



HAL
open science

Développement et comparaison de protocoles semi-automatisés de détection d'objets géomorphologiques : les surfaces d'aplanissement

Adrien Le Cadre

► **To cite this version:**

Adrien Le Cadre. Développement et comparaison de protocoles semi-automatisés de détection d'objets géomorphologiques : les surfaces d'aplanissement. Sciences de l'ingénieur [physics]. 2019. dumas-02899768

HAL Id: dumas-02899768

<https://dumas.ccsd.cnrs.fr/dumas-02899768v1>

Submitted on 15 Jul 2020

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

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

CONSERVATOIRE NATIONAL DES ARTS ET METIERS
ECOLE SUPERIEURE DES GEOMETRES ET TOPOGRAPHES

MEMOIRE

présenté en vue d'obtenir

le DIPLOME D'INGENIEUR CNAM

SPECIALITE : Géomètre et Topographe

par

LE CADRE Adrien

Développement et comparaison de protocoles semi-automatisés de détection
d'objets géomorphologiques : les surfaces d'aplanissement.

Soutenu le 04 septembre 2019

JURY

Monsieur Stéphane DURAND	Président du jury
Monsieur Paul BESSIN	Maître de stage
Madame Elisabeth SIMONETTO	Enseignant référent

Remerciements

Tout d'abord, je souhaite adresser mes sincères remerciements à Monsieur Paul BESSIN pour m'avoir encadré tout au long de ce travail. Je le remercie pour ses nombreux conseils et sa disponibilité.

Je remercie Madame Élisabeth SIMONETTO d'avoir suivi mon travail notamment pour ses remarques et ses conseils ainsi que pour la correction de ce rapport.

Je tiens également à remercier Madame Katherine KRAVITZ, Monsieur Jean BRAUN, du centre de recherche en géosciences allemand (GFZ Postdam), pour leurs esprits critiques et l'aide qu'ils ont pu m'apporter.

Pour sa réactivité et son aide, je remercie Monsieur Vincent HABCHI.

Je tiens également à remercier mes collègues de bureau et les membres du laboratoire avec qui j'ai pu échanger notamment Audrey et Mathilde.

Pour finir, j'ai une pensée toute particulière pour ma famille et les proches qui m'ont soutenu tout au long de ce travail.

Liste des abréviations

ACP : Analyse en Composantes Principales

CU : Courbure

H : Élévation

IPS NS : Identification des Surfaces d'Aplanissement Non Supervisée

IPS S : Identification des Surfaces d'Aplanissement Supervisée

MNT : Modèle numérique de Terrain

NASA : National Aeronautics and Space Administration

OBIA : Analyse d'Image Orientée Objet

PS : Surface d'aplanissement

RH : Hauteur relative

RU / TRI : Rugosité

SIG : Système d'information géographique

SH : Sans le réseau hydrographique

SL : Pente

SRTM : Shuttle Radar Topography Mission

TE : Texture

TPI : Indice de position topographique

USGS : United States Geological Survey

ZA : Zone d'étude africaine

ZAR : Zone d'étude argentine

ZT1 : Zone Test 1

ZT2 : Zone Test 2

Glossaire

Altération : ensemble des processus physiques et chimiques de surface/sub-surface qui entraînent la désagrégation des roches des surfaces continentales.

Backwearing : processus de formation d'une surface par recul horizontal de la surface, par recul des versants.

Cuirasse latéritique : carapace formée par le durcissement du sol.

Downwearing : processus de formation d'une surface par abaissement du niveau moyen des profils de rivière.

Etching : processus de formation d'une surface par érosion.

Inselberg : relief isolé qui domine significativement une surface plane (surface d'aplanissement, plaine, plateau).

Régolithe : partie recouvrant la roche mère.

Saprolite : c'est une roche meuble résultant de l'altération chimique d'une roche mère.

Table des matières

REMERCIEMENTS	2
LISTE DES ABREVIATIONS	3
GLOSSAIRE	4
TABLE DES MATIERES	5
INTRODUCTION	7
I GÉOLOGIE / GEOMORPHOLOGIE	9
I.1 LES SURFACES D'APLANISSEMENT	9
I.1.1 Les surfaces quasi planes : pénélaines, pédiments, et pédiplains	9
I.1.1.1 Les pénélaines	9
I.1.1.2 Les pédiments/pédiplains	10
I.1.2 Les surfaces de corrosion : « etchplains » et « etch-surface »	11
I.2 ÉTAT DE L'ART DES METHODES DE CARTOGRAPHIES.....	13
I.2.1 Méthode manuelle.....	13
I.2.2 Méthodes d'identification semi-automatique orientée pixel	14
I.2.2.1 Méthode utilisant l'indice de position topographique	14
I.2.2.2 Méthode de classification floue selon 4 paramètres	14
I.2.3 Autres méthodes	16
I.3 DOMAINES D'APPLICATION EN GEOSCIENCES.....	17
I.3.1 Géologie, réservoir et hydrologie.....	17
I.3.2 Traçage de l'évolution de la déformation des surfaces continentales	18
I.4 CHOIX D'UNE ZONE TEST : LES ABORDS DU FLEUVE ORANGE (FRONTIERE AFRIQUE DU SUD/NAMIBIE).....	18
II RAPPELS SUR LES METHODES DE SEGMENTATION ET DE CLASSIFICATION	20
II.1 LOGICIELS ET METHODES INTEGREES	20
II.2 FONCTIONNEMENT DES ALGORITHMES.....	21
II.2.1 Segmentation	21
II.2.2 Classification	23
II.3 COMPARAISON DES ALGORITHMES	25
II.3.1 Segmentation	25
II.3.2 Classification	25
II.3.3 Choix des algorithmes	28
III DONNEES UTILISEES ET PROTOCOLES ELABORES	29
III.1 CHOIX DES DONNEES	29
III.1.1 Namibie/Afrique du Sud.....	29
III.1.2 Argentine	30
III.2 PROTOCOLES ELABORES.....	31
III.2.1 Protocole 1 : Identification des PS par classification non supervisée (IPS-NS)	31
III.2.2 Protocole 2 : Identification des PS par classification supervisée (IPS-S)	31
IV RESULTATS	33
IV.1 ZONE DE DEVELOPPEMENT DES PROTOCOLES : LA NAMIBIE/AFRIQUE DU SUD.....	33
IV.1.1 Détermination de K.....	33
IV.1.2 Identification des PS par degré d'appartenance	34
IV.1.3 Choix des échantillons d'apprentissage	35

IV.1.4	Résultats.....	36
IV.1.4.1	IPS NS.....	37
IV.1.4.2	IPS S.....	40
IV.2	APPLICATION A UNE ZONE DE CONTROLE : REGION NORD-PATAGONIENNE (ARGENTINE)	43
IV.3	ÉVALUATION DE LA PRECISION DES RESULTATS POUR COMPARAISON AVEC LES CARTOGRAPHIES EXISTANTES.....	45
IV.3.1	Méthode des pourcentages d'occupation par classe.....	45
IV.3.2	Matrice de confusion	45
V	DISCUSSION	47
V.1	PERFORMANCES GLOBALES.....	47
V.1.1	Interprétation des résultats	47
V.1.2	Choix du paramètre K.....	49
V.1.3	Précisions obtenues.....	50
V.1.4	Objectivité et subjectivité des protocoles.....	50
V.2	LIMITATIONS	51
V.2.1	IPS NS et IPS S : Besoin d'interprétation ensuite.....	51
V.2.2	IPS S : Définition des échantillons d'apprentissage	51
V.2.3	IPS S : Application d'arbre d'une zone à une autre	51
V.3	PERSPECTIVES D'AMELIORATIONS.....	52
	CONCLUSION.....	53
	BIBLIOGRAPHIE	54
	Annexe 1 Résultats complémentaires pour l'Afrique.....	58
	Annexe 2 Résultats complémentaires pour l'Argentine	66
	Annexe 3 Statistiques et précisions pour l'Argentine	71
	Annexe 4 Comparaison RU à différentes résolutions	72
	LISTE DES FIGURES	72
	LISTE DES TABLEAUX.....	74

Introduction

Depuis la fin du XIXe siècle et les travaux classiques de W.M. Davis (1889), de nombreux chercheurs en géosciences travaillent sur des éléments géomorphologiques majeurs des reliefs émergés du globe : les surfaces d'aplanissements. Ce sont des surfaces topographiques quasi planes qui tronquent les structures géologiques sous-jacentes et peuvent s'étendre sur des dizaines de kilomètres carrés (Guillocheau et al., 2018 et références incluses). L'identification de ces surfaces d'aplanissement intervient dans un contexte plus large. En effet, leur analyse permet d'apporter de nouvelles contraintes afin de mieux caractériser les déformations des surfaces continentales au cours des temps géologiques.

L'identification des surfaces d'aplanissement est un processus long et fastidieux avec les méthodes actuelles qu'elles soient manuelles (e.g. Bessin et al., 2015) ou orientées pixel (e.g. Haider et al., 2015). De plus, c'est un travail subjectif puisqu'il dépend de l'interprétation de chacun. L'objectif principal est ainsi d'y remédier en développant un protocole d'identification semi-automatisé.

Les surfaces d'aplanissement ont des caractéristiques topographiques propres : pente, courbure, etc. (Guillocheau et al., 2018, et références incluses). Il apparaît intéressant de se diriger vers les données numériques d'élévation, ce qui sera notre point de départ. Ces données sont disponibles gratuitement et à différentes résolutions pour l'ensemble du globe, notamment grâce aux campagnes d'acquisition satellite de la NASA. Les modèles numériques d'élévation de la « Shuttle Radar Topography Mission » (STRM) sont disponibles sur le site de l'institut d'études géologiques des États-Unis (United States Geological Survey – USGS, <https://earthexplorer.usgs.gov/>). Un deuxième type de donnée qui peut s'avérer utile est les images satellites, qui permettent d'avoir un premier aperçu de la zone et des différents reliefs. Cette donnée est également accessible gratuitement et recouvre le monde entier avec des missions spatiales comme Landsat et Sentinel (NASA, <https://www.nasa.gov/> ; ESA, <https://www.esa.int/ESA>). Cette donnée est d'autant plus utile que dans les deux zones étudiées la présence de végétation est quasiment inexistante.

Aujourd'hui, dans un nouveau projet, l'objectif est d'identifier les surfaces d'aplanissement dans le bassin de Neuquén situé dans les Andes à la frontière entre l'Argentine et le Chili. Pour identifier ces surfaces d'aplanissement, un travail antérieur nous servira d'appui : la cartographie des surfaces d'aplanissement de la vallée du fleuve de

l'Orange situé à la frontière entre la Namibie et l'Afrique du Sud (Dauteuil et al. , 2015) (Figure 1).

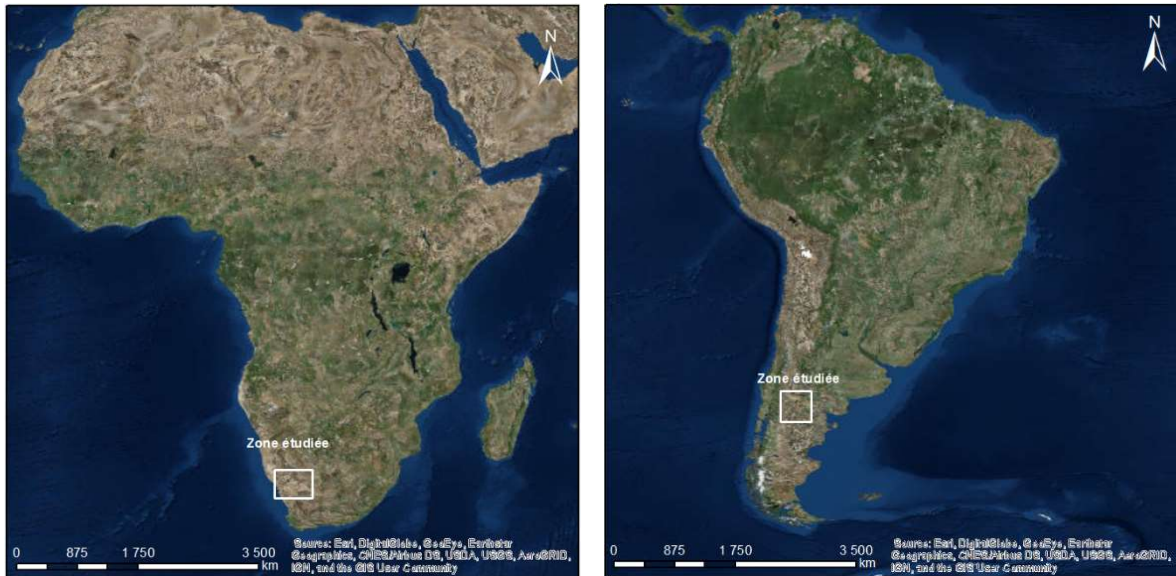


Figure 1 — Zones étudiées, Afrique (à gauche), Argentine (à droite) (source : ESRI)

La première partie de travail consistera donc à s'appuyer sur les travaux précédemment réalisés sur la vallée du fleuve Orange située à la frontière en Namibie et l'Afrique du Sud (Dauteuil et al., 2015) pour y développer une méthodologie semi-automatisée d'identification des surfaces d'aplanissement.

Le travail d'automatisation de la cartographie des surfaces d'aplanissement sera basé sur le développement d'un protocole d'analyse d'image orienté objet (OBIA – « *Object-Based Image Analysis* », Blaschke et al., 2014). Il s'appuiera sur des données de type image / raster et fera appel à des notions de traitement d'image, modélisation numérique de terrain et système d'information géographique.

Il s'agit d'un travail pionnier et innovant dans ce domaine puisque c'est la première fois qu'une méthodologie orientée objet sera utilisée pour identifier ce type d'objet géomorphologique.

C'est ainsi que dans un premier temps, nous aborderons une partie purement géologique et géomorphologique pour définir précisément les surfaces d'aplanissement. Puis, dans une deuxième partie, les différents aspects méthodologiques permettant de développer un protocole semi-automatisé seront présentés avant de détailler les différents protocoles en chapitre trois. Ensuite, les différents résultats seront présentés en quatrième partie avant d'être interprétés et discutés dans une cinquième partie. Enfin, une conclusion terminera ce rapport et présentera quelques perspectives sur ce travail.

I Géologie / Géomorphologie

Dans ce premier chapitre, nous nous intéresserons principalement aux aspects géologique et géomorphologique indispensables pour la définition des surfaces d'aplanissement et leur identification ultérieure. Un état de l'art des méthodes d'identification actuellement utilisées sera aussi présenté.

I.1 Les surfaces d'aplanissement

Les surfaces d'aplanissement sont, des surfaces topographiques quasiment planes. Celles-ci peuvent intégrer des éléments résiduels (petits monts ou collines appelés « *Inselbergs* ») et peuvent également être tronquées par un réseau hydrographique. Elles s'étendent sur plusieurs dizaines, voire plusieurs centaines de kilomètres carrés et ont été formées dans des conditions arides (Bessin, 2014).

Il existe parmi les surfaces d'aplanissement différents types synthétisés dans Bessin (2014) et Guillocheau et al. (2018) catégorisés ci-après en deux types : les surfaces quasi planes et les surfaces de corrosion. Bien que ces définitions soient basées sur le processus génétique de ces surfaces (downwearing, backwearing, etching), il est régulièrement compliqué de déterminer ces processus et nous retiendrons principalement les volets morphologiques plutôt que génétiques de ces définitions.

I.1.1 Les surfaces quasi planes : pénéplaines, pédiments, et pédiplaines

I.1.1.1 Les pénéplaines

Les pénéplaines n'ont pas de caractéristiques particulières si ce n'est qu'elles sont quasi planes et peuvent comporter quelques irrégularités de faibles amplitudes (quelques dizaines de mètres ; Figure 2). Elles peuvent aussi être tronquées par un réseau hydrographique.

Pénéplaine (péné = «presque»):

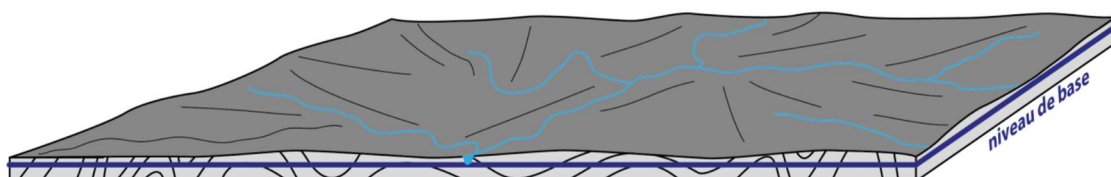


Figure 2 — Illustration d'une pénéplaine (Bessin, 2014)

Les pénéplaines sont formées par « downwearing », processus de formation associé à l'abaissement des profils de rivières.

I.1.1.2 Les pédiments/pédiplaines

Ce deuxième type de surface, pédiments/pédiplaines (Figure 3) a des caractéristiques différentes :

- Une pente faible (inférieure à 4 %)
- Un profil longitudinal concave ou rectiligne
- L'absence de réseau hydrographique
- L'absence de couverture sédimentaire

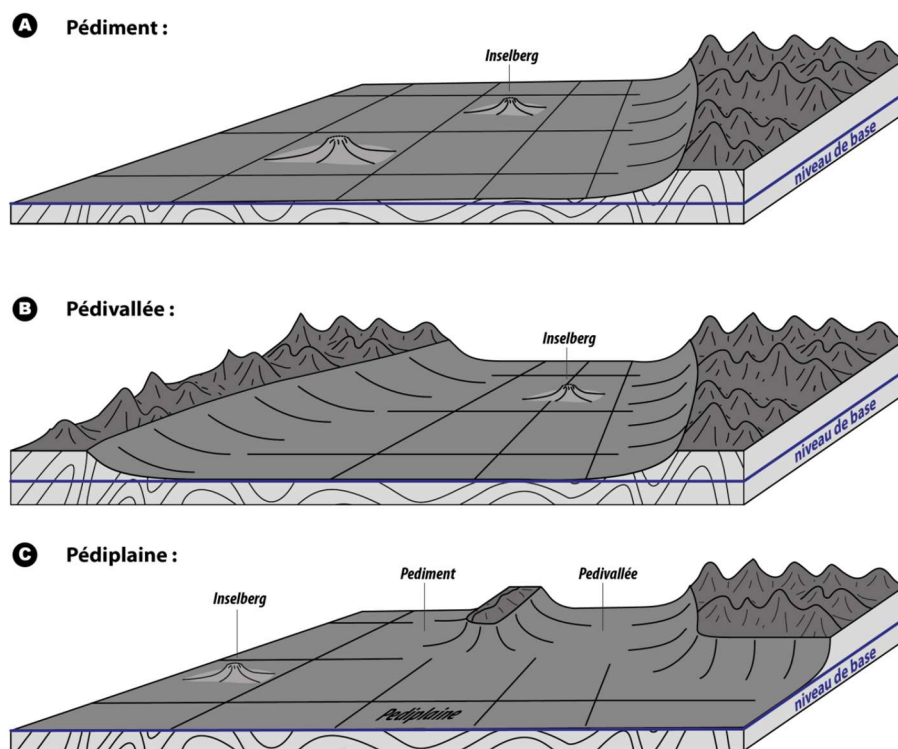


Figure 3 — Les différentes surfaces quasi planes (Bessin, 2014)

Son processus de formation est également différent, car il est associé au recul des versants, « backwearing ». Ces différentes surfaces peuvent elles aussi comporter des éléments résiduels appelés « inselbergs ».

I.1.2 Les surfaces de corrosion : « etchplains » et « etch-surface »

Les surfaces de corrosion, aussi appelées « etchplains », sont associées à un profil d'altération au sommet duquel se trouve des cuirasses latéritiques composées de fer et d'alumine leur donnant une couleur rouge.

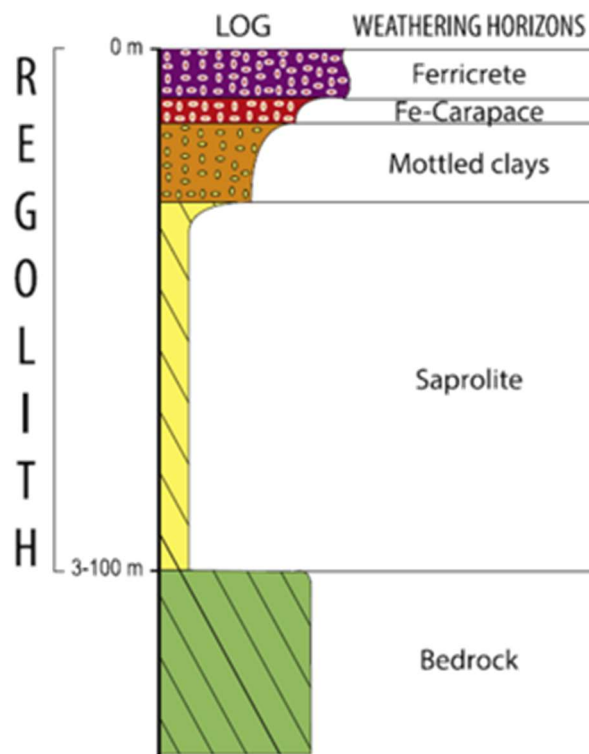


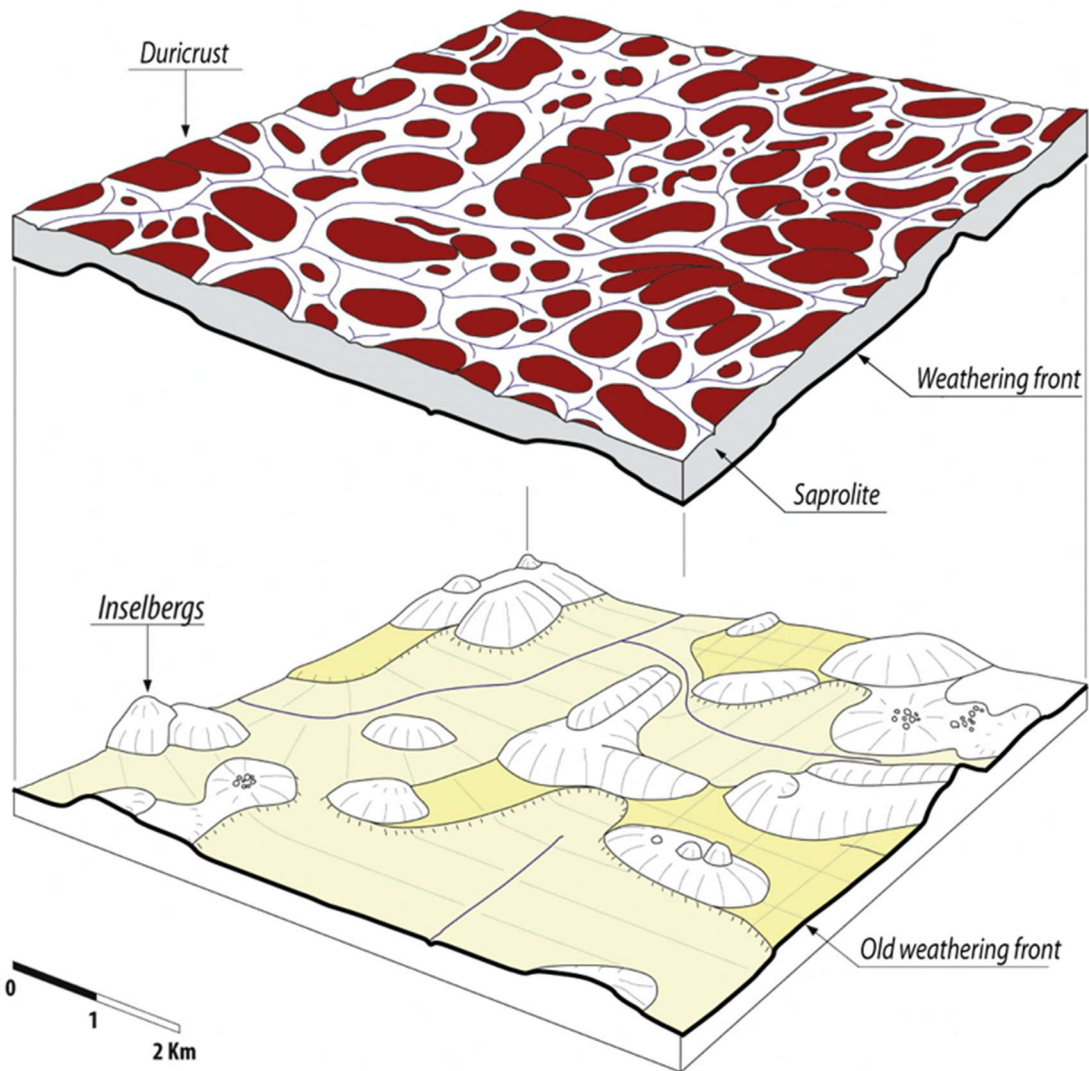
Figure 4 — Profil d'altération (Chardon et al., 2018)

En dessous de cette cuirasse et jusqu'au front d'altération se trouve une partie meuble composée de la carapace, des argiles tachetées ("Mottled clays", Figure 4) et du Saprolite (principalement composée de kaolinite).

Parmi ces surfaces de corrosion, plusieurs types qui correspondent à différents états d'érosion se distinguent (Figure 5).

- Les profils d'altération préservés, « *mantled etchplain* » ;
- Les profils d'altération partiellement décapés, « *partly stripped etchplain* » ;
- Les profils d'altération totalement décapés, « *stripped etchplain* ».

MANTLED ETCHPLAIN



Inselbergs

Weathering front

Saprolite

Old weathering front

0
1
2 Km

vertical exaggeration: x10

STRIPPED ETCHPLAIN

Figure 5 — Les différents types de surfaces de corrosions (Guillocheau et al., 2018)

I.2 État de l'art des méthodes de cartographies

Aujourd'hui, différentes méthodes existent pour cartographier les surfaces d'aplanissements (PS). Cependant, il s'agit de méthodes qui demandent beaucoup de temps (la méthode manuelle) ou bien qui ne donnent pas un résultat optimisé (méthode orientée pixel et de Haider et al., 2015).

I.2.1 Méthode manuelle

Cette première méthode consiste à cartographier les PS manuellement en dessinant les contours de ces surfaces sur un logiciel de dessin assisté par ordinateur ou dans un système d'information géographique. Pour cartographier ces surfaces, le géologue s'appuie sur les données d'élévation de la zone d'étude, les dérivées des modèles numériques de terrain (MNT) (pente, courbure, rugosité) et sur l'image satellite lorsque cela est possible. En effet, pour utiliser l'image satellite, il est nécessaire que la zone d'étude soit dégagée de toute végétation. Cela permet de distinguer plus aisément les formes du relief et d'identifier les différentes surfaces d'aplanissement.

Cependant, cette méthodologie est chronophage et des incertitudes peuvent persister comme le fait de ne pas être sûr de ce qui a été identifié. Pour une même zone, des résultats différents peuvent être ainsi obtenus en fonction de l'interprétation du cartographe comme le montre l'exemple suivant (cf. Figure 10).

I.2.2 Méthodes d'identification semi-automatique orientée pixel

Deux méthodologies semi-automatiques basées sur une démarche orientée pixel sont utilisées pour identifier les PS :

- L'indice de position topographique (TPI) ;
- La classification floue selon quatre paramètres.

I.2.2.1 Méthode utilisant l'indice de position topographique

Dans cette méthode, le TPI est calculé pour chaque pixel du MNT. Il correspond à la différence entre la valeur d'élévation du pixel et la moyenne des élévations des pixels se situant dans un voisinage déterminé. Le TPI dépend donc du rayon du voisinage.

La valeur du TPI obtenu nous indique alors dans quel type de relief nous nous trouvons, en fonction du rayon utilisé pour déterminer le voisinage (Figure 6). Un TPI nul ou quasi nul nous indique alors que nous avons une surface plane.

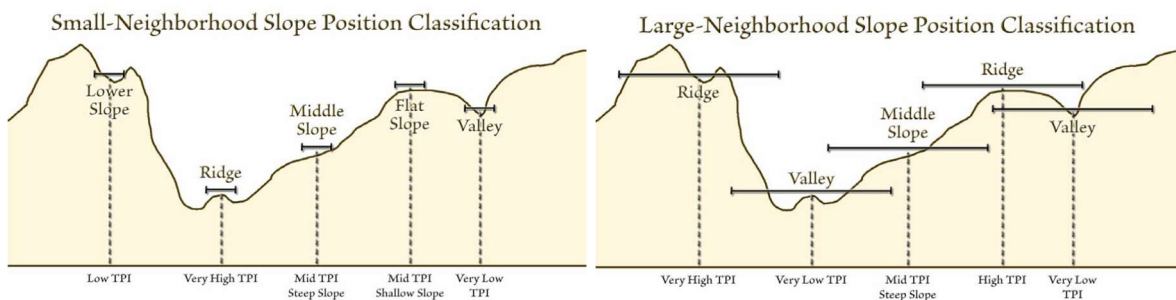


Figure 6 — Type de relief en fonction de la valeur du TPI et de la dimension du voisinage (Jenness, 2006)

Pour améliorer les résultats, l'écart type peut également être ajoutée (Bosh, 2016).

I.2.2.2 Méthode de classification floue selon 4 paramètres

Une autre méthode d'identification des PS est proposée dans un article scientifique par Haider (Haider et al., 2015). Cette méthode est basée sur une analyse morphométrique (sur la forme) à partir de MNT de la mission spatiale SRTM. C'est une méthode orientée pixel qui permet de donner le pourcentage d'appartenance à une surface d'aplanissement pour chaque pixel.

Cette méthode propose d'identifier les surfaces à partir de leur pente (SL), courbure (CU), rugosité (TRI) et leur hauteur relative (RH). Ces paramètres sont calculés à partir du MNT sous SIG (Figure 7).

Ensuite, pour chaque paramètre le degré d'appartenance à une PS est défini en fonction de sa valeur. Pour appartenir à 100 % à une PS :

- La pente (SL) du pixel doit être comprise entre 0° et 10°
- La courbure (CU) entre $-0,14$ et $0,14 \text{ m}^{-1}$
- La rugosité (TRI) entre 0 et 80 m
- La hauteur relative (RH) entre 100 et 600 m

Le pourcentage d'appartenance à une PS est calculé pour chaque paramètre. Pour terminer, les degrés d'appartenance de chaque paramètre sont multipliés pour obtenir le degré d'appartenance de chaque pixel à une PS.

