Guide Technique :

Constitution d'une base de données de référence dédiée à l'apprentissage et la validation des classifications d'images satellites : la vérité-terrain

Montpellier, mars 2021

Camille Lelong, Cirad UMR TETIS, Montpellier, France

Stéphane Dupuy, Cirad UMR TETIS, Montpellier, France

Ce rapport a été élaboré dans le cadre du projet Territoires Productifs Résilients en Haïti (TPR).

Référence à citer :

Camille Lelong, Stéphane Dupuy, 2021. Guide technique - Constitution d'une base de données de référence dédiée à l'apprentissage et la validation des classifications d'images satellites : la vérité-terrain, Montpellier (France) : Cirad, 16 pages.

© Cirad, Département ES - UMR TETIS, 2021.

SOMMAIRE

Table des matières

1. Introduction 3
2. Nomenclature 3
3. Matériel terrain (GPS) 4
4. Projections 5
5. Digitalisation 5
6. Images 9
7. Table attributaire 11
8. Densité de polygones 12
9. Conclusion 14
10. Remerciements 15

1. INTRODUCTION

La création d'une base de données de référence consiste à relever la réalité du terrain, sans interprétation préalable et sans a priori, et à l'enregistrer sous forme de couche vectorielle, composée de polygones délimitant des surfaces homogènes bien renseignées. Cette base doit être suffisamment fiable pour être considérée comme la « vérité », car elle servira autant à calibrer les modèles de représentation cartographiques qu'à en évaluer les résultats.

Ce document décrit donc la méthodologie de construction de cette base de données de vérité-terrain, consigne les différentes étapes de cette construction, et insiste sur les différents points de vigilance à apporter à chacune de ces étapes afin de constituer une base robuste et valide.

2. NOMENCLATURE

En amont de toute cartographie (occupation du sol, usage du sol, systèmes de cultures, ou encore nature du sol ou état sanitaire des cultures, etc…) se trouve une question : « que souhaite-t-on représenter sur la carte » ? Il s'agit de clarifier l'information qui est attendue par l'utilisateur de la carte. Il en découle alors **un thème (ou titre) de carte**.

Ensuite une liste des différents types de surfaces/d'objets souhaités être reconnus au travers de la carte doit être exprimée par les utilisateurs, en cohérence avec le thème de la carte uniquement. Par exemple, on ne peut pas mélanger dans une même carte différents niveaux d'information comme la discrimination des surfaces recouvertes de végétation de celles laissées en sol nu, d'une part, et la détection des cultures atteintes de maladie, d'autre part. Cette liste est la **nomenclature** de base de la carte.

Cette nomenclature doit toutefois être ensuite confrontée à ce qui est réellement distinguable sur les données de télédétection disponibles, afin de dresser conjointement avec les producteurs de la carte une **liste des classes** qui pourront être reconnues. Cette liste de classes sera la **nomenclature définitive** de la carte.

C'est véritablement la nomenclature qui est l'élément fondamental de toute cartographie :

- Elle détermine le type d'information que l'on souhaite cartographier
- Elle détermine aussi le niveau précision souhaité pour cette information
- Elle liste tous les types possibles dans la zone d'étude au regard de cette information
- Elle permet d'établir des groupes d'objets similaires, sur la base de caractéristiques communes au regard de cette information

La nomenclature doit donc être définie avec soin, et en collaboration étroite entre les différents utilisateurs potentiels de la carte et ses producteurs.

Dans le cas de la **cartographie de l'occupation des sols dans la région de Nippes en Haïti**, thème qui nous concerne au sein du TPR, la nomenclature a été discutée entre les agents du PTPR (J. Rambao, A. Médard, C. Mondé), du CNIGS (B. Piard, P. Mondésir, M.-E. Manasse), de la Banque Mondiale (C. Grosjean) et du CIRAD (C. Lelong, S. Dupuy, R. Gaetano).

Dans le tableau 1, la deuxième colonne indique les classes d'occupation du sol qu'il serait souhaitable de discriminer, du point de vue des utilisateurs de la carte (le PTPR en particulier), et qui sont potentiellement distinguables par télédétection à très haute résolution spatiale (images de type Pléiades). Toutefois, le projet prévoit d'établir ces cartes à l'échelle de toute une région agroécologique, i.e. les Nippes, bien trop étendue pour prétendre être cartographiée à cette échelle de manière

opérationnelle. L'objectif étant plutôt d'établir cette cartographie à l'échelle des données de type Sentinel 2, à des fins de reproductibilité dans le temps et d'étendue dans l'espace, certaines classes ne pourront pas être discernées les unes des autres : elles ont donc été fusionnées sous le même type d'occupation du sol. C'est cette liste de classes, listée dans la première colonne, qui sera retenue pour la cartographie sous IOTA2. Il est aussi important de signaler que les classes 2 et 3 ont peu de chances d'être distinguables au moyen de données Sentinel-2 et devront probablement être fusionnées pour éviter les erreurs d'affectation de classe : ce type de classes fait partie de celles à tester lors des traitements afin de décider de leur bien fondé.

Tableau 1 : Nomenclature retenue pour la cartographie de l'occupation des sols dans la région des Nippes à Haïti

Classes de niveau 1 pour des données haute résolution spatiale (Sentinel 2)		Classes de niveau 1 pour des données très haute résolution spatiale (Spot6/7 ou Pléiades)	
Codes	**Noms**	**Codes**	**Noms**
1	Zones agricoles (cultures annuelles ou pluriannuelles)	11	Cultures annuelles ou pluriannuelles ans arbres
		12	Cultures annuelles ou pluriannuelles avec arbres
2	Zone d'agroforesterie ou d'arboriculture	21	Système agroforestier
		22	Arboriculture mono-spécifique
3	Zone naturelle boisée	31	Végétation naturelle arborée
		32	Végétation naturelle arbustive
4	Zone naturelle herbacée	4	Végétation naturelle à dominance herbacée
5	Zone humide	51	Zone humide intérieure
		52	Zone humide côtière et marine
6	Zone de sols nus	6	Espace ouvert sans ou avec très peu de végétation
7	Zone bâtie	71	Zone bâtie
8	Eaux continentales	8	Eaux continentales
9	Eaux maritimes	9	Eaux maritimes

3. MATERIEL TERRAIN (GPS)

Le matériel de base sur le terrain est un **GPS** permettant d'enregistrer des points en stockant leurs coordonnées géographiques ainsi qu'au moins un code de classe. La collecte de tout autre type d'information permettant de comprendre la nature de l'occupation du sol, son hétérogénéité, son évolution, son environnement, etc…. est toujours un gain non négligeable.

