Rivière Thomonde	Principale	Permanent
313	Rivière Terre Neuve	Principale	Temporaire
111	Rivière Tapion	Principale	Permanent
258	Rivière Taluy	Principale	Temporaire
70	Rivière Source Alera	Principale	Permanent
277	Rivière Siane	Principale	Temporaire

Figure 7. Sélection par attributs d'une couche Eau sous QGis.

- 5) Suppression des entités vides. La sélection des entités vides avant leur suppression a été réalisée avec l'outil « Sélection par expression » de QGis avec l'expression :

`$geometry is NULL` (1)

Le script *vector_functions.py* option -e sert également à supprimer les entités vides d'une couche.

- 6) Vérification et correction de la géométrie (Figure 8). Ceci permet d'identifier les géométries invalides (auto-intersections, nœuds dupliqués) puis de les corriger ou supprimer si la correction n'est pas possible. Ce traitement a été réalisé sous QGis avec l'outil « Vérifier les géométries » ou avec le script *vector_functions.py* option -v depuis le terminal Linux lorsque les couches étaient trop lourdes pour pouvoir être gérées par QGis. Notamment les couches issues d'OpenStreetMap qui pouvaient avoir plusieurs centaines de milliers d'entités (~800000 entités dans la couche Bâti découpée à l'emprise d'Haïti).

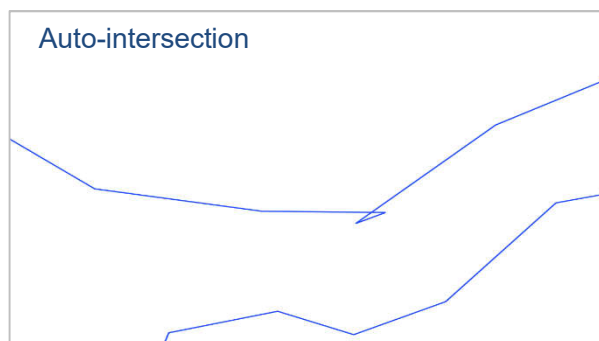


Figure 8. Exemple d'auto intersection dans un échantillon.

- 7) Simplification des géométries. Cette étape n'est pas indispensable mais elle peut s'avérer très utile lorsque l'on travaille avec des couches très lourdes. La

simplification des géométries permet d'éliminer des détails inutiles tout en gardant l'exactitude topologique des polygones et en réduisant la taille des fichiers. Elle a été appliquée à toutes les couches bien que dans certains cas cela n'ait pas apporté de changement. L'outil « simplifier la géométrie » de QGis a été utilisé mais il est possible aussi d'utiliser le script *SimplifyPoly.py*.

- 8) Suppression des doublons (Figure 9). Ce traitement et le suivant, en particulier, ont représenté une grande difficulté notamment pour les grands jeux de données comme la couche « Bâti_osm ». En effet, les couches provenant d'OpenStreetMap ne sont pas de très bonne qualité en termes de propreté du contenu. On y trouve un grand nombre de doublons ainsi que des chevauchements entre les polygones au sein d'une même couche. Ces couches étant trop lourdes, la suppression des doublons avec QGis a été impossible malgré de multiples tentatives. La division des couches en sous-couches de ~50000 entités afin de réduire le nombre d'entités à traiter et l'application des outils « *v.clean* » et « *v.clean advanced* » de la librairie GRASS (*Geographic Resources Analysis Support System*) ont été testés en vain avant d'avoir eu accès à la librairie *Vector-tools* et plus particulièrement au script servant à la suppression de doublons *DeleteDuplicateGeometries.py*
- 9) Suppression des chevauchements (Figure 9). Pour ce faire, deux alternatives à QGis ont été trouvées : i) le script *vector_functions.py* option *-i* qui est censé supprimer la partie commune entre des polygones qui se chevauchent au sein d'une même couche. Néanmoins, cette option n'a pas fonctionné systématiquement ; et ii) le script *remove_duplicates.py* créé dans le service dans lequel ce stage se déroule. Contrairement à la première option, en plus de supprimer des doublons, ce script supprime un des polygones chevauchés et non seulement la partie commune. Dans un contexte de données limitées cela pourrait poser un problème. Cependant, dans le cas de ce travail les couches contenant des chevauchements étaient des gros jeux de données et la quantité d'entités supprimées pouvait être négligée.

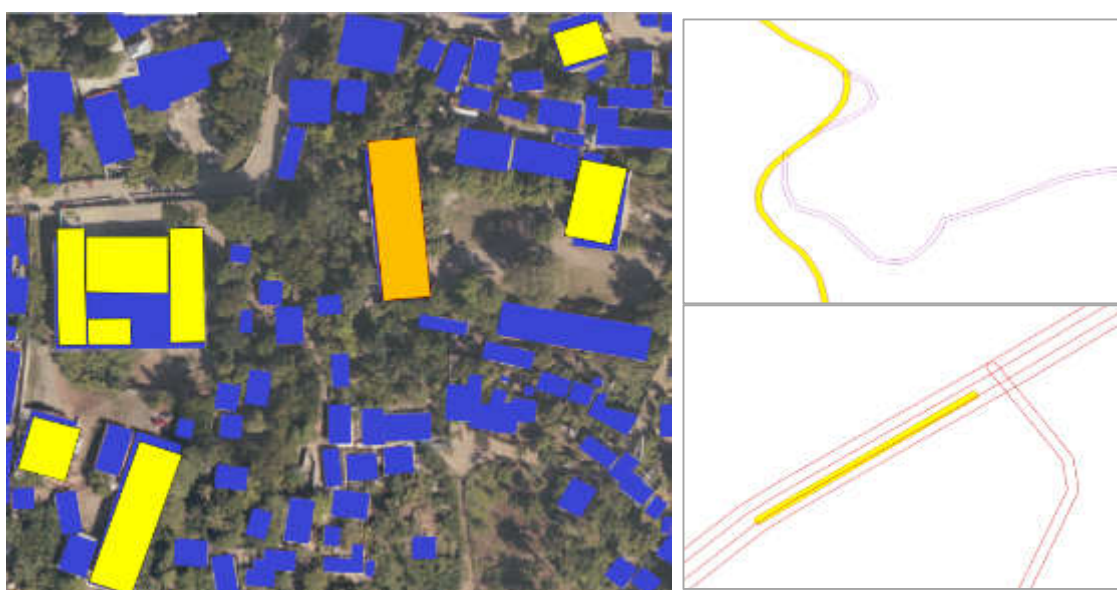


Figure 9. Exemple de doublons (entité en orange) et de chevauchements (entités en jaune) au sein d'une couche polygone et une couche polyligne.

10) Érosion (Figure 10). Dans certains cas, réaliser l'érosion plus tôt dans le traitement des couches peut servir à éliminer des chevauchements entre les polygones. Néanmoins, cela peut ne pas être suffisant. L'érosion permet aussi d'affiner les échantillons et d'éliminer les contours des objets qui correspondent souvent à des pixels mixtes. Elle a été appliquée en fonction de la taille des polygones pour la plupart des classes (Tableau 2) afin de préserver un maximum de polygones. En raison de la petite taille des échantillons de la classe Bâti l'érosion appliquée a été différente.

Tableau 2. Critères appliqués aux polygones pour effectuer l'érosion.

	Condition	Érosion [m]
	Aire $\geq 1000 \text{ m}^2$	-10
	Aire $< 1000 \text{ m}^2$	-5
Bâti	Aire $\geq 30 \text{ m}^2$	-1
	Aire $< 30 \text{ m}^2$	0



Figure 10. Érosion appliquée à une couche Culture. En violet, les polygones avant l'érosion ; en bleu, les polygones érodés superposés aux polygones initiaux.

11) Morceaux multiples vers morceau unique. L'érosion et le découpage des polygones peuvent créer des entités uniques constituées de plusieurs parties parfois de très petite taille. Ce traitement permet de gérer individuellement chacun de ces morceaux et les supprimer si nécessaire. L'outil « Morceaux multiples vers morceau unique » de Qgis a été utilisé.

12) Suppression en fonction de la taille. Typiquement le seuil correspond à la taille du pixel des images (100 m^2). Cependant, une manière d'alléger les gros jeux de données utilisés a été de choisir un seuil de 150 m^2 . Tous les polygones avec une superficie inférieure au seuil ont été supprimés sauf dans la classe Bâti où le seuil a été défini à 30 m^2 en raison de la petite taille de la majorité des polygones de cette classe. L'outil « Sélection par expression » de QGis a été utilisé avant la suppression des entités.

13) Découpage entre les couches (Figure 11). Les superpositions entre les différentes couches sont traitées avec des découpages. Généralement, les couches avec des

grands polygones ont été découpées par celles avec des petits polygones (Bâti et Route). Ce traitement a été réalisé avec l'outil « Différencier » de QGis.

14) Réitération des étapes 10, 11 et 12 (Figure 11). Le découpage est censé résoudre les superpositions entre les couches. Toutefois, cela n'est pas toujours vrai. Après ce traitement, QGis continue à identifier des superpositions, obligeant à réitérer les traitements des étapes 10, 11 et 12. En cette occasion, une petite érosion (-0,5 m) a été appliquée sur la couche découpée.