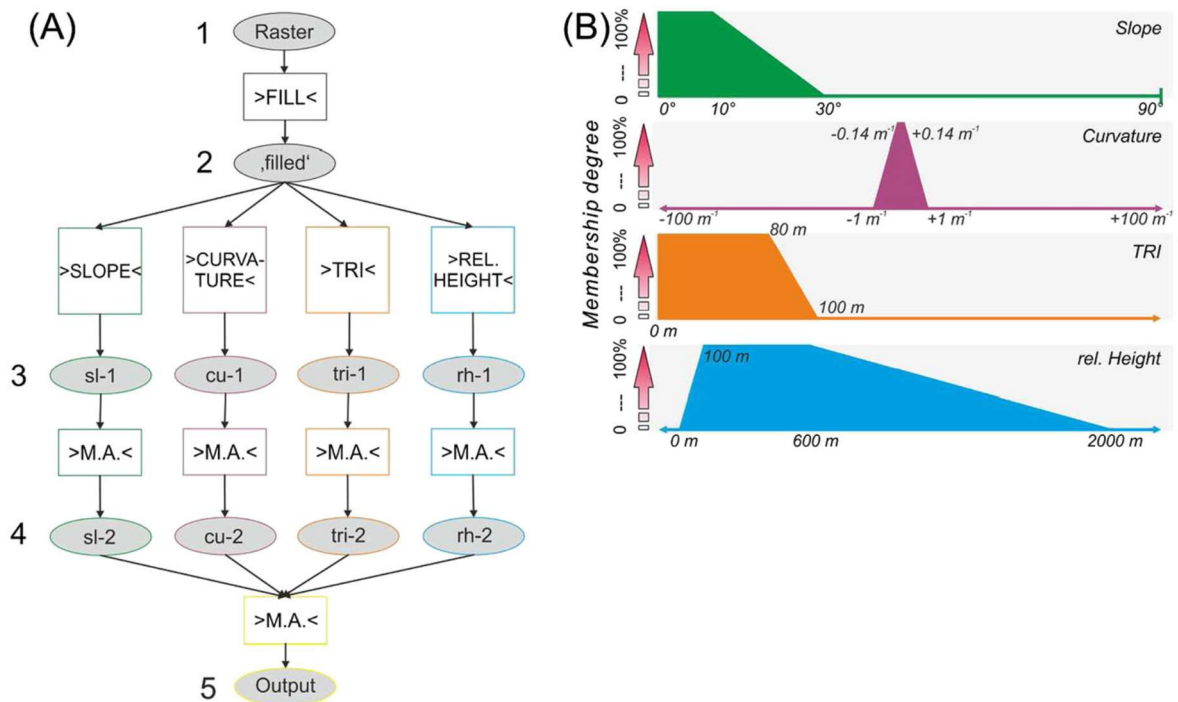


Figure 7 — Schéma illustrant la méthode d'Haider (Haider et al., 2015)

(A) Illustration de l'algorithme, 1 – entrées, 2 – remplissage des trous dans le raster, 3 – calculs de la pente ..., 4 – conversion des données en rasters appropriés avec des méthodes logiques floues et mathématiques, 5 – sortie finale après multiplication des rasters, (B) Degrés d'appartenance définis

En représentant les paramètres les uns en fonction des autres pour différents types de reliefs (montagne, péninsule ou surface d'aplanissement, et plateau), sous forme de nuage de

points (« *scatterplot* » ; Figure 8), les caractéristiques particulières permettant de différencier chaque type de relief permettent de définir les valeurs bornant l'appartenance à une PS.

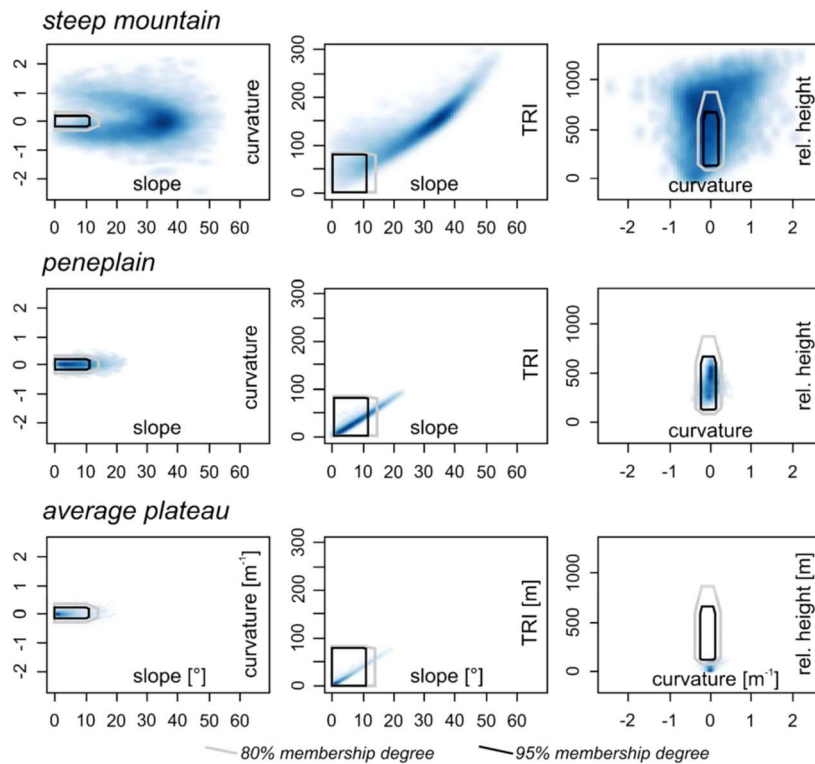


Figure 8 — Représentation des paramètres définissant différents types de relief extrait d'une zone d'étude au Tibet (Haider et al., 2015)

I.2.3 Autres méthodes

Deux autres méthodes sont proposées par Jarsve et al., 2014 (Jarsve et al., 2014), pour identifier les PS. La première méthode consiste à calculer, par moindres carrés, la surface qui s'ajuste le mieux à la zone étudiée à partir des coordonnées x , y , z du MNT. Ensuite les PS sont identifiées en fonction de la valeur de leur pente.

La seconde méthode utilise un algorithme de croissance de régions à partir de points choisis aléatoirement qui formeront les germes des régions. Les pixels voisins sont ajoutés à la région lorsque des critères sont respectés. Deux critères sont utilisés : (i) la différence de pente entre pixels voisins est inférieure à une valeur seuil T_a et (ii) la différence de pente entre un pixel et le germe de la région est inférieure à un seuil T_s . Les PS correspondront alors aux régions générées.

I.3 Domaines d'application en géosciences

L'étude des PS présente des applications dans deux domaines principaux. Le premier concerne les propriétés réservoirs des roches sous-jacentes aux PS tandis que le second concerne la reconstruction de l'évolution de la déformation des surfaces continentales.

I.3.1 Géologie, réservoir et hydrologie

Les PS occupent une grande partie des surfaces continentales et sont régulièrement associées ou affectées par des profils d'altération.

Ces profils d'altération sont organisés en couches de compositions minéralogiques (Figure 4) différentes présentant des capacités de perméabilité et de porosité différentes (Figure 10). Ces couches étant perméables, elles peuvent capturer l'eau et entraîner la création d'un réservoir hydrologique (Figure 9 – Zone 'd').

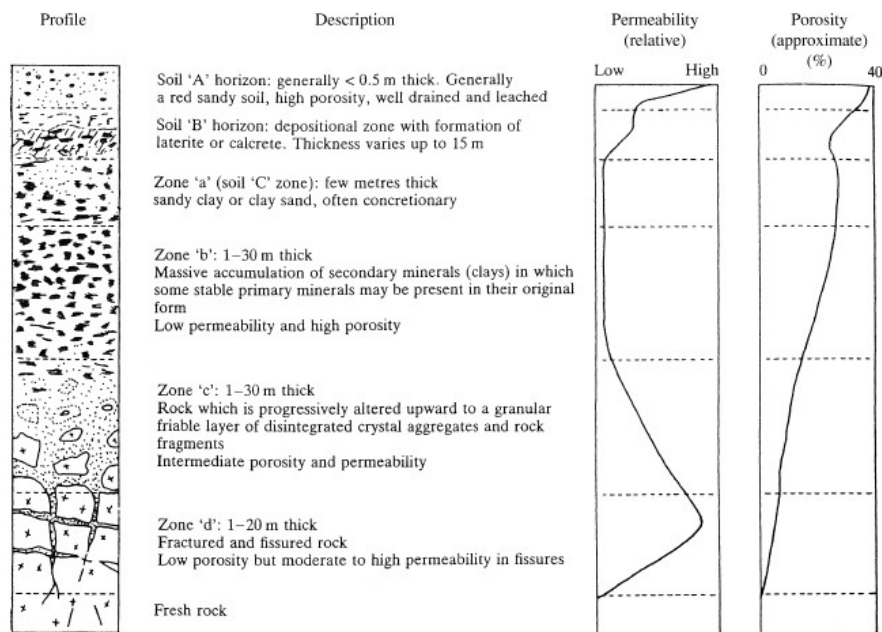


Figure 9 — Profil d'altération, perméabilité et porosité associée (Lachassagne et al., 2011)

D'un autre point de vue, cette perméabilité du substratum sous-jacent à certaines PS pourrait permettre de capter et retenir du pétrole. Il y a donc un intérêt économique d'identifier les surfaces d'aplanissement et de comprendre leurs conditions de formation et déformation.

I.3.2 Traçage de l'évolution de la déformation des surfaces continentales

L'identification des pédiments permet de tracer l'évolution de la déformation des surfaces continentales (Chardon et al., 2018).

La présence de pédiment a fortement contribué à l'érosion d'un épais manteau de régolithe (la partie du sol recouvrant la roche) engendrant ainsi un nivellement de ces surfaces et créant une diversité dans les pédiments. Cette diversité s'expliquant par la nature et l'intensité de l'érosion ainsi que du durcissement. De plus, la cartographie ainsi que la reconstitution des différentes étapes paysagères des pédiments fournissent un guide permettant d'interpréter les anomalies géochimiques des pédiments et permettent de retracer leurs sources potentielles.

L'étude des zones climatiques passées et présentes suggère un gradient dans le processus de pédimentation (formation des pédiments) entre une zone aride et une zone humide. Ce qui explique pourquoi les pédiments ont été négligés dans les zones humides. Les pédiments ont également permis la conservation d'éléments résiduels qui nous donnent des informations sur la déformation de la surface continentale et sont aussi indicateur d'une érosion très lente. Ils sont aussi des marqueurs de la morphogénèse passée et un outil pour comprendre la dynamique de surface.

I.4 Choix d'une zone test : les abords du fleuve Orange (frontière Afrique du Sud/Namibie)

Pour développer une nouvelle méthode d'identification des PS nous nous appuyerons sur les travaux d'identification des PS déjà réalisés sur les abords du fleuve Orange, à la frontière entre l'Afrique du Sud et la Namibie. En effet, les PS de cette zone sont connues depuis de nombreuses années (Dauteuil et al., 2015 ; Figure 10 - A) et une carte de ces surfaces a été récemment publiée (Baby, 2017 ; Figure 10 - B). De plus, cette zone présente l'avantage de ne pas être recouverte de végétation, ce qui facilite l'identification des PS.

Cette zone d'étude soulève également les problèmes des différentes méthodes de cartographie actuelle. Avec un œil averti, il est apparent que les deux cartes (Figure 10) ne se superposent pas, les surfaces d'aplanissement identifiées ne sont pas les mêmes. Ce qui montre la difficulté d'identifier les surfaces d'aplanissement.

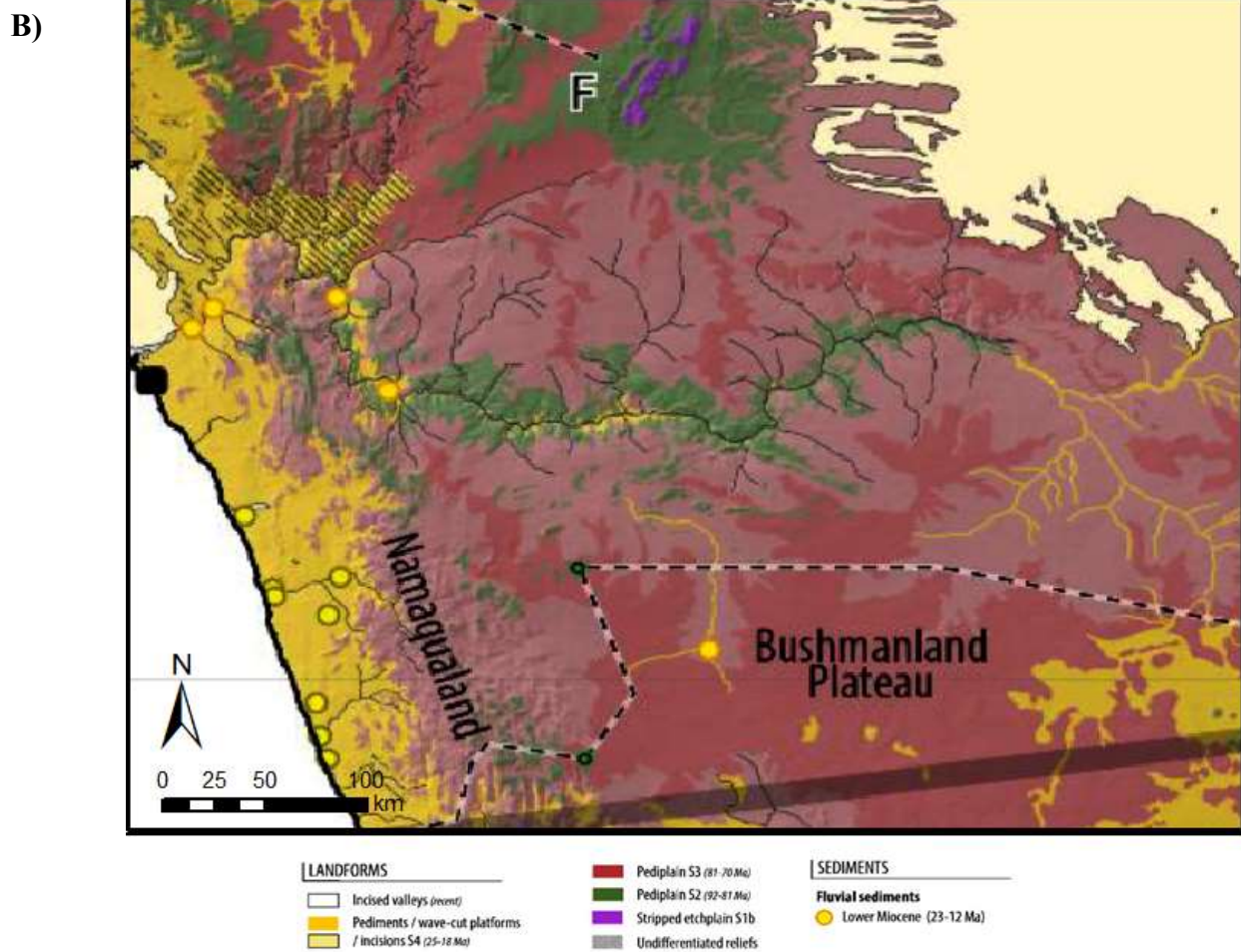
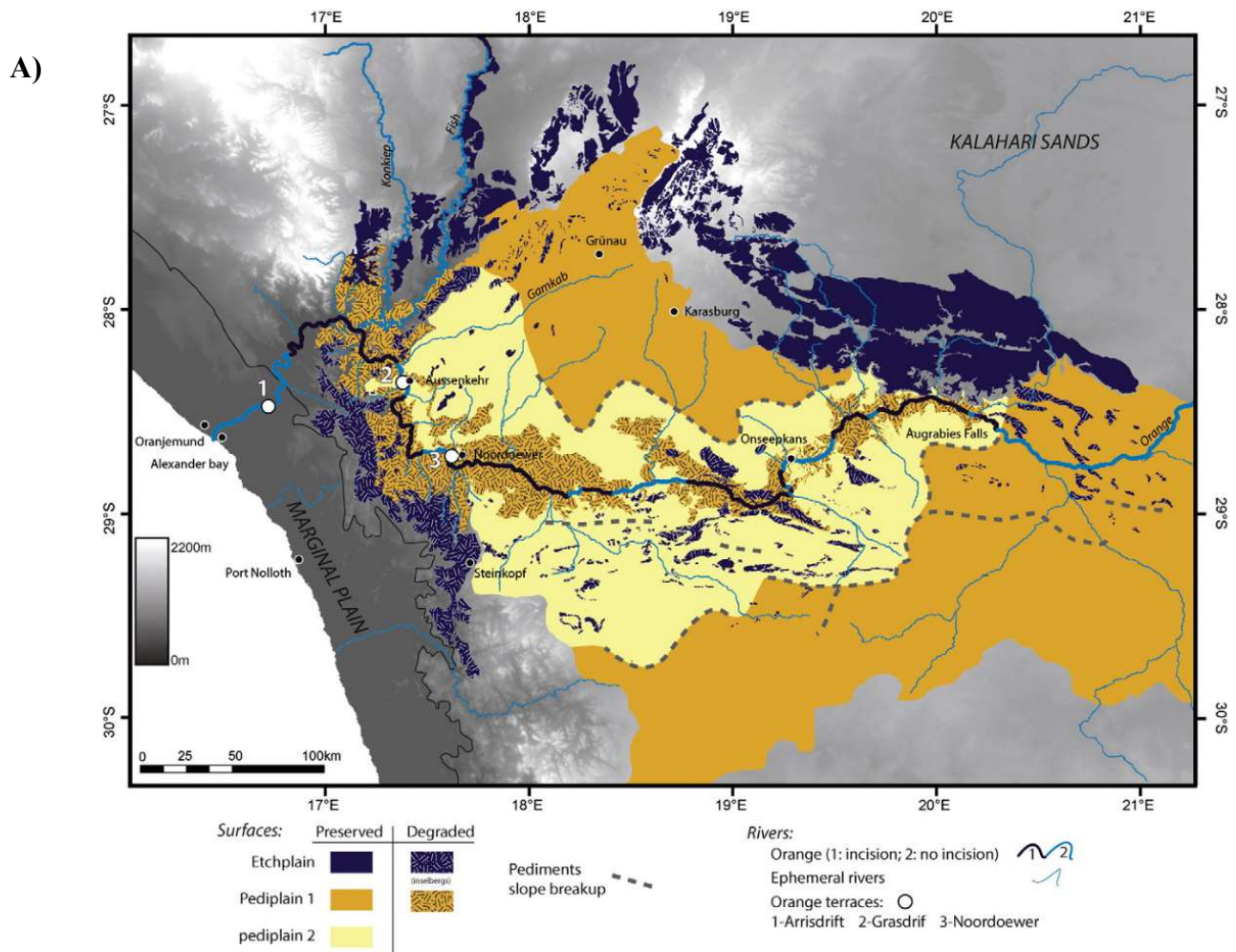


Figure 10 — Surfaces d'aplanissement au sud de l'Afrique (A – Dauteuil et al., 2015 ; B – Baby et al, 2017)

II Rappels sur les méthodes de segmentation et de classification

Dans cette partie un rappel sur différentes méthodes de segmentation et classification est fait. Tout d'abord, je vais commencer par présenter les différents logiciels et méthodes qu'ils intègrent. Dans un second temps, je vais expliquer le fonctionnement des différents algorithmes et ensuite réaliser une comparaison de ces derniers.

II.1 Logiciels et méthodes intégrées

De nombreux logiciels ont des méthodes de traitement d'image intégrées telles que la segmentation et la classification. Des logiciels payants comme ENVI, eCognition de Trimble, ou encore ArcGIS de ESRI, existent mais aussi des logiciels gratuits, libres de droits ou en open source comme SAGA GIS, ImageJ, GRASS et QGIS.

En raison du coût de certains logiciels et du temps de prise en main, les logiciels ArcGIS, ENVI, et SAGA GIS ont été privilégiés, combinés à l'utilisation du langage de programmation Python et de ses différentes bibliothèques dont Scikit-learn, pour réaliser ces traitements d'image.

La segmentation d'une image consiste à regrouper des pixels, ayant des caractéristiques similaires, en objet (Figure 11). Les groupes de pixels forment des objets qui constituent une partition de l'image et différents logiciels proposent cette opération en utilisant différents algorithmes (Tableau 1).

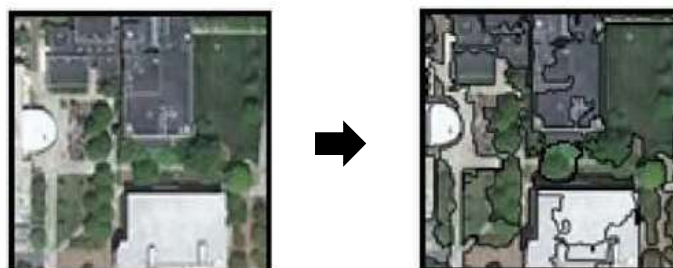


Figure 11 — Illustration de la segmentation (Chen et al.,2018)

Logiciel	ArcGIS	ENVI	SAGA GIS
Algorithme	Mean Shift	Watershed	Seeded Region Growing

Tableau 1 — Algorithmes de Segmentation pour chaque logiciel

La classification attribue à chaque objet un label, les objets ayant le même label appartiennent ainsi à la même classe (Figure 12). Beaucoup d’algorithmes de classification existent, j’ai fait le choix de me restreindre à quelques-uns, tout en m’assurant de choisir tant des algorithmes de classification supervisée que des algorithmes de classification non supervisée. J’ai retenu les algorithmes que les logiciels choisis proposent, des algorithmes simples et performants (Tableau 2).



Figure 12 — Illustration de la classification (Chen et al., 2018)

Logiciel	ArcGIS	ENVI	SAGA GIS	Python
Algorithmes	Random Forest	K-Means Isodata	K-Means Isodata Random Forest	K-Means Random Forest

Tableau 2 — Algorithmes de classification pour chaque logiciel

II.2 Fonctionnement des algorithmes

II.2.1 Segmentation

Mean Shift (Comaniciu & Meer, 2002)

L’algorithme "Mean Shift" fonctionne sur le principe d’une fenêtre glissante et un processus itératif.

Un ensemble de fenêtres est distribué sur le raster en entrée. Puis, les étapes suivantes sont réalisées pour chaque fenêtre (Figure 13):

1. Calcule la moyenne des valeurs de la fenêtre.
2. La fenêtre glisse jusqu’à cette valeur moyenne.
3. Ces deux étapes sont répétées jusqu’à convergence de la fenêtre.

Ensuite, les fenêtres ayant la même moyenne sont regroupées au sein d'un même objet.

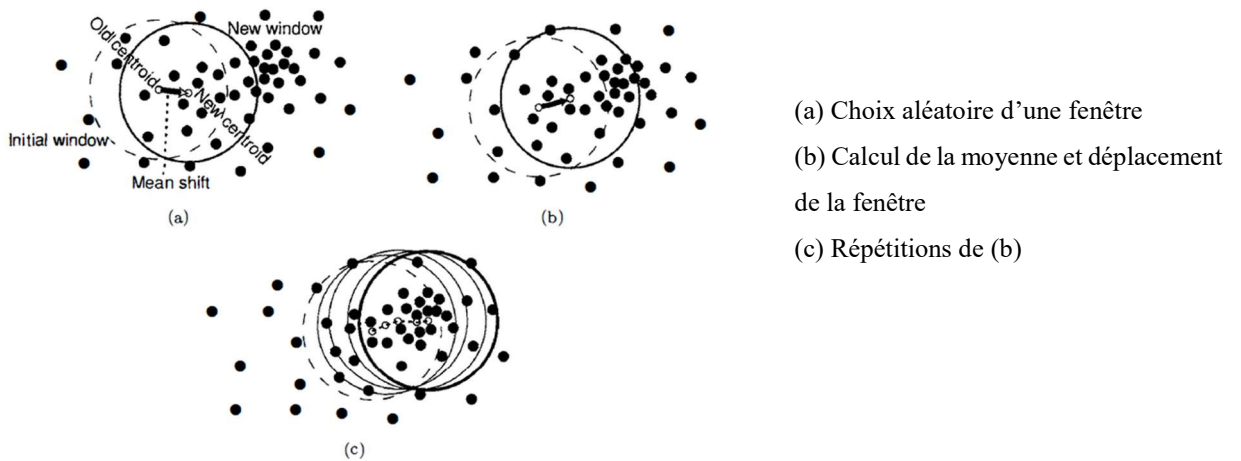


Figure 13 — Illustration de l'algorithme Mean Shift (Abrishami, 2015)

Seeded Region Growing (Adams & Bischof, 1994)

Il s'agit de régions qui grandissent lorsqu'un prédicat est vrai (Figure 14). Tout d'abord, des pixels sont choisis aléatoirement, ce sont des germes (« seed »). Ensuite, pour chaque pixel dans le voisinage du germe, la différence de valeur entre ce dernier et le germe est calculée. En fonction de cette différence et du prédicat, les pixels voisins à la région du germe sont attribués ou non à cette région. Ces étapes sont ensuite répétées jusqu'à ce que chaque pixel soit attribué à une région.

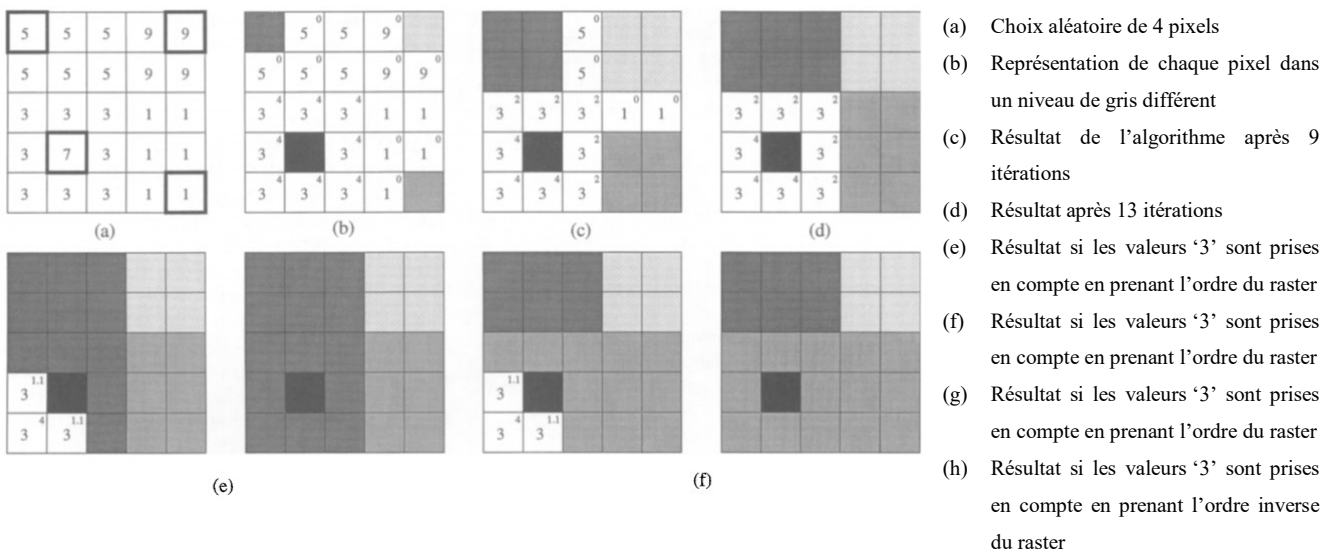


Figure 14 — Illustration de l'algorithme Seeded Region Growing (Mehnert, 1997)

Watershed (bassin versant) (Vincent & Soille, 1991)

Ici, la segmentation part de ce principe de bassin versant et cherche à les créer (Figure 15). Dans cet algorithme l'image est considérée comme des niveaux de gris, chaque pixel a une valeur que l'on assimile à une altitude. L'objectif est de calculer les lignes de partage des eaux de ce relief en « inondant » ce dernier, cela formera ensuite des bassins faisant apparaître les lignes de partage des eaux, qui deviendront nos différents objets.

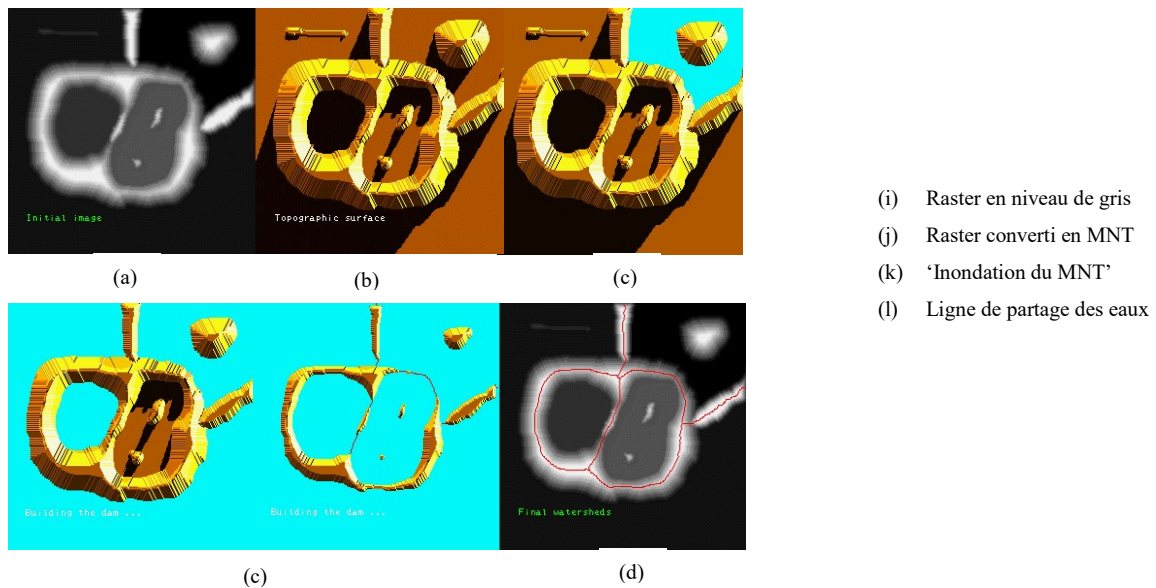


Figure 15 — Illustration de l'algorithme Watershed (Beucher, 2010)

II.2.2 Classification

Il existe deux types de classification, la classification non supervisée et la classification supervisée. Dans la classification non supervisée, l'algorithme détermine lui-même les différentes classes (K-Means et Isodata). Tandis que dans la classification supervisée, c'est l'utilisateur qui définit les classes grâce à des échantillons d'apprentissage (Random Forest).

