

ENSG
Géomatique

ÉCOLE NATIONALE
DES SCIENCES
GÉOGRAPHIQUES




UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE
UFR GÉOGRAPHIE

 **Cerema**

Rapport de stage 

Licence professionnelle Géomatique et environnement

Apport de levés topo-bathymétriques pour l'identification des ouvrages situés dans les fonds côtiers



Claude Guillet, Cerema 2010

Loïc GENTILHOMME

le 05/09/2019

 Non confidentiel

Président(e) du jury :

Jean-François HANGOUËT , ENSG

Commanditaire :

Cerema Normandie Centre – Pôle littoral
10 Chemin de la Poudrière, 76120 Le Grand-Quevilly

Encadrement de stage :

Pierre VIGNÉ, Cerema, Maître de stage
Sébastien SAUR, ENSG/IGN, Professeur référent

Responsables pédagogiques du cycle :

- pour Paris1, UFR08 :

Brice ANSELME (Maître de conférences)

- pour l'ENSG :

Francis DHÉE (Responsable des enseignements en cartographie)

Patricia PELLARDI (Enseignante photogrammètre)

Nicolas PAYET (Enseignant cartographe)

Stage du 03/06/2019 au 23/08/2019

Diffusion Web :  Internet

Situation du document : rapport de stage présenté en fin de 3^{ème} année de Licence du cycle Licence Professionnelle MPGE, parcours imagerie.

Nombre de pages : 40 dont 3 d'annexes

Remerciements

Je tiens tout d'abord à remercier mon maître de stage Pierre VIGNÉ pour son accompagnement et ses conseils tout au long du stage.

Je remercie aussi Hervé DUSSART mon colocataire de bureau qui a eu la patience de répondre à toutes les petites questions sur un milieu littoral que je connaissais peu, ainsi que toute l'équipe du DADT et du Cerema qui m'ont beaucoup appris pendant ces trois mois.

Je termine en remerciant les personnes impliquées en dehors du Cerema, représentants de l'Université de Caen, du Réseau d'Observatoire du Littoral Normandie Hauts de France ou de l'ENSG.

Table des matières

Introduction.....	8
I État de l'art et présentation des données.....	9
I.a) Présentation des données LIDAR.....	9
A. le nuage de point.....	9
B. les attributs spécifiques au LIDAR.....	9
C. les produits principaux du LIDAR.....	10
I.b) Inspiration bibliographique.....	10
A. Classification des bâtiments par imagerie optique spatiale.....	10
B. Classification d'objets en utilisant les données LIDAR Bathymétrique.....	11
C. Classification Random-Forest et Automatisation pour l'identification de zone humide en s'appuyant sur des descripteurs dérivés du MNT.....	11
I.c) Résumé des données fournies.....	12
A. Format des données.....	12
B. Limites des données.....	12
C. Précision et Densité.....	13
D. Zone d'étude et taille des données.....	14
II Traitement des fichiers las et Classification.....	16
II.a) Principe du traitement.....	16
II.b) Programme et Automatisation.....	23
A. Prérequis du programme.....	23
B. Fonctions principales.....	23
II.c) Classification.....	25
III Résultat et ouverture.....	27
III.a) Autre application du LIDAR sur le milieu côtier.....	27
A. Suivi du trait de côte.....	27
B. Suivi de la Conchyliculture.....	28
III.b) Qualité des résultats.....	30
A. Pertinences des données LIDAR pour la photo-interprétation.....	30
B. Classification.....	32
C. Temps de calcul et taille des données.....	33
Conclusion.....	35

Index des illustrations

Exemple de nuage de point.....	9
Multi-écho illustré dans la thèse de Thomas RISTORCELL pour l'Université de Toulouse.....	9
Illustration de la méthode décrite dans l'article.....	10
Tableau des résultats de la classification dans l'étude citée. Les 80-20% correspondent respectivement au taux d'apprentissage et de contrôle.....	11
Spécifications annoncées du produit.....	13
Exemple de paysage de la première zone. Credit Alain Freytet, Paysagiste.....	14
Pointe du Roc rendu par un des traitements. On peut y voir un fichier corrompu qui a été écarté de l'étude.....	14
Carte de la zone 3. Le point jaune est la plage Utah Beach, située à l'ouest de la baie de Vey. Source IGN et SHOM.....	14
Extrait du plateau du Calvados avec en rouge les éléments artificiels sur l'estran. Source: Cerema.....	15
Répartition des .las dans la zone de Dunkerque. En rose les dalles LIDAR et en rouge la zone choisi originalement. Source orthophoto: Cerema.....	15
Carte des 5 Zones étudiées.....	15
Tous les exemples sont sur la même zone ici montrée en photo aérienne. Source : Cerema.....	17
MNS calculé par CloudCompare.....	17
Image de contour par soustraction.....	19
Image d'intensité.....	19
Filtre de contour kernel.....	20
Filtre de contour par soustraction.....	20
Filtre de tortuosité.....	21
Image de sortie convexe RVB.....	22
En rouge, les lignes détectées sur le filtre kernel, en vert sur le filtre morphologique convexe.....	22
Une des premières classifications, la falaise n'est pas encore présente dans les zones d'apprentissages	26
Zone d'exemple pour la classification. Fausse couleur par bande de contour et tortuosité.....	26
La classification finale, avec encore quelques erreurs difficiles à corriger.....	26
OrthoPhoto, MNS et MNT de falaise complexe au sud de Dieppe. Source orthophotographie Cerema.	27
Suivis d'une ligne de crête par le traitement.....	28
Parc de conchyliculture.....	28
Conchyliculture dans la zone 3 avec en orange les concessions enregistrés dans les bases de données du CEREMA.....	28
Résumé du traitement sur la conchyliculture.....	29
Épave du Videcoq, au large de Granville, a plus de 7 km des côtes et par 8m de fond.....	31
Balise permanente située en mer. Elle est visible sur l'image en fausse couleur et sur la carte maritime du SHOM.....	31
Épave de la seconde guerre mondiale dans le plateau du Calvados.....	31
Répartition de 3 classes d'objets sur les zones d'études. 2 des 5 zones n'en présentait aucun.....	31
Exemple d'une classification test de la zone 4, en zone sous-marine principalement.....	32

Index des tableaux

Tableau 1: Statistiques de la numérisation manuelle.....	38
Tableau 2: Statistiques de la classification sur Dieppe.....	39
Tableau 3: Quelques statistiques du temps de calcul pour les 5 zones. Tous les pourcentages sont exprimés sur le temps total du traitement.....	41
Tableau 4: Statistiques de temps écoulé par zones.....	42

Glossaire et Sigles

LIDAR: Light Detection and Ranging. Principe de détection d'objets et de mesure de distance en utilisant un laser.

ROL : ~~Réseau d'observatoire du littoral. Ce sont des organismes qui visent l'acquisition de données pertinentes et systématiques sur l'évolution des plages et du trait de côte, de capitalisation de ces informations, et de leur mise à disposition auprès d'un large public.~~ Le ROL cité est toujours celui de Normandie et des Hauts-de-France

Shom : Le Service Hydrographique et Océanographique de la Marine est un établissement public administratif (EPA) sous tutelle du ministère des Armées. Il est l'un des fournisseurs les plus importants de données sur les fonds côtiers.

Las : format de fichier de nuage de point, souvent associé au LIDAR.

MNT : Modèle numérique de terrain. C'est une donnée géographique qui permet d'appréhender la forme du terrain sur un lieu. Souvent présenté comme une image, ou raster à maille carrée, il enregistre pour chaque pixel l'élévation du sol, sans prendre en compte le bâti ou la végétation.

MNS : Modèle numérique de surface. Proche d'un MNT, cette donnée enregistre sur chaque pixel l'élévation de l'objet le plus haut, y compris la végétation ou les bâtiments. Les câbles suspendus n'y sont par contre pas pris en compte.

Raster : Format de donnée matriciel encodé en bits. Dans le langage courant, un raster est une image, par exemple une photographie ou un dessin.

Temps GPS : Heure commune à tous les satellites GPS. Cette valeur extrêmement précise et homogène sur toute la surface du globe permet une échelle de temps commune pour les données temporelles.

Conchyliculture : Élevage de coquillages, comme les huîtres ou les moules.

Estran : Étendue côtière qui est à la fois immergée et émergée en fonction des marées. C'est un espace de contact et d'échange entre le milieu marin et terrestre.

Bibliothèque de fonction : Ensemble de fonctions déjà codées mis à disposition afin de faciliter le développement de programmes informatiques.

Variable d'environnement Windows : Chemin des différentes fonctionnalités disponibles pour Windows pour de multiples applications. Sert entre autre de chemin pour que la console Windows puisse connaître et utiliser des nouvelles fonctions.

VRT : Format Virtual Raster. Ce format géoréférencé est seulement la liste des images à afficher et leur chemin d'accès. C'est un moyen pratique de produire un grand raster sans pour autant avoir une image de plusieurs Gigaoctets.

Épi : Ouvrage perpendiculaire au trait de côte, servant à retenir le sable. Il se présente souvent sous la forme de long mur en béton.

Platiers : zone de l'estran principalement rocheux, c'est un lieu privilégié de la biodiversité de l'estran.

Qgis : Logiciel de gestion de données spatiales. Ce type de logiciel est appelé SIG (système d'information géographique.)

Introduction

Adoptée en 2008, la Directive cadre stratégie pour le milieu marin (DCSMM) met en place un cadre permettant aux États membres de l'Union Européenne de prendre toutes les mesures destinées à atteindre ou maintenir un bon état écologique du milieu marin. Dans cette perspective, pour le suivi de l'artificialisation des côtes, l'Union Européenne a introduit en 2018 des indices d'artificialisations avec la vocation de les mettre à jour tous les six ans. Un de ces indices est l'artificialisation des fonds côtiers situés entre 0 et 10 m de profondeur.

Mais pour obtenir les données pour cette zone d'étude sous-marine peu de méthodes sont disponibles, et les données déjà existantes sont dispersées, peu précises voire inexistantes.

L'utilisation du LIDAR bathymétrique a alors été envisagée, et en réponse à un appel à projet du Réseau d'Observatoire du Littoral (ROL) de Normandie et des Hauts-de-France pour l'utilisation de données LIDAR bathymétriques produites en partenariat avec le Service Hydrographique et Océanographique de la Marine (Shom), l'Agence de l'Eau de Seine-Normandie, l'Agence de l'Eau Artois-Picardie, le parc marin estuaire picard mer d'opale, les régions Normandie et Hauts-de-France, l'État et l'Europe, un stage fut mis en place par le CEREMA Normandie-Centre pour évaluer la pertinence de ces données pour l'aide à la création de cet indicateur et pour le suivi du littoral en général.

Ce stage de trois mois au pôle littoral du CEREMA Normandie-Centre est donc une évaluation des possibilités de ces données pour la détection d'ouvrages artificiels.

Pour cela, deux axes ont été mis en avant : l'aide à la photo-interprétation, avec une donnée LIDAR transformée afin de faciliter l'identification visuelle, et l'automatisation à l'aide d'une classification pour retrouver les différents ouvrages.

Dans ce rapport nous reviendrons sur les données LIDAR en général et sur les données fournies, ainsi que sur les différentes études qui ont aidé à choisir la méthode utilisée lors du stage. Nous verrons ensuite les différents indices qui ont été réalisés pour accompagner la photo-interprétation et la classification, puis les fonctions et la programmation afin de rendre fonctionnels ces indices sur un grand nombre de données et enfin sur la classification qui a été utilisée pour l'automatisation de la détection. Pour finir quelques pistes d'approfondissements seront présentées, et nous ferons un retour sur les résultats des différentes parties.