Dans la mesure où aucun autre moyen d'enregistrement d'informations complémentaires n'est possible, il faudra écrire dans un carnet de terrain tous les renseignements en prenant soin de les identifier avec un numéro de point identique au point GPS enregistré, et si possible en réécrivant les coordonnées à des fins de contrôle.

A l'heure actuelle, de nombreuses tablettes tactiles offrent la fonctionnalité GPS. Equipées d'un logiciel dédié, il est alors possible d'y afficher une image satellite et/ou une carte pré-chargées, des couches vectorielles produites en amont, de s'y repérer par rapport à sa position sur le terrain, et bien entendu d'y enregistrer des points ou des polygones, voire même des déplacements sous forme de traces. Le stockage des informations se fait alors sous forme de donnée vectorielle, et les coordonnées des points enregistrés se voient attribuer toute une liste d'informations (table attributaire), saisies à la main ou sélectionnées dans des listes déroulantes prédéfinies. C'est le mode opératoire à privilégier dans la mesure où il est rapide à mettre en œuvre sur le terrain, permet d'enregistrer un grand nombre et une grande variété d'informations, et se récupère sous forme numérique exploitable dans un SIG sans saisie post-terrain et échangeable entre les participants au projet.

Pour les tablettes sous Android, il est préférable d'installer et utiliser **QField**, un logiciel libre et gratuit équivalent à QGIS sur PC, donc totalement compatible, et facile à manipuler. Mais d'autres applications sont possibles, comme le Kobotoolbox utilisé par le TPR.

En supplément de ces saisies d'informations, des **photographies** prises pour illustrer le milieu du point de mesure sont bienvenues, à condition d'être correctement référencées par rapport à l'identifiant du point de mesure en question.

4. PROJECTIONS

Une grande attention doit être portée à la projection dans laquelle toute information géographique est stockée, que ce soit au bureau comme au terrain. Il est en effet essentiel que tous les fichiers échangés soient projetés dans le même système de référence ; sans cela on s'expose à de gros problèmes de cohérence et de superposition. Or le principe fondamental du système d'informations géographiques, manipulant des fichiers vectoriels et des fichiers images au format raster de différentes sources, est de pouvoir superposer parfaitement ces différentes sources d'information.

La projection à privilégier est la projection standard UTM, associée au géoïde WGS84. En Haïti, la zone à considérer est UTM 18N.

Une projection locale risque de ne pas être reconnue, ou d'être différemment paramétrée dans un logiciel différent et donc d'engendrer des erreurs.

Enfin, lorsqu'on exporte le fichier shapefile, en particulier depuis ArcGis, il faut s'assurer de conserver les informations de projection afin que les autres utilisateurs puissent facilement la récupérer et utiliser le fichier sans risque d'erreurs, y compris sous un autre logiciel (exemple : à l'import sous QGIS d'un fichier ArGis, il faut corriger le SCR). Pour cela il est utile de bien préciser dans le nom du fichier le numéro EPSG de la projection utilisée. Il sera ainsi plus facile de paramétrer la bonne projection sur un nouveau logiciel. Par exemple, le code EPSG de la projection WGS84-UTM 18N est : 32618.

5. DIGITALISATION

La digitalisation consiste à construire la couche vectorielle en créant les polygones correspondant aux objets de référence. Un objet de référence correspond à une **surface uniforme et homogène** dans son affectation de classe. Il lui est attribué une classe et une seule dans la nomenclature. La délimitation de ces polygones se fait par photo-interprétation, sur une image suffisamment résolue pour en apprécier les détails à l'échelle de précision de la nomenclature. Mais c'est l'information recueillie sur le terrain qui fait foi de la réalité, en matière de caractérisation de l'objet.

Il est donc important de prendre soin de plusieurs critères lorsque l'on délimite un polygone :

- *C1 :* La surface considérée est homogène au regard de la classe qui sera attribuée au polygone. Par exemple, s'il s'agit d'une culture il ne doit pas y avoir de bâtiment, de portion route ou de tout espace en sol nu, ni d'eau, ni même de portion de forêt, à l'intérieur du polygone. Si la zone considérée utile est trop hétérogène, on ne peut pas la considérer comme un objet de référence et on ne crée pas de polygone à cet endroit. Ceci est illustré au bloc A de la figure 1 : le polygone d'origine, détouré en rouge, englobe à la fois des espaces bâtis et de l'agroforêt, ce qui est à proscrire ; les contours verts délimitent les parcelles d'agroforêt en évitant le bâti.