K-Means

Pour classer les objets, K-Means choisit aléatoirement K objets, K étant défini par l'utilisateur, qui seront les centres de classes. Ensuite, l'algorithme calcule la « distance » de chaque objet à chaque centre de classe. L'objet est ensuite attribué à la classe dont il est le plus proche. Le centre de chaque classe est recalculé. Puis cette étape est réitérée pour chaque objet jusqu'à ce qu'ils soient tous classés (Figure 16).

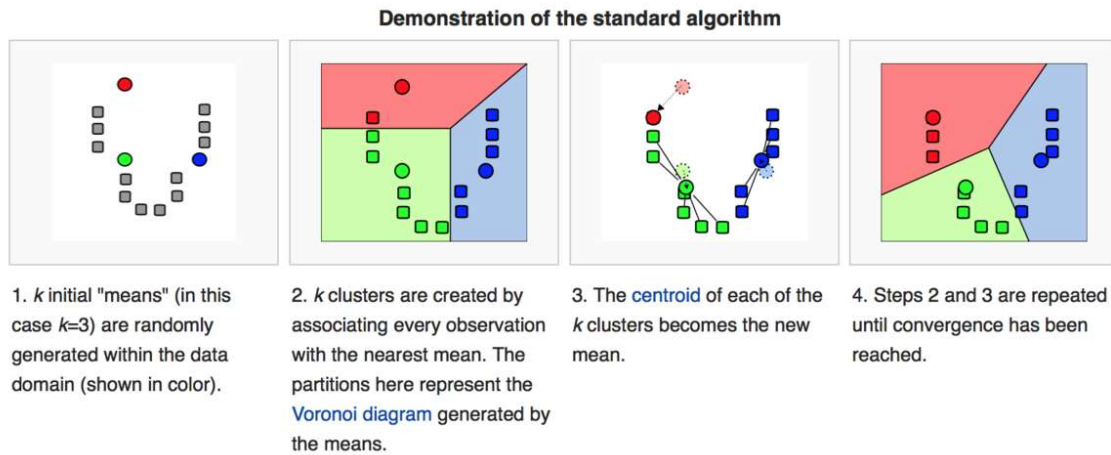


Figure 16 — Illustration de l'algorithme K-Means (Landman et al., 2019)

Isodata (Memarsadeghi et al., 2007)

Isodata est un algorithme basé sur K-Means, mais il comporte des étapes supplémentaires. Ici le nombre K n'est pas donné par l'utilisateur, mais il donne K initial, ensuite l'algorithme va déterminer K . Après avoir utilisé K-Means, l'algorithme utilise deux traitements. Un traitement de fusion, il fusionne les classes qui sont similaires à partir d'un certain seuil. Et à l'inverse, il divise les classes dont le seuil de ressemblance n'est pas assez élevé. C'est ainsi que l'algorithme détermine lui-même le nombre de classe K .

Random Forest (Breiman, 2012)

Random Forest est un algorithme de classification supervisée basé sur des arbres décisionnels. À partir des échantillons d'apprentissage, l'algorithme crée un ensemble d'arbres de décision (Figure 17). À chaque objet est associée la classe pour laquelle les arbres de décisions donnent la majorité.

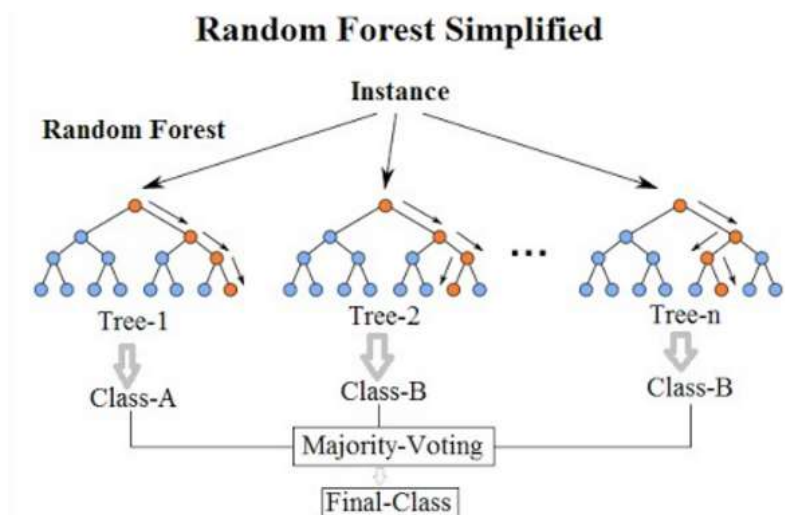


Figure 17 — Illustration simplifiée de l'algorithme Random Forest (Koershen,

II.3 Comparaison des algorithmes

Maintenant que nous avons vu les différents logiciels et algorithmes de segmentation/classification, il est important de comparer ces derniers afin de déterminer leur potentiel et leur pertinence pour la cartographie des surfaces d'aplanissement.

II.3.1 Segmentation

Pour comparer les différents algorithmes de segmentation, je me suis donc basé sur la même donnée raster en entrée, le MNT d'une partie de ma zone d'étude (Figure 18 – A a). Le Tableau 3 – A résume les différentes entrées, sorties et les paramètres utilisés par les logiciels pour chaque algorithme.

La différence entre les résultats se trouve principalement dans la taille des segments. Après plusieurs tests, j'ai remarqué que la dimension des segments peut varier en fonction des paramètres d'entrée, mais les résultats restent similaires. Pour le résultat d) beaucoup de polygones sont présents, ce qui correspond à un découpage régulier de l'image. Dans les résultats b) et c) les polygones correspondent à des zones uniformes du MNT, ce qui correspond à la morphologie de résultat recherché.

II.3.2 Classification

Comme pour la segmentation les paramètres en entrée pour la classification diffèrent. Ici, pour réaliser la comparaison, j'ai classifié une image satellite d'un extrait de la zone d'étude par rapport à ses couleurs (Figure 18 – B a). Cela permet de mieux observer la classification que sur un MNT où les classes obtenues correspondraient à des zones de mêmes altitudes. La classification est basée sur les couleurs, d'un côté les couleurs claires/oranges et de l'autre côté les couleurs sombres/noires. C'est de cette manière que les échantillons d'apprentissage ont été définis. Le Tableau 3 — B résume les paramètres utilisés pour chaque algorithme. Les résultats sont assez similaires, à part pour Isodata (c), une classe supplémentaire est obtenue. Les différences entre ces résultats sont principalement localisées au niveau de la limite des clusters. Il est donc difficile de choisir quel algorithme utilisé. Ce sont les aspects pratiques, le fonctionnement ainsi que la précision qui vont être déterminants (cf V Discussion).

Segmentation (A)

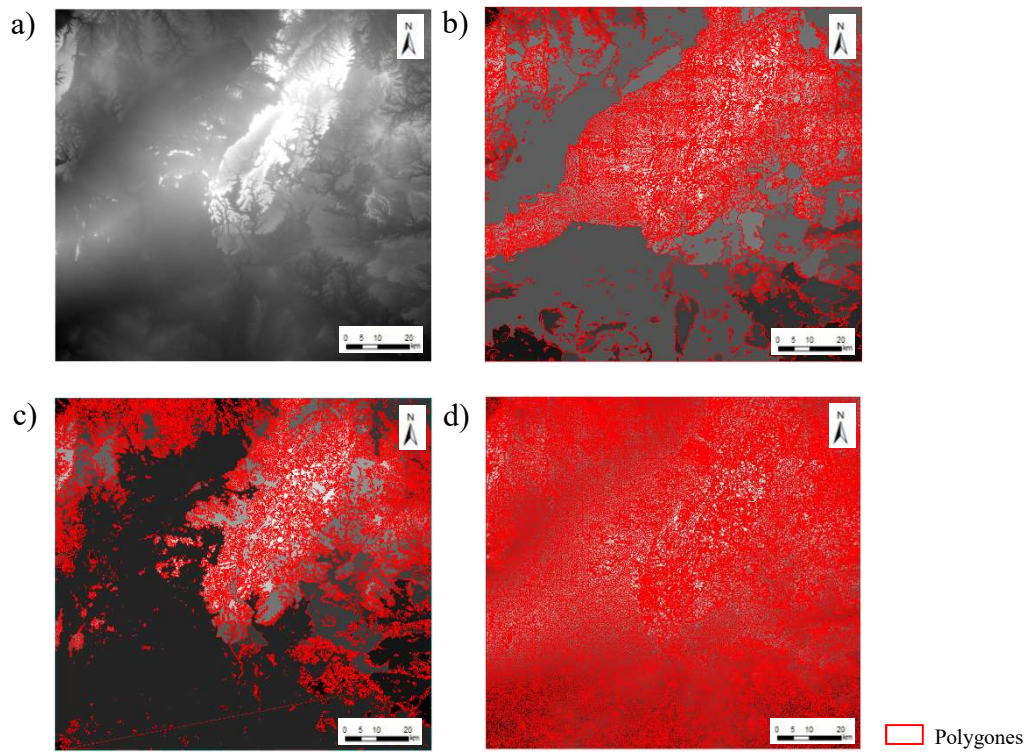
Algorithme	Mean Shift	Watershed	Region Seeded Growing
Logiciel	ArcGIS	ENVI	Saga GIS
Entrée	Raster		
Paramètres	Détail spectral : niveau d'importance donné aux différences spectrales des entités dans votre imagerie Détail spatial : niveau d'importance donné à la proximité entre des entités dans votre imagerie Nombre minimum de pixels dans un objet	Facteur d'échelle Niveau de fusion des polygones	Nombre de germes
Sortie	Raster	Raster Polygones	Polygones
Figure 19 — A	(b)	(c)	(d)

Classification (B)

Algorithme	K-Means	Isodata	Random Forest
	Non Supervisé		Supervisé
Logiciel	Saga GIS Python	Saga GIS	Arcgis Python
Entrées	Raster K (nombre de classe)	Raster $K_{initial}$ K_{max}	Raster Échantillons d'apprentissage (d)
Sorties	Raster classé		
Figure 19 — B	(b)	(c)	(e)

Tableau 3 — Comparaison des algorithmes

Segmentation (A)



Classification (B)

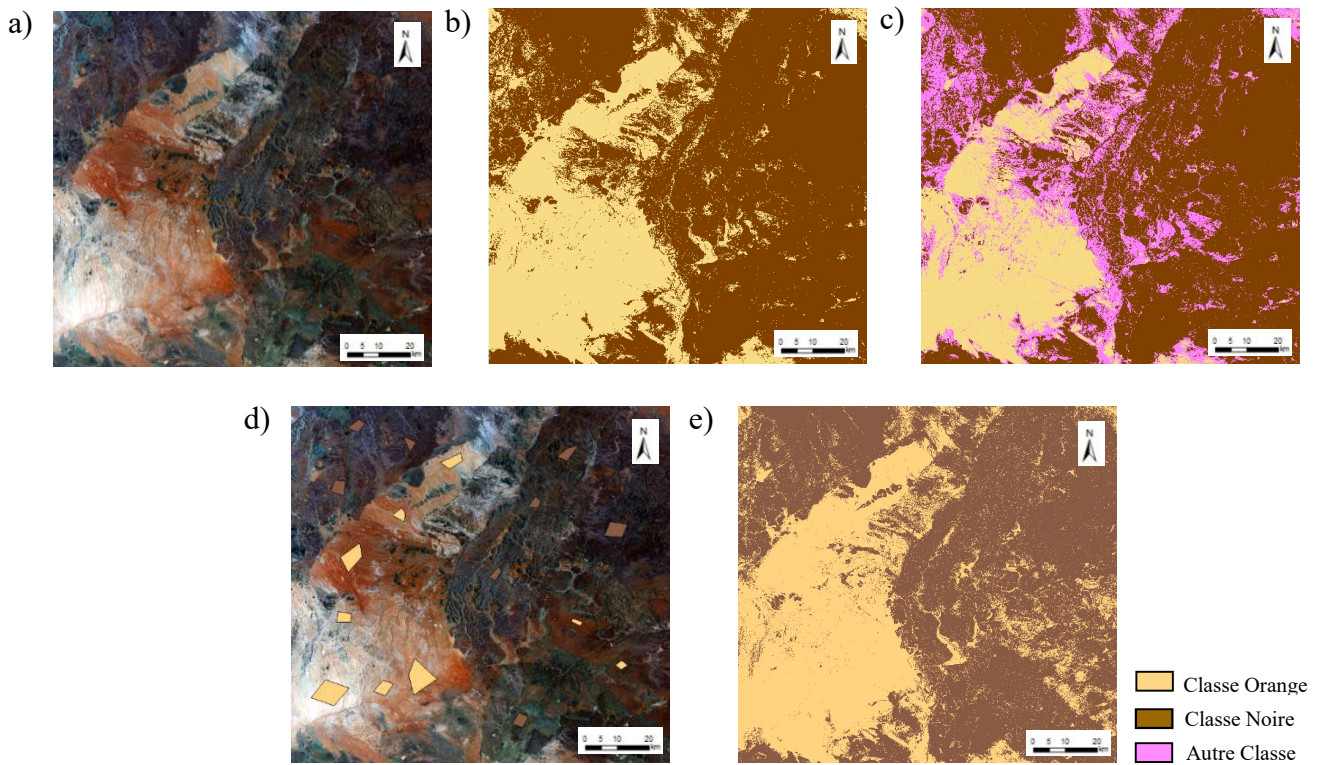


Figure 18 — Comparaison des résultats de segmentation et classification
 (A) – a) MNT avant segmentation, b) segmentation avec Mean Shift, c) segmentation avec Watershed, d) segmentation avec Region Seeded Growing
 (B) – a) Image avant classification, b) classification par K-Means, c) classification par Isodata, d) échantillons d'apprentissages définis, e) classification par Random Forest

II.3.3 Choix des algorithmes

Pour déterminer les algorithmes utilisés par la suite, j'ai dressé un tableau des avantages et inconvénients pour chaque algorithme de segmentation (Tableau 4) et de classification (Tableau 5).

Algorithme	Mean Shift	Watershed	Region Growing
Avantage(s)	Présent dans SIG Présent dans ArcGIS Sortie raster Segmentation attendue du MNT	Segmentation attendue du MNT	Présent dans SIG
Inconvénient(s)		Dans un logiciel de traitement d'image Sortie vecteur Nombreux bugs	Sortie vecteur Beaucoup de petits segments
Temps de calcul	De 5 à 60 minutes	< 30 minutes	De 5 à 30 minutes

Tableau 4 — Avantages et inconvénients pour la segmentation

Algorithme	K-Means	Isodata	Random Forest
Avantage(s)	Présent dans un SIG	Présent dans un SIG Détermine K	Dans ArcGIS Présent dans un SIG Basé sur le machine learning
Inconvénient(s)	Besoin d'indiquer un K	Besoin d'un K initial et un K max	Besoin d'échantillons d'apprentissage
Temps de calcul	< 5 minutes	< 15 minutes	< 1 minute

Tableau 5 — Avantages et inconvénients pour la classification

Pour établir les protocoles d'identification des PS, j'ai ainsi retenu Mean Shift pour la segmentation. L'avantage majeur de cet algorithme est qu'il est présent dans les logiciels de SIG, notamment dans ArcGIS, logiciel utilisé dans le laboratoire et plus largement dans la communauté des géosciences.

Pour la classification, les algorithmes K-Means et Random Forest sont retenus, ce sont ceux qui présentent le plus d'avantages. Cela permettra la comparaison des deux classifications, une classification non supervisée et la deuxième basée sur l'apprentissage.

III Données utilisées et protocoles élaborés

III.1 Choix des données

Pour déterminer les données qui seront utiles, je me suis basé sur les données déjà utilisées par les géologues et géomorphologues. Ainsi j'ai retenu le MNT avec une résolution de 90 m et ses dérivées : pente (SL), courbure (CU), ainsi que la rugosité (RU) et hauteur relative (RH). De plus, une autre donnée a été ajoutée, il s'agit de la texture (TE), ce descripteur permet de mettre en avant la présence de motifs à partir du MNT. Cette donnée n'a pas été utilisée auparavant, mais pourrait se trouver intéressante à prendre en compte. En effet, les PS pourraient avoir une texture particulière qui permettrait de les identifier.

Avant d'appliquer les protocoles définis ci-après deux zones ont été définies et serviront de zone test afin de limiter les temps de calcul et de tester les protocoles élaborés.

La zone d'étude sur laquelle je me suis principalement appuyé est située à la frontière entre Namibie et Afrique du Sud. J'ai aussi travaillé sur une zone de contrôle située dans le nord de la Patagonie argentine pour confronter mes résultats.

III.1.1 Namibie/Afrique du Sud

Au sein de cette zone d'étude africaine (ZA ;Figure 19), j'ai découpé deux zones tests plus petites. La Zone Test 1 (ZT1) qui se situe au nord de la zone d'étude et la Zone Test 2 (ZT2)

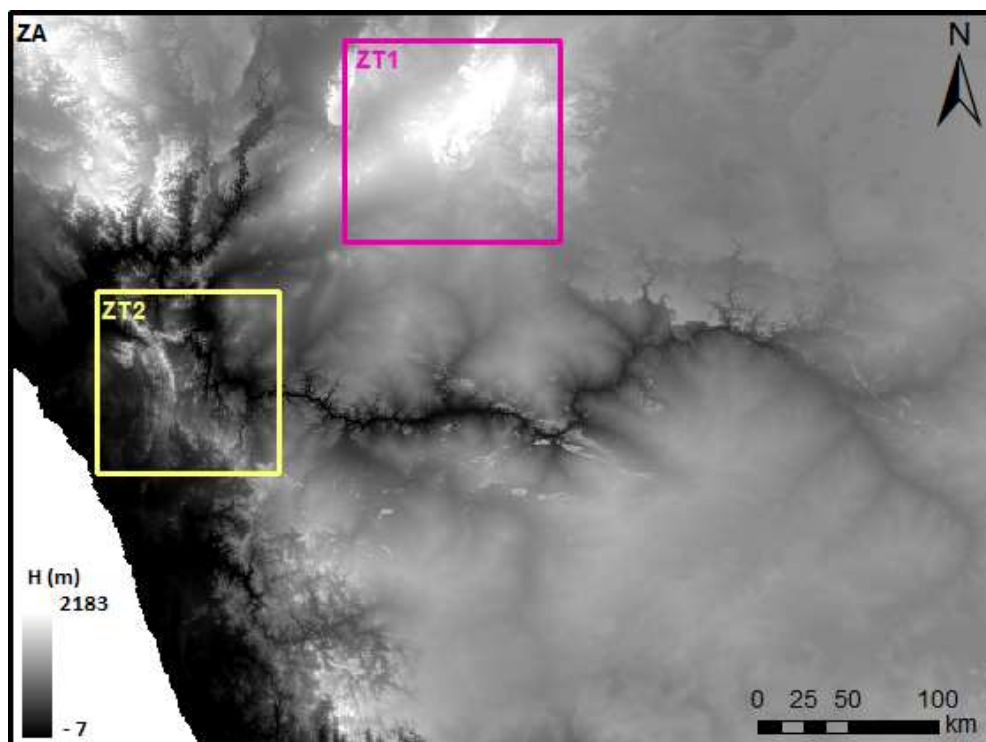


Figure 19 — Zone d'étude africaine (ZA)

se situant à l'est de la zone d'étude. Ces deux zones tests sont composées chacune des deux types de surfaces que nous cherchons à identifier : pédiments et etchplains (Figure 10 – A).

III.1.2 Argentine

La zone d'étude argentine (ZAR ; Figure 20) a été utilisée entièrement, sans création de zones de tests.

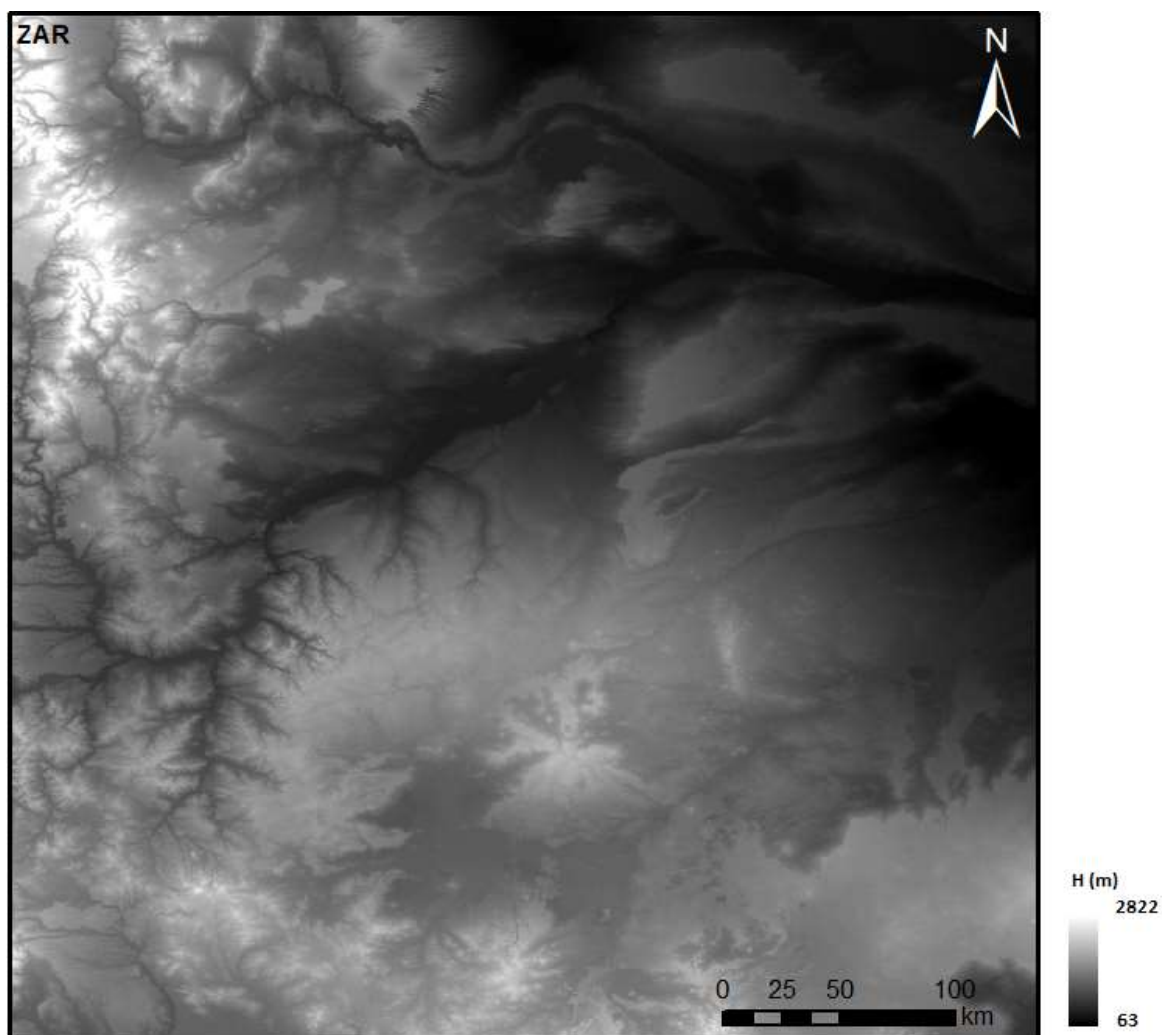


Figure 20 — Zone d'étude argentine (ZAR)

III.2 Protocoles élaborés

Pour élaborer mes protocoles, je me suis basé sur l'OBI (Blaschke et al., 2014) qui correspond à une chaîne de traitement. Tout d'abord, il y a un prétraitement des données puis les données sont segmentées avant d'être classifiées. Enfin, les résultats sont analysés. Deux protocoles ont été élaborés : un premier protocole utilisant une classification non supervisée et un deuxième protocole basé sur une classification supervisée. Les deux protocoles ont le même « squelette » et seule l'étape de classification diffère.

Ainsi, la première étape de ces deux protocoles consistera à calculer les données utiles : la pente, la courbure, la rugosité et la hauteur relative. Ces descripteurs sont calculés à partir des logiciels de SIG (ArcGIS et/ou QGIS). (E1). Puis dans une deuxième étape, chaque donnée est segmentée individuellement grâce à l'algorithme Mean Shift (ArcGIS). (E2). Une fois toutes les segmentations réalisées, ces dernières formeront une image multi bandes (E3) avant l'étape de classification et d'analyse propre à chaque protocole.

III.2.1 Protocole 1 : Identification des PS par classification non supervisée (IPS-NS)

Dans ce protocole, les étapes citées précédemment (E1 à E3) sont réalisées. La classification, non supervisée est appliquée à l'image multibande, avec l'algorithme K-Means. K, correspondant au nombre de classes souhaitées que l'utilisateur détermine lui-même (Figure 21 - A).

La classification nous donne K classes, chacune avec des statistiques différentes. Ce qui nous permet d'identifier les PS grâce à la réalisation de nuage de points.

III.2.2 Protocole 2 : Identification des PS par classification supervisée (IPS-S)

Dans ce deuxième protocole (IPS-S) les étapes E1 et E3 sont exécutées comme dans le protocole précédent. La classification supervisée sera réalisée à partir de l'algorithme Random Forest (Figure 21 – B). Contrairement au protocole précédent, il faudra créer des échantillons d'apprentissage en parallèle de cette chaîne de traitement afin de pouvoir utiliser Random Forest.

La classification nous donne alors des entités classifiées avec les classes données aux échantillons d'apprentissage. Les entités ont des propriétés similaires aux échantillons. Les calculs de matrices de confusion vont alors nous aider à analyser les résultats.

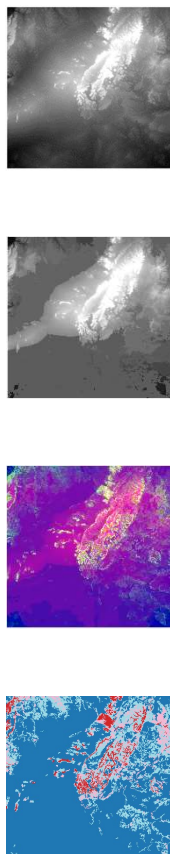
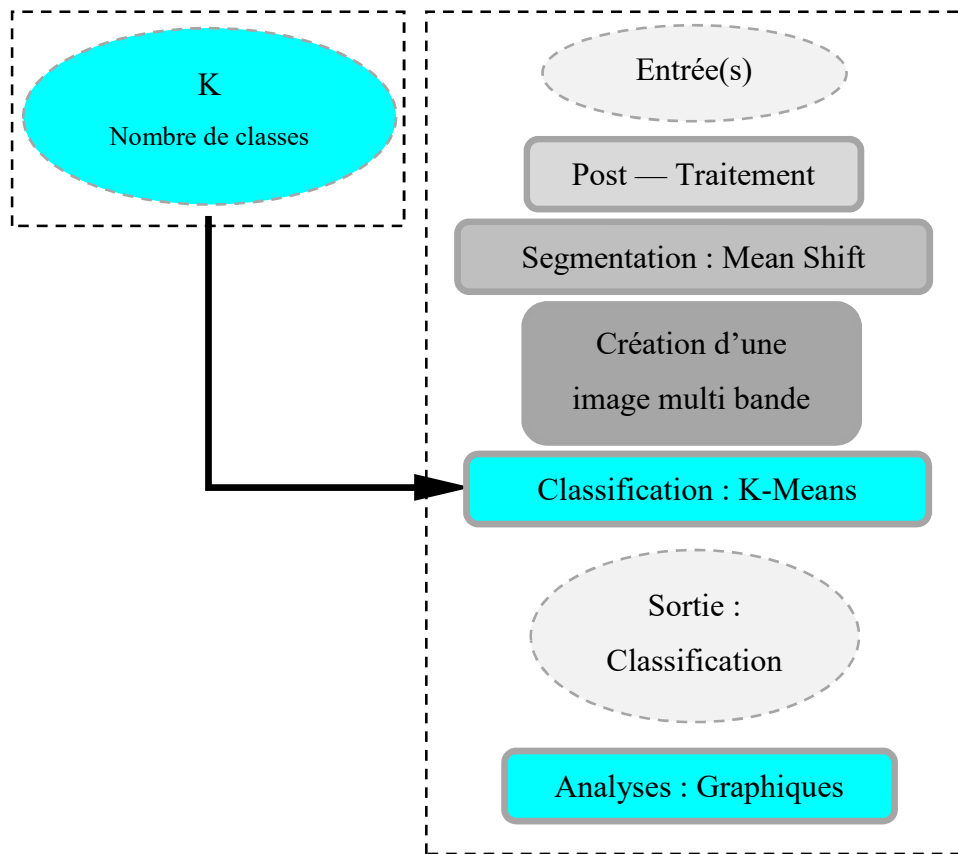
Note : *Un tutoriel permettant de mettre en œuvre ces deux protocoles a été rédigé et est disponible sur demande.*

Protocole non supervisé

A)

IPS NS

Méthode objective



Protocole supervisé

B)

IPS S

Méthode d'apprentissage automatique

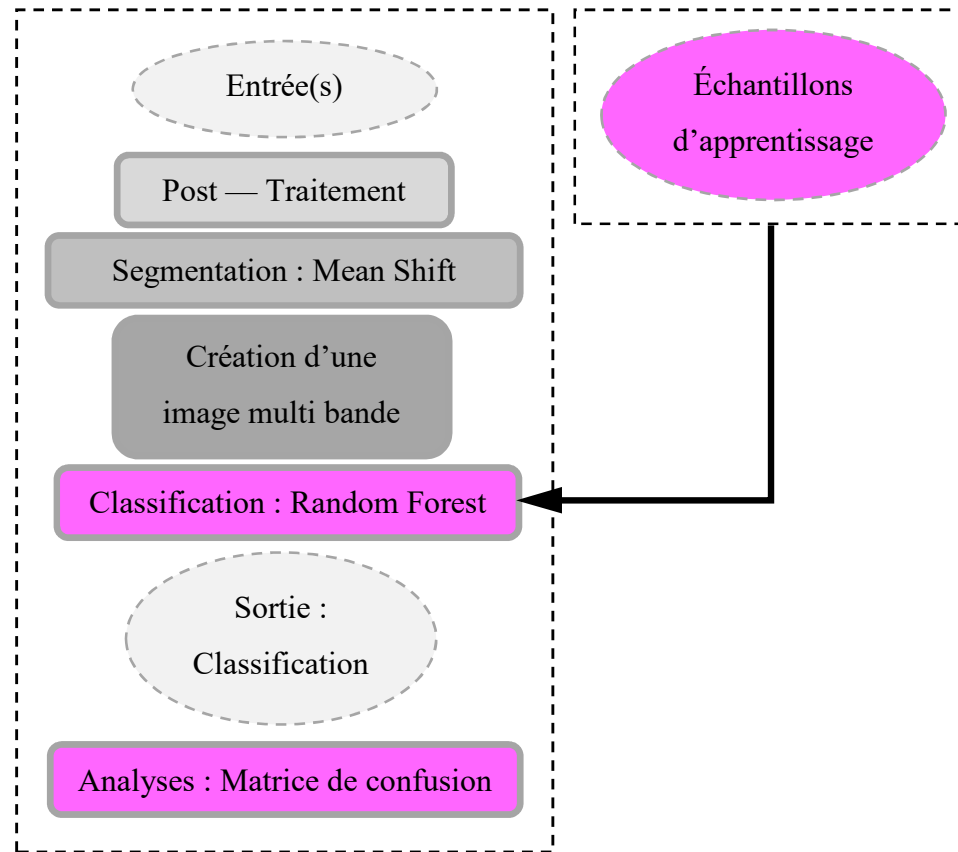


Figure 21 — Illustration des protocoles élaborés

IV Résultats

Dans cette avant-dernière partie, les résultats des différentes étapes relatives aux protocoles élaborés seront présentés. Les deux protocoles ont été développés et testés tout d'abord sur les zones tests (ZT1 et ZT2) extraites de la zone d'étude africaine (résultats présentés en Annexe 1) puis sur la zone d'étude africaine entière (ZA) puisqu'il s'agit d'une zone bien connue des géologues ou de nombreux écrits ont été publiés. Le même travail est ensuite appliqué sur l'Argentine pour confronter les solutions proposées.