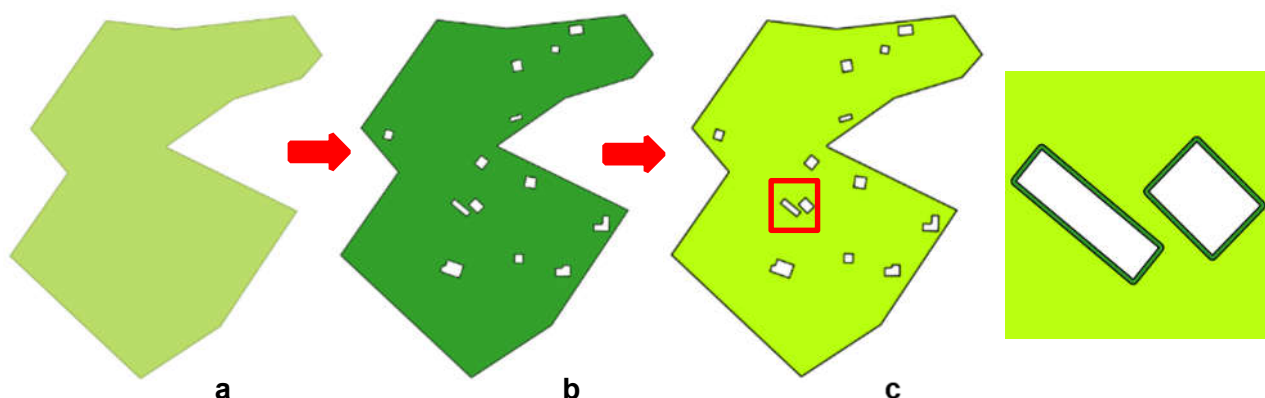


Figure 11. Découpage d'une couche Végétation boisée par une couche Bâti puis érosion. a. Couche initiale ; b. Couche découpée ; c. Couche érodée et zoom sur l'érosion.

15) Fusion. Lorsqu'il n'y a plus de géométries invalides, doublons, chevauchements, ou superpositions entre les couches la fusion peut enfin se faire. Elle a été faite avec l'outil « Fusionner les *Shapefiles* en un seul » de QGis mais il est possible d'utiliser le script *merge.py*.

16) Nomenclature (Tableau 3). Chaque entité de la couche fusionnée a été identifiée par le code de la classe à laquelle elle appartient. Le code a été renseigné dans un champ de type Entier (int) dans le tableau d'attributs de la couche. La nomenclature utilisée est une adaptation de la nomenclature des cartes d'occupation du sol OSO du CESBIO.

Tableau 3. Nomenclature.

Classe	Code
Culture	10
Végétation boisée	33
Bâti	41
Route	44
Surface minérale	46
Eau	51
Mer	52
Zone Inondable	53
Prairie	211
Sol nu	255

2.4.2. Données de type polyligne

Lorsque les données à traiter étaient en polyligne, des étapes supplémentaires s'imposaient entre la Simplification des géométries et la Suppression des doublons :

- 1) Suppression par la longueur. Les entités ayant une longueur inférieure à 30 m ont été supprimées. Ce seuil a été choisi en fonction de la taille du pixel des images Sentinel-2 (10 m) et de l'échelle du pays à laquelle on travaille.
- 2) Création des zones tampon. Afin de convertir les échantillons de polygones en polygones des zones tampon ont été créées. Elles ont été définies en fonction du type de route allant de 1,5 m pour les routes sans catégorie renseignée jusqu'à 2,5 m pour les routes du réseau primaire. La zone tampon pour les échantillons des cours d'eau a été définie par défaut à 2 m car aucune information relative à la largeur n'était associée aux données. La conversion de polyligne en polygone a créé des chevauchements entre les entités notamment dans les croisements et lorsque les lignes étaient trop rapprochées (Figure 9). La suite du traitement est donc la suppression des doublons et des chevauchements décrite précédemment. Néanmoins, comme déjà mentionné, ces derniers échantillons ont finalement été abandonnés au profit du masque d'eau *Global Surface Water* présenté ci-dessous.

2.4.3. Masque d'eau *Global Surface Water*

Une alternative aux données de type polyligne de la classe Eau a été le masque d'eau *Global Surface Water* (au format raster). Ce masque d'occurrence d'eau de surface dans le monde renseigne pour chaque pixel le pourcentage de fois pour lequel on y a détecté de l'eau. Sur Haïti le masque a été créé à partir d'un historique d'environ 400 images à haute résolution (30 m) depuis 1984 (Pekel *et al.*, 2016). Particulièrement dans ce travail, Il a permis d'identifier des zones d'eau permanente et des zones d'inondation sur Haïti qui ont été intégrées dans les classifications en tant que données de référence de ces deux classes.

L'extraction des échantillons a été réalisée en utilisant l'OTB (<https://www.orfeo-toolbox.org/download/>) à partir du raster « occurrence_80W_20N » (Figure 12) téléchargé depuis <https://global-surface-water.appspot.com/download>. Après avoir balayé le masque et l'avoir comparé aux images Sentinel-2 et aux données vectorielles de la classe Eau, les critères de sélection des pixels ont été définis. Les pixels répondant à deux conditions ont donc été extraits séparément avec l'outil « *Band Math* » :

- i) Occurrence >90% pour l'eau permanente,
- ii) 10% < Occurrence < 62% pour les zones d'inondation.

Les rasters générés ont été ensuite découpés à l'emprise du pays puis vectorisés avec les outils « Découper » et « *GDAL poligonize* » de QGis, respectivement. Cette vectorisation a produit des couches incluant des nombreuses erreurs de géométrie tel que des auto-intersections. Des doublons ainsi que des entités vides ont été également générés. Ces

deux couches ont été traitées comme décrit dans la section 2.4.1 puis intégrées dans les derniers tests de classification.

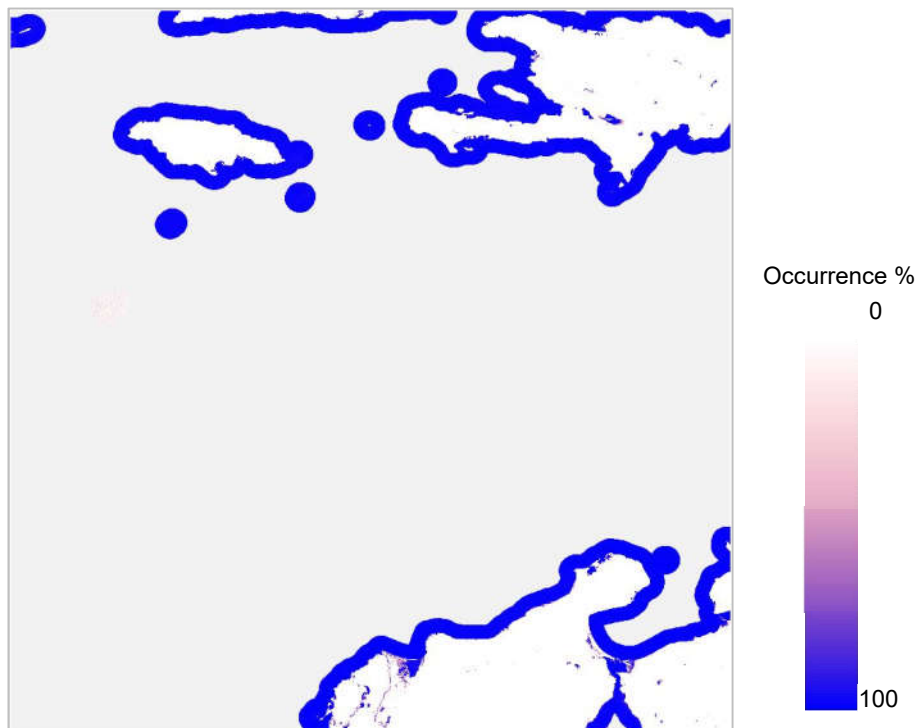


Figure 12. Masque d'eau *Global Surface Water* « Occurrence_80W_20N ». Haïti en haut à droite.

2.5. Lancement de *iota*²

Le travail de classification demande des allers-retours entre les étapes de préparation des données et de lancement de la chaîne. Cette manière d'opérer sert à identifier des points faibles dans les données de référence et des zones problématiques à prendre en compte afin d'améliorer les résultats. Des multiples lancements de la chaîne ont été réalisés avant d'obtenir la carte d'occupation du sol sur l'ensemble d'Haïti. Deux mois après le début du stage les premiers lancements ont commencés alors que le traitement de données d'entraînement continuait en parallèle. Ces premiers lancements ont été réalisés sur une seule tuile Sentinel-2 (T18QYG) avec des données de quatre classes d'occupation du sol : Bâti, Végétation boisée, Route et Eau. La quantité d'images, le nombre de données de référence et de classes ont augmenté au fur et à mesure de lancements (Tableau 4). Dans le cas d'Haïti, *iota*² n'a pas pu être exécutée avec l'approche d'occupation du sol par stratification éco-climatique, utilisée dans la production de la carte OSO, car aucune carte des régions climatiques haïtiennes n'était disponible.