I État de l'art et présentation des données

I.a) Présentation des données LIDAR.

Le LIDAR est un appareil qui mesure le temps de retour d'une onde laser émise afin d'en déduire la distance entre l'appareil et l'objet qui a réfléchi le laser. Ici le LIDAR utilisé utilise une bande verte contrairement aux longueurs d'onde infrarouge utilisées la plupart du temps, et ce afin de pouvoir avoir un retour sous le niveau de l'eau. Cela nous permet d'avoir une modélisation des fonds sous-marins de basses profondeurs (moins de 15 m).

Nous allons effectuer ici un tour d'horizon de ce que le LIDAR peut offrir, ce qui ne sera pas forcément adapté aux données disponibles.

A. le nuage de point

Les données les plus importantes du LIDAR sont les nuages de points 3D géoréférencés. Ils sont obtenus en géoréférencement direct, c'est-à-dire calculé à l'aide des données de trajectographie de l'avion et des données d'angle et distance calculées par l'appareil. Cela fournit un nuage où chaque point possède des coordonnées x, y et z.

Le format de ces données est standardisé en un .las, où chaque point possède des attributs de géoréférencement mais aussi de nature comme l'angle d'acquisition ou le temps GPS d'acquisition. À noter que le .las possède des points repartis de façon non régulière et non ordonnée sur la zone d'acquisition, rendant la lecture des données plus longue qu'un format de type raster comme un MNT.

Ce nuage peut être livré classifié avec une série de classe standardisée comprenant entre autre le sol, le bâti et la végétation.

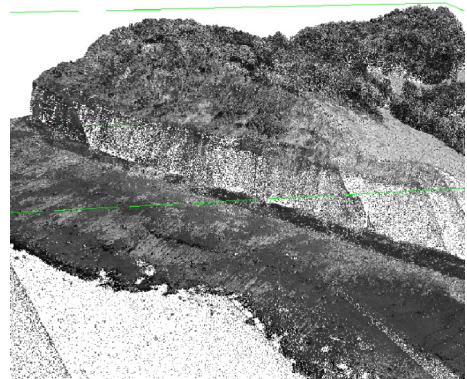


Figure 1: Exemple de nuage de point

B. les attributs spécifiques au LIDAR

Dans les données des points d'acquisitions se trouve d'autres informations spécifiques au LIDAR.

Une particularité du LIDAR est la possibilité de multi-écho sur un envoi de signal. La zone d'acquisition n'étant pas un point mais un cercle, les éléments comme les bords de bâtiment, les obstacles fins comme les branches d'arbres ou les zones de multi-

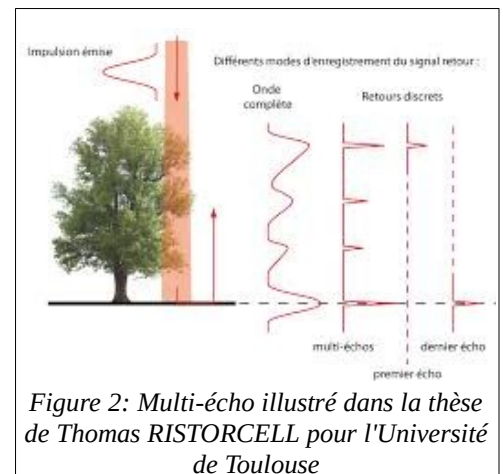



Figure 2: Multi-écho illustré dans la thèse de Thomas RISTORCELL pour l'Université de Toulouse

trajets donnent plusieurs retours sur un signal, permettant ensuite une meilleure classification des objets comme la végétation.

Ce multi-écho peut être complété par le Retour d'Onde Complète (ROC), permettant de sauvegarder non seulement les pics d'intensités mais aussi la forme des pics souvent modélisés par une courbe gaussienne. Cette forme permet d'obtenir des informations de rugosité et de réflectance de l'objet observé.

C. les produits principaux du LIDAR

La classification obtenue sur le nuage permet d'obtenir facilement un MNT raster à maille carrée par interpolation du MNT triangulé. Un MNS raster à maille carrée est aussi facilement produit, et dans notre cas les deux produits sont très proches comme nous le verrons juste après. 

I.b) Inspiration bibliographique.

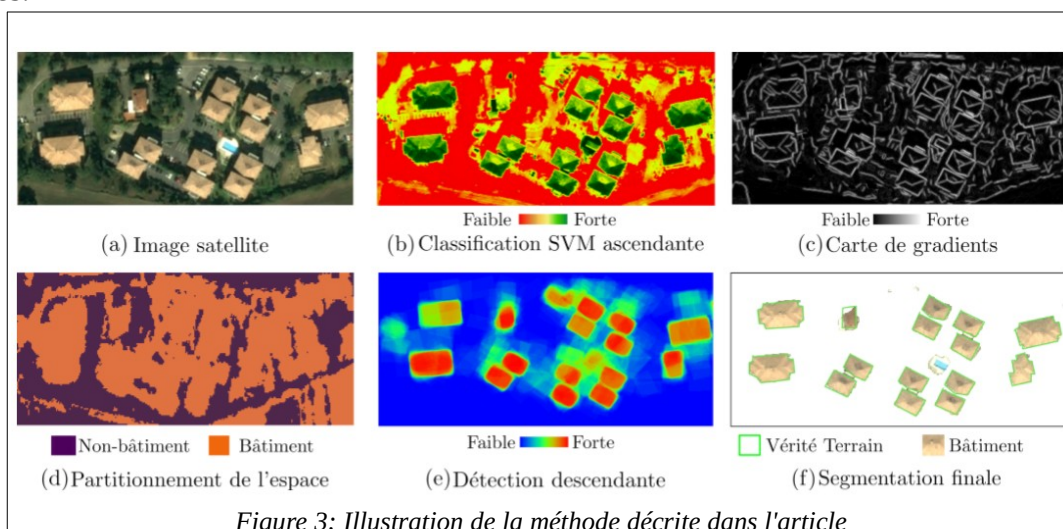
Ce stage étant un travail de recherche, il a fallu s'inspirer de travaux réalisés sur des données semblables ou avec une méthodologie proche afin d'orienter le sujet.

Le travail de recherche bibliographique a donc été une priorité durant les deux premières semaines. Cela coïncidait à peu près avec la période de livraison des données.

A. Classification des bâtiments par imagerie optique spatiale.

Article : <https://hal.archives-ouvertes.fr/hal-00989026/document>

Cette étude porte sur la numérisation de bâtiment automatisé à l'aide d'image satellite. Même si le format de données est différent du sujet traité, la méthode est très complète et fait intervenir beaucoup de contraintes géométriques pour identifier les bâtiments et ouvrages. On y retrouve des outils comme la détection de ligne droite, élément caractéristique des structures artificielles. L'étude utilise aussi un algorithme de détection de forme rectangulaire, outil qui paraît puissant mais qui n'a pas pu être maîtrisé durant le stage de par sa complexité et les formes variées qui sont prises par les ouvrages recherchés.



B. Classification d'objets en utilisant les données LIDAR Bathymétrique

Article : https://rfiap2018.ign.fr/sites/default/files/ARTICLES/CFPT2018/Oraux/CFPT2018_paper_leguennec.pdf

Cette étude porte sur l'intérêt d'un LIDAR bathymétrique pour l'identification et la classification d'objets.

Elle utilise des descripteurs (ou indices) afin d'avoir des informations géométriques pour la classification supervisée. La partie bathymétrie est peu exploitée mais les bons résultats d'une classification type *random-forest* à l'aide d'indices géométriques sont une piste d'inspiration pour l'identification des ouvrages sur les fonds côtiers.

C. Classification Random-Forest et Automatisation pour l'identification de zone humide en s'appuyant sur des descripteurs dérivés du MNT.


Article : https://www.researchgate.net/publication/305385776_Random_Forest_Classification_and_Automation_for_Wetland_Identification_based_on_DEM_Derivatives


Si la méthode est très proche de l'étude présenté juste avant, d'autres indicateurs sont utilisé ici, comme les pentes. De plus, les résultats prometteurs valident les choix qui ont été faits par la suite.

				Random Forest		
Type	Project	Data	Total Records	0-1 Error Rate	1-0 Error Rate	Total Error Rate
RCP	R2554	80%	31469	0.00%	0.00%	0.00%
		20%	7868	2.33%	6.41%	3.69%
		100%	39337	0.47%	1.29%	0.74%
	B3654	80%	626	0.00%	0.00%	0.00%
		20%	157	17.39 %	1.49%	3.82%
		100%	783	2.56%	0.32%	0.77%
	R2823	80%	37094	0.00%	0.00%	0.00%
		20%	9274	5.35%	1.34%	2.89%
		100%	46368	1.06%	0.27%	0.58%
	wy	80%	14594	0.00%	0.00%	0.00%
		20%	3649	3.68%	6.88%	4.66%
		100 %	18243	0.73%	1.39%	0.93%
Group			104731	0.72%	0.68%	0.70%

Figure 4: Tableau des résultats de la classification dans l'étude citée. Les 80-20% correspondent respectivement au taux d'apprentissage et de contrôle

I.c) Résumé des données fournies.

Les données fournies ont été ~~produites récemment par le Shom dans le cadre d'un partenariat entre le Service Hydrographique et Océanographique de la Marine (Shom), le Réseau d'Observation du Littoral (ROL) Normandie, Hauts de France, l'AESN, l'AEAP, le parc marin estuaire picard-mer d'opale, les régions Normandie et Hauts de France, l'État et l'Europe. C'est le ROL qui banquardise et met à disposition les données au Cerema pour ce stage.~~ 

C'est l'un des premiers projets qui utilise ces données, ce qui a entraîné une vérification plus approfondie afin de pouvoir remonter ces informations au ROL pour les projets futurs. Ce temps de contrôle qualité et d'échange avec le ROL n'était pas envisagé au départ du stage 

A. Format des données

Les données sont des fichiers las (nuages de points) découpés en dalle de 1 km². Aucun MNT n'a été fourni avec celles-ci.

Les nuages de points possèdent 3 champs scalaires en complément des coordonnées X Y et Z.

Le champ GPSTime donne le temps GPS de l'acquisition du point. Le format d'enregistrement du temps est variable, créant une incohérence entre les tuiles (le temps peut varier significativement entre deux tuiles voisines)

Le champ Intensity donne l'intensité du retour enregistré dans une unité inconnue.

Le champ Classification donne une classification des points. Cette classification est par contre totalement erronée (+ de 90 % des points sont classés en végétation dense) et est donc inexploitable en tant que telle.

D'autres champs existent dans le .las mais ils sont tous mis à une valeur NO_DATA (0 ou 1 suivant les champs)

B. Limites des données

Certaines informations n'apparaissent pas dans les données fournies.

Le numéro de la bande n'est pas renseigné, donnant un nuage de points où les bandes sont visibles mais non renseignées dans les données. Il est cependant possible d'extraire cette information depuis le temps GPS, car les bandes sont espacées dans le temps de plusieurs minutes, voire jours. Une procédure qui n'a pas été mise en place sauf pour quelques vérifications manuelles.

Les multi-échos ne sont pas non plus renseignés. Ils sont probablement enregistrés car la densité de points grimpe par moment à des niveaux trop élevés pour un appareil classique. Ces valeurs extrêmes sont par ailleurs constatées au-dessus des terres et sont probablement le fait de multi-échos créés par la végétation. Cette information n'est pas entièrement pertinente pour le traitement, car il y a peu de

multi-écho sur la partie sous-marine mais empêche une classification classique du nuage de points sur les terres.