Pour vérifier les résultats obtenus, ceux-ci seront comparés avec la cartographie existante proposée par Dauteuil et al. (2015). De plus, les précisions obtenues sur les résultats seront présentées.

IV.1 Zone de développement des protocoles : la Namibie/Afrique du Sud

Les deux protocoles proposés ont été testés pour différentes combinaisons de descripteurs en entrée : (1) H SL CU RU, (2) SL CU RU RH, (3) SL CU RU, (4) SL CU, (5) SL CU Texture (TE) et (6) SL CU sans le réseau hydrographique (SH). L'objectif de tester ces combinaisons de descripteurs est de déterminer ceux qui sont optimaux pour identifier les PS.

Pour chaque combinaison, les descripteurs ont été calculés, segmentés puis l'image multibande générée (cf. Figure 21). Ensuite, les paramètres nécessaires pour chaque algorithme de classification ont été déterminés : (1) K pour IPS NS, (2) les classes correspondant à des PS suite à la classification et (3) les échantillons d'apprentissage pour IPS S.

IV.1.1 Détermination de K

Pour déterminer, le nombre de classes, K, la classification a été réalisée pour différents K, K allant de 2 à 10. Avec les différents résultats obtenus (Figure 22), une tendance s'affiche : les délimitations des classes sont similaires. Lorsque K est grand, de nouvelles classes sont créées à l'intérieur des classes, induisant un effet « courbes de niveau ». Le K choisi sera alors le K pour lequel les frontières des classes se fixent par rapport à K-1 et avant le K+n pour lequel le phénomène de « courbes de niveau » apparaît.

Pour chaque combinaison, le Tableau 6 indique le nombre de classes déterminées.

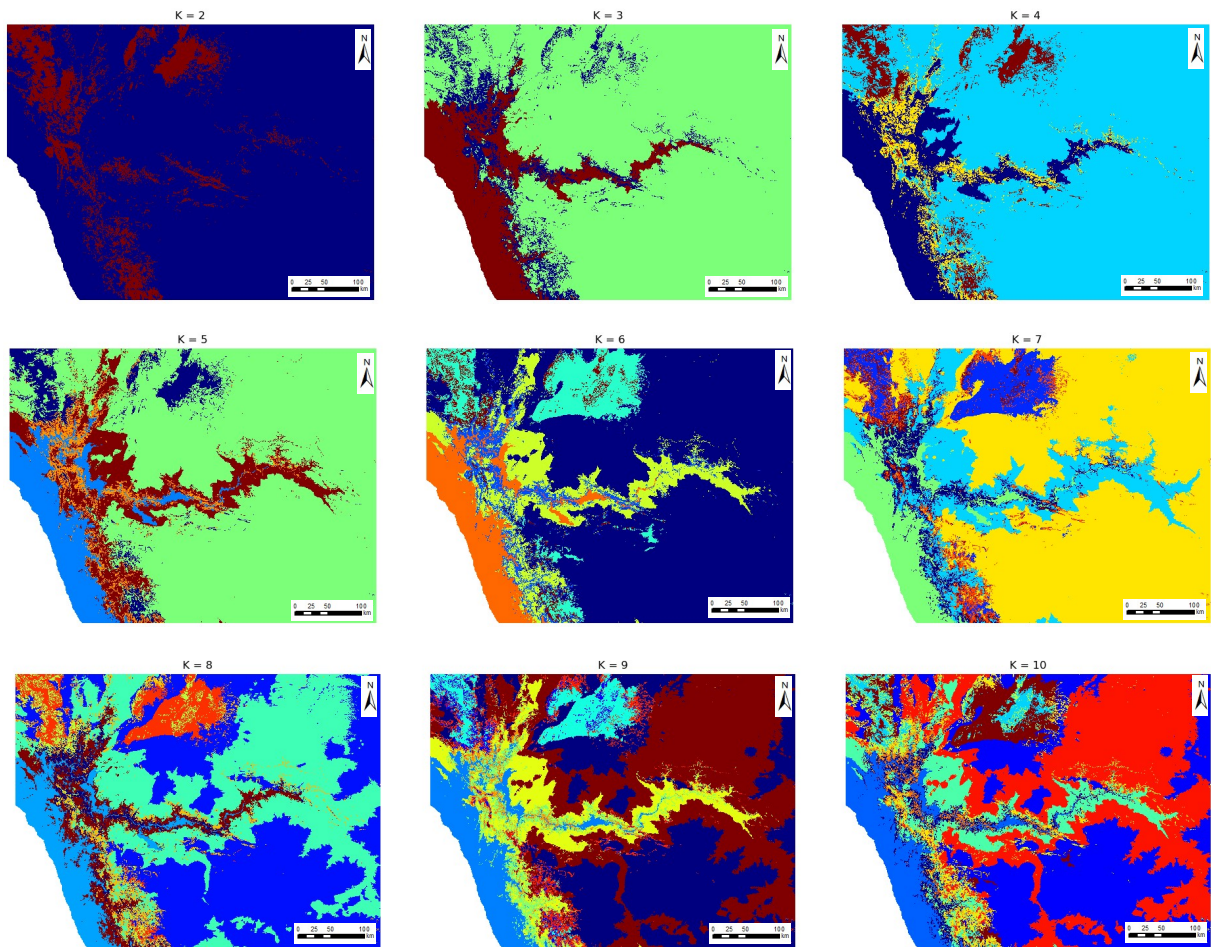


Figure 22 — Détermination de K pour H SL CU RU

	H SL CU RU	SL CU RU RH	SL CU RU	SL CU	SL CU TE	SL CU SH
K	4	4	4	4	4	3

Tableau 6 — K déterminé pour chaque combinaison

IV.1.2 Identification des PS par degré d'appartenance

En s'inspirant de la méthode de détermination du degré d'appartenance proposé par Haider et al. (2015), une fois K déterminé, les classes correspondant aux PS peuvent être identifiées. Des nuages de points sont créés en fonction des descripteurs : H/SL, SL/CU, SL/RU, pour chaque classe (Figure 23).

Grâce aux propriétés morphométriques des PS, un degré d'appartenance est défini pour chacun des paramètres calculés et leur combinaison par multiplication permet de définir le

degré d'appartenance à la catégorie PS. Ainsi, pour être identifiée comme une PS, en se basant sur la littérature existante (Haider et al., 2015 ; Liu et al., 2018), la classe doit avoir $SL < 2^\circ$, $CU < 0,14 \text{ m}^{-1}$ et $RU < 75 \text{ m}^{-1}$ (représenté par les contours noirs).

Le Tableau 7 résume les classes identifiées comme des PS pour chaque combinaison de descripteurs. L'Annexe 1 reprend l'ensemble des graphiques permettant de déterminer les classes.

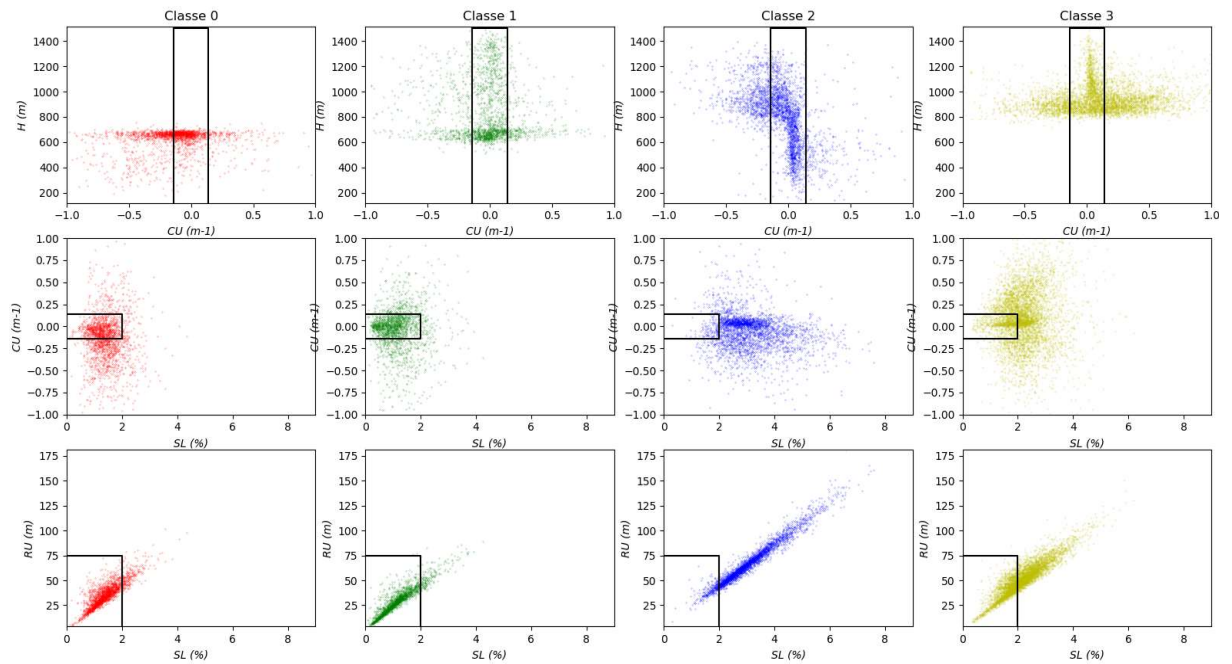


Figure 23 — Exemple de représentation des descripteurs par classes 100 % d'appartenance

	H SL CU RU	SL CU RU RH	SL CU RU	SL CU	SL CU TE	SL CU SH
Classes	0 ; 1	1	0	0 ; 2	3	0

Tableau 7 — Numéros des classes identifiées comme PS

IV.1.3 Choix des échantillons d'apprentissage

Pour chaque zone, les échantillons d'apprentissage ont été déterminés par rapport à la cartographie des PS existante. Trois classes ont été définies : pédiments (PS1), etchplains (PS2) et divers (Figure 24).

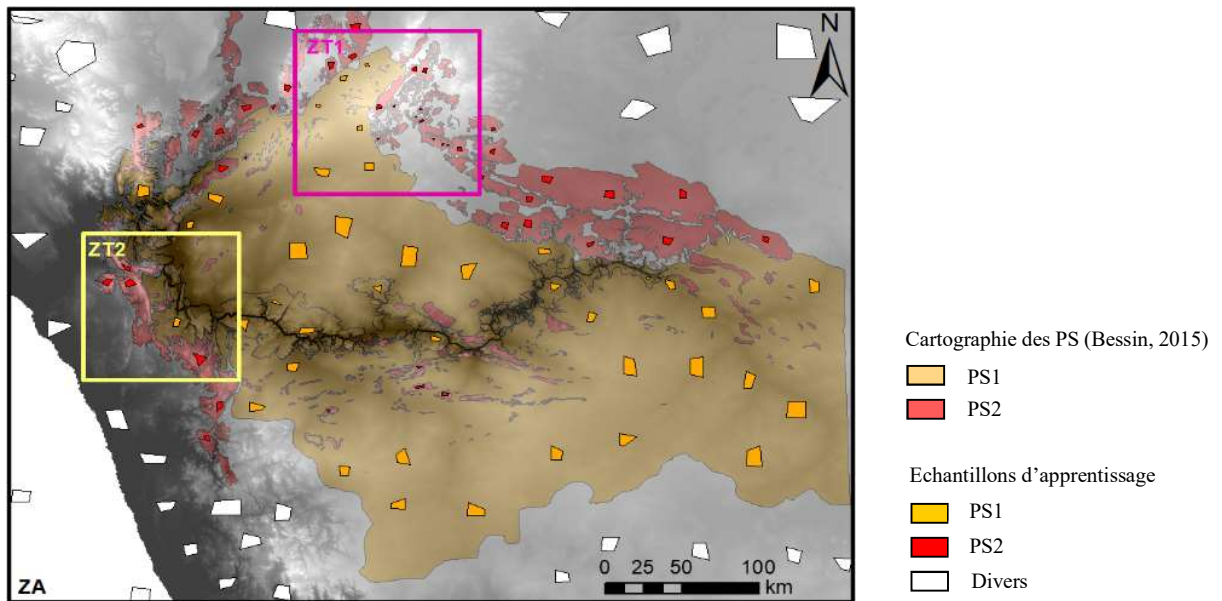


Figure 24 — Échantillons d'apprentissage définis

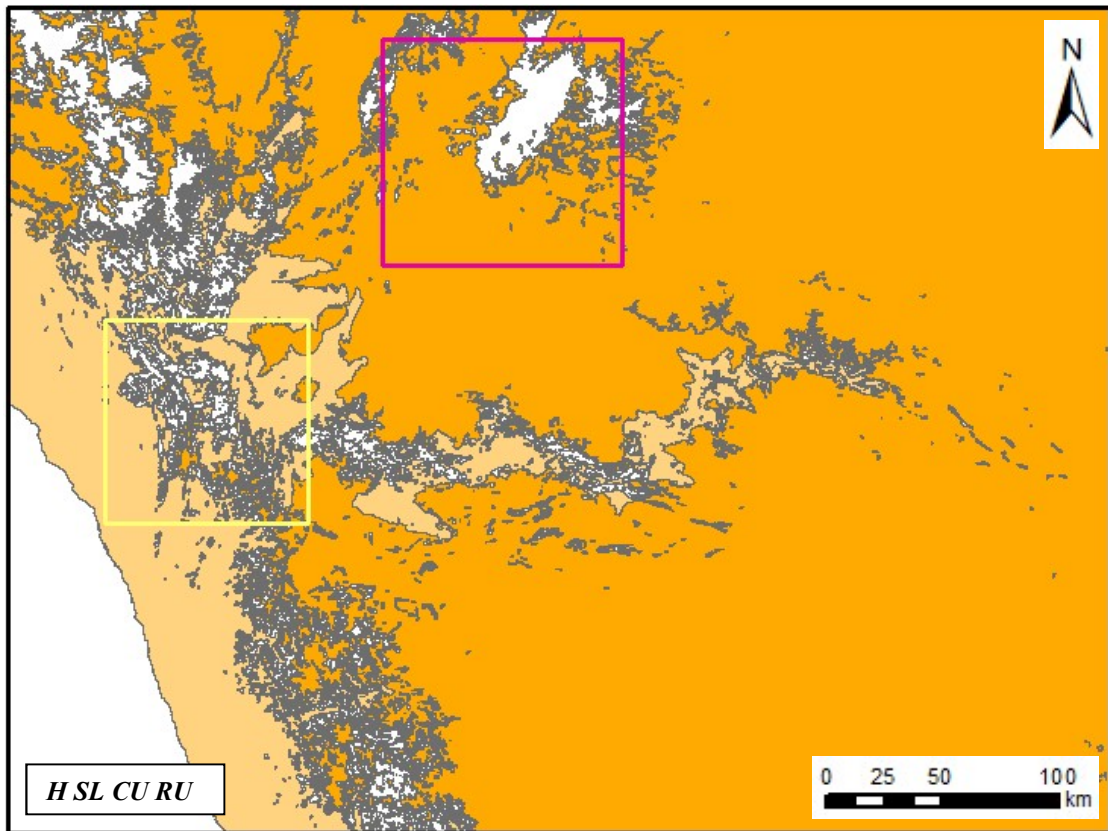
IV.1.4 Résultats

Les figures suivantes reprennent les résultats obtenus pour chaque combinaison de paramètres. Dans un premier temps pour le protocole IPS NS et dans un second temps pour le protocole IPS S.

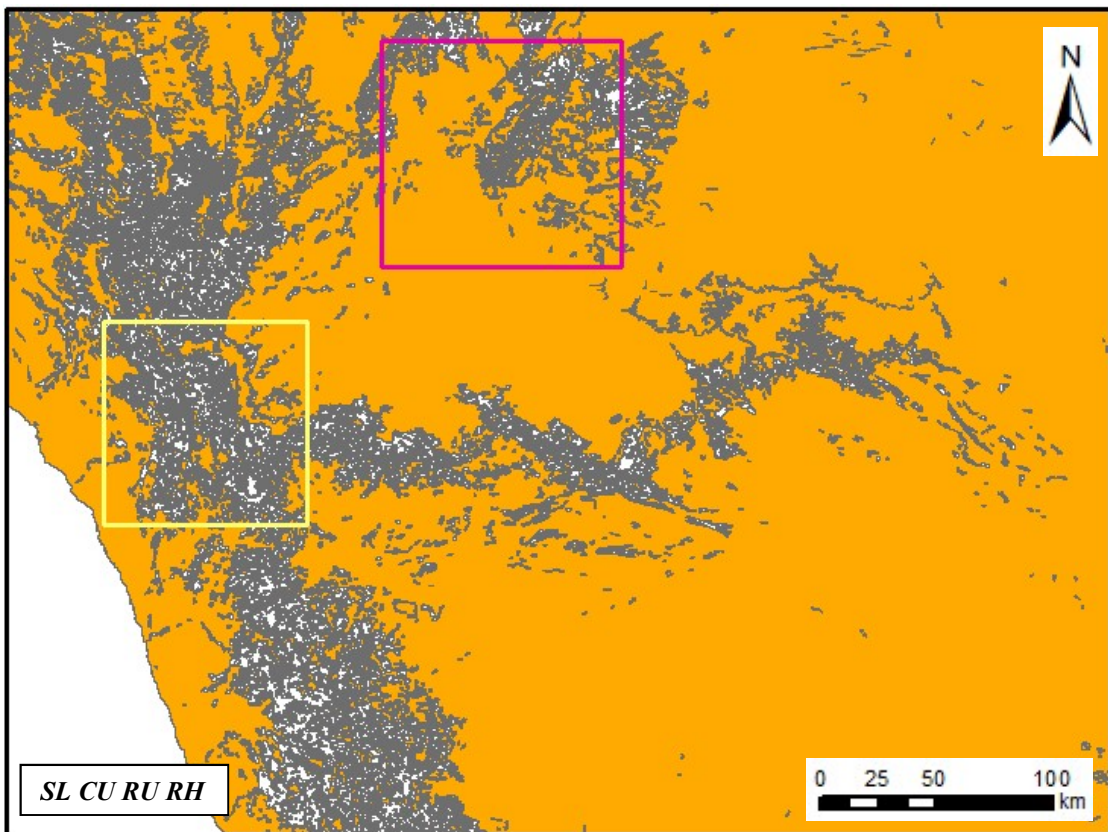
Le premier protocole, IPS NS, identifie des PS mais ne permet pas de différencier s'il s'agit de pédiments (PS1) ou de etchplains (PS2). Les surfaces seront alors nommées PS0.X (X permettant de différencier les classes identifiées comme PS).

IV.1.4.1 IPS NS

A)



B)



Type de surfaces




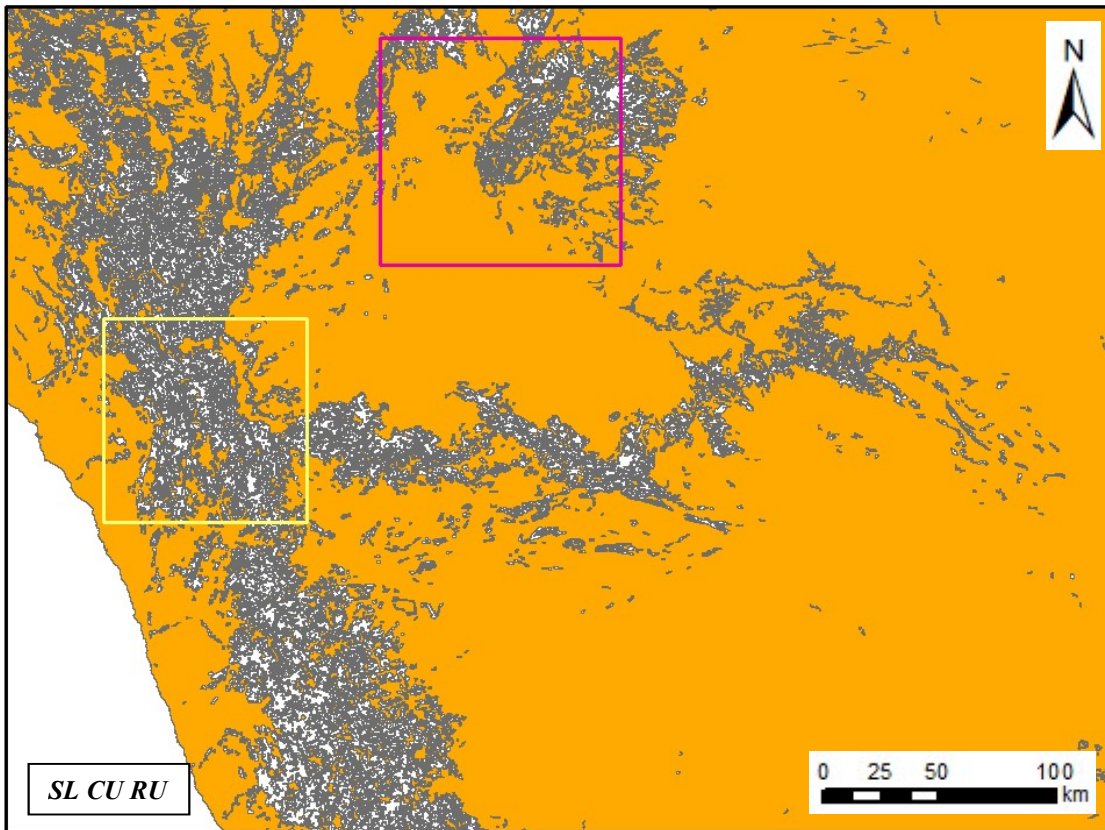
-  PS0.0
-  PS0.1
-  Divers

Figure 25 — Résultats protocole IPS NS

C)



D)

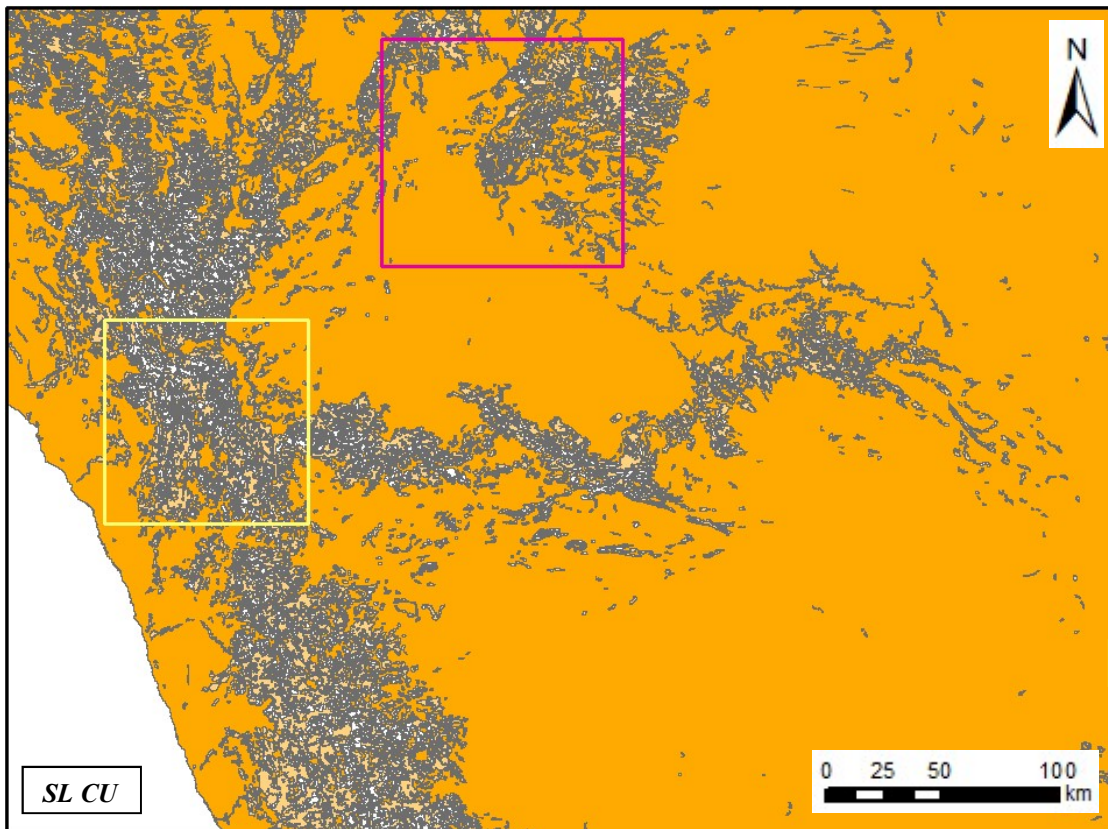
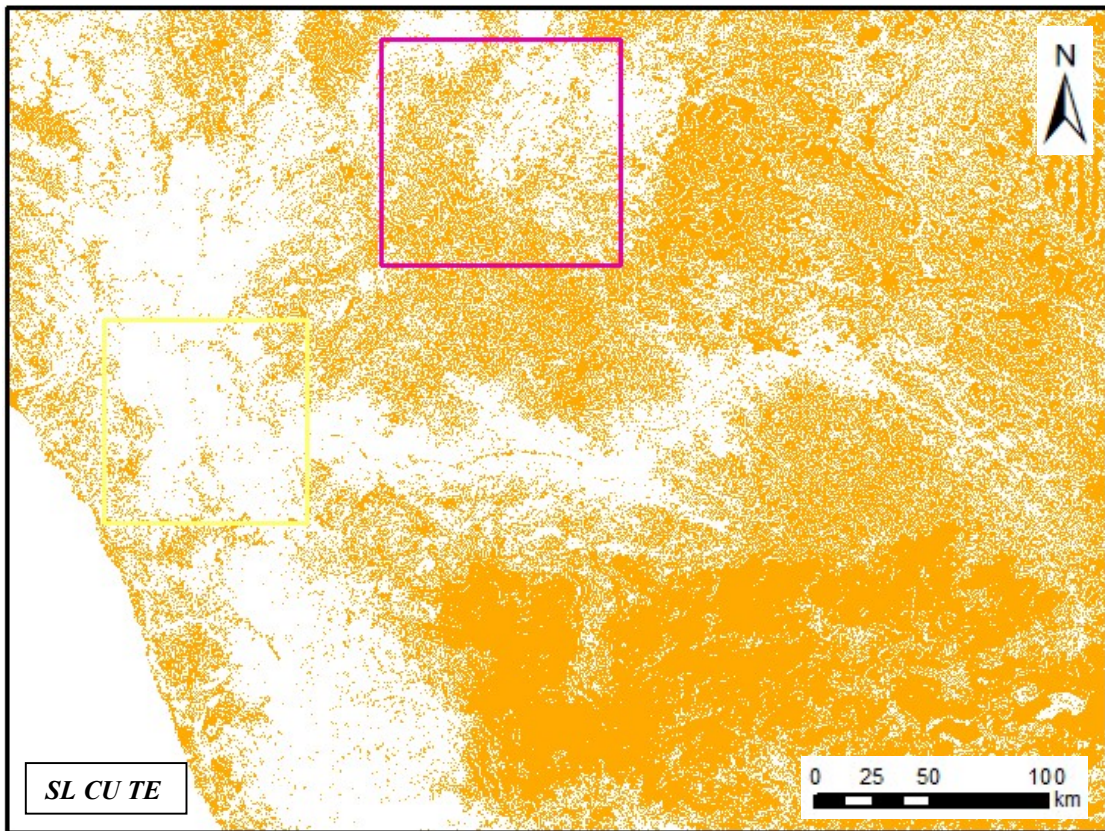
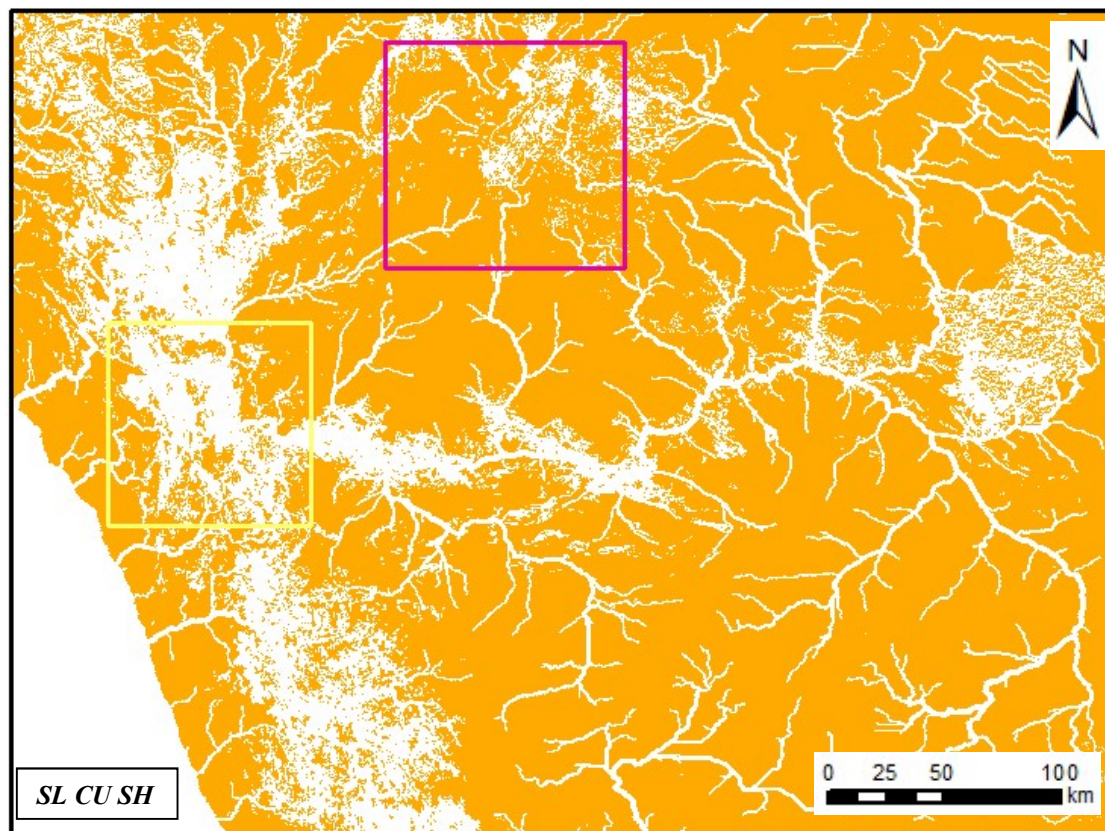