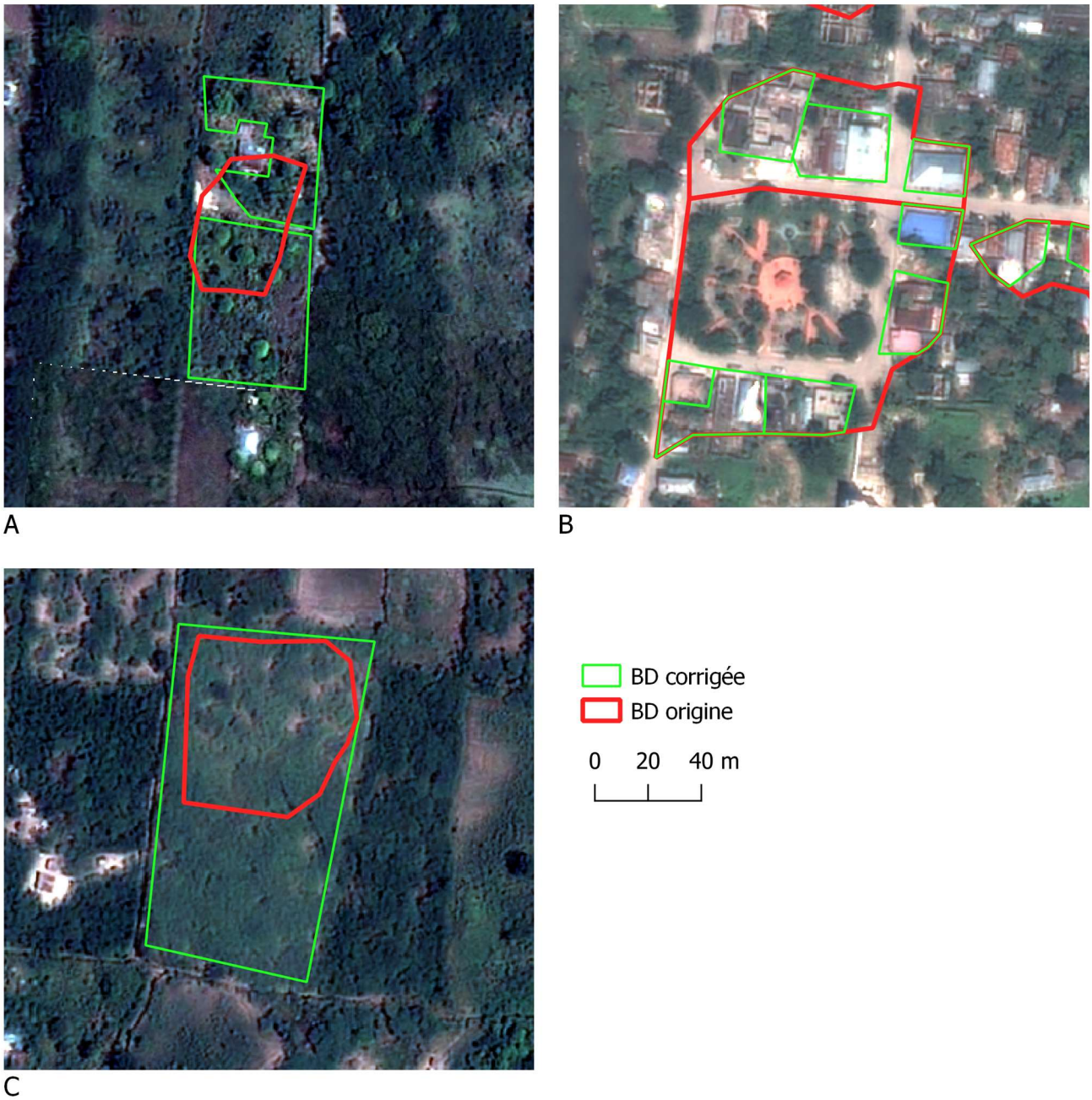

Figure 1: Illustration des erreurs de digitalisation à éviter avec en rouge les contours d'origine et en vert les contours corrigés. A : correspond au critère C1, B correspond au critère C2 et C correspond au critère C3 (Contient des informations © CNES 2020, © Distribution Airbus DS, tous droits réservés)

- *C2 :* Il vaut mieux privilégier plusieurs polygones distincts de la même classe dans le même voisinage qu'un seul gros polygone qui contiendrait des surfaces à exclure de la classe. Ceci est illustré au bloc B de la figure 1 : les polygones d'origine, détourés en rouge, englobent à la fois des espaces bâtis et de la végétation ; les contours verts délimitent uniquement les parcelles de bâti. Même si les parcs et autres espaces verts entourant le bâti peuvent être considérés comme de la végétation dite urbaine, ce sont des surfaces végétalisées, avec des propriétés spectrales bien distinctes de celles du bâti. En outre, un algorithme de classification supervisé ne prend pas en compte les critères de contexte pour reconnaitre des espaces selon leur environnement. On ne doit donc pas mélanger ces différentes surfaces.

- *C3* : La surface considérée est suffisamment grande pour englober la variabilité locale au sein de la classe. Le polygone doit donc être le plus grand possible, tout en respectant la clause d'homogénéité en regard de la classe. Il peut donc y avoir au sein du même polygone, par exemple, des zones plus claires ou plus sombres, du moment qu'elles appartiennent bien à la classe qui sera affectée au polygone. De même dans le cas d'une culture arborée, qui présente une structure de plantation avec des alternances d'arbres et d'inter-rangs herbacés ou en sol nu, c'est l'ensemble de cette structure

qui caractérise la plantation donc tout doit être inclus dans le polygone. Ceci est illustré au bloc C de la figure 1 : le polygone d'origine, détouré en rouge, ne couvre pas toute la parcelle d'agroforêt, pourtant bien homogène en structure et en type d'occupation du sol ; le contour vert délimite une surface plus importante et donc plus représentative de cette parcelle d'agroforêt.

- *C4 :* Il est toujours intéressant de visualiser le polygone sur l'image qui servira à la classification (Sentinel-2) pour vérifier cette cohérence en taille et en homogénéité. La figure 2 ci-dessous illustre la digitalisation d'une parcelle superposée à une image Pléiades dans le bloc A et à une image Sentinel-2 dans le bloc B, qui montre la bonne adéquation du polygone aux différentes résolutions.
- *C5 :* Il est préférable d'éviter les bordures de parcelles afin d'éviter les transitions de classes dans des pixels plus grands que ceux utilisés pour la photo-interprétation, et les décalages éventuels entre l'image à très haute résolution utilisée pour la photo-interprétation (ex : Pléiades) et les images à plus faible résolution utilisées pour la classification (ex : Sentinel-2). La figure 2 illustre aussi la digitalisation d'une parcelle qui évite de prendre en compte les contours.

Figure 2 : Illustration des critères de digitalisation 4 et 5 expliqués ci-dessus. Le polygone rouge correspond au contour de la parcelle digitalisée superposé à une image Pléiades en A et à une image Sentinel-2 en B (Contient en A des informations © CNES 2020, © Distribution Airbus DS, tous droits réservés ; contient en B des données à valeur ajoutée traitées par le CNES pour le pôle de données Theia www.theia.land.fr à partir de données Copernicus. Les traitements utilisent les algorithmes développés par les Centres d'Expertise Scientifique de Theia).

- *C6 :* Pour les classes faciles à identifier par photo-interprétation, il est préférable d'éviter de digitaliser des polygones de très grande taille par rapport à celle des polygones habituellement détourés pour les autres classes, surtout si la classe concernée n'est pas représentative dans le paysage. En effet, cela engendrerait un déséquilibre en termes de superficie échantillonnée par rapport aux autres classes. La figure 3 illustre par exemple la digitalisation des classes « eaux maritimes » et « zones humides » : dans le bloc A, les polygones détourés en rouge sont trop étendus ; ils ont été corrigés dans le bloc B (en vert) afin de couvrir une superficie du même ordre de grandeur que celle des polygones des autres classes.