Le lancement de *iota*² nécessite 3 fichiers de paramétrage liés : i) le fichier de configuration (.cfg) qui sert à définir les réglages et à personnaliser les lancements ; ii) le fichier de nomenclature (.csv) dans lequel les noms et les codes des classes d'occupation du sol sont renseignés ; et iii) le fichier de couleur (.txt) qui contient le code couleur RVB (Rouge, Vert,

Bleu) assigné à chaque classe. Dans ce stage, des fichiers exemples dans le dépôt Git de *iota*² ont été modifiés et adaptés en fonction de lancements (ANNEXE 3).

Tableau 4. Récapitulatif des lancements de *iota*².

Lancement	Tuiles	Images Sentinel-2	Classes	Données de référence
1	1	2	4	1392
2	1	17	4	1392
3	1	17	4	115580
4	1	57	7	119261
5	2	46	9	54362
6	9	46	9	88493
7	9	46	9	105303

3. Résultats et discussion

Tel que proposé par Inglada *et al*, 2017, les principaux résultats sont analysés quantitativement sur la base des indices cités dans la section 2.3, l'indice Kappa et la Précision globale ; et qualitativement avec une évaluation visuelle des classifications pour mettre en évidence des erreurs qui ne peuvent pas être détectées dans l'évaluation statistique globale. Notamment, l'analyse des premiers résultats s'est basée sur l'évaluation qualitative car les indices statistiques ne sont pas fiables lorsque les classifications sont réalisées avec très peu de données de référence et/ou nombre de classes.

Les classifications obtenues avant la classification finale sur tout le territoire haïtien ont servi à l'amélioration des données de référence des lancements suivants. Particulièrement, les données de la classe Eau ont été retirées et remplacées à l'issue de la troisième classification (Figure 13a) à cause de la confusion de cette classe provoquée par le décalage des données de référence de type polyligne ; les données de la classe Route ont été définitivement abandonnées pour la même raison après la quatrième classification. Dans ce premier résultat (Figure 13a) on observe une sur-détection des surfaces Boisées, des surfaces en Eau et de la Route qui est tout à fait normale car les données de référence ne contiennent pas certaines classes présentes sur le territoire traité. Les classes sont donc définies par un espace spectral très large lors de la construction du modèle par le classifieur. Pour ce qui est du Bâti, on peut déjà identifier sur cette classification au moins cinq zones denses représentant des villes (cercles rouges sur la Figure 13a). Cependant, aucune surface bâtie isolée n'est visible. Ceci est peut-être dû à la petite taille des échantillons de cette classe qui finissent par se confondre aux autres classes lorsqu'ils sont isolés. Ce résultat a mis en évidence le besoin de données des classes Culture, Sol nu et particulièrement des Zones d'inondation. Des données de référence pour cette dernière classe, obtenues à partir du masque d'eau *Global Surface Water*, ont été intégrées dans le lancement suivant (Lancement 4. Figure 13b).

La structure du paysage a commencé à être visible à partir du résultat du lancement 4. Cette classification réalisée avec plus d'une année d'images et des échantillons de 7 classes (Figure 13b) a une précision globale de 0,828 et un indice Kappa de 0,770. D'après la

matrice de confusion, la classe la moins bien classifiée est la Route avec la Précision la plus faible : 0,240. Cette valeur indique que seulement 24% des pixels identifiés comme Route correspondent vraiment à de la Route. De plus, elle se confond avec toutes les autres classes causant la sur-classification visible dans la Figure 13b. Ces confusions sont essentiellement dues aux données de référence de la classe Route. D'une part, la majorité du réseau routier représenté par les échantillons sont des routes non goudronnées ce qui explique que des surfaces, visiblement en sol nu sur les images Sentinel-2, aient été classifiées en Route. Et d'autre part, le décalage des échantillons, fait que souvent ils débordent sur d'autres classes. Pour ces raisons, ces données n'ont plus été utilisées dans les lancements de la chaîne et à défaut d'une alternative, la classe Route a été supprimée. Statistiquement, la Végétation boisée a donné les meilleurs résultats (Précision = 0,908), tandis que l'analyse visuelle permet d'identifier qu'une grande partie de ces surfaces correspondent en réalité à des prairies ou zones de culture. En revanche, les classes Culture et Zones d'inondation avec des Précisions respectives de 0,843 et 0,835, illustrent mieux la réalité du paysage. La détection des Zones d'inondation à la place d'Eau dans le fleuve l'Artibonite (ellipse rouge sur la Figure 13b), par exemple, suggère des lits de rivières très larges avec des écoulements variables au cours de l'année. Ce résultat pourrait s'avérer intéressant dans une analyse de disponibilité de l'eau car il reflète la réalité de l'alternance entre des périodes de sécheresse, centrées sur l'hiver boréal, et des fortes crues récurrentes lors des précipitations intenses et qui provoquent des inondations tout au long du fleuve notamment autour de mai et septembre (MEF, 2016 ; Barrette et Daleau, 2012). Cependant, pour des raisons de rendu cartographique à la grande échelle ici travaillée, les données de référence de la classe Zone d'inondation associées aux

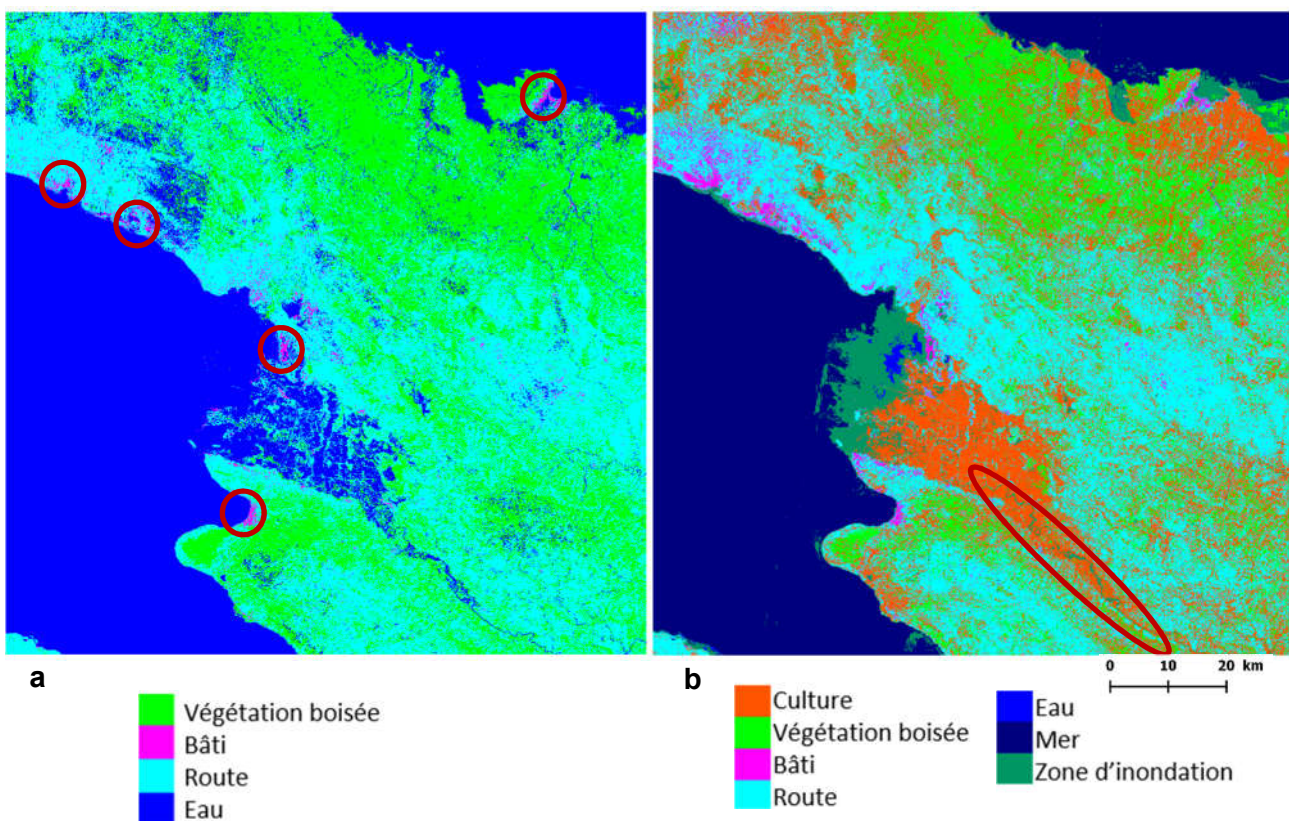


Figure 13. Classifications d'occupation du sol sur une tuile à partir des images Sentinel-2.
a) Lancement 3 avec 17 images et 4 classes ; **b)** Lancement 4 avec 57 images et 7 classes.

principaux cours d'eau ont été déplacées vers la classe Eau avant de réaliser les lancements suivants.