Figure 25 — Résultats protocole IPS NS (suite)

E)



F)



Type de surfaces




-  PS0.0
-  PS0.1
-  Divers

Figure 25 — Résultats protocole IPS NS (suite)

IV.1.4.2 IPS S

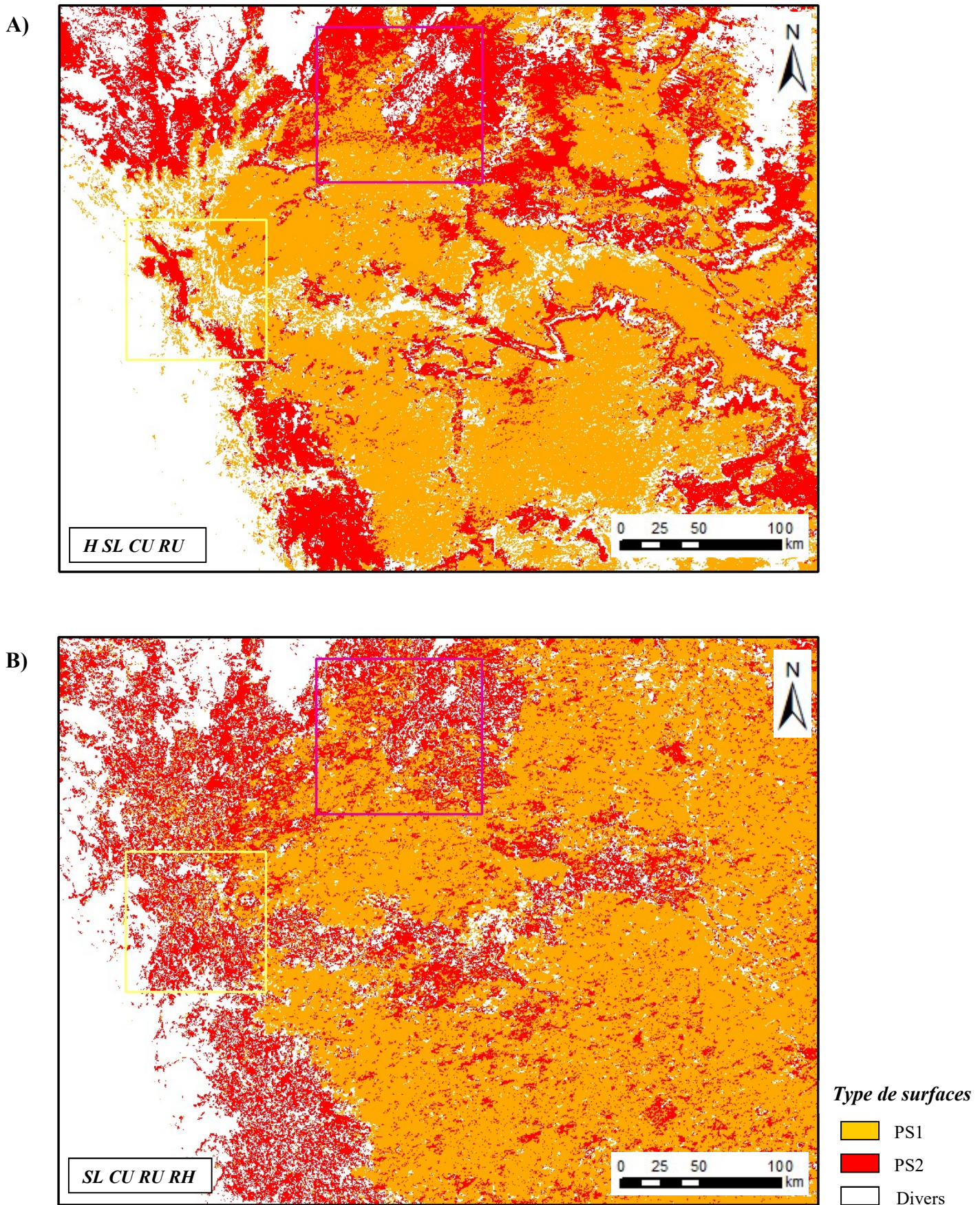
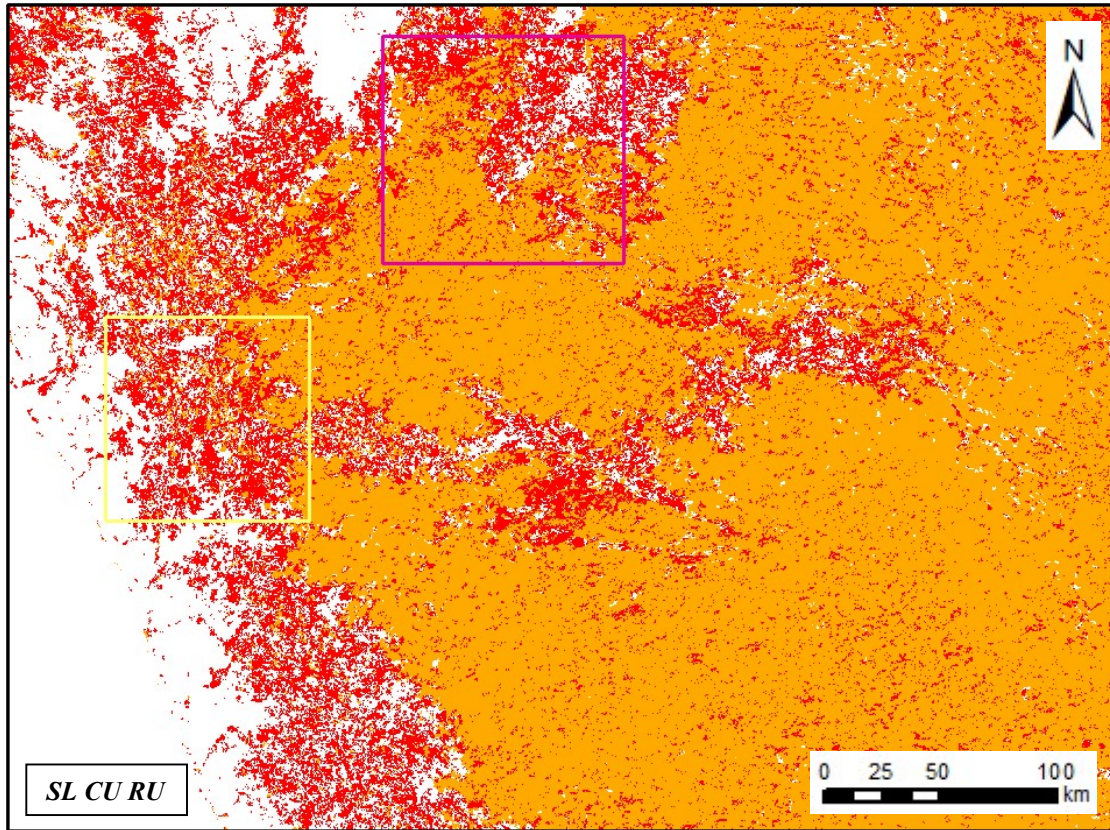
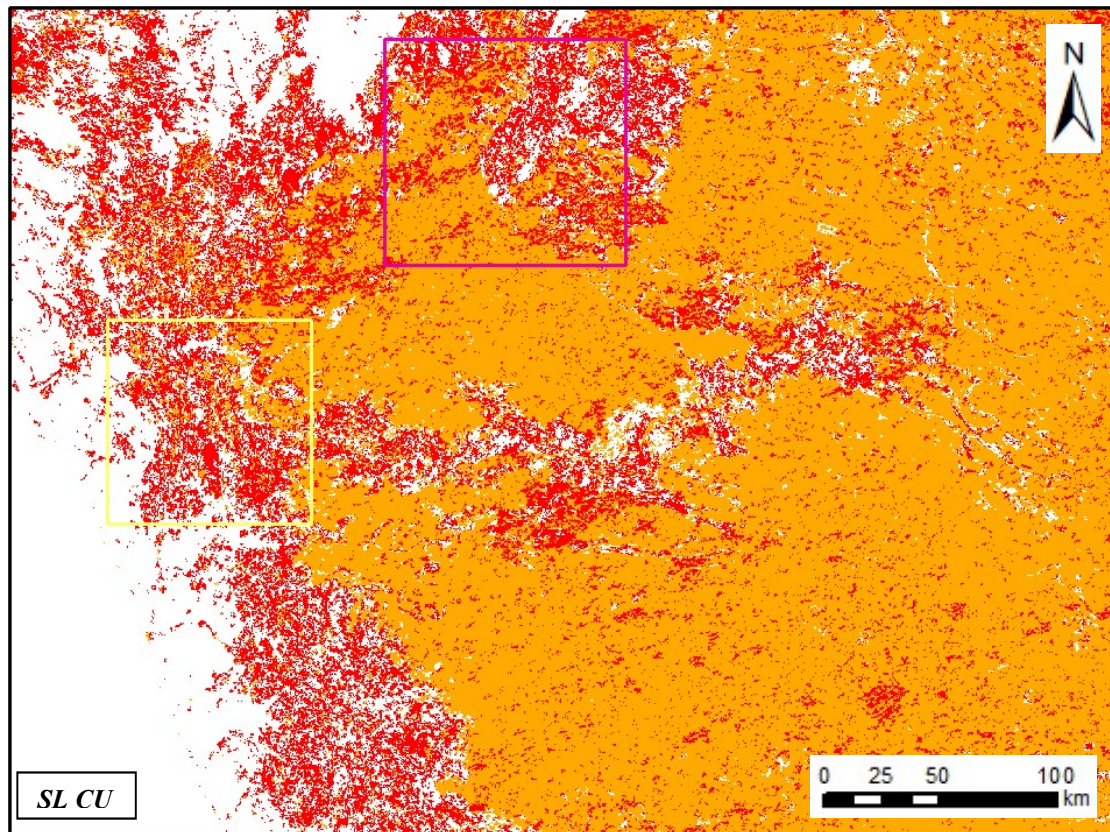


Figure 26 — Résultats protocole IPS S

C)



D)



Type de surfaces




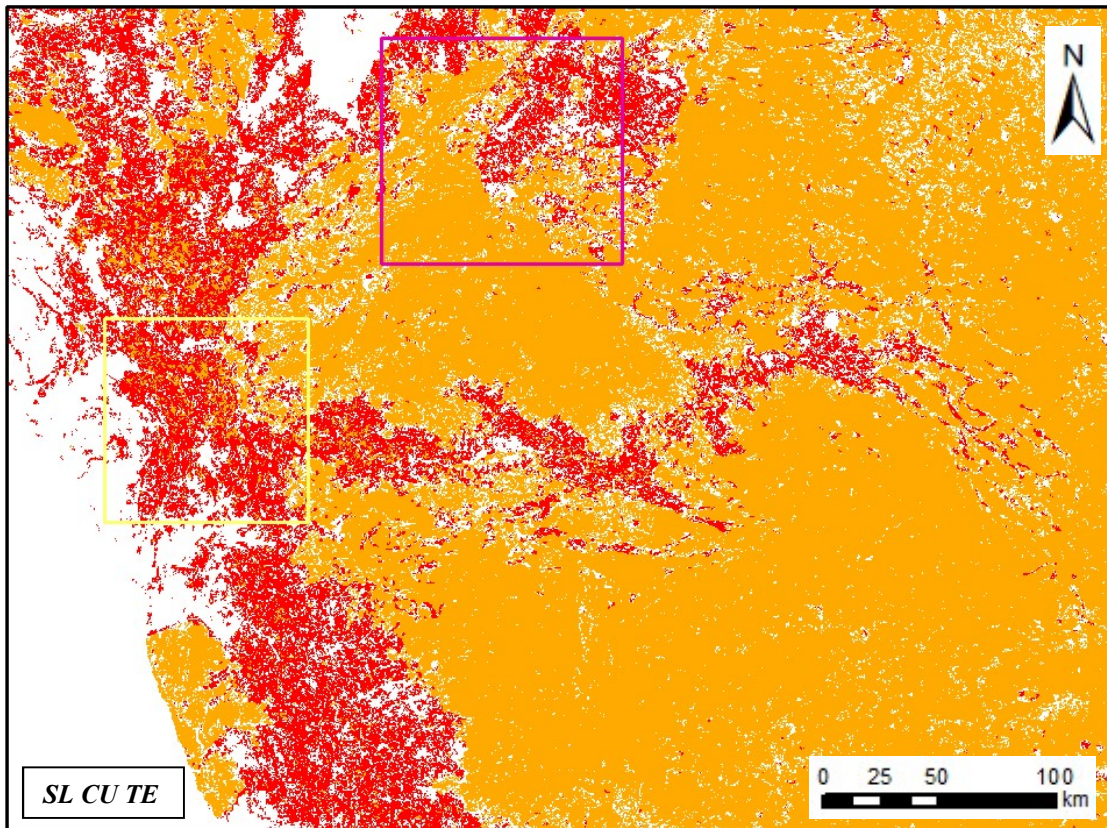
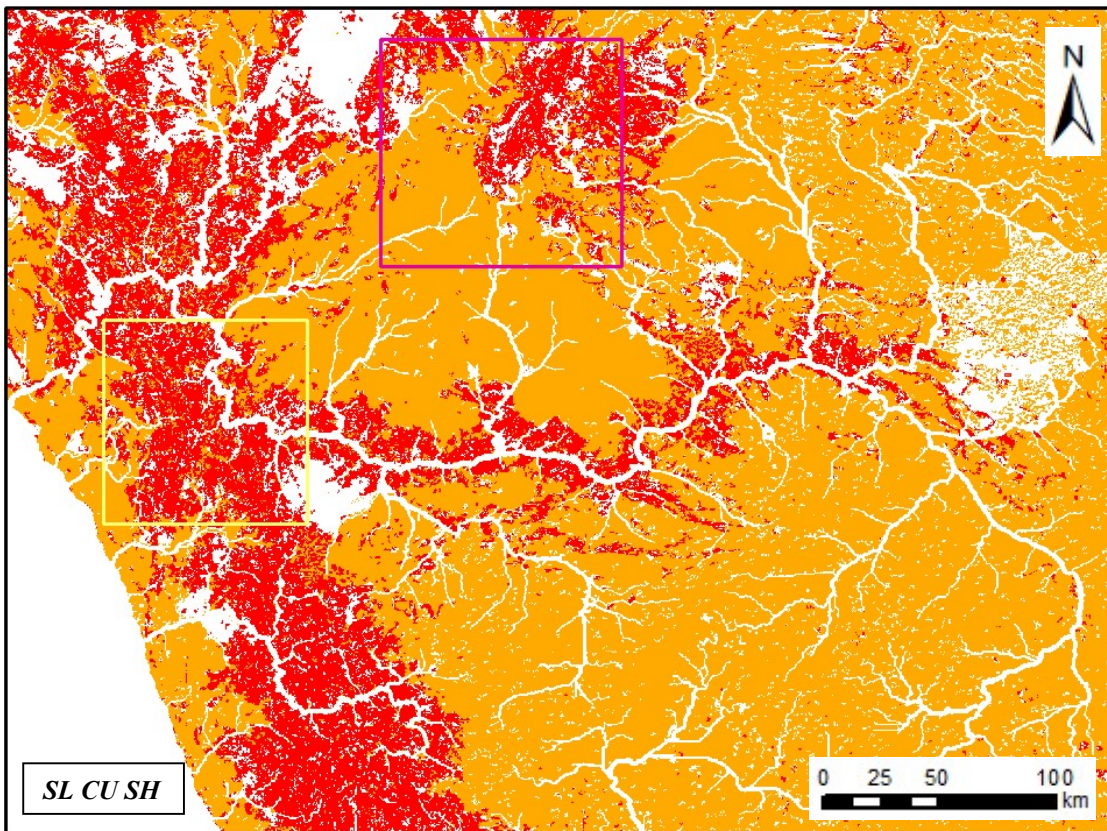
-  PS1
-  PS2
-  Divers

Figure 26 — Résultats protocole IPS S (suite)

E)



F)



Type de surfaces

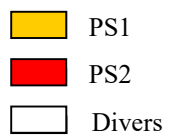


Figure 26 — Résultats protocole IPS S (suite)

IV.2 Application à une zone de contrôle : région Nord-Patagonienne (Argentine)

Le même travail que précédemment a été reproduit sur l'Argentine pour confirmer les résultats obtenus sur l'Afrique :

- Travail préliminaire pour définir K, les classes assimilables à des PS et la définition des échantillons d'apprentissage à partir des cartographies des PS existantes (Aguilera & Rabassa, 2010; Gonzales Diaz & Di Tommaso, 2011)
- Application des deux protocoles pour H SL CU RU, SL CU RU RH, SL CU RU, SL CU, SL CU TE

L'Annexe 2 reprend les différents résultats. Le résultat ci-dessous montre les PS identifiées pour la région située en Argentine Nord-Patagonienne.

A)

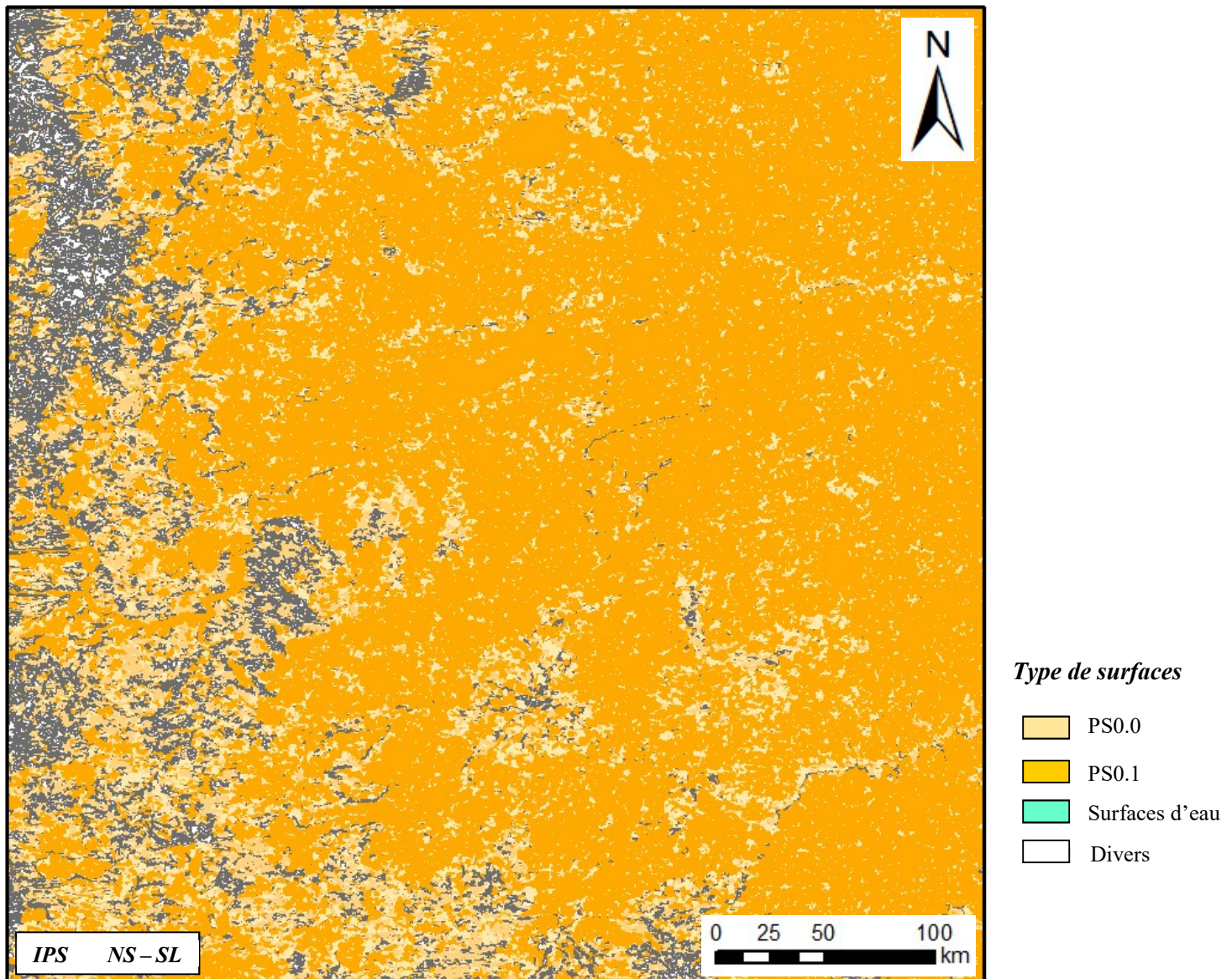
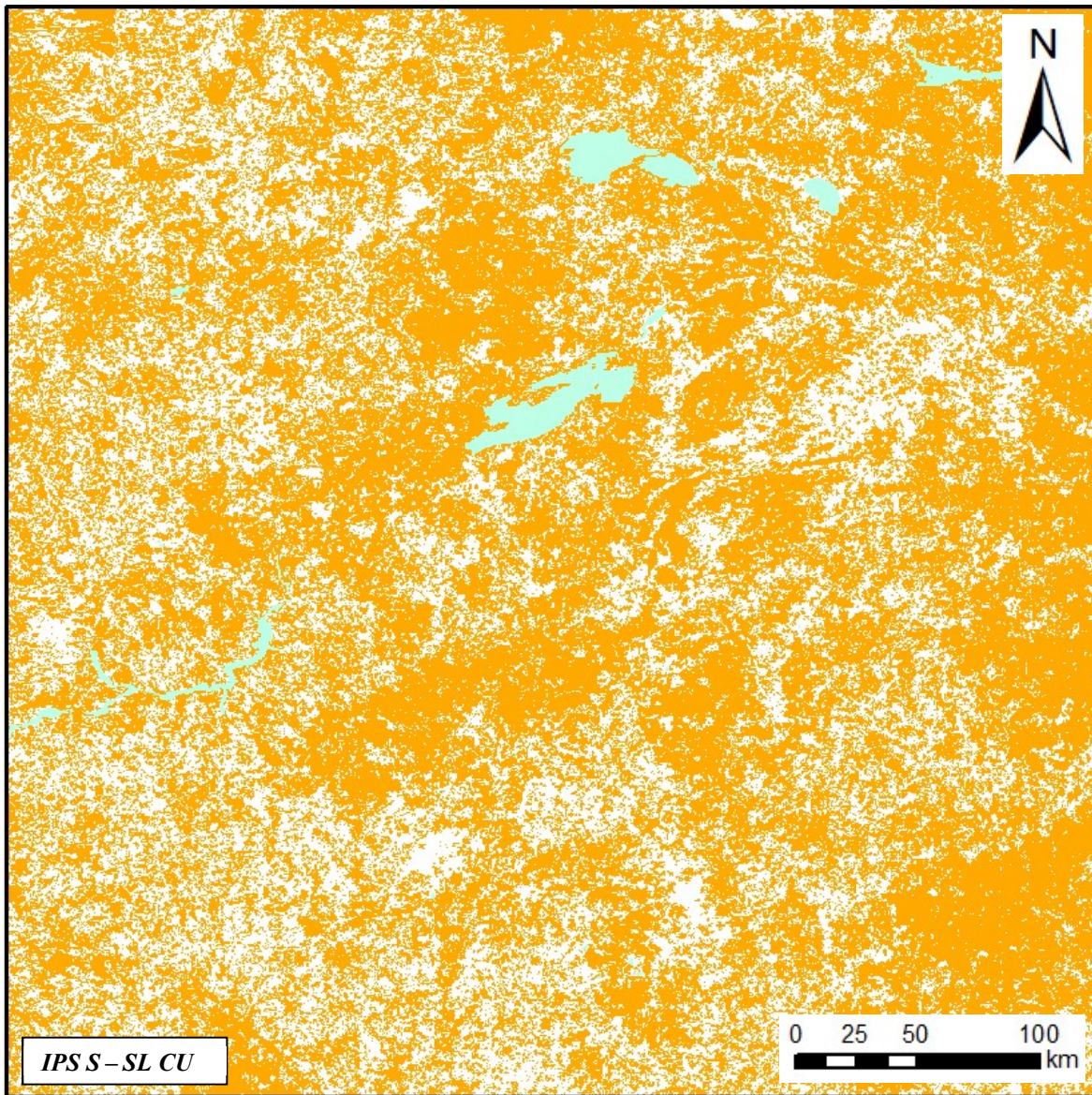


Figure 27 — Résultats protocole IPS NS avec SL, CU pour ZAR

B)



Type de surfaces





-  PS0.0
-  PS0.1
-  Surfaces d'eau
-  Divers

Figure 28 — Résultat protocole IPS S avec SL CU pour ZAR

IV.3 Évaluation de la précision des résultats pour comparaison avec les cartographies existantes

Pour comparer les résultats entre eux et avec les cartes existantes, j'ai utilisé deux méthodes qui s'appliquent aux données et résultats obtenus. La première consiste à comparer les pourcentages d'occupation de chaque classe d'un résultat à un autre. La deuxième méthode est basée sur les matrices de confusion.

IV.3.1 Méthode des pourcentages d'occupation par classe

Avec cette méthode, le pourcentage d'occupation de la carte pour chaque classe, pour ZA la zone d'étude africaine, est déterminé. Cela permet de comparer les résultats entre eux, mais aussi avec la cartographie existante, bien que la cartographie existante ne recouvre pas toute la zone d'étude (Tableau 8).

Protocole	IPS NS		IPS S		
Classes	PS	Divers	PS1 (Ped)	PS2 (Etch)	Divers
H SL CU RU	0,75	0,25	0,44	0,25	0,39
SL CU RU RH	0,82	0,18	0,53	0,25	0,22
SL CU RU	0,82	0,18	0,58	0,21	0,21
SL CU	0,93	0,07	0,61	0,20	0,19
SL CU TE	0,43	0,57	0,62	0,17	0,21
SL CU SH	0,70	0,30	0,58	0,19	0,23
Ancienne carte	0,47	0,53	0,39	0,08	0,53

Tableau 8 — Comparaison des pourcentages d'occupation par classe pour chaque test pour ZA

Pour le premier protocole, IPS NS, comme pour le deuxième protocole, IPS S, les pourcentages d'occupation de chaque classe sont assez similaires. La proportion de PS et divers sont du même ordre de grandeur, excepté pour SL CU TE avec le protocole IPS NS où les pourcentages sont différents des autres résultats.

IV.3.2 Matrice de confusion

Les matrices de confusion permettent de mesurer la qualité d'une classification. Elles comparent deux ensembles : la « vérité » terrain et la classification réalisée. Ici, la carte des PS de Dauteuil et al., 2015 est considérée comme la « vérité » terrain (elle contient seulement

les classes PS1 et PS2, pas de classe divers) et les résultats du protocole IPS S pour ZA comme la classification réalisée. Il est difficile de calculer les matrices de confusion pour le protocole IPS NS puisque les classes obtenues sont différentes de la cartographie existante.

La lecture de chaque matrice nous indique combien d'objets sont bien classés et combien sont mal classés pour une classe donnée. L'abscisse représente la « vérité » terrain et l'ordonnée la classification. Par exemple : 60 056 points de la classe C_1 classés sont issus de la classe C_1 de la « vérité » terrain pour H SL CU RU.

		IPS S					
H SL CU RU	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	60056	3807	0	63863	0,940388	0
	C_2	10965	10331	0	21296	0,485115	0
	C_3	11710	3131	0	14841	0	0
	Total	82731	17269	0	100000	0	0
	P_Accuracy	0,725919	0,59824	0	0	0,70387	0
	Kappa	0	0	0	0	0	0,319052
SL CU RU RH	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	58510	5784	0	64294	0,910038	0
	C_2	16490	7749	0	24239	0,319691	0
	C_3	7731	3736	0	11467	0	0
	Total	82731	17269	0	100000	0	0
	P_Accuracy	0,707232	0,448723	0	0	0,66259	0
	Kappa	0	0	0	0	0	0,208387
SL CU RU	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	63985	6824	0	70809	0,903628	0
	C_2	12627	6674	0	19301	0,345785	0
	C_3	6119	3771	0	9890	0	0
	Total	82731	17269	0	100000	0	0
	P_Accuracy	0,77341	0,386473	0	0	0,70659	0
	Kappa	0	0	0	0	0	0,22961
SL CU	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	66703	7470	0	74173	0,899289	0
	C_2	11996	6576	0	18572	0,354081	0
	C_3	4032	3223	0	7255	0	0
	Total	82731	17269	0	100000	0	0
	P_Accuracy	0,806264	0,380798	0	0	0,73279	0
	Kappa	0	0	0	0	0	0,245782
SL CU TE	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	65443	7501	0	72944	0,897168	0
	C_2	7072	5791	0	12863	0,450206	0
	C_3	10216	3977	0	14193	0	0
	Total	82731	17269	0	100000	0	0
	P_Accuracy	0,791034	0,335341	0	0	0,71234	0
	Kappa	0	0	0	0	0	0,231501
SL CU SH	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	59040	7150	0	66190	0,891978	0
	C_2	8495	6717	0	15212	0,441559	0
	C_3	6098	2617	0	8715	0	0
	Total	73633	16484	0	90117	0	0
	P_Accuracy	0,801814	0,407486	0	0	0,729685	0
	Kappa	0	0	0	0	0	0,267408

C_1 = PS1
C_2 = PS2
C_3 = Divers

Tableau 9 — Matrices de confusion

Ces matrices de confusion nous permettent de dire que les descripteurs optimaux seraient SL et CU puisque c'est pour ces paramètres que nous obtenons la meilleure précision globale : 73 %.