Les angles et l'altitude de vol ne sont pas non plus renseignés. Ils auraient peut-être pu être utiles pour la normalisation de l'intensité, qui a donc été écartée des traitements.

De plus sur les 800 fichiers 2 sont corrompus, et donc inutilisables. Ils sont principalement situés dans les terres et ont donc juste été écartés du traitement.

Ces limites ne sont pas si gênantes pour notre étude, mais sont en contradiction avec les promesses lors de la livraison. Un temps important (plus d'une semaine) a été consacré pour faire remonter l'état général de la qualité des données à la demande du ROL.

C. Précision et Densité

La précision a été contrôlée en vérifiant l'écart entre les altitudes de 2 vols se recouvrant. Cet écart est de l'ordre de 0,1m avec des pics maximaux autour de 0,3m. C'est suffisant pour identifier les ouvrages artificiels, mais cette précision n'a pas été contrôlée sur toutes les données et la méthode de vérification n'est pas rigoureuse. Il faut donc faire confiance aux spécifications techniques du produit.

Caractéristiques	Partie terrestre		Partie marine	
	Litto3D®	Stratégie de Suivi	Litto3D®	Stratégie de Suivi
Précision planimétrique	< 50cm	< 20cm	<40cm	<40cm
Précision altimétrique	< 20cm	<20cm	2,80m	2,80m
Densité	Jusqu'à 2 points/m2	8 à 15 points/m2	1 point/25m2	1 point/9m2
Couverture	Altitude +10m jusqu'à 2km dans les terres	BMVO à 400m après TCH à minima	En Manche : isobathe 5m	
Altitude d'acquisition	1500m	400m	400m	
Forme d'onde complète	non	oui	oui	
Livrables	MNT, Nuage de points bruts (points hauts), MNE	MNT, Nuage de points bruts, MNE, données sortie capteur, Imagerie, Classification SOL, LAS	MNT, Nuage de points bruts (dont tables conchyliques), MNE, Données sortie capteur, Imagerie, Classification SOL, LAS	

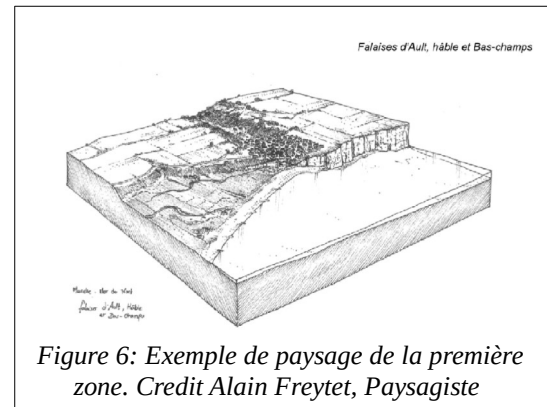
Figure 5: Spécifications annoncées du produit.

La densité du nuage varie beaucoup suivant le nombre de passage au-dessus d'une zone avec la présence de multi-écho comme évoqué dessus. Elle semble avoir été fixée autour de 1pt/m² par bande, et la densité sur la bande des 0-10m est d'au pire 0,65pt/m² (écart entre les points de 1,3m).

D. Zone d'étude et taille des données

Les zones choisies pour l'étude sont au nombre de 5 :

La première zone est située à l'ouest de Dieppe. Le type de terrain est une alternance entre des falaises et des basses vallées. Les fonds sont diversifiés, avec des amas rocheux et des étendues sableuses. Elle est assez réduite (d'environ 57 km²) et possède principalement des épis comme artificialisation. Elle a été livrée plus tôt que les autres et c'est celle sur laquelle le plus de traitements ont été effectués.



La deuxième zone est située dans les environs de Granville, situé au nord de la baie du mont Saint Michel. La côte est constituée de longues plages sableuses parfois artificialisées, ainsi que de falaises. C'est la limite des données actuellement produites, et elle comporte une centaine de fichiers. L'artificialisation y est faible mais on note quelques balises de navigation et quelques épaves.

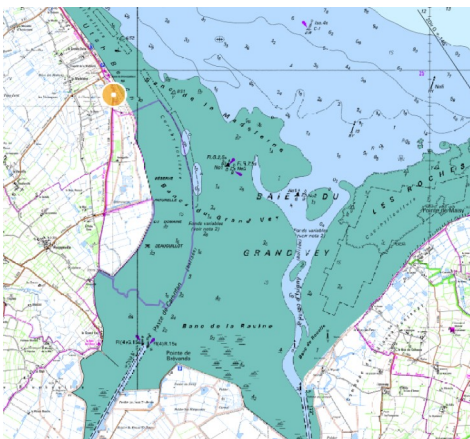
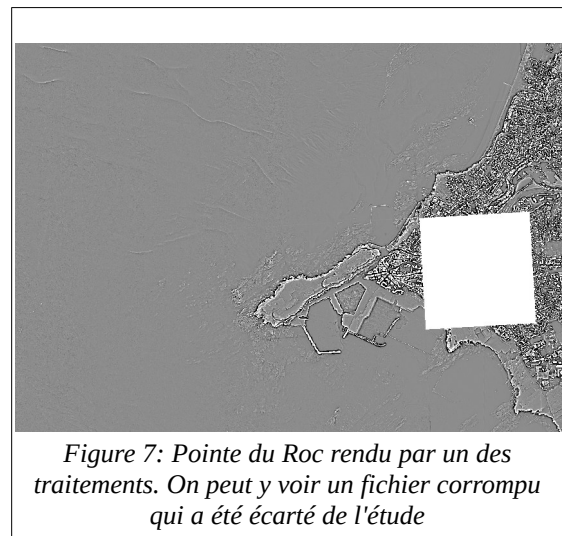


Figure 8: Carte de la zone 3. Le point jaune est la plage Utah Beach, située à l'ouest de la baie de Veys. Source IGN et SHOM

La troisième zone est la plage Utah Beach et ses environs. Quelques restes du débarquement de 1944 y sont présents, ainsi que des plages aménagées et artificialisées. Située dans un estuaire, elle possède de nombreux lieux de conchyliculture situés sur l'estran. Elle s'étend sur 200 km² environ ce qui en fait la deuxième plus grande zone d'étude.

La quatrième zone est la plus grande et suit la côte tout le long du plateau du Calvados. C'est un endroit très artificialisé avec plus de 290 épis sur l'estran. Les 294 km² qui la composent sont par contre un problème en terme de temps de calcul et d'affichage. Les fonds marins y sont diversifiés, et on peut y voir des épaves et des restes de la seconde guerre mondiale. Cette zone est celle qui sera privilégiée pour les tests en condition réel de par la diversité des situations qu'elle présente.



Figure 9: Extrait du plateau du Calvados avec en rouge les éléments artificiels sur l'estran. Source: Cerema.

La cinquième et dernière zone est le port de Dunkerque. Ici les données sont parcellaires car l'acquisition au-dessus des ports n'était pas assurée. Elle a été choisie afin d'essayer de détecter les câbles sous-marins. Elle possède 166 fichiers, mais l'étendue de la zone est plus grande car de nombreux trous sont présents comme illustré ci-contre.



Figure 10: Répartition des .las dans la zone de Dunkerque. En rose les dalles LIDAR et en rouge la zone choisie initialement. Source orthophoto: Cerema

Pour finir l'ensemble des données est composé de plus de 800 fichiers pour un total de 230 Go. Les problématiques de temps de calcul et d'espace de stockage ont donc été très vite envisagées.

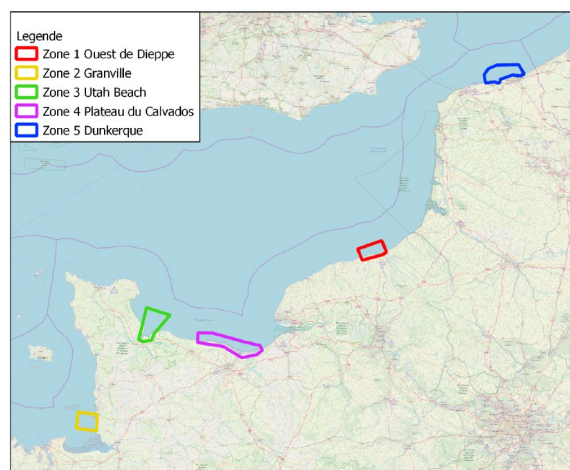


Figure 11: Carte des 5 Zones étudiées

II Traitement des fichiers las et Classification.

Tout d'abord il faut préciser les ouvrages qui vont être recherchés dans nos images. Les ouvrages artificiels utilisés par l'indice sont des objets créés par l'homme, fonctionnels ou non, permanents et bâtis. J'entends par là que les objets qui n'ont pas été bâtis sur les lieux, comme les épaves, ne sont pas des ouvrages artificiels mais sont considéré comme des obstructions marines. Les bouées temporaires et les zones de cultures de coquillages sont aussi exclues car non pérennes. Cependant, des objets comme les épaves sont mal référencés, et il a été choisi de les traiter aussi, afin d'avoir un aperçu de l'état des données sur la zone d'étude.

II.a) Principe du traitement.

On a choisi pour la simplicité des traitements de transformer les nuages de points en raster, en se basant sur un MNS (quasiment identique au MNT sur l'estran et la partie immergée.) produit par CloudCompare.

Afin de limiter la charge des données produites, tous les traitements sont regroupés dans une seule image contenant actuellement 21 bandes de données. La numérotation des bandes est arbitraire, et il est préférable d'utiliser les fonctions de sélection de données vues plus loin pour l'accès aux données.

La liste des bandes par ordre de numérotation est la suivante :

- bande 1 : filtre kernel
- bande 2 : filtre de contour par soustraction de filtre gaussien
- bande 3 : indice de tortuosité
- bande 4 : indice de rugosité sur une zone de 5px par 5px
- bande 5 : indice de rugosité sur une zone de 21px par 21px
- bande 6,7 et 8 : filtre morphologique convexe de l'orfeo toolbox
- bande 9,10 et 11 : filtre morphologique concave de l'orfeo toolbox
- bande 12,13 et 14 : filtre morphologique dit 'leveling' de l'orfeo toolbox
- bande 15 : filtre contour par soustraction de 2 MNS CloudCompare
- bande 16 : algorithme de détection de ligne appliqué sur la bande 1
- bande 17 : algorithme de détection de ligne appliqué sur la bande 6
- bande 18 : algorithme de détection de ligne appliqué sur la bande 15
- bande 19 : Intensité du retour LIDAR avec interpolation
- bande 20 : MNS obtenu par CloudCompare
- bande 21 : masque grossier obtenu par limite d'altitude

Les bandes sont approfondies ici dans un ordre plus logique dans la progression du programme.



Figure 12: Tous les exemples sont sur la même zone ici montrée en photo aérienne. Source : Cerema

MNS calculé par CloudCompare (bande 20)

Cloud Compare possède une fonction 'projeter un nuage sur un raster'. Plusieurs options sont importantes et modifiables en ligne de commande.

Voici la ligne de commande utilisé par le programme pour la version 'maximum' :

```
CloudCompare -silent -o fichier.las -rasterize -grid_step
résolution -proj max -empty_fill interp -output_raster_z -
clear'
```

Les options sont :

- **silent** : permet d'éviter d'avoir une box à cliquer à la fin du traitement, mais rend aussi tout le traitement silencieux (aucune info dans la fenêtre de commande)
- **o** : O comme open et non output, cette commande ouvre le fichier dans CloudCompare. Plus d'options sont disponibles avec cette commande sur la doc.
- **rasterize** : la commande pour projeter le nuage de points. La majorité des paramètres après seront des options pour cette commande.