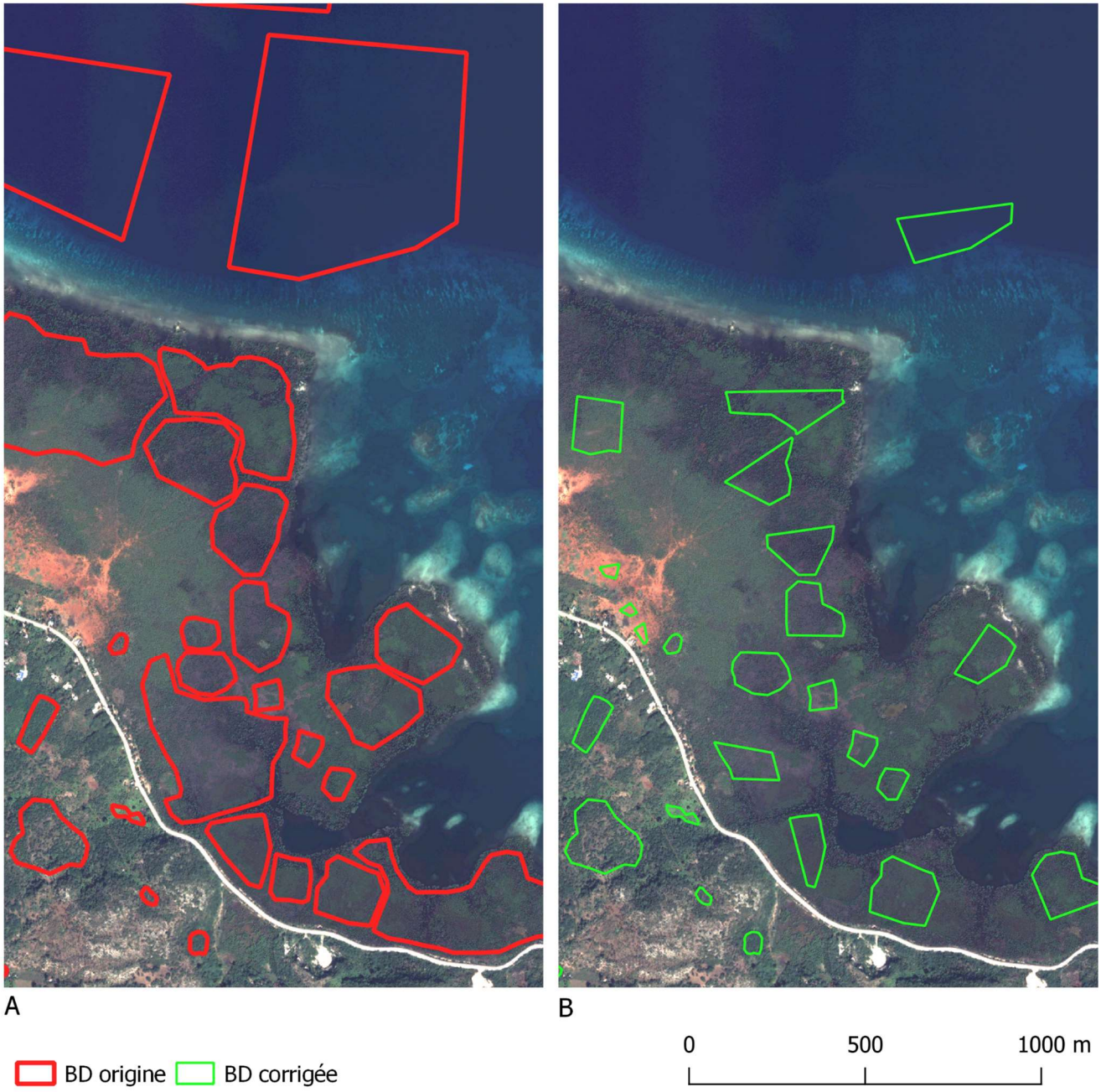

Figure 3: Illustration du critère de digitalisation C6 avec en rouge les contours d'origine et en vert les contours corrigés, superposés à une image Pléiades. (Contient en A des informations © CNES 2020, © Distribution Airbus DS, tous droits réservés).

Par ailleurs, la photo-interprétation est un travail délicat, qui requiert une grande expérience du terrain et de la manipulation d'images de télédétection, afin d'appréhender la réalité « déformée » au travers de l'image. Il est donc particulièrement utile de faire vérifier la base d'échantillons par une tierce personne afin de vérifier que les interprétations convergent ou s'il y a lieu d'une discussion, voire d'une correction, pour certains polygones.

6. IMAGES

Comme pour la manipulation des données vectorielles, il est primordial de s'assurer que toutes les données images (raster) utilisées pour la digitalisation sont **dans la bonne projection (UTM, WGS84, zone 18N)**.

Même si la carte finale repose sur des données à plus faible résolution, il est plus intéressant de **digitaliser les polygones sur la base d'une image à très haute résolution** : cela permet une meilleure précision des contours, mais aussi d'éviter des erreurs d'interprétation. Dans le même temps, on cherchera toujours à vérifier que l'emprise du polygone digitalisé est cohérente avec la donnée de traitement (Sentinel-2), c'est à dire, par exemple, que la taille du polygone doit toujours couvrir plusieurs pixels Sentinel-2, sinon il ne sera pas pris en compte. Il est aussi pratique de s'assurer que le polygone est un objet homogène sur la Sentinel-2. Il faut donc constamment basculer la visualisation entre les deux types de résolutions.