La vallée de l'Artibonite, caractérisée par une vaste culture de riz et la plus importante superficie irriguée du pays (FAO, 2016^b ; MARNDR, 2013), est clairement détectée par *iota*². Dans la Figure 14, on compare une image Sentinel-2 utilisée dans la classification à la carte produite dans le lancement 4 avec un zoom sur cette zone. On y distingue facilement la vallée, le système parcellaire, la ville de Saint-Marc ainsi que la grande forêt de mangrove de Gonaïves/Grande Saline classée comme Zone d'inondation. Quant au Bâti, avec une Précision de 0,359 et un Rappel de 0,491 (le plus faible de toutes les classes), on est face à une grande erreur de classification par commission. C'est-à-dire, des surfaces d'autres classes, principalement de sol nu, ont été classifiées comme du Bâti (ellipses rouges dans la Figure 14). L'erreur de commission est calculée à partir du Rappel de la classe ($[1 - \text{Rappel}] * 100\% = 50,09\%$) et indique une sur-classification. Dans le cas du Bâti, elle signale donc que seulement la moitié des pixels a été bien classifiée. Cette difficulté à classifier correctement le Bâti a souvent été constatée dans d'autres études quelle que soit l'étendue de la zone étudiée, la résolution spatiale des images ou le classifieur utilisé. Soit sur la totalité du territoire d'un pays comme la France métropolitaine (535000 km²) avec des

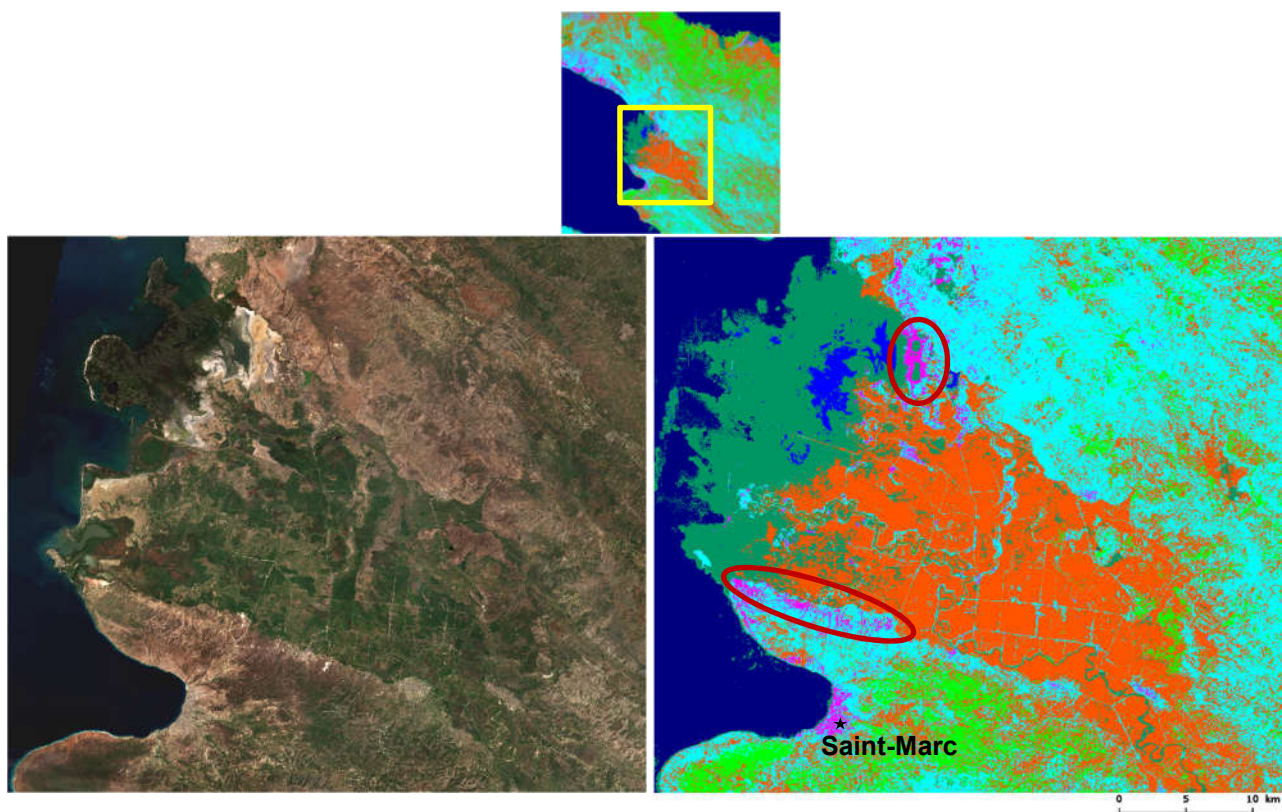


Figure 14. Comparaison entre l'image Sentinel-2 du 25 mars 2017 et la classification sur la vallée de l'Artibonite.

images Landsat à 30 m de résolution spatiale et le classifieur *Random Forest* (Inglada *et al.*, 2017), soit sur une emprise de ~5 km² du centre-ville de Baden en Suisse avec des photographies aériennes hyper-spectrales à 1,8 m de résolution spatiale et le classifieur *Support Vector Machine* (Chen *et al.*, 2018). Plusieurs raisons pourraient expliquer cette difficulté : i) la petite taille des bâtiments, notamment lorsqu'ils sont plus petits que le pixels des images utilisées ; ii) Les matériaux divers des toits tel que le béton, le bitume, des tôles

ou divers carreaux colorés qui peuvent avoir des réponses spectrales très différentes et par conséquent rendre difficile la caractérisation de la classe ; et iii) l'ombre des bâtiments rapprochés qui peuvent être fréquentes dans des environnements urbains.

Enfin, l'ajout d'images pour la classification mais surtout de données de nouvelles classes a joué un rôle majeur dans l'amélioration de la classification bien que la classe Sol nu soit toujours absente. Les résultats du lancement 5 étant similaires en termes de précision et rendu visuel à ceux du lancement 6, ils ne sont donc pas présentés dans ce document (ANNEXE 4).

La Figure 15 montre la première et la seule classification obtenue sur l'ensemble d'Haïti au moment de la rédaction de ce rapport (lancement 6). Les résultats d'autres lancements en cours (lancement 7) et à venir constitueront le rendu final du stage et seront montrés lors de la présentation finale. La Précision globale de 0,964 et l'indice Kappa de 0,887 de cette classification sont supérieurs à ceux de la classification précédente bien que visuellement le résultat semble moins satisfaisant. La Précision, le Rappel et le F-Score par classe sont présentés dans la matrice de confusion (Tableau 5).

La Mer a obtenu la Précision la plus élevée (0,998) avec une légère confusion avec l'Eau (Rappel = 0,9999) qui pourrait être due à la similarité de leur signatures spectrales. L'Eau avec une Précision de 0,777 est surtout confondue aux Zones d'inondation et à la Culture (Précisions respectives de 0,788 et 0,806). La confusion entre l'Eau et les Zones d'inondation souvent trouvée dans d'autres études (Inglada *et al.* 2017 ; Ma *et al.* 2017 ; Mahdianpari *et al.* 2017) est parfois le résultat de la présence de végétation flottante sur les bords des surfaces en eau, ce qui provoque un degré de confusion entre les classes. Ceci est également produit par la variation temporelle des zones d'inondation qui fait qu'à certaines périodes de l'année elles se trouvent complètement inondées et donc peuvent être considérées par le classifieur comme des surfaces en Eau. Ce résultat met en évidence l'intérêt des séries chronologiques denses d'images qui permettent le suivi à haute fréquence du comportement des surfaces d'eau et du stade/état de la végétation (Yesou *et al.*, 2016).

Dans la matrice de confusion, la classe Surface minérale est caractérisée par la Précision la plus faible (0,314). Elle est principalement confondue avec les Zones d'inondation. La mauvaise performance de la classification sur cette classe provient du nombre insuffisant d'échantillons utilisé pour la classification (153 échantillons = 3,68 km²). Il en est de même pour la classe Sol nu avec à peine 29 échantillons faisant au total 0,784 km². Malgré le grand nombre d'échantillons de référence de la classe Bâti, la Précision atteinte dans la classification reste moyenne (0,677). Ceci est en accord avec les conclusions de Ma *et al.* (2017) qui ont trouvé une corrélation positive entre les valeurs de Précision et la taille des échantillons. La taille moyenne des échantillons de cette classe dans notre étude était de 41,74 m², ce qui est même en-dessous de la taille du pixel (100 m²). Si l'on prend en compte la densité du Bâti dans les centres urbains, cela pourrait compenser la taille des échantillons garantissant un minimum de « pureté » dans le pixel, ou autrement dit, de représentativité en terme spectral. Cependant, dans les zones rurales avec une densité de Bâti très faible, la taille des échantillons reste limitante pour la classification. Environ la moitié de la

population d'Haïti habitant en milieu rural, des difficultés de classification de la classe Bâti sont compréhensibles.

Tableau 5. Matrice de confusion de classification d'occupation du sol d'Haïti.