L'indice kappa indique un pourcentage d'accord entre les deux jeux de données. 24,5 % est la valeur maximale obtenue pour SL et CU.

Les précisions obtenues pour l'Argentine sont présentées en Annexe 3.

V Discussion

Les résultats précédents montrent la validité des deux protocoles à identifier des PS puisque nous obtenons des résultats cohérents avec les cartographies existantes. Bien que les résultats des deux protocoles soient différents à cause de la méthode de classification utilisée, non supervisée ou supervisée, les résultats sont exploitables et répondent à la problématique posée. Dans cette partie, nous reviendrons sur les différents résultats obtenus, mais aussi sur les limitations que les protocoles élaborés font surgir et les perspectives d'améliorations.

V.1 Performances globales

V.1.1 Interprétation des résultats

Les protocoles permettent d'utiliser des combinaisons de descripteurs différents en entrée. Les résultats ont montré que les différentes combinaisons de descripteurs donnent différentes identifications des PS.

Pour la combinaison initialement utilisée, H SL CU RU, les résultats font apparaître un phénomène de « courbes de niveau » (Figure 25 – A). Que ce soit pour l'Afrique ou pour l'Argentine, ce phénomène apparaît nettement pour les deux protocoles proposés. L'attribut H aurait une importance trop forte dans l'identification des PS ? Une analyse en composante principale (ACP) montre qu'un des descripteurs a un degré d'importance élevé, 48 % contre 26 %, 24 % et 2 % pour les autres descripteurs. Ce paramètre serait donc la donnée H. En effet l'ACP réalisée ici ne permet pas de savoir quel pourcentage est associé à quel descripteur. Ainsi, H pourrait être exclue des données initiales pour l'identification des PS. RH, utilisée par Haider et al. (2015) peut être une donnée de substitution de H. En effet, la donnée RH représente la différence d'altitude et peut se substituer à l'altitude et d'éliminer le phénomène de courbe de niveau. De plus, l'ACP (Figure 29) montre une corrélation entre SL et RU qui sera discutée plus loin.

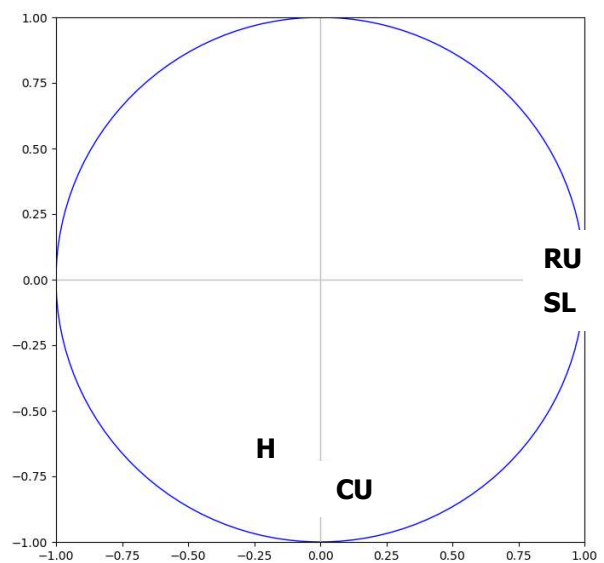


Figure 29 — ACP pour H SL CU RU

Les résultats obtenus en remplaçant H par RH dans la combinaison des descripteurs précédents ne présentent pas de phénomène de « courbes de niveau ». Cependant, le résultat pour le protocole IPS S présente de nombreuses surfaces identifiées comme PS2 (Figure 26 – B). L'ACP avec ces quatre attributs montre ici qu'un descripteur a un fort degré d'importance, 74 %, et un paramètre a un degré d'importance très faible de 2 %. L'ACP réalisée ne donnant pas le degré d'importance associé à chaque attribut, les différents tests permettront de déterminer l'association des descripteurs et de leurs degrés d'importance. De plus, l'ACP montre une corrélation entre RH, SL et RU (Figure 30).

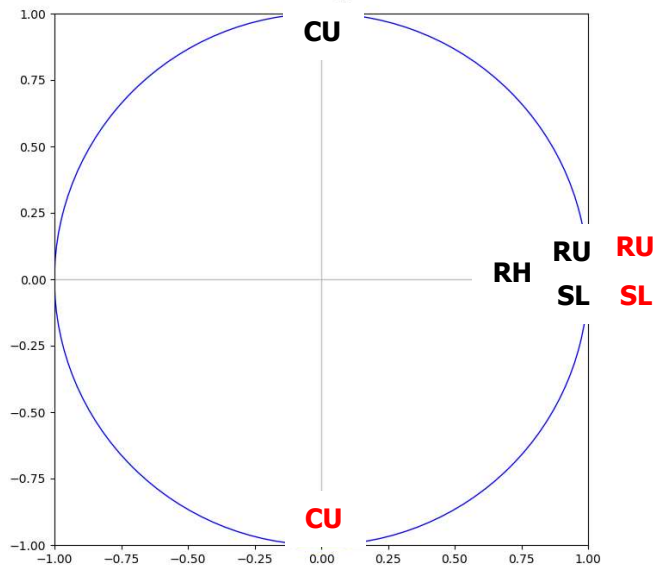


Figure 30 — ACP pour SL CU RU RH en noir,
ACP pour SL CU RU en rouge

Avec les trois descripteurs, SL, CU et RU, les résultats sont assez similaires à ceux obtenus avec SL CU RU RH. En comparant les cercles de corrélation (Figure 30), le descripteur RH serait donc celui qui a le moins d'importance. Ce dernier pourra être négligé par la suite. L'ACP sur ces trois descripteurs montre toujours que SL et RU sont corrélés. Les degrés d'importance pour les descripteurs de cette combinaison sont de 64 %, 33 % et 3 %. Comme SL et RU sont corrélés, un des deux paramètres peut être mis de côté. SL est le paramètre principal caractérisant les PS, RU permet lui de distinguer différents types de PS, il pourra être négligé. Avant de négliger ce paramètre, RU a été calculé pour d'autres résolutions (30 m, 180 m, 360 m et 720 m) et avec des méthodes de calcul différentes. En effet, le résultat de RU dépend de la résolution, l'identification des PS avec SL CU et RU à des résolutions différentes donne des résultats différents dont un exemple est présenté en Annexe 4. Ainsi le paramètre RU peut être exclu pour une résolution de 90 m, mais il pourrait être utilisé avec d'autres résolutions.

Les résultats obtenus pour SL et CU sont eux aussi similaires aux deux résultats précédents, démontrant ainsi que RU et RH ont une importance relativement faible dans l'identification des PS, ce qui est en accord avec les résultats des ACP (Figure 30). Les deux paramètres les plus propices à l'identification des PS sont donc SL et CU. Pour les deux protocoles, les résultats obtenus avec ces deux paramètres sont satisfaisants. L'ACP sur ces deux paramètres

montre que leurs importances est presque égale dans l'identification et que leurs comportements respectifs sont décorrélé.

Les résultats obtenus avec les paramètres SL CU TE sont très différents. Ils sont composés d'objets beaucoup plus petits, principalement liés à la texture. Le protocole supervisé identifie lui de nouvelles PS non présentes dans les résultats précédents. Plusieurs PS au Nord-ouest et une PS au sud-ouest de la zone d'étude (Figure 26 – E). Ces PS pourraient correspondre à des PS identifiées par Baby, 2017 sur la plaine marginale (zone non étudiée dans Dauteuil et al., 2015). L'ACP

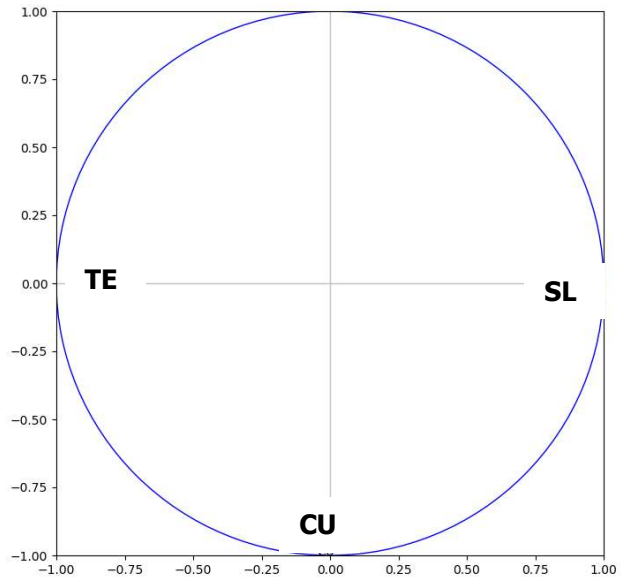


Figure 31 — ACP pour SL CU TE

sur ces trois composantes (Figure 31) montre qu'elles ne sont pas corrélées, avec des degrés d'importance de 52 %, 33 % et 15 %. La texture joue donc un rôle non négligeable dans l'identification des PS. Les PS auraient donc une texture propre qu'il serait intéressant d'étudier plus en détail.

Les tests réalisés sans le réseau hydrographique donnent des résultats très similaires à ceux obtenus sans le réseau hydrographique, il est donc peu pertinent d'enlever le réseau hydrographique pour identifier les PS.

Avec les différents résultats obtenus, les combinaisons de paramètres les plus adéquats pour identifier les PS semblent être SL CU et la combinaison SL CU TE. H, RH et RU peuvent ainsi être exclus puisqu'ils ont soit un degré d'importance trop ou peu significatif soit une corrélation avec un autre paramètre.

V.1.2 Choix du paramètre K

Le choix du paramètre K n'est pas un choix aisé. En effet, le choix de K ici est fait de manière arbitraire, en se basant et en comparant les résultats obtenus pour différents K notamment sur l'évolution de la frontière des différentes classes obtenues. La méthode de détermination du K pourrait être améliorée.

V.1.3 Précisions obtenues

Les précisions globales, obtenues avec les matrices de confusion, sont assez proches pour les différentes combinaisons de paramètres avec une précision minimale de 66 % et une précision maximale de 73 % (Tableau 9). La moyenne des précisions globales obtenues étant de 70,2 % pour l'ensemble des résultats.

Il est important de noter que ces précisions sont issues des matrices de confusion utilisant une cartographie des PS réalisée manuellement et pouvant comporter quelques biais liés à la subjectivité.

La meilleure précision obtenue, 73 % avec SL CU, semble être une précision relativement acceptable lorsqu'elle est comparée à la précision obtenue dans le cadre de développement de méthodes OBIA sur d'autres types d'objets géomorphologiques. Par exemple, dans ses travaux de recherche, Lewington (Lewington et al., 2019) développe et applique un protocole OBIA pour identifier des morphologies sous-glaciaires indicatrices de flux de glaces dans de paléo-calottes glaciaires. Ses tests de précision globale révèlent une valeur de 42 % de surfaces cartographiées manuellement identifiées par son protocole OBIA. La précision obtenue dans notre étude, supérieure à celle obtenue par Lewington, semble donc satisfaisante. Ces précisions supérieures pourraient en partie être expliquées par la différence de taille entre les objets identifiés (des surfaces plusieurs centaines de km² pour les PS et des couloirs de quelques dizaines de km de long pour les morphologies sous-glaciaires).

V.1.4 Objectivité et subjectivité des protocoles

Les deux protocoles élaborés diffèrent par leur étape de classification. Non supervisée pour un et supervisée pour l'autre. Le premier protocole, IPS NS, peut être qualifié de protocole « plus objectif ». En effet, le résultat est objectif dans le sens où le protocole détermine lui-même les différentes classes sans apports de l'utilisateur. Le deuxième protocole, IPS S, lui peut être qualifié de « plus subjectif », le résultat obtenu est basé sur les échantillons d'apprentissage définis par l'utilisateur et ses connaissances. Ainsi les deux protocoles permettent une identification différente en fonction des connaissances assimilées sur la zone d'étude. Pour une zone d'étude peu étudiée avec peu de connaissances, le protocole IPS NS sera plus adapté. Alors que pour une zone où de nombreux travaux existent déjà le protocole IPS S pourra être envisagé en fonction des connaissances.

V.2 Limitations

V.2.1 IPS NS et IPS S : Besoin d'interprétation ensuite

Une fois les résultats obtenus par les deux protocoles, le travail d'identification n'est pas terminé. Les résultats nécessitent d'être interprétés et plus particulièrement pour le protocole IPS NS qui donne un résultat objectif. Les résultats obtenus doivent ensuite être confrontés à d'autres données et connaissances géologiques. Par exemple, la confrontation avec une carte géologique permettra d'exclure ou non certains objets identifiés comme des PS. En effet, l'identification des PS identifie des objets morphologiques qui ont des caractéristiques similaires aux PS, mais qui n'en sont pas, comme des coulées volcaniques pour l'Argentine. Ces dernières peuvent alors être écartées grâce à une carte géologique. De plus, la carte géologique permet de discriminer les différentes générations de PS.

V.2.2 IPS S : Définition des échantillons d'apprentissage

Une première limitation aux protocoles élaborés est la définition des échantillons d'apprentissage pour le protocole IPS S. En fonction des échantillons d'apprentissage définis, les résultats vont être différents. Ainsi pour une même zone d'étude, les échantillons d'apprentissage définis par deux utilisateurs peuvent être différents. Les résultats ne seront alors pas nécessairement les mêmes (30 % de différence Figure 32).

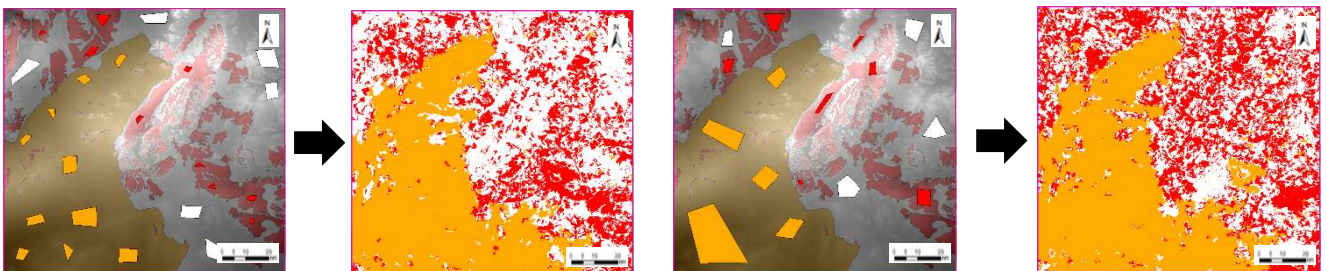


Figure 32 — Illustration des résultats obtenus pour des échantillons d'apprentissages différents

V.2.3 IPS S : Application d'arbre d'une zone à une autre

Les arbres de décisions créés à partir des échantillons d'apprentissage pour une zone ne sont pas forcément applicables à une autre zone. Par exemple, l'arbre de décision généré pour l'Afrique ne fonctionne pas sur l'Argentine. En effet, pour l'Afrique il y a deux types bien différents de PS alors qu'en Argentine seulement un type de PS est identifié (Aguilera & Rabassa, 2010; Gonzales Diaz & Di Tommaso, 2011).

V.3 Perspectives d'améliorations

Pour améliorer le processus d'identification, plusieurs améliorations sont envisageables. Tout d'abord la résolution utilisée, ici une résolution de 90 m peut être remise en question comme l'a montré le test pour RU avec des résolutions différentes. Est-ce la résolution optimale pour identifier les PS ? Des résolutions plus faibles (180 m, 360 m, 720 m) pourraient être utilisées comme les objets que nous cherchons à identifier ont de grandes superficies. Cela permettrait de mieux traduire la rugosité des PS qui s'étendent sur des kilomètres. Au contraire, des résolutions plus élevées (30 m, 10 m) pourraient être utilisées pour identifier les PS pour avoir une texture plus précise. Ainsi chaque paramètre utilisé avec la bonne résolution donnerait de meilleurs résultats.

Une autre voie d'amélioration envisageable est l'utilisation de nouvelles données en entrée ou en post traitement. L'image satellite, Landsat, ou des données géologiques, comme la carte géologique, pourraient être exploitées dans le processus d'identification. En effet, pour les zones non végétalisées (comme c'est le cas ici avec les zones d'études africaine et argentine), l'image satellite donne une première distinction des différentes surfaces présentes dans la zone d'étude. La carte géologique, elle, permettrait d'éviter de l'utiliser en post traitement et d'obtenir un résultat avec une première discrimination des objets.

Pour améliorer les résultats, les deux protocoles pourraient être combinés et modifiés. Après une première étape de classification non supervisée une deuxième étape de classification supervisée serait réalisée ou bien les résultats des deux protocoles seraient combinés et un traitement appliqué à cette combinaison. Cela permettrait d'affiner les résultats, mais aussi d'écarter les objets mal classés. Enfin, des étapes de post-traitement pourraient être appliquées sur le résultat. Par exemple, le filtrage des objets pour supprimer ceux de très petites dimensions.

Conclusion

L'identification des surfaces d'aplanissements est aujourd'hui un processus long et fastidieux. Pour répondre à ce problème, deux méthodes innovantes d'identification semi-automatique de ces surfaces ont été développées.

Les surfaces d'aplanissement ayant des caractéristiques topographiques précises, pente, courbure et rugosité, c'est sur ces données initiales que le développement des méthodes d'identification semi-automatisé est basé.

Les deux méthodes élaborées ont été inspirées par l'analyse d'image orientée objet (OBIA, Blaschke, 2013). Elles sont composées d'une chaîne traitement associant dans un premier temps une étape de segmentation puis une étape de classification. Chaque méthode étant différenciée par l'étape de classification, non supervisée (IPS NS) pour une et supervisée (IPS S) pour l'autre apportant ainsi un caractère plus objectif à la première et plus subjectif à la deuxième.

Afin d'identifier ces PS, à partir de ces deux méthodes, l'interrogation principale a été le choix de paramètres à utiliser. Pour répondre à cette question, plusieurs combinaisons de paramètres ont été testées à partir des paramètres initiaux définissant les PS, chacune donnant des résultats plus ou moins différents. L'analyse des données a montré que les paramètres H, RH et RU pouvaient être mis de côté tandis que TE s'est révélé être un paramètre intéressant à intégrer. De plus, le calcul des précisions, grâce aux matrices de confusion, a confirmé le choix des données retenues : la combinaison SL + CU et la combinaison SL + CU + TE. Avec des précisions respectives de 73 % et 71 % par rapport aux cartographies existantes.

Bien que ces deux méthodes répondent à la problématique posée qui était de développer une méthode semi-automatique d'identification des surfaces d'aplanissement, des perspectives d'amélioration apparaissent. En effet, une fois les surfaces d'aplanissement identifiées, une étape d'analyse doit être envisagée afin de contrôler les résultats obtenus à partir d'une carte géologique. Une autre voie d'amélioration serait de déterminer quelles sont les résolutions optimales à utiliser pour chaque paramètre.

Ainsi, l'objectif initialement posé est donc atteint. Deux méthodes ont été élaborées et comparées dans le but d'identifier les surfaces d'aplanissement et plusieurs perspectives d'améliorations sont proposées.

Bibliographie

- Abrishami, H., Xiao, X., Karteek, A., & Wei, H. (2015). *Adaptive Video-based Vehicle Classification Technique for Monitoring Traffic*. (134874).
- Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641–647.
- Aguilera, E. Y., & Rabassa, J. (2010). Origin of the Northern Patagonian Massif regional paleosurface. *Geociências (São Paulo)*, 29(4), 467–477.
- Baby, G. (2017). *Mouvements verticaux des marges passives d’Afrique australe depuis 130Ma, étude couplée: stratigraphie de bassin - analyse des formes du relief*.
- Baby, G., & Guillocheau, F. (2017). Mouvements verticaux des marges passives d’Afrique australe depuis 130 Ma, étude couplée : stratigraphie de bassin - analyse des formes du relief. *Géosciences Rennes, PhD*, 363.
- Bessin, P. (2014). *Geomorphologique du Massif armoricain Paul Bessin Évolution géomorphologique du Massif armoricain approche terre-mer*. Université de Rennes 1.
- Bessin, P., Guillocheau, F., Robin, C., Schroëtter, J. M., & Bauer, H. (2015). Planation surfaces of the Armorican Massif (western France): Denudation chronology of a Mesozoic land surface twice exhumed in response to relative crustal movements between Iberia and Eurasia. *Geomorphology*, 233, 75–91. <https://doi.org/10.1016/j.geomorph.2014.09.026>
- Beucher, S. (2010). The Watershed Transformation. Retrieved from Centre for Mathematical Morphology website: <http://cmm.ensmp.fr/~beucher/wtshed.html>
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., ... Tiede, D. (2014). Geographic Object-Based Image Analysis - Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>
- Bosh, G. (2016). *Pénéplation et dynamique de la lithosphère dans les Pyrénées*. Université de Rennes 1.
- Breiman, L. (2012). Random forests. *Ensemble Machine Learning: Methods and Applications*, 157–175. https://doi.org/10.1007/9781441993267_5
- Chardon, D., Grimaud, J. L., Beauvais, A., & Bamba, O. (2018). West African lateritic pediments: Landform-regolith evolution processes and mineral exploration pitfalls. *Earth-Science Reviews*, 179(January), 124–146. <https://doi.org/10.1016/j.earscirev.2018.02.009>

- Chen, G., Weng, Q., Hay, G. J., & He, Y. (2018). Geographic object-based image analysis (GEOBIA): emerging trends and future opportunities. *GIScience and Remote Sensing*, 55(2), 159–182. <https://doi.org/10.1080/15481603.2018.1426092>
- Comaniciu, D., & Meer, P. (2002). Mean Shift: Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.
- Dauteuil, O., Bessin, P., & Guillocheau, F. (2015). Topographic growth around the Orange River valley, southern Africa: A Cenozoic record of crustal deformation and climatic change. *Geomorphology*, 233, 5–19. <https://doi.org/10.1016/j.geomorph.2014.11.017>
- Gonzales Diaz, E. F., & Di Tommaso, I. (2011). Geomorfologia in Relatorio del XVIII congreso geologico Argentino - Neuquen. *Geomorfologia*, 421–438.
- Guillocheau, F., Simon, B., Baby, G., Bessin, P., Robin, C., & Dauteuil, O. (2018). Planation surfaces as a record of mantle dynamics: The case example of Africa. *Gondwana Research*, 53, 82–98. <https://doi.org/10.1016/j.gr.2017.05.015>
- Haider, V. L., Kropáček, J., Dunkl, I., Wagner, B., & von Eynatten, H. (2015). Identification of peneplains by multi-parameter assessment of digital elevation models. *Earth Surface Processes and Landforms*, 40(11), 1477–1492. <https://doi.org/10.1002/esp.3729>
- Jarsve, E. M., Krøgli, S. O., Etzelmüller, B., & Gabrielsen, R. H. (2014). Automatic identification of topographic surfaces related to the sub-Cambrian peneplain (SCP) in southern Norway - Surface generation algorithms and implications. *Geomorphology*, 211, 89–99. <https://doi.org/10.1016/j.geomorph.2013.12.032>
- Jenness, J. (20086). *Topographic Position Index (TPI)*.
- Koehrsen, W. (2017). Random Forest Simple Explanation. Retrieved from Medium.Com website: <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>
- Lachassagne, P., Wyns, R., & Dewandel, B. (2011). The fracture permeability of Hard Rock Aquifers is due neither to tectonics, nor to unloading, but to weathering processes. *Terra Nova*, 23(3), 145–161. <https://doi.org/10.1111/j.1365-3121.2011.00998.x>
- Landman, N., Pang, H., & Williams, C. (2019). K-means clustering.
- Lewington, E. L. M., Livingstone, S. J., Sole, A. J., Clark, C. D., & Ng, F. S. L. (2019). An automated method for mapping geomorphological expressions of former subglacial meltwater pathways (hummock corridors) from high resolution digital elevation data. *Geomorphology*, 339, 70–86. <https://doi.org/10.1016/j.geomorph.2019.04.013>
- Liu, F., Gao, H., Li, Z., Su, H., & Pan, B. (2018). Quantitative analysis of planation surfaces of the upper Yangtze River in the Sichuan-Yunnan Region, Southwest China. *Frontiers*

- of Earth Science*, 13(1), 55–74. <https://doi.org/10.1007/s11707-018-0707-y>
- Mehnert, A., & Jackway, P. (1997). An improved seeded region growing algorithm. *Pattern Recognition Letters*, 18(10), 1065–1071. [https://doi.org/10.1016/S0167-8655\(97\)00131-1](https://doi.org/10.1016/S0167-8655(97)00131-1)
- Memarsadeghi, N., Mount, D. M., Netanyahu, N. S., & Le Moigne, J. (2007). A Fast Implementation of the Isodata Clustering Algorithm. *International Journal of Computational Geometry & Applications*, 17(01), 71–103. <https://doi.org/10.1142/s0218195907002252>
- Vincent, L., & Soille, P. (1991). Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 583–598.

Table des annexes

Annexe 1 Résultats complémentaires pour l’Afrique	58
Annexe 2 Résultats complémentaires pour l’Argentine.....	66
Annexe 3 Statistiques et précisions pour l’Argentine	71
Annexe 4 Comparaison RU à différentes résolutions	72

Annexe 1

Résultats complémentaires pour l'Afrique

Détermination du K

Les K obtenus pour chaque combinaison sont résumés dans le tableau suivant.

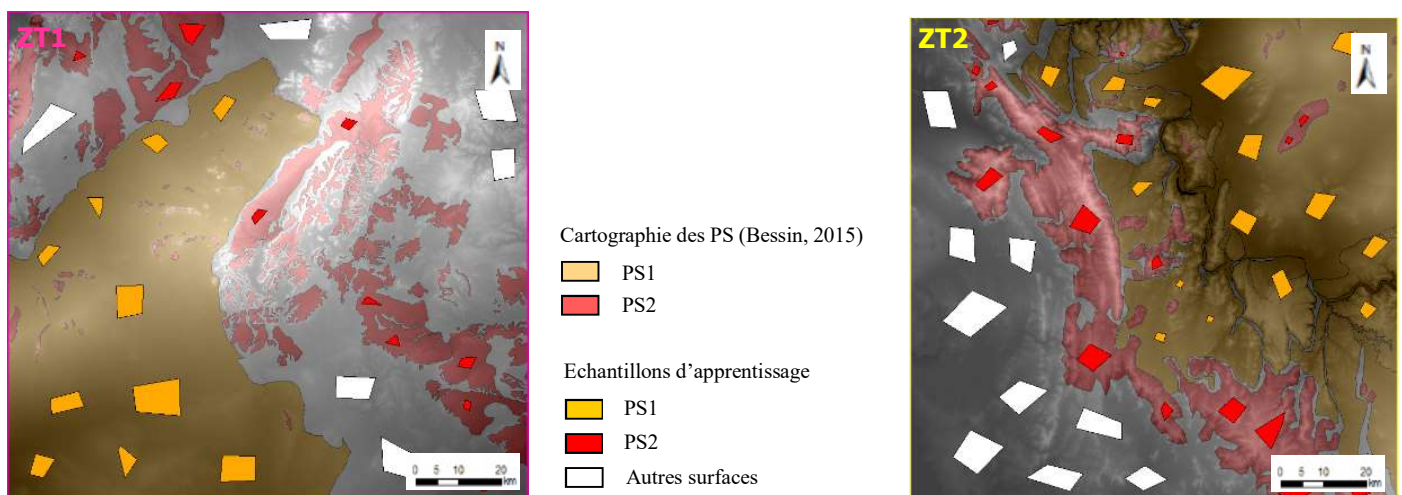
	H SL CU RU	SL CU RU RH	SL CU RU	SL CU	SL CU TE
K ZT1	4	4	4	4	3
K ZT2	4	4	4	4	3

Détermination des classes

Le tableau suivant résume les classes identifiées comme PS pour chaque combinaison.