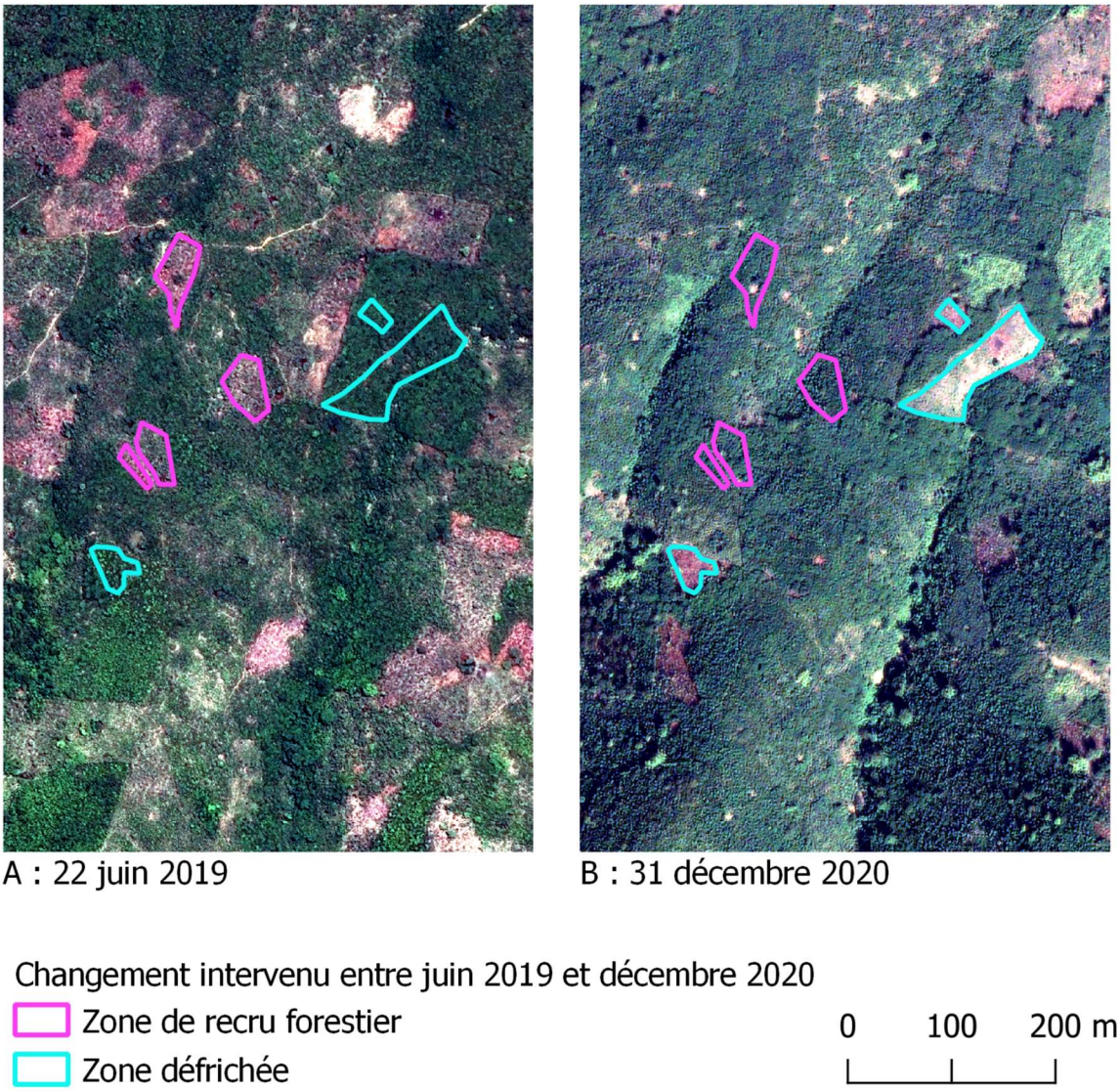

Figure 4 : Exemple de fortes variations de l'occupation du sol entre les images Pléiades acquises en juin 2019 (bloc A, à gauche) et décembre 2020 (bloc B, à droite). (Contient des informations © CNES 2020, © Distribution Airbus DS, tous droits réservés).

L'idéal est de pouvoir bénéficier d'une image de type Pléiades, acquise à une **date proche de celle où le terrain a été fait**. Dans certaines zones, la dynamique est telle que des erreurs peuvent rapidement être engendrées par l'utilisation d'une image acquise à une date trop antérieure ou

postérieure au terrain et fausser la compréhension du milieu. Il est donc important de prêter attention aux possibles variations de l'occupation du sol entre la date du terrain et celle d'acquisition de l'image, et de préférer ne rien enregistrer que de déclarer une information erronée. La figure 4 illustre clairement les fortes variations de l'occupation du sol que l'on peut rencontrer entre 2 images Pléiades acquises dans les Nippes, respectivement en juin 2019 (à gauche dans le bloc A) et décembre 2020 (à droite dans le bloc B). Des zones de forêt ont été totalement défrichées entre les deux dates (exemples au contour bleu), alors que le recru forestier est nettement visible dans d'autres zones (exemples au contour rose).

Si on ne dispose pas de données THRS disponibles (le coût de ces données étant souvent trop élevé), il est possible d'utiliser le fond d'images THRS disponible par exemple dans le logiciel QGIS. Le problème dans ce cas est qu'il n'est pas possible de connaitre la date d'acquisition de l'image visualisée. Ceci peut poser des problèmes quand on constate les évolutions d'occupation du sol qui interviennent en 18 mois (Cf. figure 4). Cependant, ce type de source peut être utile pour identifier des surfaces particulières en l'absence d'autre source à très haute résolution, la digitalisation du polygone reposant alors plutôt sur la Sentinel-2.

Dans la mesure du possible, il est par contre préférable d'éviter l'utilisation de GoogleEarth qui n'est pas rigoureusement bien calé donc pas toujours superposable aux données de traitement. Par contre, l'utilisation de GoogleEarth peut pallier à l'écart de date entre une éventuelle Pléiades (sur laquelle se base la digitalisation du polygone) et la date du terrain et ainsi permettre de vérifier que l'occupation du sol n'a pas changé. Il est en effet possible dans GoogleEarth de connaitre la date d'acquisition de l'image.

Enfin, une image de type SPOT6/7, peu idéale pour la photo-interprétation, sera toujours un meilleur compromis qu'une Sentinel-2 seule si ces données sont disponibles pour une date proche du terrain (Cf. Figure 5).