		Classification								Rappel	F-Score	
		Culture	Végétation boisée	Bâti	S. minérale	Eau	Mer	Zone d'inondation	Prairie			Sol nu
Référence	Culture	365178	41605	603	57	139	0	2271	1498	26	0,888	0,845
	Végétation boisée	28231	284966	320	30	138	124	1501	1419	2	0,900	0,865
	Bâti	2502	3713	3441	31	0	0	563	73	2	0,333	0,447
	S. minérale	218	218	116	480	15	0	947	2	3	0,240	0,272
	Eau	9227	625	31	3	132974	5179	33678	98	1	0,731	0,753
	Mer	0	0	0	0	313	5541785	0	0	0	0,999	0,999
	Zone d'inondation	32796	6016	341	891	37657	3441	145714	83	2	0,642	0,708
	Prairie	14704	4595	178	1	0	0	66	31791	0	0,619	0,705
	Sol nu	224	180	50	34	0	0	193	3865	102	0,022	0,043
Précision		0,806	0,833	0,677	0,314	0,777	0,998	0,788	0,819	0,739		

Précision globale : 0,964 Indice Kappa : 0,887

La Figure 15 montre que l'ajout des trois nouvelles classes (Surface minérale, Prairie et Sol nu) n'a pas eu un effet positif sur le rendu visuel de la classification. En revanche, la suppression de la classe Route semble avoir occasionné des sur-classifications de la Végétation boisée au détriment des Prairies et du Sol nu. Bien que la Précision, le Rappel et le F-Score de la Végétation boisée soient relativement bons, il ne faut pas perdre de vue que ces indicateurs dépendent des données de référence utilisées pour la classification et qu'ils peuvent être biaisés. C'est pourquoi l'on recommande de faire l'évaluation visuelle de la carte produite. La Végétation boisée de la dernière classification en est un exemple. On y observe une prédominance de ces surfaces sur l'ensemble de la carte, tandis que la réalité à Haïti est à l'opposé de ce résultat. La FAO rapporte pour l'année 2015 une surface boisée de 5914,58 km² (cultures ligneuses, zones arborées, mangroves et zones arbustives) pour tout le pays, soit 21,46% de son territoire (FAO, 2017), dont 970 km² correspondait à la végétation forestière, l'équivalent de 3,5% de la surface du territoire (Groupe de la Banque Mondial, 2018). Toutefois, la surface de la classe Végétation boisée calculée à partir de notre classification est d'environ 18000 km², soit ~3 fois la surface estimée par la FAO. En ce qui concerne la Culture, en 2014, 13500 km² de terres étaient cultivées, soit 48,65% du territoire (FAO, 2016^a). En revanche, avec la classification on obtient une surface de ~10000 km², l'équivalent de 37,50% du territoire. Ce dernier résultat semble plus cohérent que celui de la Végétation boisée compte tenu de l'effet dévastateur de l'ouragan Matthew en 2016 sur les cultures, lors duquel, environ 95% des récoltes ont été détruites dans la Péninsule

de Tiburon dans le sud du pays. Il est possible qu'à l'heure actuelle la totalité de ces terres ne soient toujours pas remises en culture.

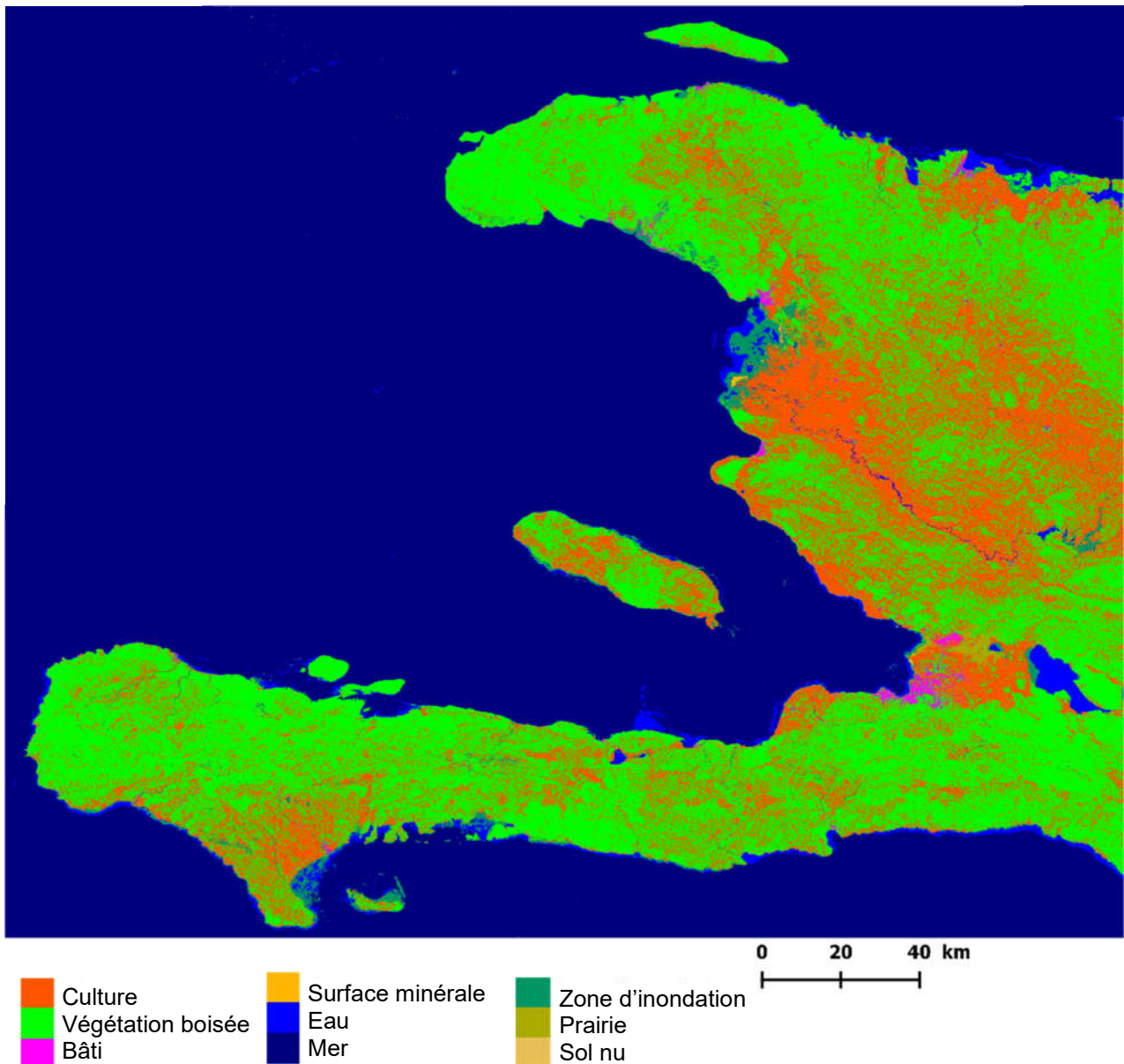


Figure 15. Classification d'occupation du sol sur Haïti à partir de 46 images sentinel-2, et 9 classes (lancement 6).

On constate également l'effet du relief sur la classification. Les zones avec un relief très accidentés sont plus difficiles à classifier que les zones avec des reliefs plus homogènes car les ombres dues à la pente peuvent provoquer des confusions entre les classes (ANNEXE 5), typiquement, classifiées en tant qu'Eau. Les résultats de Zhang *et al.* (2018) indiquent que les surfaces imperméables à l'obscurité, le sol nu et les pixels ombrés étaient plus facilement classés par erreur dans d'autres classes.

4. Conclusions

iota². L'objectif du stage étant d'identifier le potentiel de *iota²* avec des données de référence limitées, on peut constater que les performances de la chaîne sont d'autant plus satisfaisantes que les données de référence pertinentes sont disponibles. Ce travail a permis de tester le potentiel de l'approche proposée par *iota²* dans un contexte différent pour lequel elle a été conçue. Les résultats sont favorables du point de vue de la chaîne car elle s'est avérée opérationnelle sur le territoire haïtien avec des productions en accord avec les données de référence.

Ce travail a confirmé le besoin de l'expertise humaine dans l'évaluation visuelle des rendus des classifications. Les indices statistiques ne reflètent pas toujours la qualité des résultats et peuvent donner des appréciations très optimistes qui diffèrent de la réalité.

Données de référence. Dans le contexte haïtien, le besoin de plus de données de références de Prairie et du Sol nu est indiscutable. Étant donné qu'une grande partie du pays est couverte par ces classes, ces données ne peuvent pas être négligées dans la réalisation d'une carte d'occupation du sol pertinente par classification. Quant aux données OpenStreetMap, leur potentiel à grande échelle reste discutable compte tenu de leur disponibilité mondiale très hétérogène.

Difficulté de classification. Plusieurs causes d'erreur de classification ont été identifiées : i) Principalement, le manque des données de référence des classes prédominantes sur le territoire (Sol nu, Prairie) ; ii) la taille des échantillons. Certaines classes comme le Bâti sont représentées par une majorité d'échantillons de taille inférieure à la taille du pixel des images ce qui rend difficile la classification ; iii) l'utilisation des données en polygone (classes Routes et Cours d'eau). Ces données sont souvent décalées induisant en erreur le classifieur ; iv) la similarité des signaux spectrales de différents classes. Des classes comme le Sol nu et les Routes peuvent présenter des réponses spectrales très similaires provoquant également des confusions dans la classification ; v) et des difficultés liées au relief. Des ombres créées par le relief rendent difficile la différenciation des surfaces. Toutefois, ces erreurs peuvent être corrigées lors du post-traitement.