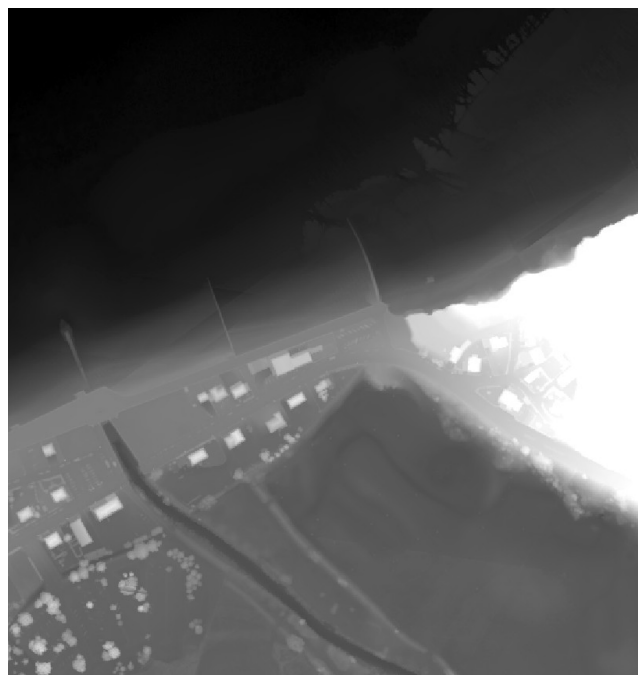


Figure 13: MNS calculé par CloudCompare

- grid_step** : Entre la résolution du fichier tif en sortie.
- proj** : Choix de la projection quand plusieurs points sont sur le pixel. 3 options : max, min, avg. Cela correspond au 3 MNS calculés dans l'étude.
- empty_fill** : Choix pour les cases vides. Plusieurs options disponibles sur la doc, mais nous avons choisi une interpolation pour ne pas avoir de trou. L'interpolation à l'air d'être une interpolation de type proche voisin. Cependant elle ne fonctionne pas pour les bords de l'image, laissant quelques vides sur les bords.
- output_raster_z** : Option permettant de préciser que c'est la valeur en z que nous voulons projeter. Dans les faits les valeurs des champs scalaires seront aussi enregistrées, mais aucune interpolation ne sera fait dessus.

-**clear** : commande permettant de décharger le nuage de points de CloudCompare. Sinon le nuage reste ouvert dans la mémoire RAM (en pratique il a l'air d'être déchargé à la fin de la commande en python). Il faut donc mettre toutes les instructions dans la même commande pour l'exécuter en python.

Interpolation pour les effets de bords :

Comme préciser plus haut, CloudCompare n'interpole pas bien les bords de l'image/nuage de points. Il faut alors trouver une astuce pour remplir ces vides. Il a été choisi ici de ne pas passer par une fonction d'interpolation « classique » qui prendrait du temps de calcul, mais d'utiliser un MNS qui possède une résolution bien moindre. Il suffit alors de remplacer les valeurs 'NaN' (not a number) par la valeur du MNS sous-résolu pour avoir une interpolation approximative. Vu que ce problème ne concerne que peu de points, cette solution est suffisante et n'interpole pas trop sur les vraies zones qui ne possèdent pas de données.

Masque d'altitude (bande 21)

Ce masque peu précis est un outil pour masquer les images.

Il est uniquement composé de deux valeurs : 0 pour les zones à garder et NaN (not a number) pour les zones masquées. Il est en pratique utile qu'en interne du programme, car ajouter le masque à n'importe quelle bande permet de mettre la valeur NaN (valeur no data par défaut) sur toutes les zones masquées. Le masque est calculé par valeur d'altitude sur le MNS lissé (afin de garder les structures basses de l'estran) et la valeur ici choisie est 6 m d'altitude.

Image de contour par soustraction de MNS (bande 15)

Vu que nous n'utilisons que le MNS avg pour la suite du traitement, nous perdons beaucoup d'informations sur les zones où la résolution est très forte (ici les zones de terres, moins pertinentes pour nous). Afin de récupérer ces informations, une image de contour est réalisée qui possède des informations à un niveau inférieur au pixel. Elle est produite par soustraction du raster AVG avec celui MAX. Cela implique que lors d'un mur par exemple, le haut du mur sera dans le raster MAX, alors que la moyenne des points sera plus basse, donnant une idée du contour des objets. Cela permet aussi de mettre en évidence la végétation haute dans une donnée ne possédant pas les indices de multi-échels. Cette image possède beaucoup d'informations, entre autres sur la terre ou l'estran.



Figure 14: Image de contour par soustraction

Image d'intensité (bande 19)

Cette image est directement issue des données LIDAR exportées avec le MNS. L'intensité n'étant pas interpolée par CloudCompare, il est nécessaire de l'interpoler en interne afin d'avoir une image utilisable. L'interpolation est réalisée par fonction cubique, ce qui implique un temps de calcul assez long en python (bien moins rapide que CloudCompare). Cette donnée peut être utile pour de l'interprétation visuelle, et semble intéressante pour les parties terrestres, mais n'apparaît pas pertinente pour les zones immergées dans notre étude.



Figure 15: Image d'intensité

Filtre Kernel (bande 1)

Le filtre Kernel est un filtre de contour appliqué au MNS.

Il est produit par convolution d'une matrice 5*5 au MNS. La matrice est ainsi composée :

-1	-1	-1	-1	-1
-1	1	2	1	-1
-1	2	4	2	-1
-1	1	2	1	-1
-1	-1	-1	-1	-1

Ce filtre efface totalement l'altitude du MNS et des objets (sur une surface plane le filtre donne 0 en résultat). De plus le sens de la pente est renseigné par le signe des valeurs, les pixels situés en bas de pente ont des valeurs négatives là où les pixels en haut sont positifs.

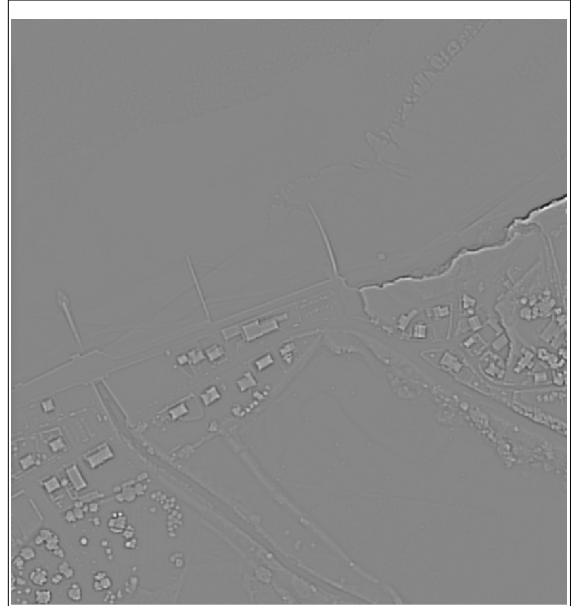


Figure 16: Filtre de contour kernel

Filtre de contour par soustraction (bande 2)

L'image de contour de ce filtre est obtenue par soustraction du MNS avec un MNS lissé par un filtre gaussien de taille 5. Cela donne à la fois les contours des objets, mais aussi leurs élévations relatives pour ceux de l'ordre de quelques mètres. Les premiers résultats ont montré que les informations données par ce filtre ne sont pas redondantes avec celles du filtre kernel.

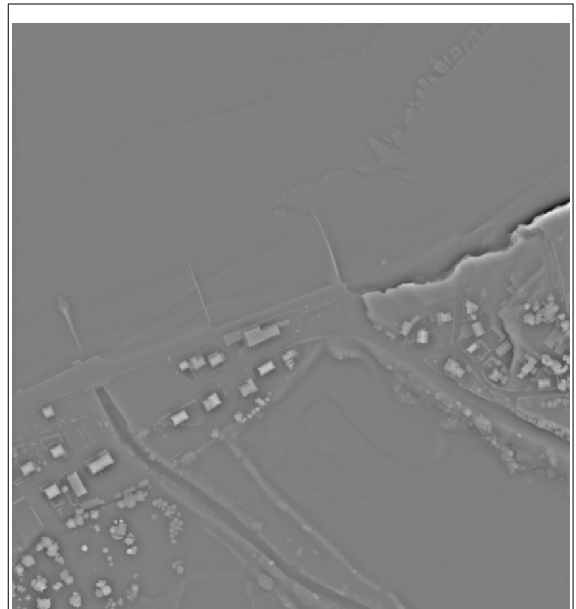
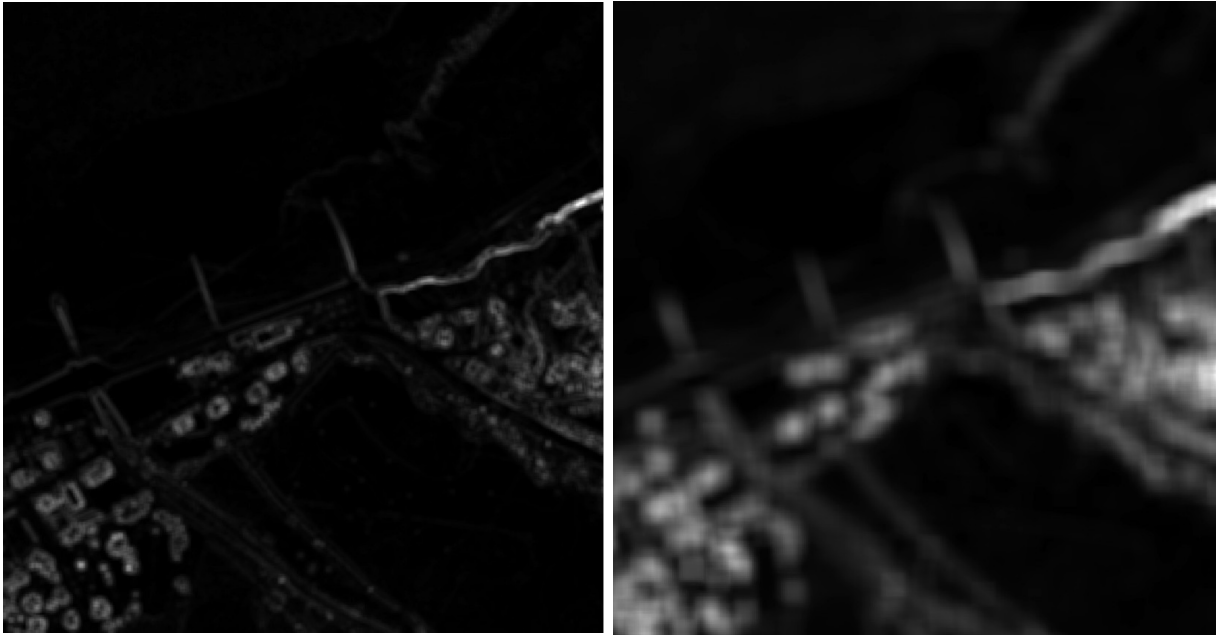


Figure 17: Filtre de contour par soustraction

Indice de rugosité (bande 4 et 5)

Cet indice est simple mais assez grossier. Il possède cependant des informations intéressantes sur la nature du terrain. Il est la somme des valeurs absolues des pentes du filtre Kernel sur une zone variant suivant le but du traitement. La taille 5 donne donc un aperçu de la rugosité de la surface étudiée, alors que la taille 21 montre le type de surface général de la zone.