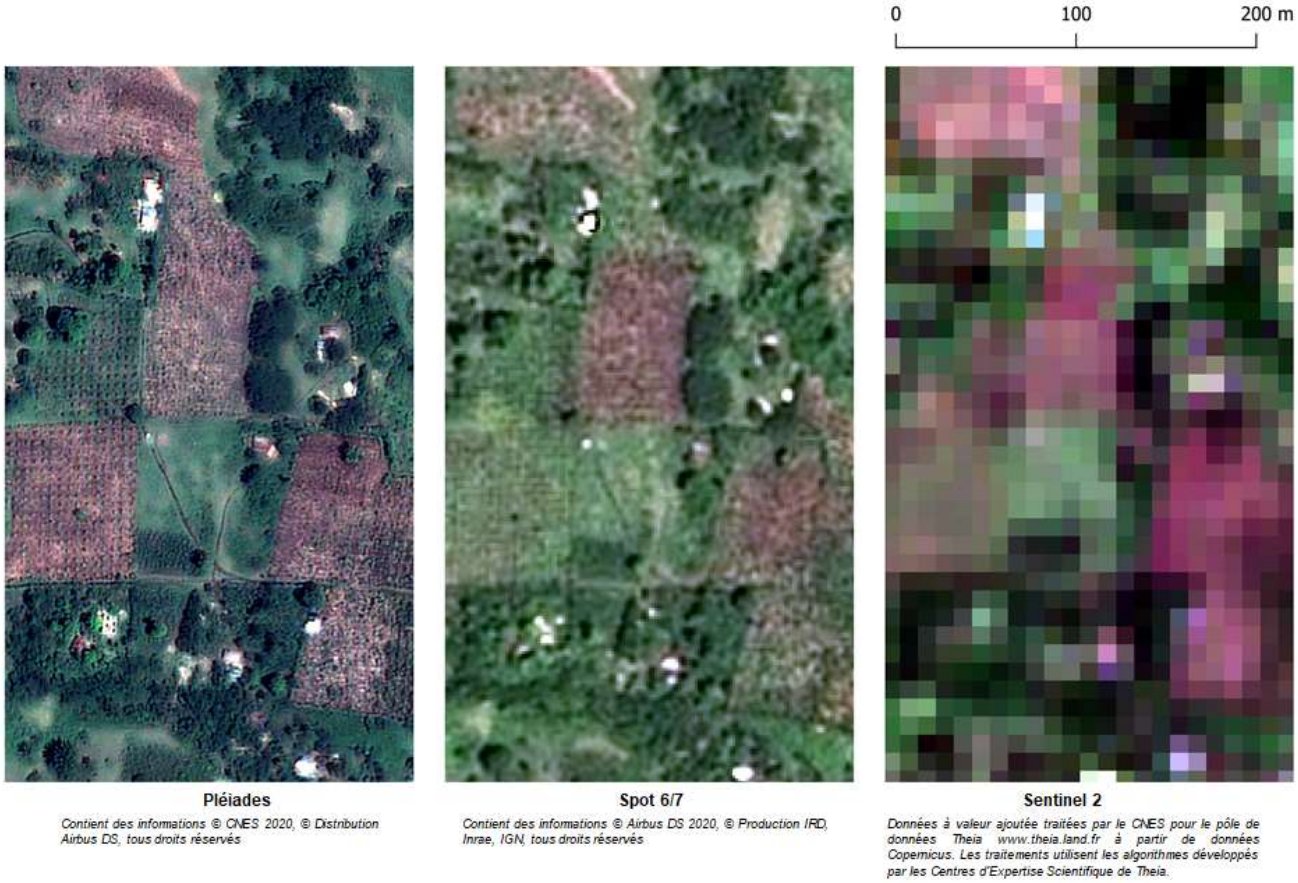

Figure 5: Variation de résolution entre les images Pléiades, Spot6/7 et Sentinel-2 (de gauche à droite).

7. TABLE ATTRIBUTAIRE

La table attributaire est la liste des informations disponibles sur chaque polygone de la couche vectorielle. Elle doit donc contenir au minimum :

- Le numéro de la classe dans la nomenclature, suivant un code totalement numérique
- Le nom de la classe en toutes lettres
- Le nom de la ou des classes de niveau hiérarchique inférieur s'il y a lieu
- La mention oui/non correspondant au fait que la classe a bien été vérifiée sur le terrain ou bien seulement photo-interprétée

Elle peut aussi contenir tout autre type d'informations recueillies sur le terrain, telles que la zone agro-climatique, l'altitude et la pente, le district ou autre département administratif, le nom du propriétaire s'il s'agit d'une parcelle, le nom du point d'eau s'il s'agit d'une rivière, mare ou autre, ou encore la liste des espèces cultivées dans la parcelle (que ce soient des cultures annuelles/vivrières ou des fruitiers ou des essences forestières) et leur structure de plantation (rangées, intercalage, disposition aléatoire, etc…). Toutes ces informations pourront être utiles à un moment ou un autre du process, notamment pour mieux comprendre d'éventuelles erreurs de classification.

Elles peuvent être aussi utiles à d'autres études sur la même zone ayant des objectifs différents et il est beaucoup plus onéreux en temps comme en argent de retourner sur le terrain ultérieurement que d'acquérir un maximum d'informations la première fois.

Une colonne indiquant l'existence de photographies prises sur le terrain dans cette parcelle est un élément supplémentaire de contrôle et de compréhension.

Enfin, une colonne portant la mention de la personne ayant digitalisé le polygone peut être utile en cas de nécessité de vérification et de discussion.

Les figures 6 et 7 donnent des exemples de tables attributaires.

Entité	Valeur
BD_ref_20210211_cirad	
Nom2	Systemes agro-forestiers
(Dérivé)	
(Actions)	
OBJECTID	10
Classes_oc	Systèmes agro-forestiers
Sentinel_I	Agrofresterie ou arboriculture
Code_cnigs	23
Code_IOTA	7
Changement	Non
Code_TPR	5p
Verificati	photointerprété et vérifié
Code1	3
Nom1	Agroforesterie et arboriculture
Code2	4
Nom2	Systemes agro-forestiers
Auteur	CNIGS

Figure 6 : Exemple de table attributaire de la base de données de référence CNIGS dans la zone des Nippes

Entité	Valeur
tpr_geospatial	
username	tpr_haiti_geospatial
(Dérivé)	
(Actions)	
start	2020-07-22T13:02:51.524-04:00
end	2020-07-22T13:08:47.556-04:00
username	tpr_haiti_geospatial
simserial	NULL
subscriber	NULL
deviceid	14:dd:a9:46:de:c2
phonenumbe	NULL
commune	anse_a_veau
section	baconnois_grand_fond
localite	Rocher Laval
classe_sen	zone_agroforesterie
classe_ple	agroforesterie
zone_agro	plaine_irrigable
pente	inf_15_pourcent
type_roche	calcaire
type_sol	sable_riche_ma
culture_an	banane_plantain banane_figue
dominance_	non
culture_pe	cachiman arbre_veritable fruit_a_pain cedre cocotier mangue chene fresne
dominanc_1	oui
gps1	18.50547832 -73.39407073 7.699999809265137 11.0
_gps1_lati	18,505478320000002
_gps1_long	-73,394070729999996
_gps1_alti	7,699999809265137
_gps1_prec	11,000000000000000
photo_zone	1595437698791.jpg
photo_zo_1	1595437710609.jpg
photo_zo_2	1595437723844.jpg
photo_zo_3	NULL
starttime	NULL
endtime	NULL
simid	NULL
devicephon	NULL
_id	60969945
_uuid	3af09793-16a6-4717-ba7b-dbd48ae03f40
_submissio	2020-07-22T20:51:16
_validatio	NULL
_index	16

Figure 7 : Exemple de table attributaire de la base de données de référence du TPR dans la zone des Nippes.