	H SL CU RU	SL CU RU RH	SL CU RU	SL CU	SL CU TE
Classe(s) ZT1	0 et 3	0 et 3	0 et 3	1	1 ; 2 et 3
Classe(s) ZT2	2	1 et 3	1 et 2	0 et 3	2

Échantillons d'apprentissage utilisés

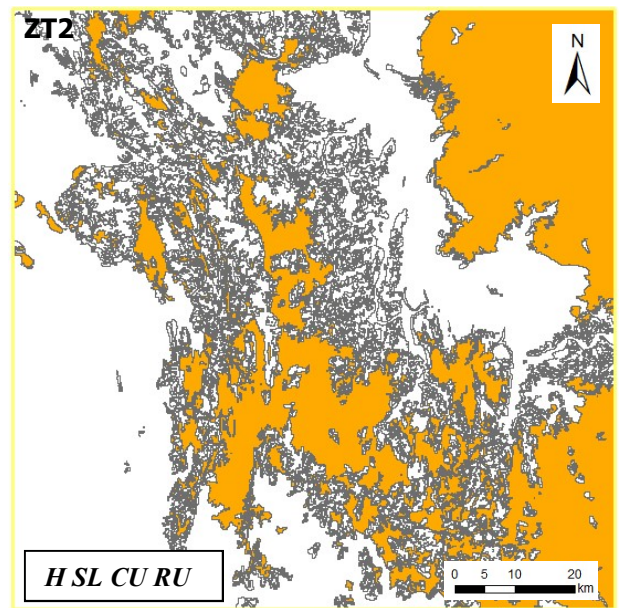
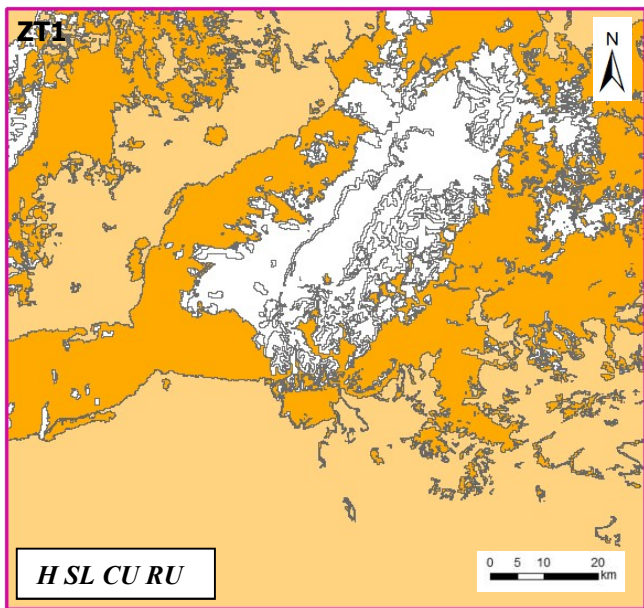


Annexe 1 Figure 1 - Echantillons d'apprentissage utilisés

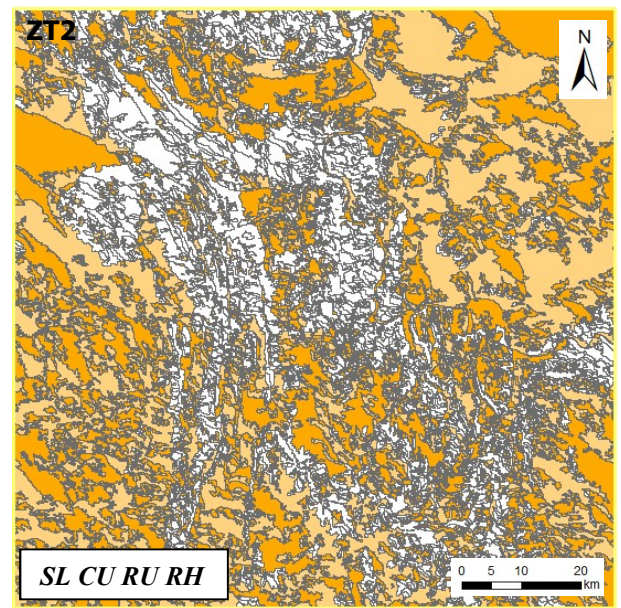
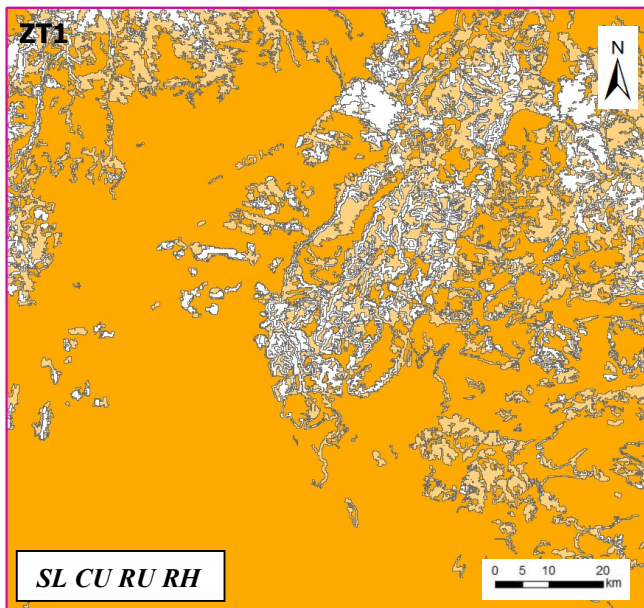
Résultats

IPS NS




A)



B)

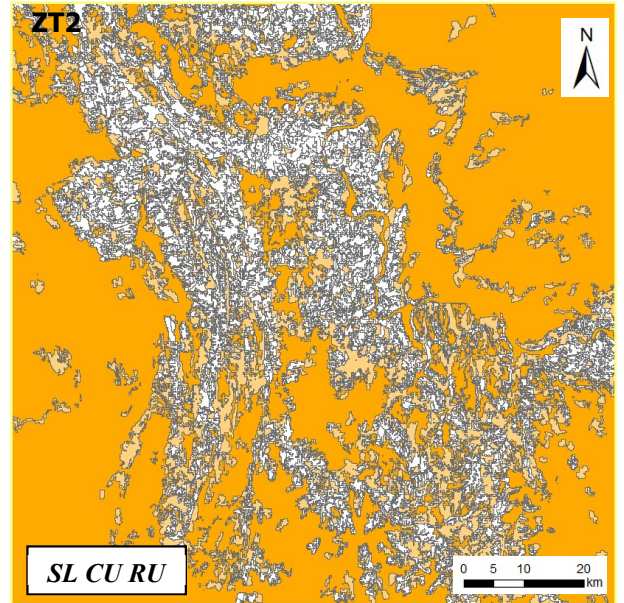
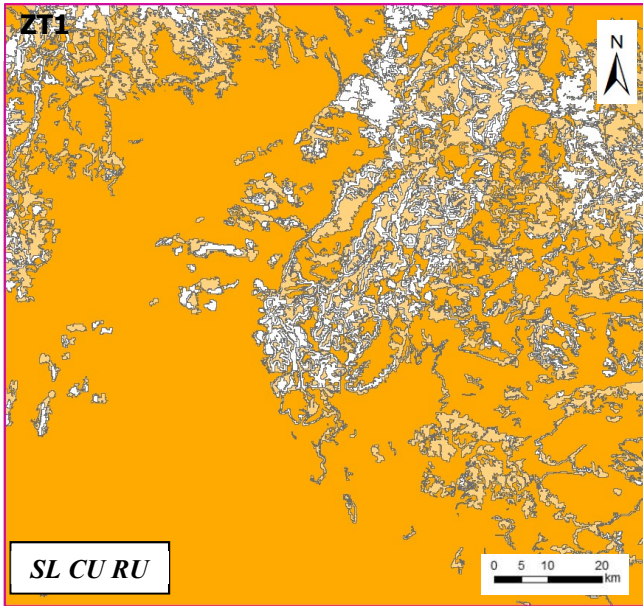


Type de surfaces

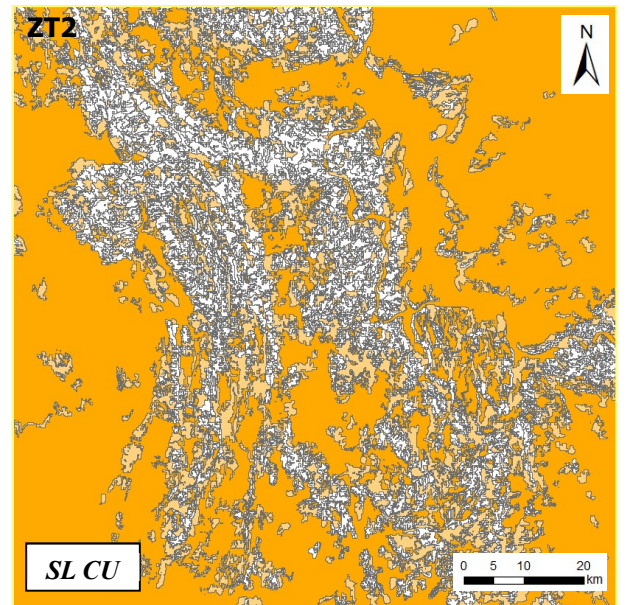
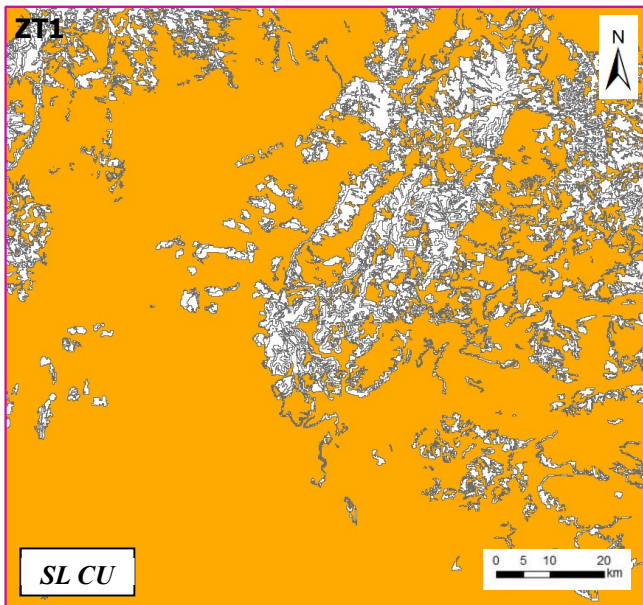
-  PS0.0
-  PS0.1
-  Autres surfaces

Annexe 1 Figure 2 - Résultats protocole IPS NS




C)



D)

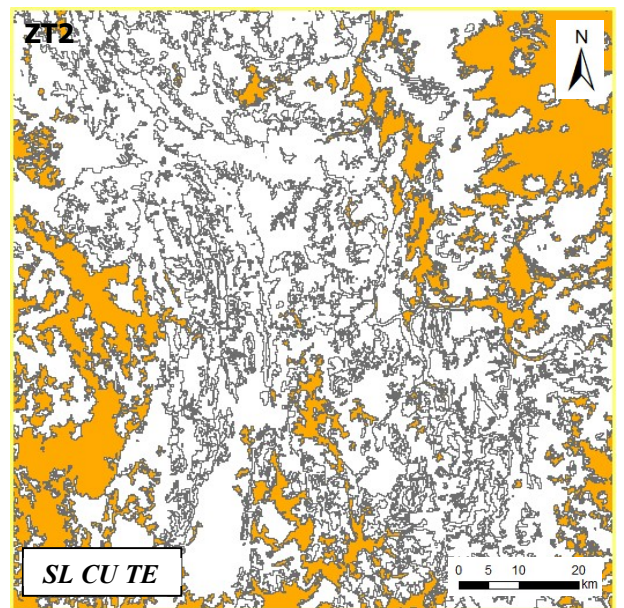
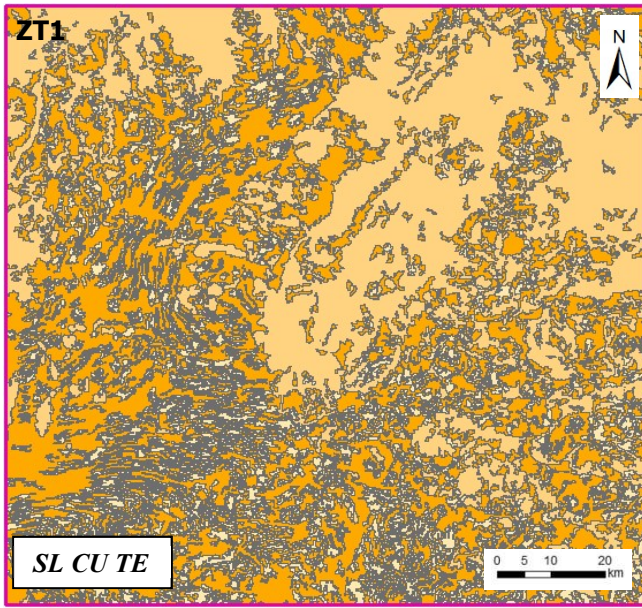


Type de surfaces




-  PS0.0
-  PS0.1
-  Autres surfaces

Annexe 1 Figure 3 - Résultats protocole IPS NS (suite)

E)

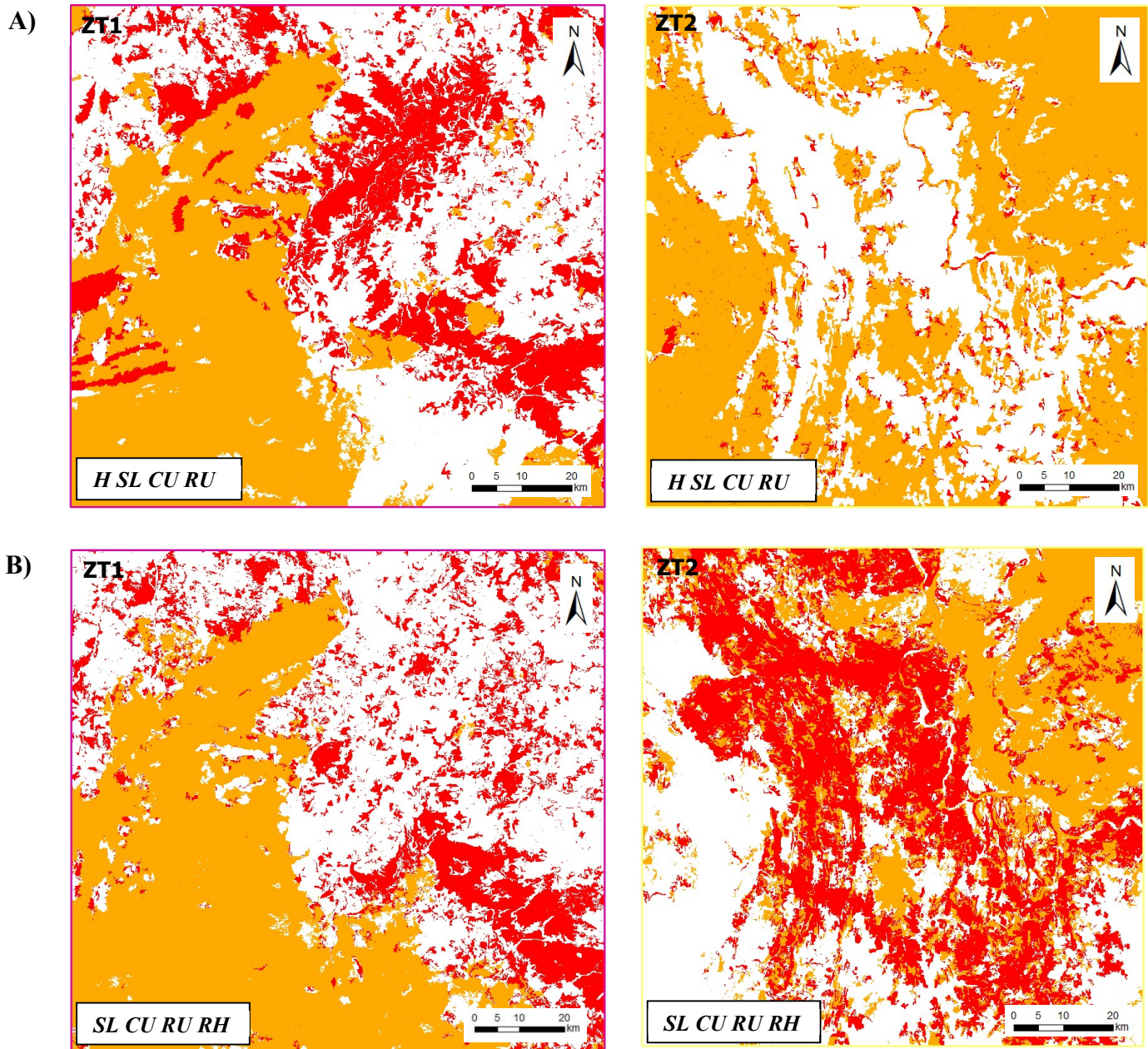


Type de surfaces




-  PS0.0
-  PS0.1
-  Autres surfaces

Annexe 1 Figure 4 - Résultats protocole IPS NS (suite)

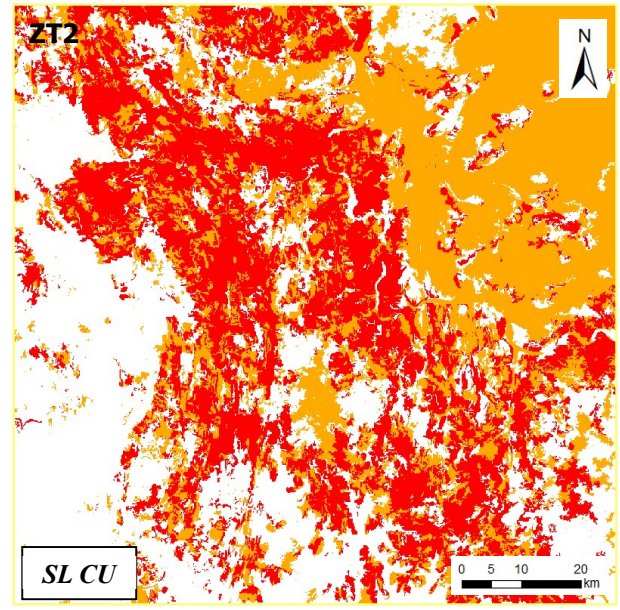
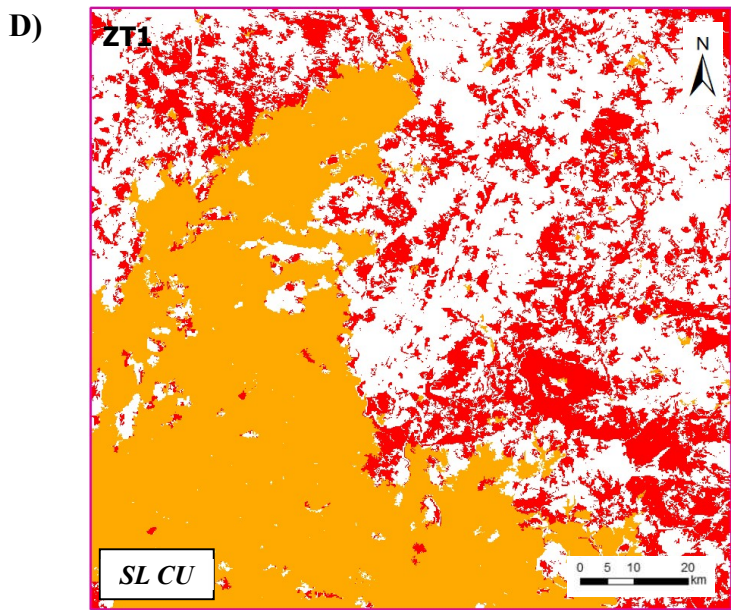
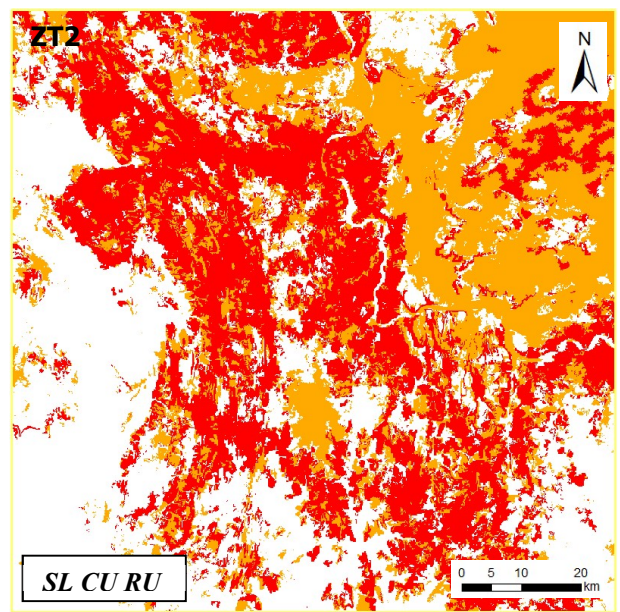
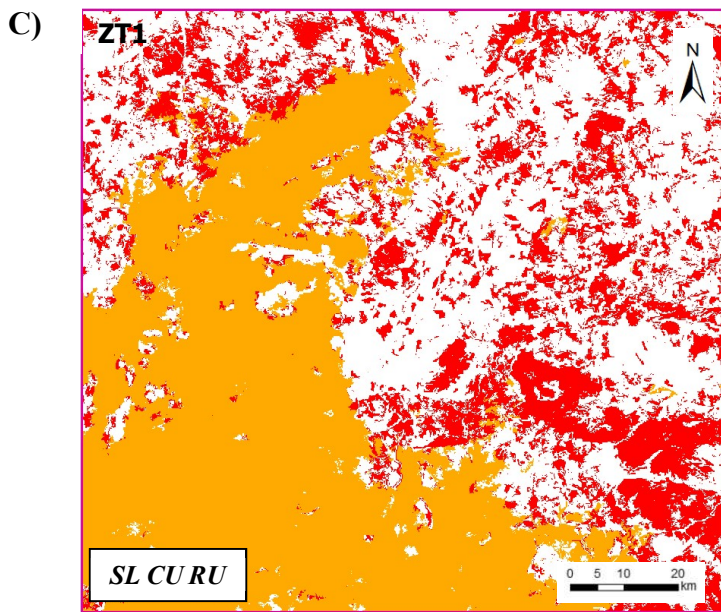
IPSS






Type de surfaces

-  PS1
-  PS2
-  Autres surfaces

Annexe 1 Figure 3 - Résultats protocole IPSS

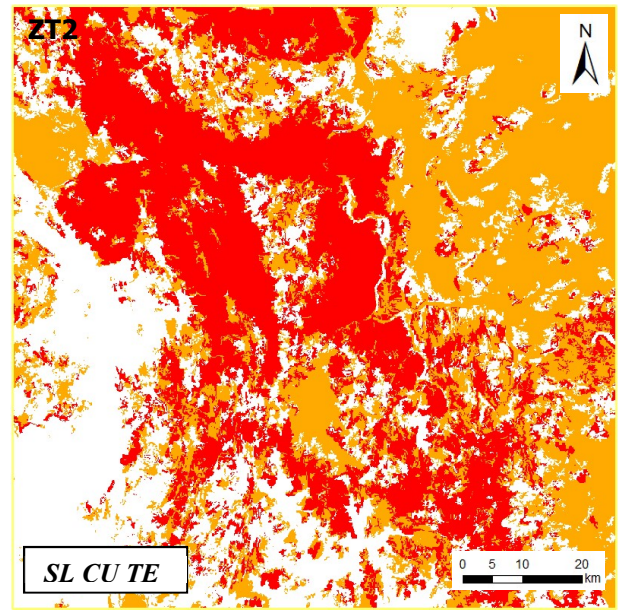
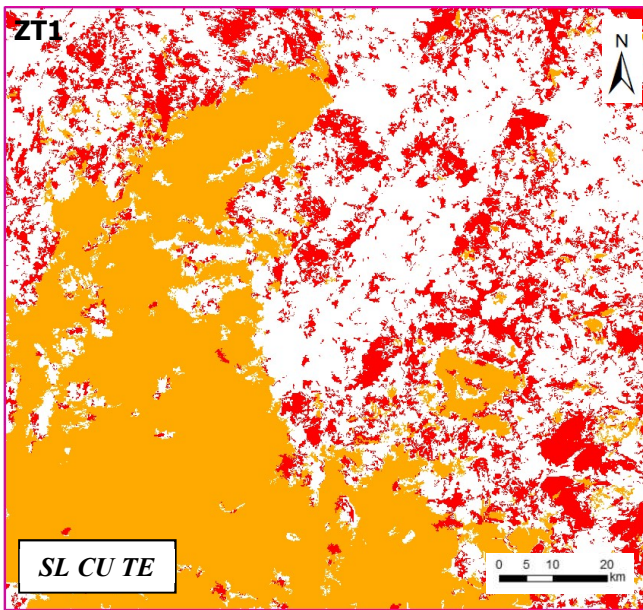


Type de surfaces




-  PS1
-  PS2
-  Autres surfaces

Annexe 1 Figure 3 - Résultats protocole IPS S (suite)

E)



Type de surfaces

-  PS1
-  PS2
-  Autres surfaces

Annexe 1 Figure 3 - Résultats protocole IPS S (suite)

Précisions obtenues pour chaque zone test

ZT1

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	3704	134	0	3838	0,965086	0
C_2	425	1259	0	1684	0,747625	0
C_3	615	961	0	1576	0	0
Total	4744	2354	0	7098	0	0
P_Accuracy	0,780776	0,534834	0	0	0,699211	0
Kappa	0	0	0	0	0	0,462806

H SL CU RU

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	970	39	0	1009	0,961348	0
C_2	423	1740	0	2163	0,804438	0
C_3	2928	127	0	3055	0	0
Total	4321	1906	0	6227	0	0
P_Accuracy	0,224485	0,912907	0	0	0,435202	0
Kappa	0	0	0	0	0	0,277048

ZT2

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	4113	85	0	4198	0,979752	0
C_2	142	940	0	1082	0,868762	0
C_3	489	1329	0	1818	0	0
Total	4744	2354	0	7098	0	0
P_Accuracy	0,86699	0,39932	0	0	0,711891	0
Kappa	0	0	0	0	0	0,480093

SL CU RU RH

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	1082	236	0	1318	0,820941	0
C_2	1184	1391	0	2575	0,540194	0
C_3	2055	279	0	2334	0	0
Total	4321	1906	0	6227	0	0
P_Accuracy	0,250405	0,729801	0	0	0,397141	0
Kappa	0	0	0	0	0	0,170249

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	4178	70	0	4248	0,983522	0
C_2	151	1053	0	1204	0,874585	0
C_3	415	1231	0	1646	0	0
Total	4744	2354	0	7098	0	0
P_Accuracy	0,880691	0,447324	0	0	0,736968	0
Kappa	0	0	0	0	0	0,516261

SL CU RU

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	1113	202	0	1315	0,846388	0
C_2	1253	1379	0	2632	0,523936	0
C_3	1955	325	0	2280	0	0
Total	4321	1906	0	6227	0	0
P_Accuracy	0,257579	0,723505	0	0	0,400193	0
Kappa	0	0	0	0	0	0,171635

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	4058	66	0	4124	0,983996	0
C_2	256	1089	0	1345	0,809665	0
C_3	430	1199	0	1629	0	0
Total	4744	2354	0	7098	0	0
P_Accuracy	0,855396	0,462617	0	0	0,725134	0
Kappa	0	0	0	0	0	0,499183

SL CU

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	1057	229	0	1286	0,821928	0
C_2	1043	1285	0	2328	0,551976	0
C_3	2221	392	0	2613	0	0
Total	4321	1906	0	6227	0	0
P_Accuracy	0,244619	0,674187	0	0	0,376104	0
Kappa	0	0	0	0	0	0,159465

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	3917	231	0	4148	0,944311	0
C_2	234	794	0	1028	0,772374	0
C_3	593	1329	0	1922	0	0
Total	4744	2354	0	7098	0	0
P_Accuracy	0,825675	0,337298	0	0	0,663708	0
Kappa	0	0	0	0	0	0,400962

SL CU TE

ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
C_1	867	217	0	1084	0,799815	0
C_2	1131	1397	0	2528	0,552611	0
C_3	2319	292	0	2611	0	0
Total	4317	1906	0	6223	0	0
P_Accuracy	0,200834	0,732949	0	0	0,363812	0
Kappa	0	0	0	0	0	0,157073

Annexe 1 Figure 5 - Matrices de confusion pour les zones tests

Annexe 2

Résultats complémentaires pour l'Argentine

Pour l'Argentine, j'ai fait le choix de seulement tester les combinaisons H SL CU RU, SL CU RU, SL CU et SL CU TE.

Détermination du K

Les K obtenus pour chaque combinaison sont résumés dans le tableau suivant.

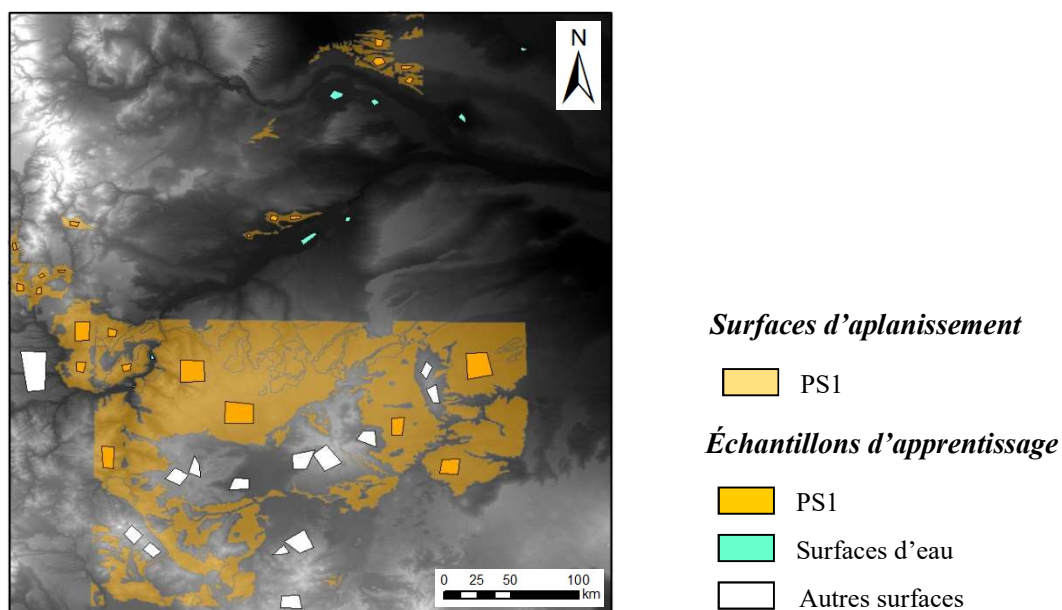
	H SL CU RU	SL CU RU	SL CU	SL CU TE
K	4	4	5	4

Détermination des classes

Le tableau suivant résume les classes identifiées comme PS pour chaque combinaison.