Indice de rugosité avec une taille de filtre de 5 (à gauche) et de 21 (à droite)

Indice de tortuosité (bande 3)

Cet indice est aussi un indice de rugosité, mais prenant plus en compte la forme générale de l'objet. C'est un rapport entre l'aire effective sur le terrain et l'aire fictive d'une surface plane de même dimension. La taille ici choisie est de 3 pixels par 3 pixels.



Figure 18: Filtre de tortuosité

Filtre Morphologique Orfeo Toolbox (bande 6 à 14)

Ces filtres sont produits par la boîte à outils d'Orfeo Toolbox. Un filtre morphologique utilise des dilations et érosions d'ensemble afin de séparer et quantifier les différentes parties d'un espace. Appliqué à une image de contour, cela permet de mettre en valeur certaines spécificités du terrain et de séparer les ensembles géodésiques. Une grande partie du choix de ces filtres a été effectué par tâtonnement, donnant au final les paramètres suivants sur la commande console `otbcli_MorphologicalMultiScaleDecomposition` :

- le paramètre `structype` est `ball` (type de structure utilisée par l'algorithme)
- le paramètre `channel` est `1` (bande sur laquelle est appliquée l'algorithme)
- le paramètre `levels` est `3` (cela nous donne des images RVB en sorties, d'où les 9 bandes obtenues)
- le paramètre `radius` est `5` (paramètre par défaut de l'application)
- le paramètre `step` est `1` (paramètre par défaut de l'application)

Le résultat final nous donne 3 images RVB, la sortie convexe, concave et leveling. Sur les 9 bandes de sorties, 7 seront retenues pour la classification finale.

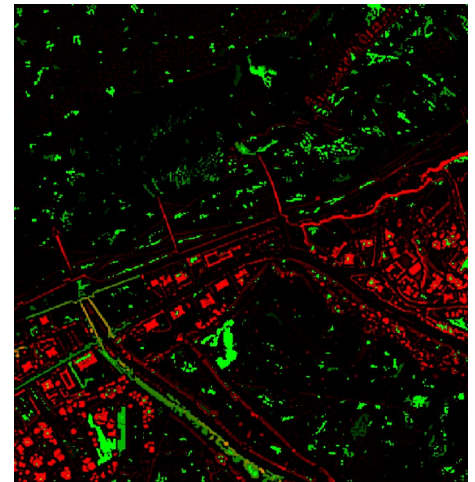


Figure 19: Image de sortie convexe RVB

Détection de ligne par l'algorithme LSD (bande 16 à 18)

Une des spécificités des ouvrages artificiels est la présence de lignes droites continues souvent assez longues. Il est donc très pertinent de chercher à détecter les lignes dans le MNS (ou plutôt le filtre de contour).

Pour cela nous utilisons l'algorithme LSD (Line Segment Detection). Il n'est pas implémenté en python mais la boîte à outil Orfeo permet de l'utiliser ici en ligne de commande. La commande ne prend aucun paramètre sinon les fichiers d'entrée et sortie et se nomme `otbcli_LineSegmentDetection`. La sortie étant un shapefile des lignes obtenues, il est ensuite rasterisé avec comme valeur de pixel la longueur de la ligne. Le raster obtenu est ensuite moyenné sur une zone de 3px par 3px afin de propager la longueur de la ligne aux pixels aux alentours (afin d'avoir la proximité avec une ligne pour notre classification).

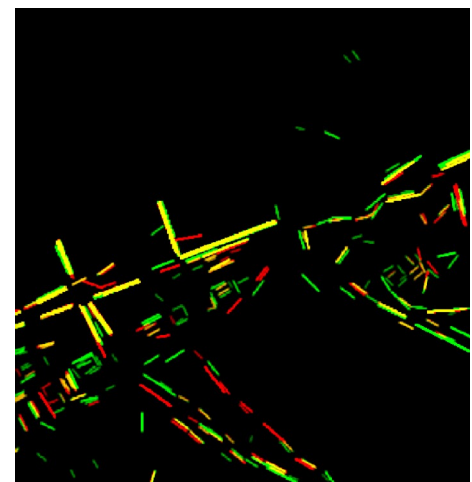


Figure 20: En rouge, les lignes détectées sur le filtre kernel, en vert sur le filtre morphologique convexe

II.b) Programme et Automatisation.

Afin d'automatiser la création des données vues ci-dessus, un programme python à été développé. Le but était non seulement de réunir les traitements mais aussi de faciliter l'utilisation du programme.

A. Prérequis du programme.

C'est le point sur lequel la facilité d'utilisation peut être largement améliorée.

Le programme tourne sur python 3, mais est adaptable facilement sur python 2.7. Il nécessite l'installation de nombreuses bibliothèques de fonctions, ainsi que de leurs dépendances respectives :

- rasterio est utilisé pour la lecture et l'écriture de fichier GeoTiff en prenant en compte le géoréférencement.
- numpy est une bibliothèque très commune pour l'utilisation de matrice en python, et donc pour la manipulation d'image sous forme matriciel.
- shutil permet des manipulations de dossiers et fichiers (copie, suppression...) afin d'avoir des dossiers propres lors du traitement.
- scipy possède de nombreuses fonctions de manipulation d'image, que ce soit pour l'écriture/lecture ou sous forme matricielle pour l'interpolation par exemple.
- gdal (et ogr, qui lui est associé) possède l'ensemble des fonctions GDAL de Qgis, par exemple la création de VRT afin de rassembler les dalles produites par le traitement.
- PyShp (importé sous le nom de shapefile) manipule les fichiers .shp, permettant la création d'index sous forme vecteur par exemple.
- laspy manipule les fichiers .las, principalement utilisé dans notre programme pour vérifier la conformité des données.

Cependant, d'autres manipulations sont pour l'instant nécessaire. En effet, le programme utilise CloudCompare et l'Orfeo Toolbox sous forme de ligne de commande console. Il est donc nécessaire que ces commandes soient disponibles dans la console Windows. Si nécessaire, il faut mettre à jour la variable d'environnement Windows pour inclure CloudCompare.exe et le fichier bin du dossier d'installation de l'OrfeoToolbox.

B. Fonctions principales.

Le programme comporte plus d'une trentaine de fonction, mais le traitement a été réduit à quelques fonctions clés présentées ici.

Une fonction de vérification des données las est disponible pour la vérification de dossiers complets. Cette fonction CheckBonFile possède un seul paramètre, le chemin du dossier (absolu ou relatif), qui est par défaut le dossier Input, utilisé pour les fonctions de traitement vues plus tard. Il vérifie que le fichier n'est pas illisible et que les points se situent bien dans les limites enregistrées dans les métadonnées. Il écrit dans la fenêtre de commande les noms des fichiers problématiques à la fin de la fonction.

La fonction principale du programme est la fonction `TraitementBrut`. Elle ne possède aucun paramètre, et traite tous les fichiers las du dossier `Input` situé dans le même répertoire que le code qui est exécuté. Le résultat de la fonction est un dossier `Output_Brut`, possédant toutes les dalles liées aux fichiers `.las` correspondants. Les Rasters produits sont des images dites « brutes », comme décrites plus haut.

Ensuite deux fonctions de sélection de données sont disponibles. `SelecData` est une fonction qui prend comme argument une liste d'éléments et deux booléens.

Le premier argument est la liste des données désirées dans l'image. Elles sont sélectionnables par mots clés, dans la liste suivant :

- `kernel` : filtre kernel
- `sous` : filtre de contour par soustraction de filtre gaussien
- `'tort'` : indice de tortuosité
- `rugofine` : indice de rugosité sur une zone de 5px par 5px
- `rugolarge` : indice de rugosité sur une zone de 21px par 21px
- `convex1,convex2,convex3` : filtre morphologique convexe de l'orfeo toolbox sélectionnable par `convexRGB` pour les trois bandes.
- `concav1,concav2,concav3` : filtre morphologique concave de l'orfeo toolbox sélectionnable par `concavRGB` pour les trois bandes.
- `lvling1,lvling2,lvling3` : filtre morphologique dit 'leveling' de l'orfeo toolbox sélectionnable par `lvlingRGB` pour les trois bandes.
- `imgcontour` : filtre contour par soustraction de 2 MNS `CloudCompare`
- `linekernel` : algorithme de détection de ligne appliqué sur la bande 1
- `lineconvex` : algorithme de détection de ligne appliqué sur la bande 6
- `linecontour` : algorithme de détection de ligne appliqué sur la bande 15
- `intensity`: Intensité du retour LIDAR avec interpolation
- `MNT` : MNS obtenu par `CloudCompare`
- `masque` : masque altimétrique

Le deux paramètres suivants sont les paramètres de masque et de séparation du vrt. Le premier paramètre permet de masquer l'image fournie avec le masque altimétrique calculé en interne.

Comme la fonction produit un VRT (virtual raster) de l'ensemble des données, il peut être intéressant de séparer le VRT pour gagner du temps d'affichage. Le paramètre suivant découpe le VRT en groupe de 20 dalles, en gardant la cohérence entre les colonnes (lorsqu'un élément d'une colonne est inclus dans le VRT, toute la colonne y est.). Le résultat final est le dossier `Output` possédant toutes les images et un VRT de celles-ci situé dans le répertoire du code python nommé `Image_total_Ouput.vrt`.

La fonction `SelecData_Sans` est la même que `SelecData`, sauf qu'elle fonctionne par exclusion plutôt que par inclusion de données. La donnée sélectionnée sera toutes les données présentes sauf celle listées dans les paramètres de la fonction. Comme cette fonction est utilisée pour la classification, le MNS et le masque n'y sont jamais inclus.

Deux dernières fonctions peuvent être utiles, la fonction `VRTDos`, qui fait un VRT de toutes les images situées dans le dossier spécifié, et la fonction `IndexTuileLas` qui produit un shapefile des emprises des différents fichiers las du dossier entré en paramètre (par défaut le dossier `Input`).

II.c) Classification

Afin de poursuivre sur l'automatisation, il a été choisi d'utiliser une classification supervisée de type Random-Forest.

Une classification supervisée fonctionne à l'aide de zone d'apprentissage permettant d'entraîner le programme à détecter les différents éléments de l'image, il a donc fallu choisir ces zones en fonction du lieu d'étude.

Un obstacle s'est présenté à ce stade ; les données de l'étude ne contiennent pas assez d'objets artificiels immergés pour effectuer une classification. En effet sur les 800 km² de côte couverte, il y a plus de 700 objets artificiels mais seulement une soixantaine sont immergés la majeure partie de l'année, dont 40 sont des restes des barges de la seconde guerre mondiale, des éléments très spécifiques et localisés sur 6 km² de données. Le choix d'utiliser l'estran et les nombreux épis qui s'y trouvent, lié au fait qu'une base de données réalisée par photo-interprétation existe pour comparer les résultats, s'est donc imposé très vite pendant le stage.

L'apprentissage s'est effectué sur la zone de Dieppe par itération successive afin d'obtenir une classification pertinente.

Quatre classes ont été choisies dans cette classification :

- La classe Sable, regroupant les sols sableux ou graveleux, avec peu de variation de rugosité à moyenne échelle.
- La classe Roche, qui regroupe les platiers et les fonds rocheux. La rugosité y est plus importante et des structures ressemblant à des épis peuvent y apparaître.
- La classe Falaise, classe permettant de mieux visualiser la côte. La falaise étant très bien caractérisée sur les images, cette classe est très bien définie. De plus, la classe contient aussi les zones où les données sont absentes, ce qui évite la confusion entre des colonnes de pixels vides et un épi.
- La classe Épi, regroupant les structures artificielles de l'estran. Elle contient sans distinction les épis, les rampes d'accès et les digues, les épis étant très majoritaire sur l'échantillon.