8. DENSITE DE POLYGONES

Afin de pouvoir utiliser une base de données de références terrain dans une classification de type Iota2 ou Moringa il est nécessaire d'avoir suffisamment de polygones pour chaque classe de la nomenclature définie.

Une estimation moyenne du nombre minimum d'échantillons d'apprentissage à fournir à la chaine Iota2 par classe est de 50, mais dépend surtout de la variabilité au sein d'une classe donnée. En effet, pour une classe à faible variabilité spectrale (la plupart des objets se ressemblent beaucoup, comme l'eau ou le sol nu), on aura besoin de moins de polygones (50 suffiront) mais il faudra augmenter drastiquement le nombre d'échantillons (à 100 voire 200) pour une classe hautement variable spectralement (comme les espaces arbustifs et/ou arborés).

Les polygones doivent être répartis spatialement le plus largement possible à travers la zone d'étude, et non regroupés dans un petit périmètre. Ceci permet de pallier à toutes variations locales, qu'elles soient d'ordre physique (topographie, nature du sol, mosaïque de tuiles, série temporelle, variations saisonnières, etc…) ou agricoles (pratiques culturales différentes par exemple).

Rappelons que la seule vérité est celle du terrain, qui doit toujours être la source d'information prise en compte en priorité. Toutefois, quand la situation n'a pas permis le recueil d'un nombre suffisant d'échantillons, il est possible d'enrichir la base de données par photo-interprétation. Cette photo-interprétation doit se fonder sur l'expérience acquise sur le terrain, en créant des clefs de

correspondance entre une classe donnée et son apparence sur l'image. Cette connaissance peut aussi être enrichie par l'expérience acquise sur d'autres terrains comparables en termes de climat, topographie et pratiques culturales. C'est l'accumulation de ces expériences qui rendent la photo-interprétation robuste et de confiance. Cette technique peut alors venir en complément du terrain pour augmenter le nombre d'échantillons dans les classes qui le nécessitent.

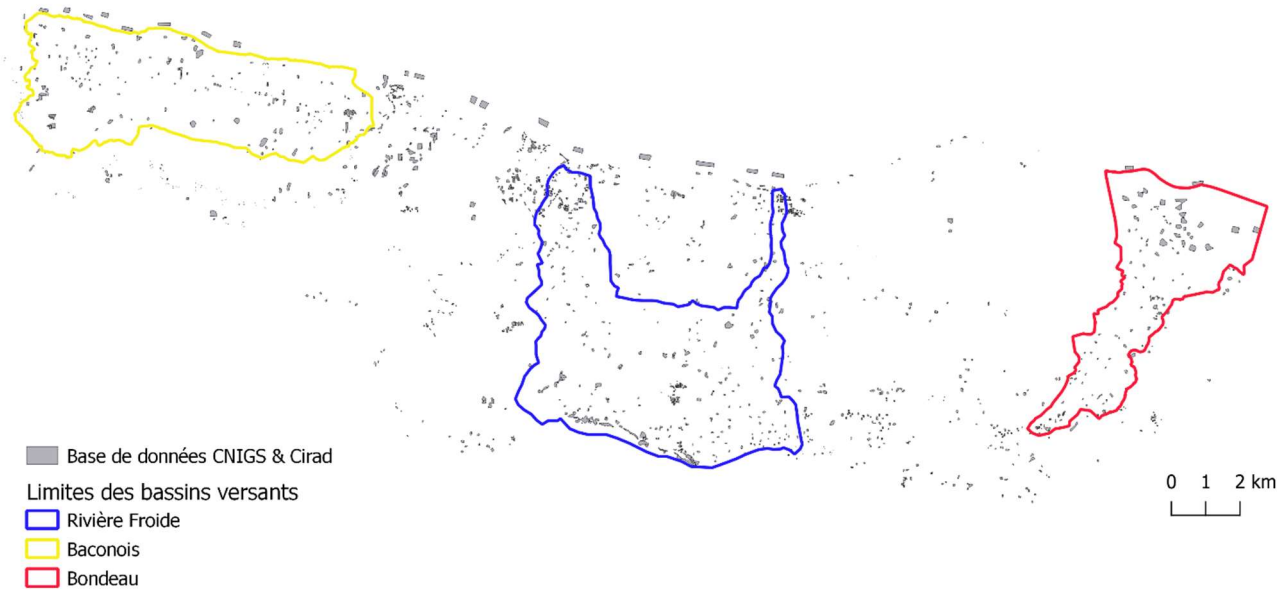

Figure 8 : Vue d'ensemble de la base de vérité terrain dans la région des Nippes, constituée en octobre/novembre 2020 par le CNIGS et le TPR et corrigée/complétée par le CIRAD en janvier 2021.

La figure 8 apporte une vue d'ensemble de la base de vérité terrain dans la région des Nippes constituée en octobre/novembre 2020 par le CNIGS et le TPR et corrigée/complétée par le CIRAD en janvier 2021. Les zones délimitées respectivement en jaune, bleu et rouge correspondent aux trois bassins versant d'intervention du TPR dans les Nippes en Haïti (d'Ouest en Est) : Baconois, Rivière Froide et Bondeau. Les tâches grisées correspondent à tous les polygones digitalisés et référencés dans la base, soit 1599 échantillons au total. Le tableau 2, informe sur la répartition et la superficie totale couverte par les échantillons de chaque classe.

Tableau 2 : Nombre de polygones et superficie totale pour chaque classe de la nomenclature (pour le niveau le plus détaillé).