	H SL CU RU	SL CU RU	SL CU	SL CU TE
Classes	0; 2 et 3	0; 1 et 2	0; 3 et 4	0 et 1

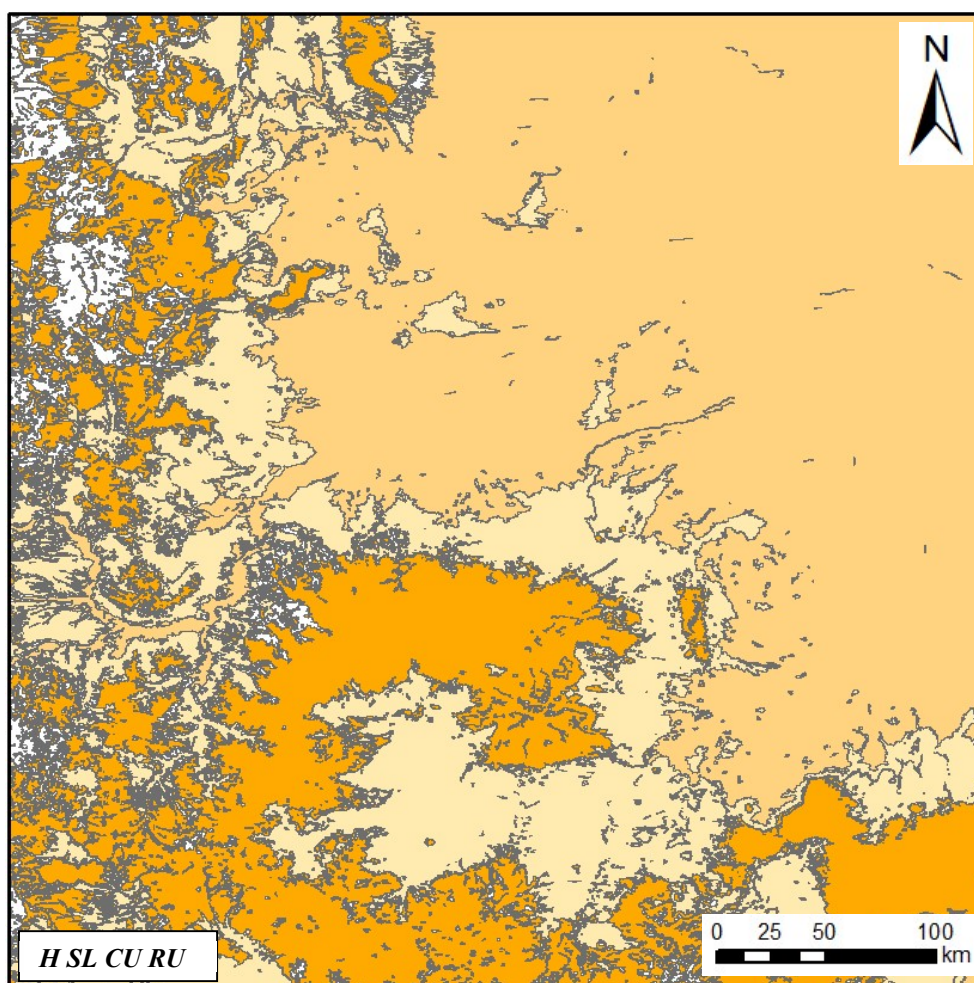
Échantillons d'apprentissage utilisés








Annexe 2 Figure 1 - Echantillons d'apprentissage utilisés pour l'Argentine

Résultats

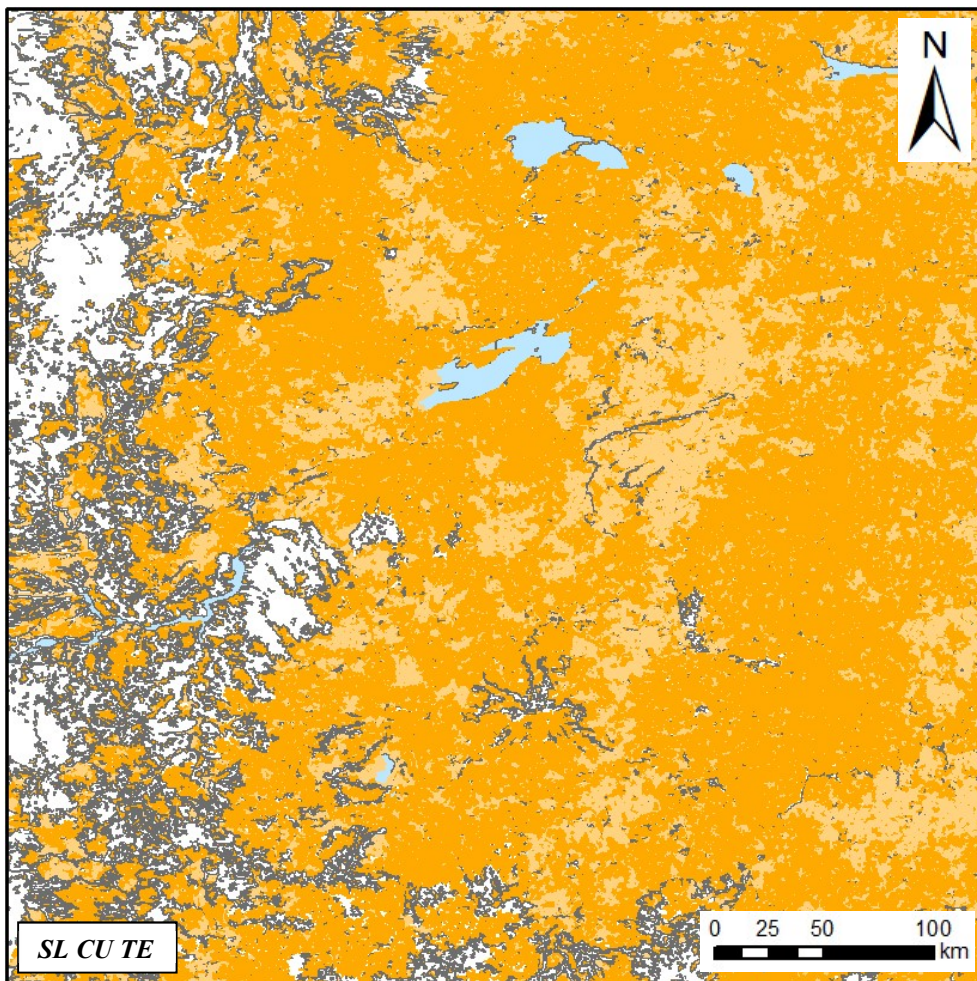
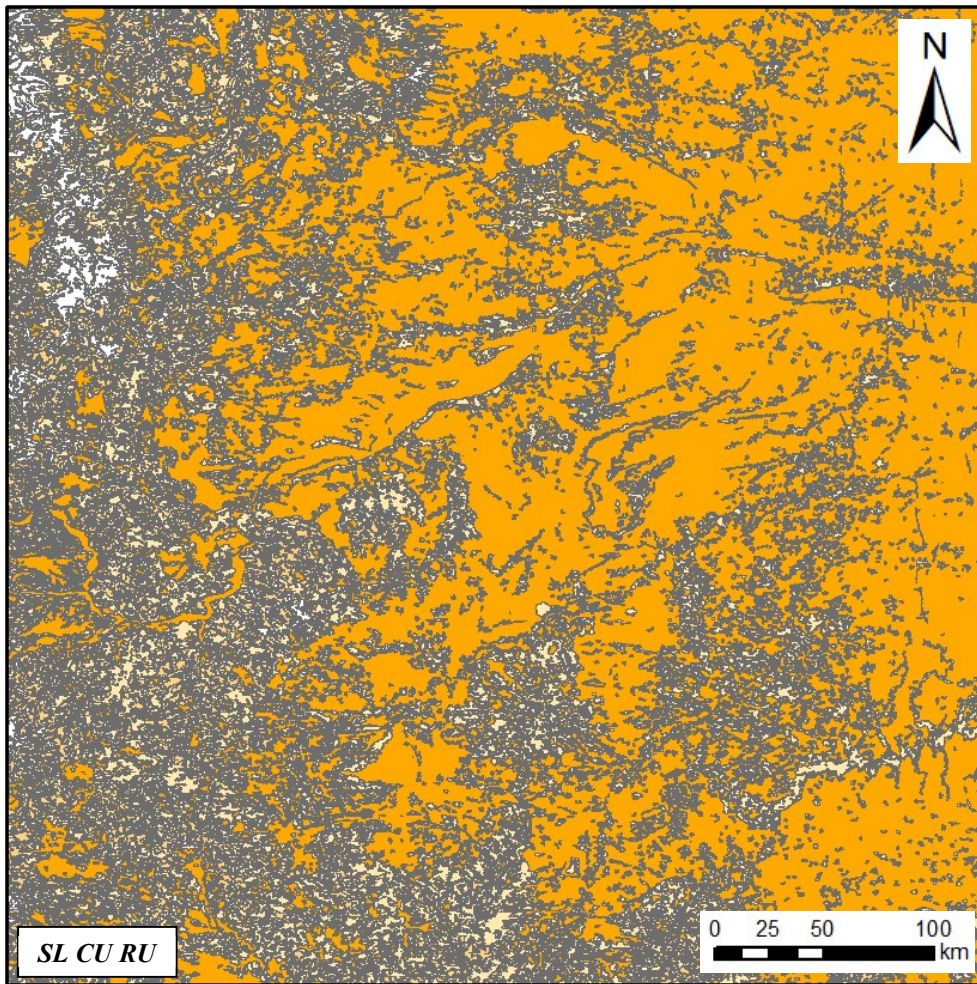
IPS NS



Type de surfaces

-  PS0.0
-  PS0.1
-  PS0.2
-  Surfaces d'eau
-  Autres surfaces

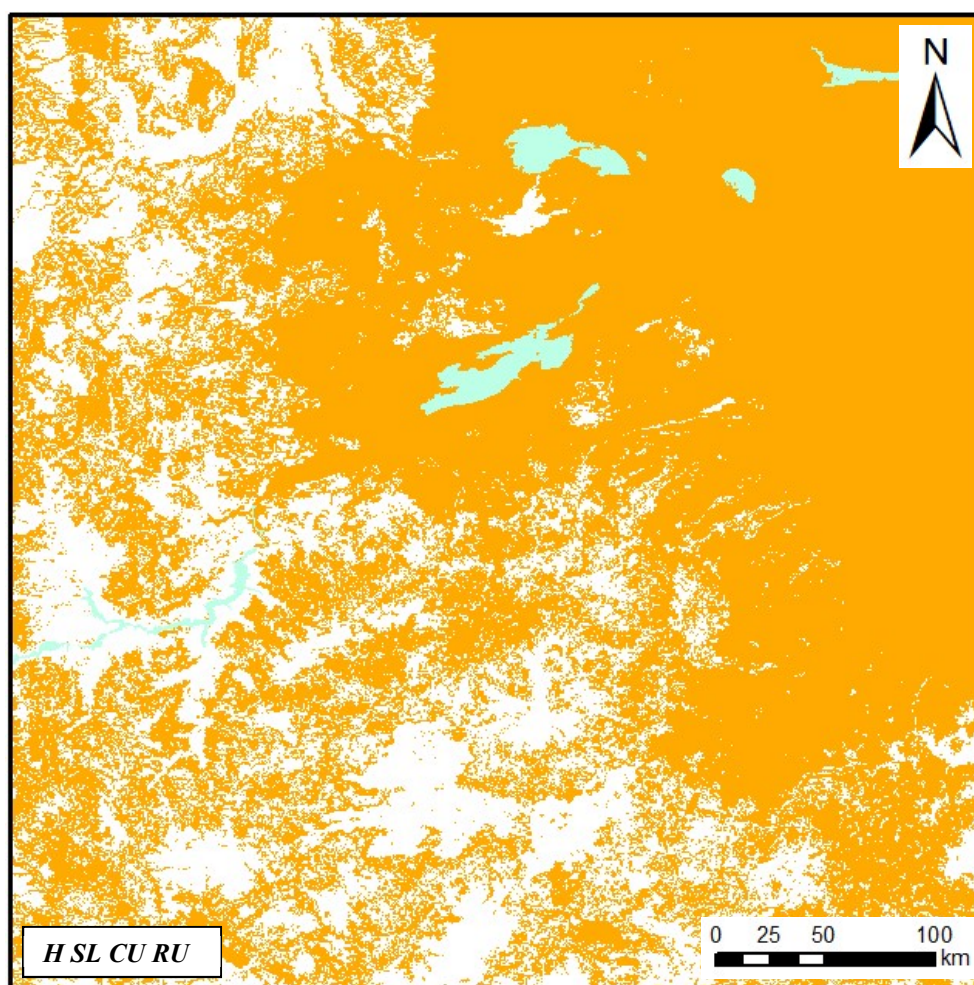
Annexe 2 Figure 2 - Résultats protocole IPS NS



Type de surfaces

- PS0.0
- PS0.1
- Surfaces d'eau
- Autres surfaces

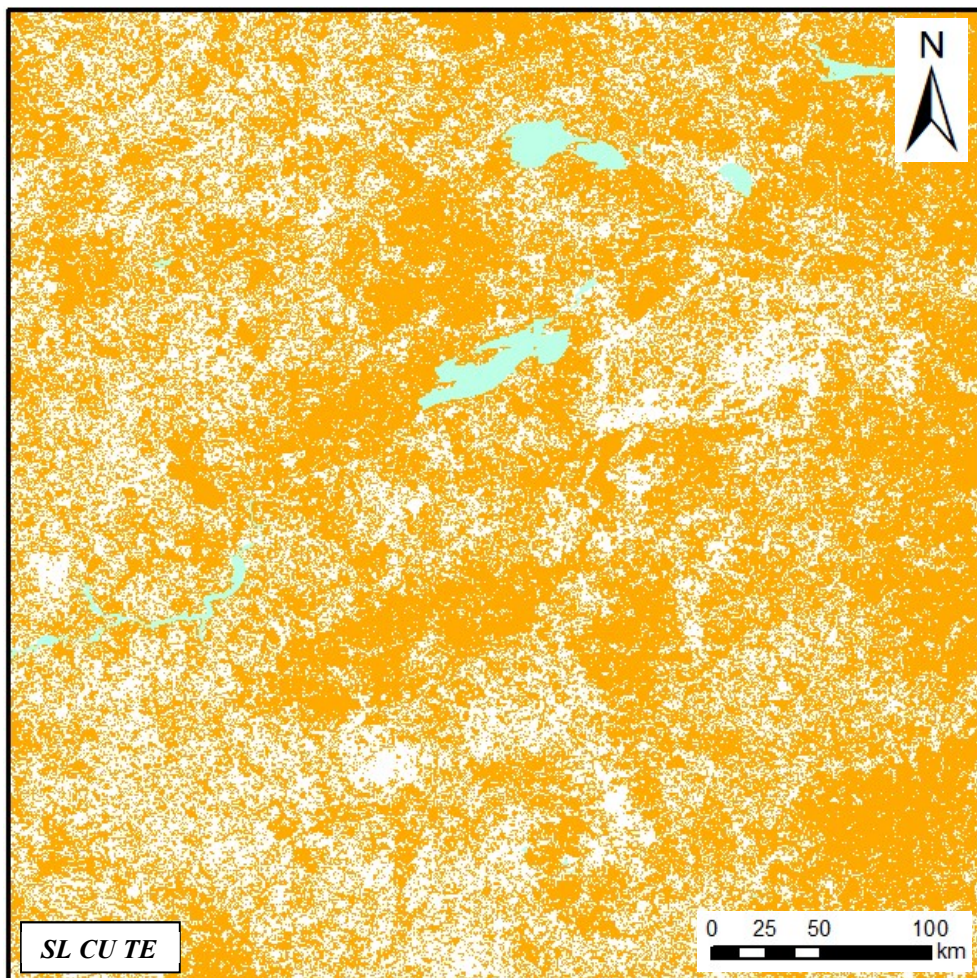
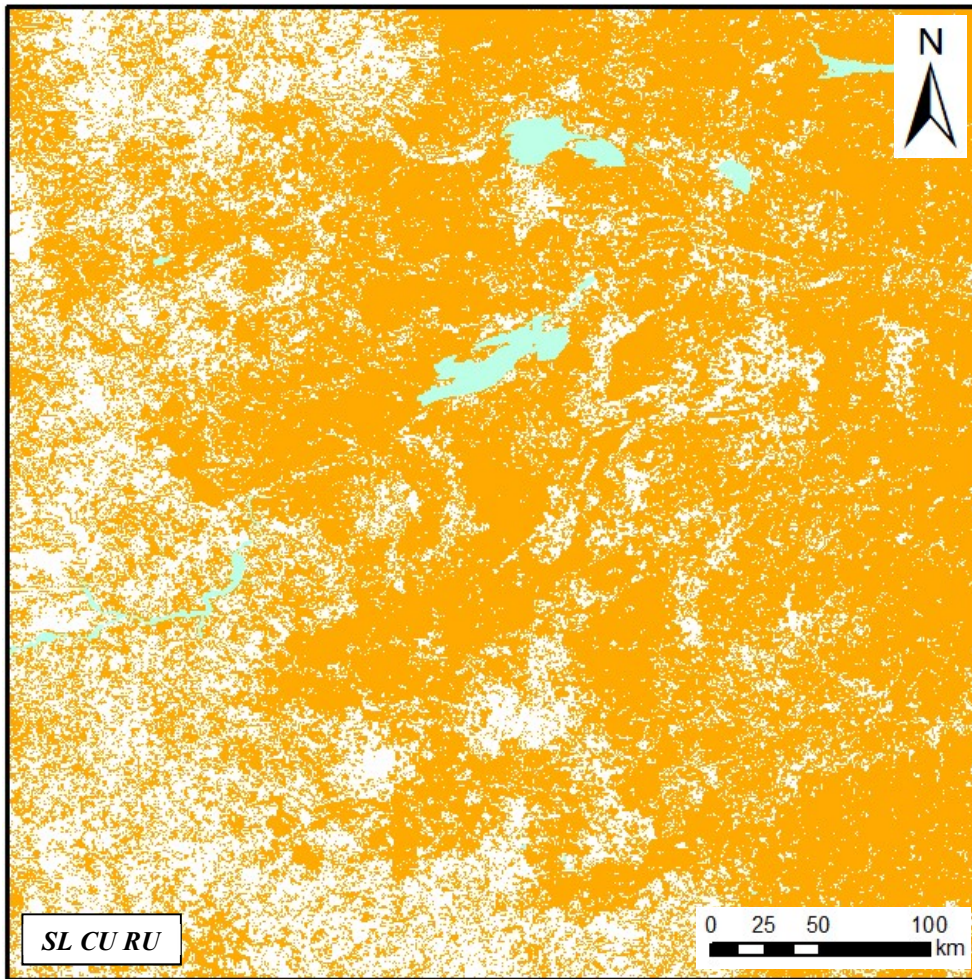
Annexe 2 Figure 3 - Résultats protocole IPS NS



Type de surfaces

- PS1
- Surfaces d'eau
- Autres surfaces

Annexe 2 Figure 3 - Résultats protocole IPS S



Type de surfaces

- PS1
- Surfaces
- Autres surfaces

Annexe 2 Figure 3 - Résultats protocole IPS S (suite)

Annexe 3

Statistiques et précisions pour l'Argentine

Pourcentages d'occupation

Protocole	IPS NS			IPS S		
	PS	Eau	Autre	PS	Eau	Autre
H SL CU RU	94%	0 %	6%	68%	1%	31%
SL CU RU	97%	0%	3%	69%	1%	30%
SL CU	90%	0%	10%	63%	1%	36%
SL CU TE	77%	3%	20%	62%	1%	37%

Matrices de confusion

IPS S							
<i>H SL CU RU</i>	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	4908	92	0	5000	0,9816	0
	C_2	158	55934	0	56092	0,997183	0
	C_3	265	31477	0	31742	0	0
	Total	5331	87503	0	92834	0	0
	P_Accuracy	0,920653	0,639224	0	0	0,655385	0
	Kappa	0	0	0	0	0	0,193668
	<hr/>						
<i>SL CU RU</i>	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	4974	100	0	5074	0,980292	0
	C_2	190	61325	0	61515	0,996911	0
	C_3	167	26078	0	26245	0	0
	Total	5331	87503	0	92834	0	0
	P_Accuracy	0,933033	0,700833	0	0	0,714167	0
	Kappa	0	0	0	0	0	0,232208
	<hr/>						
<i>SL CU</i>	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	4985	146	0	5131	0,971546	0
	C_2	177	55914	0	56091	0,996844	0
	C_3	169	31443	0	31612	0	0
	Total	5331	87503	0	92834	0	0
	P_Accuracy	0,935097	0,638995	0	0	0,655999	0
	Kappa	0	0	0	0	0	0,194971
	<hr/>						
<i>SL CU TE</i>	ClassValue	C 1	C 2	C 3	Total	U Accuracy	Kappa
	C_1	4989	114	0	5103	0,97766	0
	C_2	176	55682	0	55858	0,996849	0
	C_3	166	31707	0	31873	0	0
	Total	5331	87503	0	92834	0	0
	P_Accuracy	0,935847	0,636344	0	0	0,653543	0
	Kappa	0	0	0	0	0	0,19372

C_1 = Eau

C_2 = PS

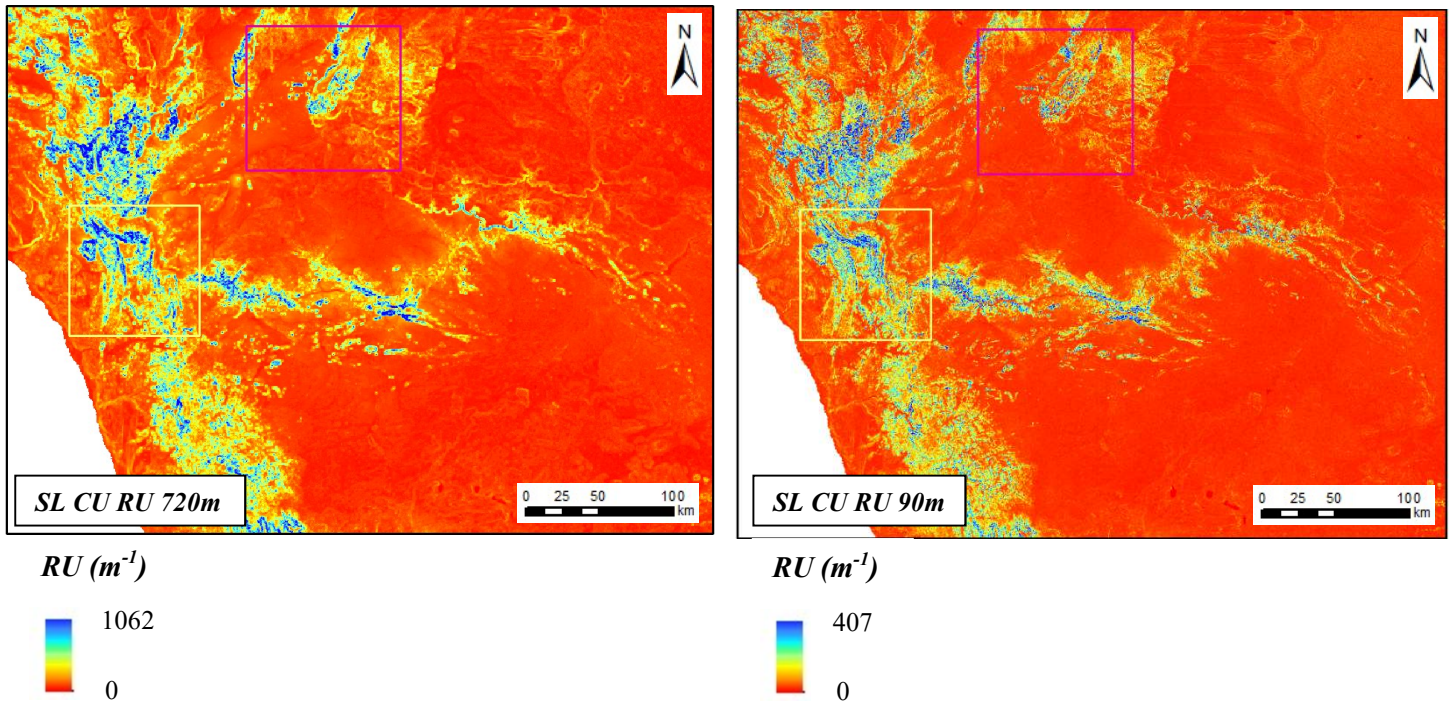
C_3 = Autre

Annexe 2 Figure 3 - Matrices de confusion pour l'Argentine

Annexe 4

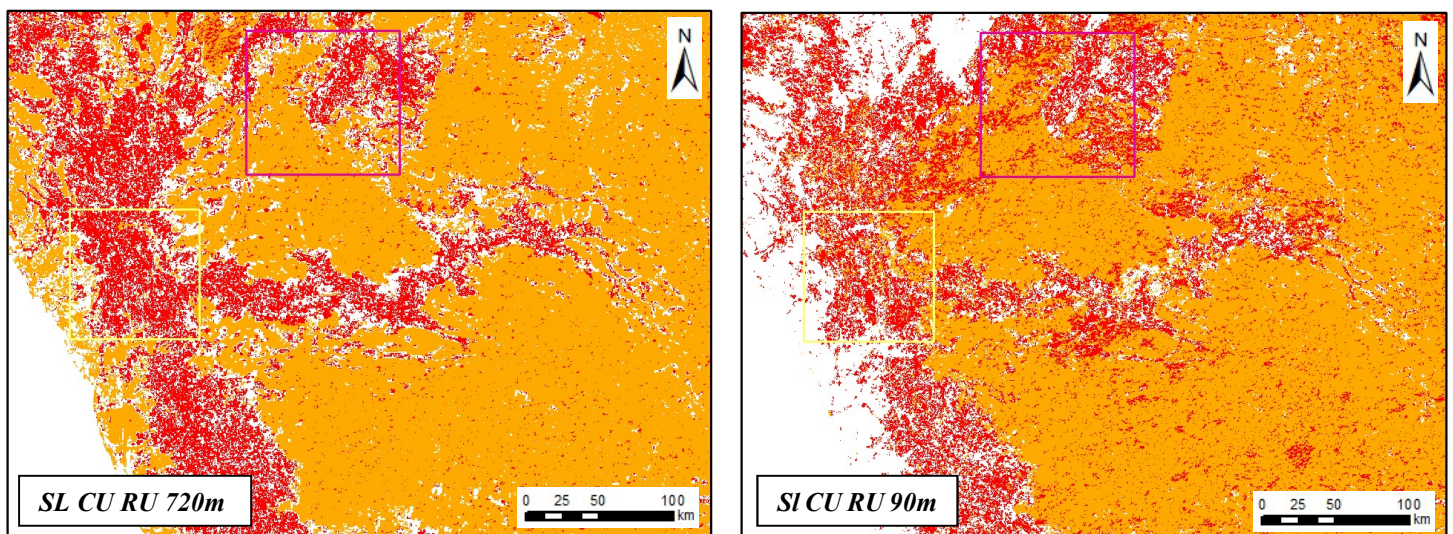
Comparaison RU à différentes résolutions

Calcul RU à différentes résolutions



Annexe 2 Figure 4 - Comparaison de RU à différentes résolution

Résultats après application du protocole IPS S



Type de surfaces

- PS1
- PS2
- Autres surfaces

Annexe 2 Figure 5 - Comparaison des résultats obtenus avec le protocole IPS S pour différentes résolution

Liste des figures

Figure 1 — Zones étudiées, Afrique (à gauche), Argentine (à droite).....	8
Figure 2 — Illustration d'une pénéplaine (Bessin, 2014).....	9
Figure 3 — Les différentes surfaces quasi planes (Bessin, 2014)	10
Figure 4 — Profil d'altération (Chardon et al., 2018).....	11
Figure 5 — Les différents types de surfaces de corrosions (Guillocheau et al., 2018).....	12
Figure 6 — Type de relief en fonction de la valeur du TPI et de la dimension du voisinage (Jeness, 2006).....	14
Figure 7 — Schéma illustrant la méthode d'Haider (Haider et al., 2015) (A) Illustration de l'algorithme, (B) Degrés d'appartenance définis.....	15
Figure 8 — Représentation des paramètres définissant différents types de relief extrait d'une zone d'étude au Tibet (Haider et al., 2015).....	16
Figure 9 — Profil d'altération, perméabilité et porosité associée (Lachassagne et al., 2011).....	17
Figure 10 — Surfaces d'aplanissement au sud de l'Afrique (A – Dauteuil et al., 2015 ; B – Baby et al, 2017).....	19
Figure 11 — Illustration de la segmentation (Chen et al.,2018).....	20
Figure 12 — Illustration de la classification (Chen et al., 2018).....	21
Figure 13 — Illustration de l'algorithme Mean Shift (Abrishami, 2015).....	22
Figure 14 — Illustration de l'algorithme Seeded Region Growing (Mehnert, 1997).....	22
Figure 15 — Illustration de l'algorithme Watershed (Beucher, 2010).....	23
Figure 16 — Illustration de l'algorithme K-Means (Landman et al., 2019).....	24
Figure 17 — Illustration simplifiée de l'algorithme Random Forest (Koershen, 2017)	24
Figure 18 — Comparaison des résultats de segmentation et classification	27
Figure 19 — Zone d'étude africaine (ZA).....	29
Figure 20 — Zone d'étude argentine (ZAR)	30
Figure 21 — Illustration des protocoles élaborés	32
Figure 22 — Détermination de K pour H SL CU RU.....	34
Figure 23 — Exemple de représentation des descripteurs par classes.....	35
Figure 24 — Échantillons d'apprentissage définis	36
Figure 25 — Résultats protocole IPS NS.....	37
Figure 26 — Résultats protocole IPS S.....	40
Figure 27 — Résultats protocole IPS NS avec SL, CU pour ZAR.....	43
Figure 28 — Résultat protocole IPS S avec SL CU pour ZAR	44
Figure 29 — ACP pour H SL CU RU	47
Figure 30 — ACP pour SL CU RU RH.....	48
Figure 31 — ACP pour SL CU TE.....	49
Figure 32 — Illustration des résultats obtenus pour des échantillons d'apprentissages différents	51

Liste des tableaux

Tableau 1 — Algorithmes de Segmentation pour chaque logiciel.....	20
Tableau 2 — Algorithmes de classification pour chaque logiciel.....	21
Tableau 3 — Comparaison des algorithmes.....	26
Tableau 4 — Avantages et inconvénients pour la segmentation.....	28
Tableau 5 — Avantages et inconvénients pour la classification.....	28
Tableau 6 — K déterminé pour chaque combinaison	34
Tableau 7 — Classes identifiées comme PS	35
Tableau 8 — Comparaison des pourcentages d’occupation par classe pour chaque test pour ZA ..	45
Tableau 9 — Matrices de confusion	46

Développement et comparaison de protocoles semi-automatisés de détection d'objets géomorphologiques : les surfaces d'aplanissement.

Mémoire d'Ingénieur C.N.A.M., Le Mans 2019

RÉSUMÉ

Les cartes de surface d'aplanissement (PS) sont principalement construites soit par numérisation manuelle, soit par classification semi-automatique de pixels, qui ne tiennent pas compte des relations entre les pixels voisins, prennent beaucoup de temps et peuvent être subjectives. Cependant, l'analyse d'image orientée objet évite ces biais en utilisant des objets. Cette méthode semi-automatisée repose sur i) une étape de segmentation suivie d'une ii) classification. Nous avons développé deux protocoles différents en utilisant les données SRTM 30arcsecond et ses dérivés. Les deux protocoles diffèrent par leur mode de classification. Le premier utilise une classification non supervisée. Le second utilise une classification supervisée basée sur de l'apprentissage automatique. Ces deux protocoles nous permettant d'identifier des PS à partir de leur pente et courbure avec une précision de 70 %.

Mots clés : surface d'aplanissement, segmentation, classification, OBIA

SUMMARY

Planation surface (PS) maps are mainly built either by manual digitalization or through semi-automated pixel classification, which does not consider relationships between neighboring pixels, can be highly time-consuming and subjective. However, Object Based Image Analysis avoid these biases by using objects. This semi-automated method relies on i) a segmentation step followed by ii) classification. We developed two different protocols using SRTM 30arcsecond data and its derivatives. Both protocols differ in their classification step. The first uses an unsupervised classification based on a clustering algorithm. The second uses a supervised classification based on a machine-learning algorithm from user-defined training samples. Both protocols allow us to identify PS from their slope and curvature with accuracy around 70%.

Key words: planation surfaces, segmentation, classification, OBIA