Les différents parasites liés aux multiples lignes de vol, au déplacement des galets entre deux prise de vue et à la similarité entre certains rochers et des épis ont mené à la réalisation de 16 classifications successives afin d'élaborer un modèle. Les zones d'apprentissage étaient redéfinies entre chaque classification afin d'exclure les zones problématiques et parfois d'inclure un épi difficile à percevoir. La décision d'arrêter l'amélioration de la classification a été prise en fonction de l'inclusivité des résultats automatiques dans les résultats manuels, c'est-à-dire qu'aucune zone ne devait être un épi si visuellement la structure n'était pas artificielle. La vérification étant faite à l'œil, certains groupes de pixel peuvent sortir de la règle s'ils sont suffisamment petits pour être filtrés ou dans des zones très problématiques.



Figure 22: Zone d'exemple pour la classification. Fausse couleur par bande de contour et tortuosité.

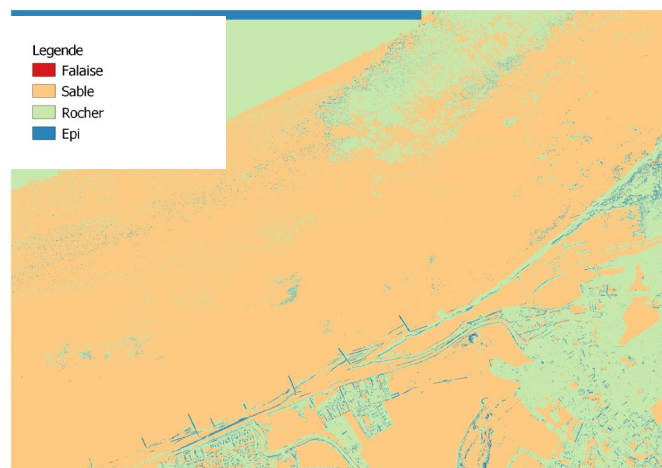


Figure 21: Une des premières classifications, la falaise n'est pas encore présente dans les zones d'apprentissages

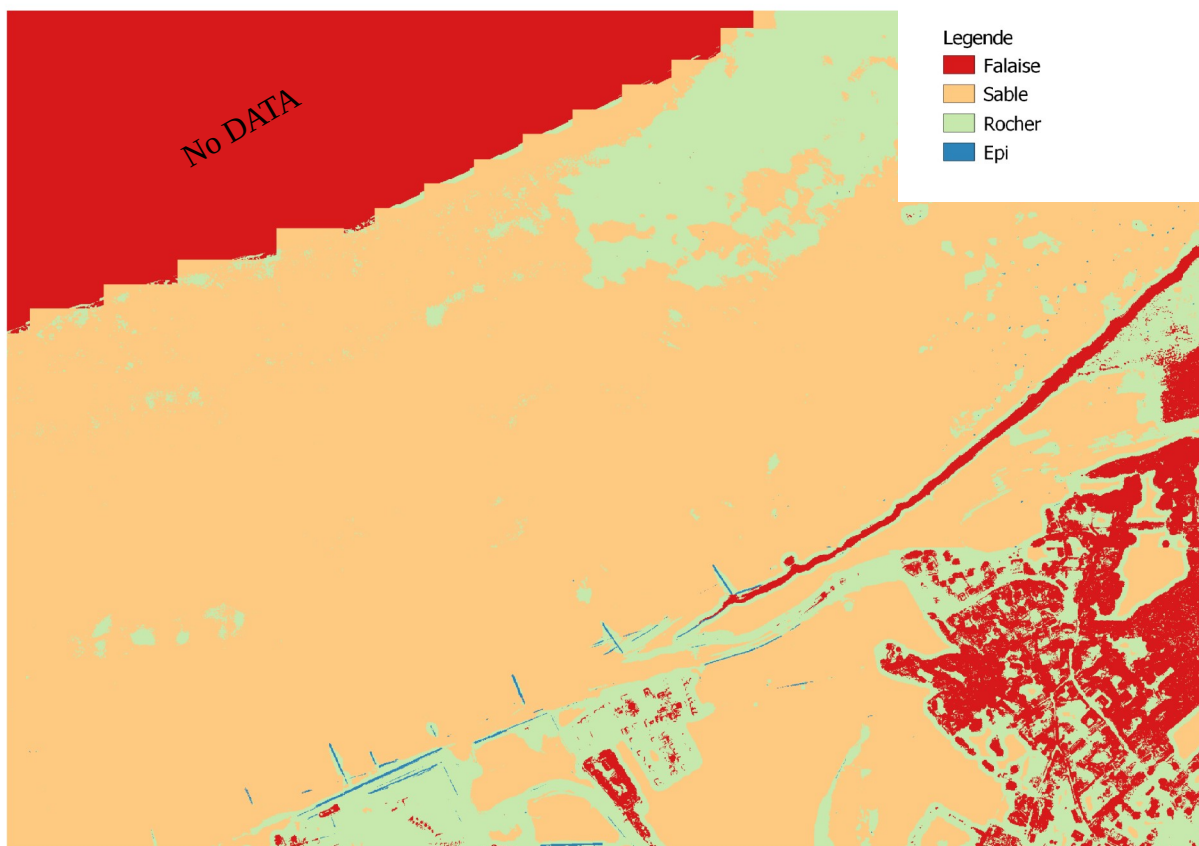


Figure 23: La classification finale, avec encore quelques erreurs difficiles à corriger.

Classification et zone d'exemple. L'image représente environ 10 % de la zone d'apprentissage.

III Résultat et ouverture.

III.a) Autre application du LIDAR sur le milieu côtier.

A. Suivi du trait de côte.

La création de MNT par LIDAR permet le suivi précis et à travers la végétation du trait de côte. En effet, le LIDAR possède à la fois des points sur la végétation et sur le sol, qui sont classifiables en utilisant des algorithmes liés à la continuité des points sols.

Afin de produire un vrai MNT, et non un MNS comme pour le reste du traitement, on utilise ici le plugin LASTools de Qgis. Ce plugin est à usage libre pour l'enseignement, et permet de traiter des images une par une sans licence particulière. Pour le traitement par lots par contre, il faut contacter le propriétaire pour une licence, ce qui a écarté ce plugin du traitement général.

Ce plugin possède entre autres l'algorithme LASGroundnew qui sépare les points sol et végétation d'un nuage, et ce même sans indication de multi-écho. Cela a permis de produire l'image suivante, sur la falaise de Varengeville-sur-Mer présentant la particularité d'avoir un palier intermédiaire entre le plateau et le pied de falaise.

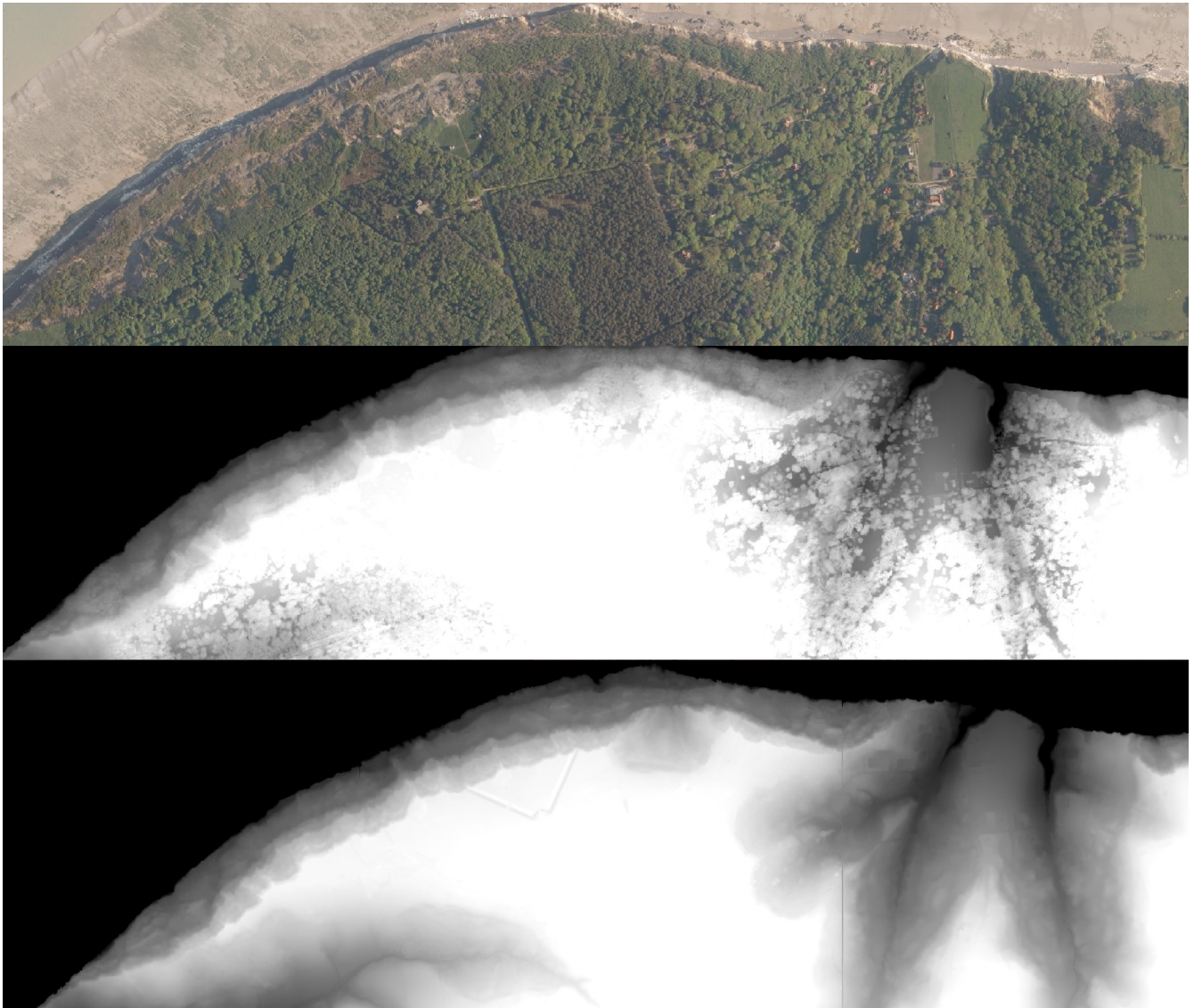


Figure 24: OrthoPhoto, MNS et MNT de falaise complexe au sud de Dieppe. Source orthophotographie Cerema.

Pour suivre une falaise, il est possible de créer un programme qui trace automatiquement la ligne de crête de celle-ci. Afin de l'illustrer j'ai utilisé le travail d'un autre stagiaire : Mehdi Elmouatadil qui cherche à suivre les fissures du plateau rocheux. Son programme, basé sur l'algorithme Fast Marching Method, suit donc les plus fortes variations de pente, ce qui correspond aussi à la présence d'une falaise pour notre MNT.