Code2	Nom	Nombre de polygones	Surface (Ha)
1	Zones urbanisées (Zones bâties)	268	9.91
2	Culture agricole sans couvert arboré	368	60.20
3	Culture agricole avec couvert arboré	90	41.86
4	Systèmes agro-forestiers	181	16.63
5	Arboriculture	34	35.90
6	Végétation arborée	109	61.13
7	Végétation arbustive	148	22.64
8	Végétation à dominance herbacée	153	9.55
9	Zone humide côtière et marine	34	4.75
10	Espace ouvert sans ou avec très peu de végétation	149	74.82
11	Eaux continentales	32	76.58
12	Eaux marines	33	84.09
	Total	**1599**	**498.06**

9. CONCLUSION

La base de données de référence dédiée à l'apprentissage et la validation des classifications d'images satellites est le prérequis indispensable à toute cartographie produite à partir de données de télédétection. C'est l'élément fondamental qui permet de calibrer les modèles de classification, lesquels vont attribuer à chaque pixel de l'image une classe et une seule.

Cette base doit donc être construite avec la plus grande attention, afin de ne pas générer des erreurs ou des confusions entre les classes et de donner toutes ses chances à l'algorithme de converger vers une solution proche de la réalité. En effet, quelle que soit la performance de l'algorithme ou de la chaine de traitements utilisés, **leur précision dépend des échantillons d'apprentissage**.

La constitution de cette base de référence doit donc suivre les règles rigoureuses présentées dans ce document afin de produire une carte de qualité.

La figure 9 illustre la base de données de référence sur la région des Nippes en Haïti, obtenue dans le cadre de la collaboration TPR/CNIGS/Cirad en 2021.

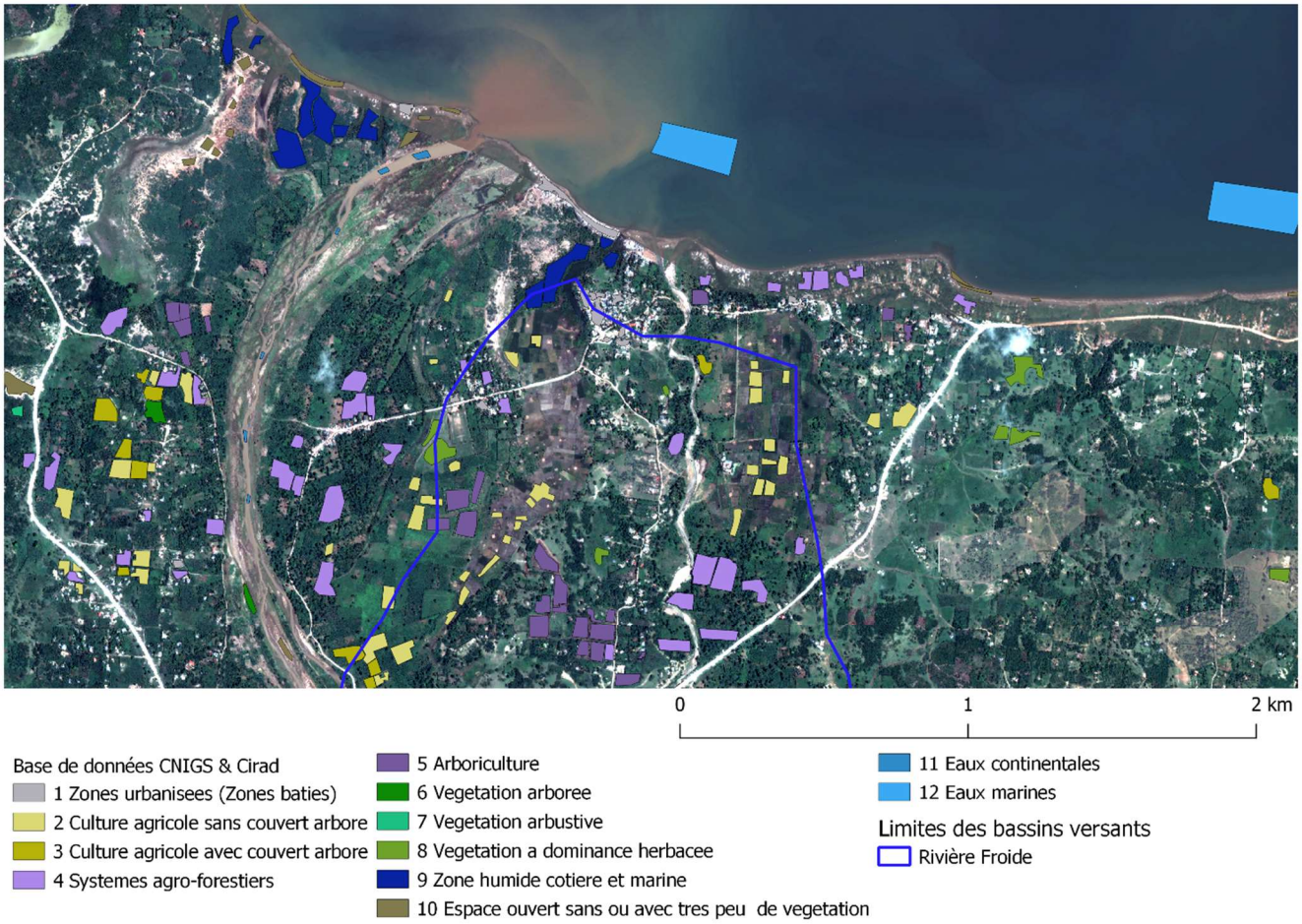

Figure 9 : Extrait de la base de données de référence dans la zone de Rivière Froide superposé à une image Pléiades acquise en décembre 2020 (Contient des informations © CNES 2020, © Distribution Airbus DS, tous droits réservés).

10. REMERCIEMENTS

Nous tenons à remercier le CNES pour la mise à disposition des images Pléiades et Spot6/7 acquises dans le cadre du projet Recovery Observatory (RO) Haïti.

Nous remercions également le CNES pour les données Sentinel-2 à valeur ajoutée traitées pour le pôle de données Theia (www.theia.land.fr) à partir de données Copernicus. Les traitements utilisent les algorithmes développés par les Centres d'Expertise Scientifique de Theia.

Département Environnement et Société (ES)
Unité Mixte de Recherche Territoire, Environnement, Télédétection et Information Spatiale (TETIS), 2021
Maison de la Télédétection - 500 rue Jean-François Breton - TA C-91/MTD 34093 Montpellier cedex 5
https://umr-tetis.fr

www.cirad.fr