Perspectives. La suite de ce travail doit consister à l'intégration des données de référence des classes manquantes ou insuffisantes dans les classifications. Nous proposons, dans l'attente des échantillons de référence prélevés sur le terrain par les partenaires haïtiens du RO, l'utilisation des données de la classe Route, dûment triées, en tant que données de référence du Sol nu. Comme expliqué précédemment, une grande majorité de ces données correspond à des routes non goudronnées. Les difficultés de classification du Bâti pourraient être palliées en ne gardant que les échantillons qui respectent un critère de densité lors de la phase de préparation des données de la classe afin de ne conserver que les aires urbaines vraiment détectables et classifiables avec les images Sentinel-2. Dans tous les cas, dans l'étape de post-traitement, le peu de Bâti isolé détecté tend à disparaître. L'utilisation d'un masque Mer pourrait également améliorer la Précision globale de la classification dans la mesure où l'Eau et les Zones d'inondations seraient mieux classifiées. Enfin, le calcul de la variance de l'erreur, à partir de plusieurs

exécutions de *iota*² sous un même ensemble de données et des paramètres, permettrait de mieux interpréter les indices statistiques.

RÉFÉRENCES

Arias, M. et Morin, D. 2017. Vector-Tools-7ab2125a. 2016. Disponible sur : http://tully.ups-tlse.fr/jordi/vector_tools

Baghdadi, N. et Zribi, M. 2017. Observation des surfaces continentales par télédétection I – Agriculture et forêt. Volume 3. Londres, Grande Bretagne : ISTE Editions Ltd, 454 p.

Barrette, N. et Daleau, L. 2012. Haïti, également terre de cyclones. Dans Rainhorn, J. (Ed.), Haïti, réinventer l'avenir. Éditions de la Maison des sciences de l'homme. doi : 10.4000/books.editionsmsh.8316

Breiman, L. 2001. Random Forests. Machine learning, 45, 5–32.

CESBIO. 2018. Vector-tools. Disponible sur : <https://framagit.org/inglada/iota2/tree/develop/scripts>

Chen, F., Jiang, H., Van de Voorde, T., Lu, S., Xua, W. et Zhou, Y. 2018. Land cover mapping in urban environments using hyperspectral APEX data: A study case in Baden, Switzerland. International Journal of Applied Earth Observation and Geoinformation, 71, 70-82.

Équipe de développement de QGIS. 2014. Système d'information géographique QGIS. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>

Drusch, M. Bello, U.D. Carlier, S. Colin, O. Fernandez, V. Gascon, F. Hoersch, B. Isola, C. Laberinti, P. Martimort, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing on Environement, 120, 25–36.

FAO. 2016^a. Base de Données Principale d'AQUASTAT, Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). Disponible sur : <http://www.fao.org/nr/water/aquastat/data/query/results.html>. Récupéré le 20 août 2018.

FAO. 2016^b. Site web AQUASTAT, Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). Disponible sur : http://www.fao.org/nr/water/aquastat/countries_regions/HTI/indexfra.stm. Récupéré le 22 août 2018.

FAO. 2017. Base de données statistiques - FAOSTAT, Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO). Disponible sur : <http://www.fao.org/faostat/fr/#data/LC>. Récupéré le 20 août 2018.

Gargani, J. 2016. Crises environnementales et crises socio-économiques. Paris, France : L'Harmattan, 149 p.

- GDTA : Groupe pour le Développement de la Télédétection Aérospatiale. 1991. Les cahiers pédagogiques du GDTA - Cahier B1 Bases physiques de la télédétection. Version 1,2. Toulouse, France : GDTA, 90 p.
- Girard, M-C. et Girard, C-M. 2010. Traitement des données de télédétection - 2e édition : Environnement et ressources naturelles. Paris, France : Dunod, 576 p.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R. 2006. Random Forests for land cover classification. *Pattern Recognition*, 27, 294–300.
- Graham, S. et Riebeek, H. 2006, Hurricanes : The Greatest Storms on Earth, Earth Observatory (NASA). Disponible sur : <https://earthobservatory.nasa.gov/Features/Hurricanes?src=features-fromthearchives>. Récupéré le 20 août 2018.
- Groupe de la Banque Mondiale. 2018. DataBank. Indicateurs du développement dans le monde. Disponible sur : <https://donnees.banquemondiale.org/indicateur/AG.LND.FRST.K2?locations=HT>. Récupéré le 20 août 2018.
- Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B. et Meygret, A. 2008. Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images. *Remote Sensing of Environment*, 112, 1689-1701.
- Hagolle, O., Huc, M., Villa Pascual, D. et Dedieu, G. 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VEN μ S, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, 114, 1747-1755.
- Hagolle, O., Huc, M., Villa Pascual, D. et Dedieu, G. 2015^a. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VEN μ S and Sentinel-2 Images. *Remote sensing*, 7, 2668-2691.
- Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V. et Poulain, V. 2015^b. SPOT-4 (Take 5) : Simulation of Sentinel-2 Time Series on 45 Large Sites. *Remote sensing*, 7, 12242-12264.
- Humboldt State University. 2015. GPS 216 Introduction to remote sensing. Disponible sur : http://gsp.humboldt.edu/olm_2015/Courses/GSP_216_Online/lesson6-1/supervised.html. Récupéré le 22 février de 2018.
- Inglada, J. 2016. OTB Gapfilling, a Temporal Gapfilling for Image Time Series Library. Disponible sur : <http://tully.ups-tlse.fr/jordi/temporalgapfilling>.
- Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P. et Koetz, B. 2015. Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote sensing*, 7, 12356-12379.

- Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., et Rodes, I. 2017. Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote sensing*, 9, 95.
- Kalideos Haïti. 2018. Tuiles Sentinel-2 pour Haïti. Disponible sur : <https://haiti.kalideos.fr/drupal/fr/content/tuiles-sentinel-2-pour-haiti-0>. Récupéré le 10 juin 2018.
- Khatami, R., Mountrakis, G., Stehman, S.V. 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. *Remote Sensing of Environment*, 177, 89–100.
- Lu, D.S., Mausel, P., Batistella, M. et Moran, E. 2004. Comparison of land-cover classification methods in the Brazilian Amazon Basin. *Photogrammetric Engineering and Remote Sensing*, 70, 723-731.
- Lu, D.S. et Weng, Q.H. 2006. Use of impervious surface in urban land-use classification. *Remote Sensing of Environment*, 102 146-160.
- Ma, L., Li, M., Ma, X., Cheng, L., Du, P. et Liu, Y. 2017. A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 277-293.
- Mahdianpari, M., Salehi, B., Mohammadimanesh, F. et Motagh, M. 2017. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band and TerraSAR-X imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 13–31.
- MARNDR (Ministère de l’agriculture, des ressources naturelles et du développement rural) – République d’Haïti. 2013. Résumé des résultats du Recensement général de l’agriculture (RGA) au niveau national. Port-au-Prince, Haïti.
- MEF (Ministère de l’économie et des finances). 2016. Étude sur les risques naturels dans la boucle Centre-Artibonite et production de cartographie des risques pour les communes de Hinche, Mirebalais, Saut d'Eau, Titanyen et St-Michel-de-l'Attalaye Programme "Développement régional de la boucle Centre -Artibonite". Echirolles, France : ARTELIA Eau et Environnement, 350 p.
- Météo France. 2018. Glossaire. Disponible sur : <http://www.meteofrance.fr/publications/glossaire/153089-ouragan>. Récupéré le 13 août 2018.
- NOAA. 2017. Frequently Asked Questions: What are the average, most, and least tropical cyclones occurring in each basin? Atlantic Oceanographic and Meteorological Laboratory, Hurricane Research Division. Disponible sur <http://www.aoml.noaa.gov/hrd/tcfaq/E10.html>. Consulté le 23 août 2018.
- © OpenStreetMap. 2018. Haïti and Dominican Republic lastest free. Récupéré le 16 mars 2018. Disponible sur : <https://www.openstreetmap.org>
- Pekel, J-F., Cottam, A., Gorelick, N. et Belward, A. S. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540, 418-422.

- 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*, 187, 156-168.
- Pauleit, S. et Duhme, F. 2000. Assessing the environmental performance of land cover types for urban planning. *Landscape and Urban Planning*, 52, 1-20.
- RNCan. 2018. Classification et analyse des images. Ressources naturelles Canada_Gouvernement du Canada. Disponible sur : <http://www.rncan.gc.ca/sciences-terre/geomatique/imagerie-satellitaire-photos-aeriennes/imagerie-satellitaire-produits/ressources-educatives/9362>. Récupéré le 10 août 2018.
- Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J. 2012. An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sensing*, 67, 93–104.
- Stewart, S. 2017. Hurricane Matthew. National hurricane center tropical cyclone report – NOAA. Les États Unis, 96 p. disponible sur : https://www.nhc.noaa.gov/data/tcr/AL142016_Matthew.pdf. Récupéré le 23 août 2018.
- Szuster, B.W., Chen, Q., Borger, M. 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography* 31, 525–532.
- Theia-land. 2018. Données satellitaires optiques. Disponible sur : <http://www.theia-land.fr/fr/produits/réflectance-sentinel-2>. Récupéré les 20 juin 2018.
- Yesou, H., Pottier, E., Mercier, G., Grizonnet, M., Haouet, S., Giros, A., Faivre, R., Huber, C. et Michel, J. 2016, juillet. Synergy of sentinel-1 and sentinel-2 imagery for wetland monitoring information extraction from continuous flow of sentinel images applied to water bodies and vegetation mapping and monitoring. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. doi : 10.1109/IGARSS.2016.7729033
- Xu, R., Zhang, H.S. et Lin, H. 2017. Urban impervious surfaces estimation from optical and SAR imagery: A comprehensive comparison. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10, 4010-4021.
- Zhang, H.S. et Xu, R. 2018. Exploring the optimal integration levels between SAR and optical data for better urban land cover mapping in the Pearl River Delta. *International Journal of Applied Earth Observation and Geoinformation*, 64, 87-95.
- Zhang, H., Li, J., Wang, T., Lin, H., Zheng, Z., Li, Y. et Lu, Y. 2018. A manifold learning approach to urban land cover classification with optical and radar data. *Landscape and Urban Planning*, 172, 11-24.