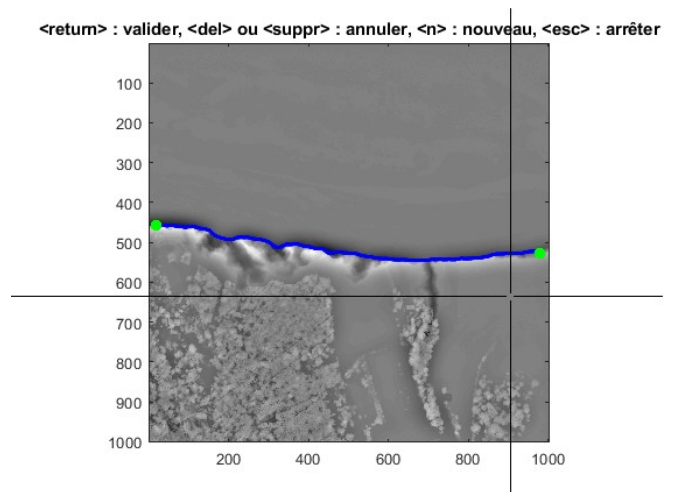


Figure 25: Suivi d'une ligne de crête par le traitement.

B. Suivi de la Conchyliculture

La conchyliculture est l'élevage de coquillages, tels que les huîtres, les palourdes etc . Pour exercer son activité, un organisme doit avoir une concession afin d'exercer sur le domaine public. Les exploitations se présentent comme de longues tables situées sur en général sur l'estran. Ces exploitations sont visibles sur les images produit par les traitements précédents sous la forme de vaguelettes superposées au MNT, comme visible dans la figure 26.

Afin de pouvoir les mettre en évidence, une piste de réflexion a été mis en place autour des transformations de Fourier. La transformée de Fourier est une opération qui transforme une fonction intégrable en une autre fonction, décrivant le spectre fréquentiel de cette dernière. En d'autres termes, elle visualise une information sous la forme des fréquences la composant, et dans notre cas des fréquences spatiales qui compose l'image. L'avantage de cet outil est que le résultat est équivalent à l'image, et que la transformée de Fourier peut être retransformée en image sans aucune perte d'information. On peut donc voir et extraire des informations fréquentielles pour le transformé en informations spatiales.

Deux traitements ont été produits avec cette méthode : un filtre à l'échelle d'une dalle et un filtre à plus basse échelle.

Le premier filtre est un filtre qui « simplifie » l'image d'origine. En pratique le filtre ne garde que 2 % des fréquences présentes sur l'image, éliminant les 98 % fréquences les plus faibles de la



Figure 26: Parc de conchyliculture



Figure 27: Conchyliculture dans la zone 3 avec en orange les concessions enregistrés dans les bases de données du CEREMA

transformée de Fourier. L'image obtenue par retransformation est une vision des plus fortes oscillations de l'image, ce qui met en évidence les discontinuités des surfaces étudiées mais aussi les lignes périodiques de la conchyliculture. Ce filtre fonctionne bien mais n'a aucun moyen de différencier la conchyliculture d'un récif naturel. Pour cela il faut une vision plus proche du terrain et qui est liée à la forme de la transformée de Fourier.

Le deuxième filtre produit une valeur de distance entre la zone étudiée et la zone idéale de la conchyliculture. Il travaille sur des carrés d'une cinquantaine de pixels de large sur l'image et est calculé sous échantillonné de 5 fois (1 pixel du résultat est une zone de 5*5 pixel sur l'image original) afin d'accélérer les temps de calcul. De plus, les emprises de calcul étant assez large, le produire pixel par pixel ne donnerait pas des résultats très différents de la version sous échantillonnée. L'algorithme compare chaque morceau de l'image à la transformée idéale montrée dans l'illustration X. Il calcule pour cela la distance des points les plus brillants de la transformée avec une droite idéale passant par le centre, ainsi les zones de conchyliculture ont des distances faibles alors que les zones de sable sont bien plus élevées.

Sur l'image présentée depuis le début une précision de plus de 80 % a été atteinte avec cette méthode en faisant varier les seuils de l'algorithme, ce qui veut dire qu'il est possible de décider d'un seuil de séparation de l'histogramme dans l'image final où 80 % des pixels classés dans conchyliculture sont bons et 80 % des pixels classés dans non-conchylicultures sont bons (statistique calculée à l'aide d'une image de résultats produite par photo-interprétation sur l'image originale).

La combinaison des deux filtres permet d'obtenir une image binaire très proche de la réalité observée, même si des erreurs sont encore visibles. Ces résultats se sont exportés avec une réussite partielle sur les autres exploitations observées, avec quelques lieux où la petite taille des exploitations et la topographie des lieux rendent les filtres moins précis (précision du deuxième filtre en dessous de 70 % par exemple).

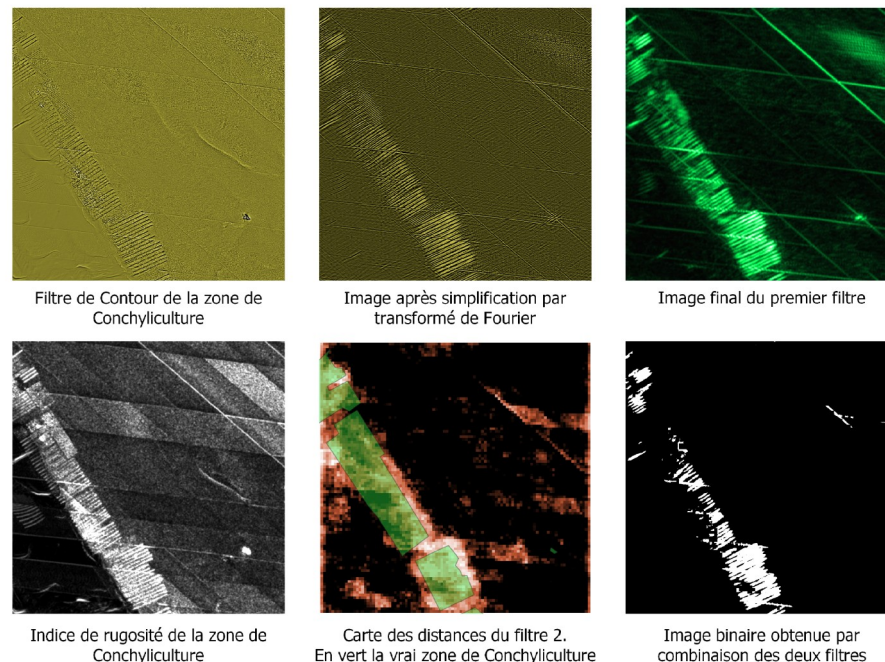


Figure 28: Résumé du traitement sur la conchyliculture

III.b) Qualité des résultats.

A. Pertinences des données LIDAR pour la photo-interprétation.

Afin de qualifier la pertinence des données LIDAR pour la détection d'ouvrage sur le front de mer, il a été choisi de comparer les informations des images extraites depuis les données LIDAR et celles de la base de données créée par photo-interprétation sur l'estran. Les statistiques ont été réalisées sur la zone 4, le plateau du Calvados, où se trouve le plus d'entités sur l'estran, et sur la zone 1, au sud de Dieppe, afin d'avoir un autre jeu de données. Les trois autres zones ne possédant pas beaucoup de structures transversales au front de mer, les statistiques n'ont pas été compilées ici, mais sont proches de celle de la zone 1.

Zone4	Saisie LIDAR manuel	Saisie Photointerpretation	rapport LIDAR/Photo
Nombre d'entité sur l'estrans	204	294	69,39 %
longueur_Total	11521,00	19047,00	60,49 %
longueur moyenne	56,48	64,79	87,17 %
nombre d'epi	151,00	225,00	67,11 %
longueur total des epis	7214,00	11739,00	61,45 %
longueur moyenne des epis	47,77	52,17	91,57 %
longueur mediane des epis	41,20	42,66	96,58 %
largeur moyenne des epis	6,40	inconnu	

Zone1	Saisie LIDAR manuel	Saisie Photointerpretation	rapport LIDAR/Photo
Nombre d'entité sur l'estrans	22	37	59,46 %
longueur_Total	1072,00	1681,00	63,77 %
longueur moyenne	48,73	45,43	107,25 %
nombre d'epi	22,00	37,00	59,46 %
longueur total des epis	1072,00	1681,00	63,77 %
longueur moyenne des epis	48,73	45,43	107,25 %
longueur mediane des epis	46,70	41,14	113,51 %
largeur moyenne des epis	7,05	inconnu	

Tableau 1: Statistiques de la numérisation manuelle

Ces statistiques montrent plusieurs choses. Déjà les deux tableaux sont proches, ce qui rassure quant à la représentativité des données que nous avons. Les longueurs moyennes des objets et des épis sont très proches, ce qui confirme que les objets trouvés sont de la même géométrie.

Le rapport d'objets vus par interprétation LIDAR est donc entre 60 % et 70 % ce qui s'explique par plusieurs raisons. Tout d'abord la numérisation a été réalisée par mes soins, dans le but d'une vérification donc sans croiser avec d'autres données. Certains éléments m'ont probablement échappé, alors qu'un expert avec une base de données pour vérification peut probablement obtenir de bien meilleurs résultats. Ensuite certains épis sont en grande partie ensablés, ce qui peut être visible par photo-interprétation mais qui est difficile à discerner avec une image possédant quasiment uniquement des informations géométriques. Les valeurs de largeur et longueur moyennes sont obtenues par enveloppe rectangulaire autour des polygones dessinés lors de la numérisation LIDAR. La base de données d'ouvrage est une shapefile linéaire, donc la longueur n'est pas un problème mais la largeur n'y est pas renseignée.

Au final, l'intérêt de ce type de données est plutôt bon, même si sur l'estran la photo-interprétation est probablement mieux adaptée pour la réalisation de données exhaustives. Cependant de par les contraintes imposées pour la création de l'ortholittoral, les données LIDAR peuvent être des informations tout aussi pertinentes pour l'acquisition de données sur l'estran.

Il est aussi important de noter que la base de données réalisées par photo-interprétation ne possède pas de valeur de largeur, ce qui est pourtant nécessaire pour la réalisation d'un indicateur surfacique. Cette étude permet donc d'en donner une estimation sur l'ensemble des épis mesurés. La largeur paraît cependant un peu élevée pour des épis dont la majorité ne font que quelques mètres. Cela s'explique probablement par plusieurs phénomènes : les petits épis sont ceux qui sont le plus mal identifiés sur les images, tirant la moyenne vers les plus larges qui sont eux toujours visibles ; les images sont à 1 m de résolution, ce qui complique l'estimation d'objets dans cet ordre de grandeur (un objet d'un mètre de large sera probablement plus épais sur l'image) ; et pour finir le traitement utilise souvent des zones plus large que les pixels, ce qui fait « baver » les épis, rendant plus compliquée la numérisation précise.

Pour la partie sous-marine, il n'y a pas assez de données ni de base de données pour faire une vraie comparaison entre terrain et réalité. On ne peut donc que donner quelques exemples sur des structures particulières.

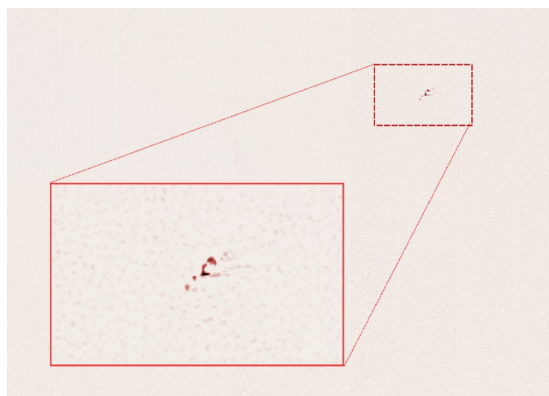


Figure 29: Épave du Videcoq, au large de Granville, à plus de 7 km des côtes et par 8m de fond

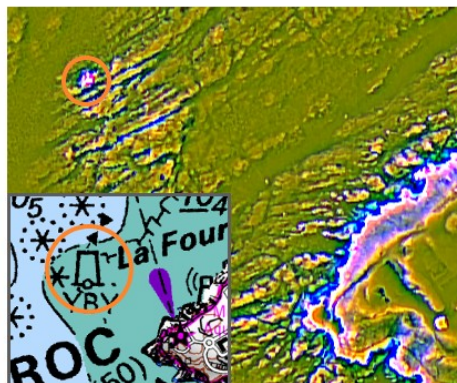


Figure 30: Balise permanente située en mer. Elle est visible sur l'image en fausse couleur et sur la carte maritime du SHOM.