ANNEXES

ANNEXE 1

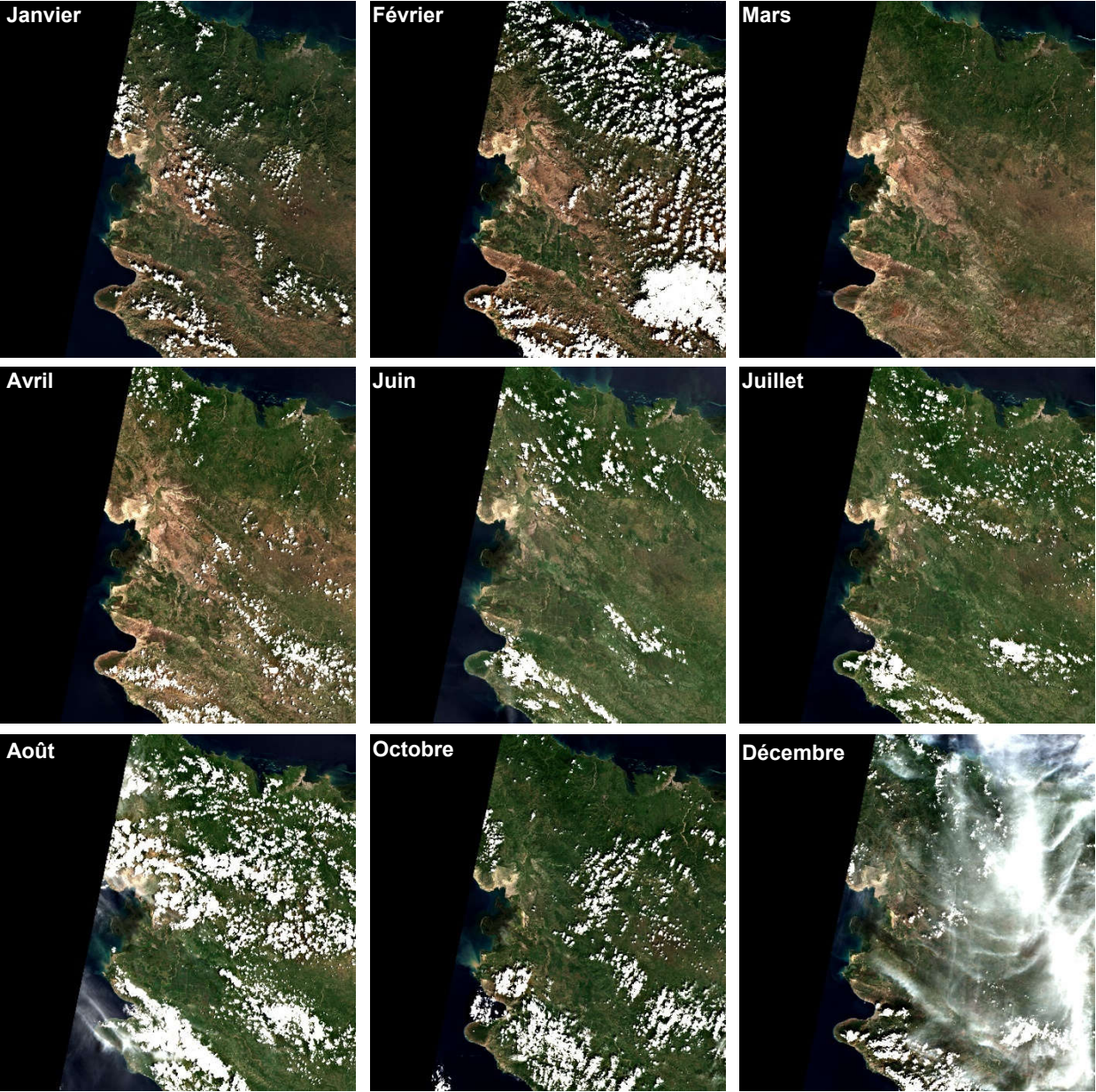


Figure 1. Série temporelle Sentinel-2 2017. Tuile 18QYG. SOURCE : Theia.

ANNEXE 2

Tableau 1. Récapitulatif des couches récupérées pour la préparation de données de référence.

Catégorie	Couche	Type de géométrie	Description	Nombre d'entités
SOURCE : DD\CNES-Haïti				
Bâti	buildings	Polygone	Bâti	37346
	buildings_2	Polygone	Bâti. La plupart des polygones sont déjà dans « buildings » ou dans la couche d'OSM. Ils sont, soit superposés à ces 2 couches, soit un peu décalés.	17741
Hydro	3Q_Reseau_hydrologique_et_endu_polyline	Ligne		31
	5Q_Reseau_hydrologique_polyline	Ligne		18
	haïti_lakes	Polygone		91
	hydrography	Ligne		12639
	Lac	Polygone		138
	Mer_Haïti	Polygone		1
	Mer_Haïti_topo	Polygone		42
	natural	Polygone		2263
	Reseau_hydrographique_principal	Ligne		367
	Reseau_hydrographique_sec_ondaire	Ligne		12184
Limites Admin	waterways	Ligne		6736
	Limite_communale	Polygone		136
	Limite_departementale	Polygone		9
	Limite_pays	Polygone		31
Macaya	Limite_sections_communales	Polygone		560
	Limite_Parc_Macaya	Polygone		1
Réseau routier	roads	Ligne		68076
	Reseau_routierNord2	Ligne		3264
Tsunammi UNESCO	Batiment_Enjeux_Caph_112014	Polygone		50
	Batiment_Enjeux_PP_112014	Polygone		44
	Zone_Impact_Tsunami_Caph	Polygone		16
	Zone_Impact_Tsunami_FL	Polygone		1
	Zone_Impact_Tsunami_PP	Polygone		2
	Zone_Inondation_Tsunami_Caph	Polygone		3

Catégorie	Couche	Type de géométrie	Description	Nombre d'entités
	Zone_Regroupement_Cap_112014	Polygone		18
Zonages	ZgeGeotech_FortLiberte	Polygone	Mangrove, zone marécageuse.	71
SOURCE : OpenStreetMap				
Bâti	Bâti	Polygone	Bâti	904250
	Water ways	Ligne	Les rivières les plus grandes sont souvent disponibles sous forme de polygones et de lignes.	12923
Hydro	Eau	Polygone	Réservoirs, piscines, rivières, zones humides, étendues d'eau	2053
	Transport	Polygone	Stations des bus, de tram...	48
Réseau routier	Traffic	Polygone	Parkings, stations de service...	1754
	Route	Ligne	Routes, chemins	186826
	Chemins de fer	Ligne	Voie ferrée, métro	154
	Points d'intérêt	Polygone	Bâti, <i>outdoors</i> :parcs, terrain de foot...	12047
Lieux d'intérêt	Lieux de culte	Polygone	Bâti	761
	Places	Polygone	Ville, village, hameau, île, ferme.	121
Nature	Nature	Polygone	Plage, falaises, arbres...	312
	Utilisation du sol	Polygone	Forêt, champ, prairie...	16906

ANNEXE 3

Fichiers de paramétrage de *iota*²

```

#####
#                               Chain
#####

chain:
(
  outputPath: '/work/OT/siaa/Stages/2018_ocs_haiti/runs/Results'
  remove_outputPath: True

  jobsPath: '/work/OT/siaa/Stages/2018_ocs_haiti/runs/jobs'
  pyAppPath: '/work/OT/siaa/Stages/2018_ocs_haiti/tools/iota2/src/iota2/scripts/common'
  chainName: 'OSO_chain_CONFID'
  nomenclaturePath: '/work/OT/siaa/Stages/2018_ocs_haiti/runs/Config/nomenclature.csv'
  outputStatistics: True

  merge_final_classifications: True
  merge_final_classifications_ratio: 0.1
  merge_final_classifications_undecidedlabel: 255
  merge_final_classifications_method: "dempstershafer"#"majorityvoting" or "dempstershafer"
  dempstershafer_mof: "precision"#"precision/recall/accuracy/kappa
  keep_runs_results: True

  listFile: 'T18QYG'
  featuresPath: '/work/OT/siaa/Stages/2018_ocs_haiti/runs/Features_T18QYG'
  L5Path: 'None'
  L8Path: 'None'
  S2Path: '/work/OT/siaa/Stages/2018_ocs_haiti/data/sentinel2'
  S1Path: 'None'
  S2_S2C_Path: 'None'
  userFeatPath: 'None' # !!

  mode: 'one_region' #one_region or multi_regions or outside
  regionPath: '/work/OT/siaa/Stages/2018_ocs_haiti/runs/limite_pays.shp'
  regionField: 'ID'
  model: 'None'

```

Figure 2. Extrait du fichier de configuration de *iota*² (.cfg).

	A
1	Culture:10
2	Végétation boisée:33
3	Bâti:41
4	Route:44
5	Surface minérale:46
6	Eau:51
7	Mer:52
8	Zone Inondable:53
9	Prairie:211
10	Sol nu:255

ColorFile - Bloc-notes				
Fichier	Edition	Format	Affichage	
0	255	255	255	
10	255	85	0	
33	0	255	0	
41	255	0	255	
44	0	255	255	
46	255	184	2	
51	0	0	255	
52	0	0	127	
53	0	150	100	
211	170	170	0	
255	230	192	86	

Figure 3. Fichier de nomenclature (.csv) et fichier de couleur (.tx).

ANNEXE 4

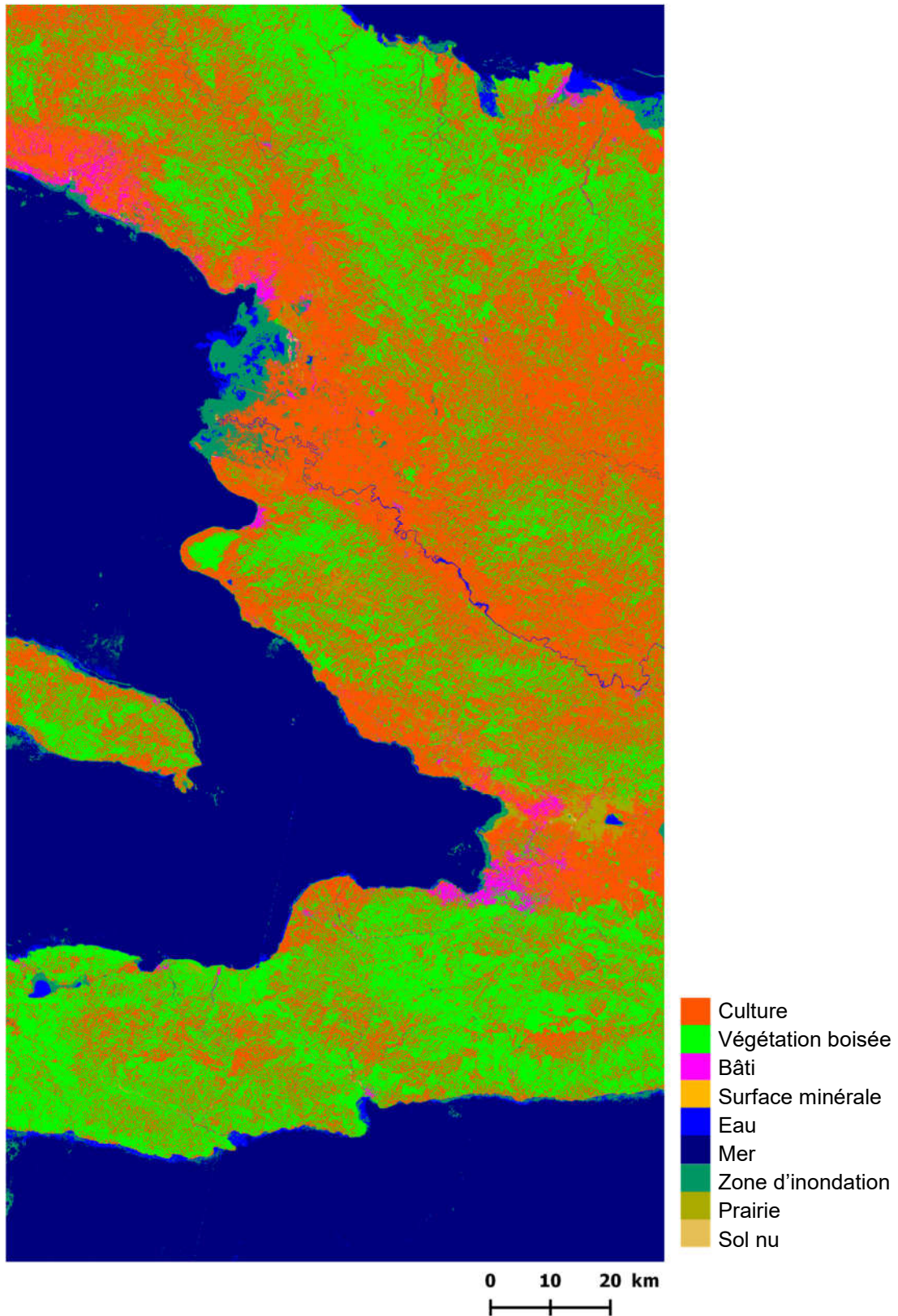


Figure 4. Classification d'occupation du sol sur Haïti à partir de 46 images sentinel-2, et 9 classes (lancement 5).

ANNEXE 5

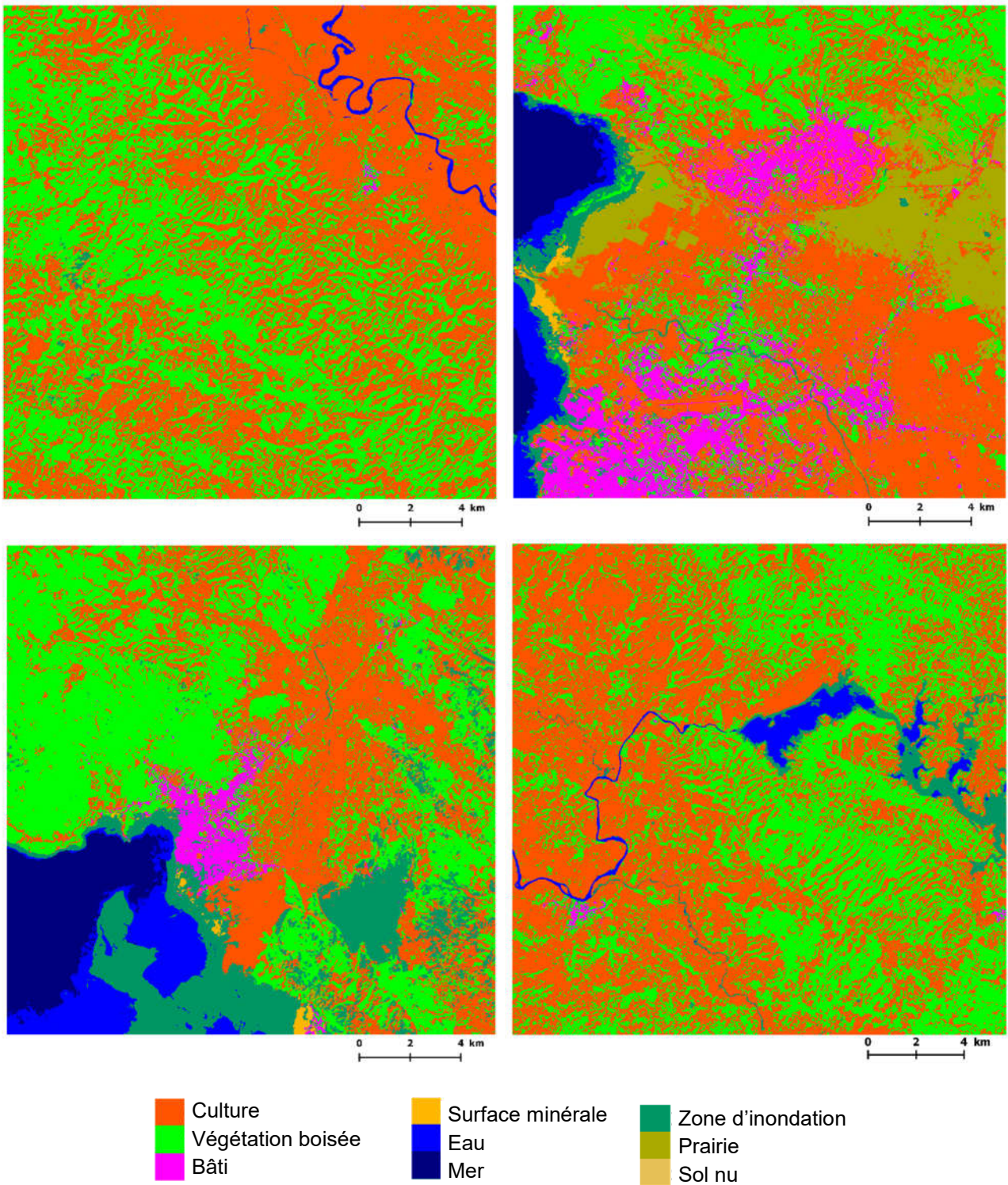


Figure 5. Extraits de la classification d'occupation du sol sur l'ensemble d'Haïti à pleine résolution (lancement 6).