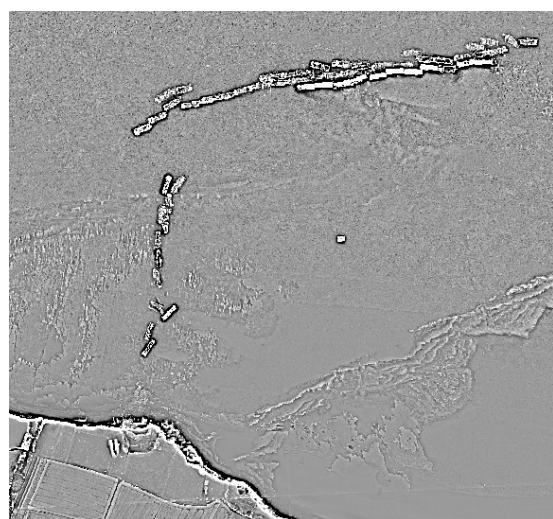


Figure 31: Épave de la seconde guerre mondiale dans le plateau du Calvados

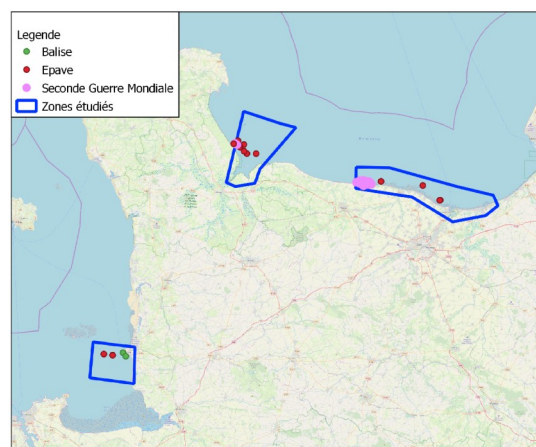


Figure 32: Répartition de 3 classes d'objets sur les zones d'études. 2 des 5 zones n'en présentait aucun

B. Classification

Le bilan de la classification supervisée est assez mitigé. Pour commencer le manque de structure sous marine ne permet pas de produire une classification pour ces ouvrages. Concernant les épaves, leur forme trop variable rend inefficace bon nombre d'indicateur réalisés, et les structures comme les balises permanentes sont facilement remarquables mais très peu nombreuses. Aucun câble n'a pu être localisé visuellement, malgré leur présence sur les cartes maritimes, cependant il y en avait peu sur la zone fournie. La classification a donc été testée sur l'estran, comme précisé précédemment, et sur la zone de Dieppe pour commencer.

Les statistiques de la classification sur la zone sont plutôt bonnes. En effet, après tri, plus de 80 % des épis ont été retrouvés avec moins de 40 % d'erreur. Ce n'est pas un résultat suffisant pour avoir une automatisation totale, mais suffisante pour permettre une vérification de données déjà existantes, par exemple pour vérifier l'existence de nouveaux éléments ou la disparition d'ancien épis.

Par contre, cette classification a nécessité de nombreuses zones d'apprentissages, avec la moitié des épis qui étaient enregistrés dans des polygones d'apprentissage.

Le modèle, cependant, est incomplet, comme le montre l'export du modèle aux autres domaines. Certaines sont bien reproduites, comme la zone 2 près de Granville, où le modèle détecte même des structures assez éloignées des objets d'apprentissage, comme des restes de bâtiments. Sur la zone 4 du plateau du Calvados, les épis sont pour la plupart invisibles sur la classification, nécessitant donc d'augmenter le nombre de polygones d'étalonnage en incorporant des épis de la zone 4.

De plus, certaines classifications menées par la suite montrent une limite importante : le dallage. En effet, le programme créant les indices fonctionnant par dalle, les différents traitements sont effectués sur chaque dalle séparée, sans aucun rapport avec ses voisines. Cela entraîne

Pourcentage de donnée saisie (par élément)	47,62 %
Pourcentage de donnée saisie (par aire)	19,36 %
Pourcentage d'épis retrouvés par classification	95,45 %
Pourcentage d'épis retrouvés par classification après tri	81,82 %
Pourcentage de l'aire total des éléments retrouvé par la classification	96,43 %

Pourcentage d'élément juste dans la classification	28,38 %
Pourcentage d'élément juste dans la classification après tri	63,64 %
Pourcentage d'aire juste dans la classification	18,86 %
Rapport des aires entre classification et données	70,96 %

Tableau 2: Statistiques de la classification sur Dieppe



Figure 33: Exemple d'une classification test de la zone 4, en zone sous-marine principalement.

des irrégularités, comme de nombreuses lignes dans des zones sableuses liées aux trajectoires de vol, mais qui ne sont pas visibles ailleurs du fait de la topographie différente.

Les classifications sont hélas trop grandes pour être montrées ici, mais des exemples sont trouvables en annexe de ce rapport.

Il est important de noter que les traitements s'opérant sur un MNS, et non sur des images radiométriques, les variations temporelles sont bien moins importantes et beaucoup plus significatives. Là où deux dates différentes entraîneront une reclassification des images radiométriques, le même modèle pourrait être utilisé pour les MNS et les variations seront en grande partie pertinentes, permettant un suivi rapide d'une zone sous réserve de l'acquisition des données (le LIDAR possède par ailleurs des conditions moins contraignantes d'acquisition que la photographie).

C. Temps de calcul et taille des données

Une problématique de ce stage était aussi le temps de calcul et l'espace utilisé pour le traitement des données.

Afin de mettre en relief les informations présentées par la suite, il a été mis à disposition pendant le stage d'un ordinateur possédant 16 Go de ram et d'un processeur Processeur Intel® Xeon® E3-1245 v2, bon processeur milieu de gamme pour les PC de calcul, mais rarement utilisé au maximum dans les traitements, la limitation étant plutôt la mémoire RAM et les accès en lecture et écriture du disque dur.

Concernant l'espace disque, les fichiers finaux font 164 Mo par kilomètre carré de données, sans variation en fonction de la densité originale des nuages. L'intégralité des 230 Go de données LIDAR est donc enregistrée en 103 Go de données raster par le traitement. A cela se rajoute la taille des fichiers créés par la sélection de données, toujours inférieure au volume des données raster brutes.

Le programme utilise par contre de nombreux fichiers temporaires durant le traitement, il faut donc prévoir au moins le volume originel des fichiers en espace libre pour pouvoir les traiter.

	Moyenne des part de calculs	Deviation Standart
temps total traitement las	39,37 %	7,60 %
temps interpolation las	5,12 %	0,98 %
temps contour las	0,22 %	0,08 %
temps total traitement image	60,39 %	7,54 %
temps préparation image mnt	0,86 %	0,45 %
temps OTB classif morpho	12,82 %	1,99 %
temps interpolation intensité	37,05 %	4,78 %
temps calcule ligne OTB	7,76 %	1,21 %

Tableau 3: Quelques statistiques du temps de calcul pour les 5 zones. Tous les pourcentages sont exprimés sur le temps total du traitement

Pour le temps de calcul, les deux plus gros traitements sont la conversion des nuages de points en raster et l'interpolation de l'intensité. Le traitement du nuage de points est long de par l'ouverture du fichier qui est lente et l'interpolation en raster qui est plus rapide mais reste significative dans le temps

total. L'interpolation de l'intensité est réalisée par une fonction qui interpole grâce à une équation cubique, qui pourrait être accélérée en baissant le degré du polynôme utilisé par cette fonction ou en changeant de fonction. Les différents rapports entre temps de calcul et volume de fichier montre que c'est le nombre de dalles qui est prépondérant dans l'estimation du temps de calcul, qui serait d'environ 50s par dalle et donc par kilomètre carré.

temps total(en s)	4955,50	15626,00	3010,5	11014,07	7477,048599
taille fichier(enGo)	17,8	86	23,3	68,5	52,1
nombre de dalles	103	294	59	200	166
temps/taille	278,40	181,70	129,21	160,79	143,51
temps/dalles	48,11	53,15	51,03	55,07	45,04
densité nuage(enGo/dalle)	0,17	0,29	0,39	0,34	0,31

Tableau 4: Statistiques de temps écoulé par zones.

Ce temps est améliorable, aussi bien par une optimisation informatique, par la création d'un programme sur une base de C plus rapide par exemple, ou par l'épuration des indices fournis (l'intensité est une information peu pertinente pour la partie sous-marine par exemple). La taille des données est elle aussi toujours un problème, mais il est possible de réaliser la quasi-totalité des traitements sur un MNT, bien plus léger et pratique à échanger que des nuages de points laser.

Il est bon de préciser que la sélection de données est elle aussi assez lente, de l'ordre de quelques secondes par dalle, du fait du temps de lecture et d'écriture du disque dur. Sélectionner les bonnes données pour une zone comme le Plateau du Calvados (300 dalles) peut prendre jusqu'à une quinzaine de minutes pour une classification par exemple.

Les différents tests sur la classification ont aussi été longs de par la taille des zones étudiées, mais aucune statistique de temps n'a été réalisée de par la trop grande variété des travaux réalisés.

Conclusion

Pour aborder la conclusion avec justesse, il faut déjà rappeler que les résultats obtenus sont très dépendants de la méthode utilisée. Elle consiste ici par la création d'indices d'origines géométriques comme la rugosité, à l'aide d'un code python puis par la classification supervisée de ces images pour l'identification automatique.

Les résultats mis en avant de cette recherche sont une richesse d'informations bien présente sur les données LIDAR et qui sont repérables visuellement grâce à quelques traitements mais une difficulté à extraire automatiquement ces informations à l'aide d'une classification supervisée. Il reste quelques pistes à explorer plus complexes et qui nécessiteraient plus de temps pour pouvoir les maîtriser, mais les limitations de données d'entrée et la variance des objets sous-marins rendent la détection automatique difficile avec les méthodes utilisées ici.

Cependant cette assertion s'est envisagé assez tôt dans le stage, et le fait que les résultats la confirment n'est pas une surprise. C'est pour cette raison qu'une partie du stage a été consacrée à la numérisation manuelle, au développement des capacités de visualisation et d'utilisation, et aux autres possibilités d'utilisation de ce type de données pour le littoral.

Il en sort un bilan riche, laissant transparaître les possibilités de cette donnée pour le suivi du milieu côtier. Le passage en MNS complété par des indices géométriques permet une visualisation et une surveillance d'étendues très vaste, en particulier pour le suivi des structures humaines, comme nous l'avons vu lors de la numérisation qui a été effectuée. Il a en effet été possible d'utiliser plus de 800 km² de surface pour le stage, et la numérisation n'a pris que quelques heures pour plus de 900 objets. Les différents tests d'automatisation montrent de vraies possibilités, même si la méthode choisie s'est heurtée à quelques difficultés. Malgré celles-ci, ce processus s'est montré pertinent pour la création d'un suivi partiellement automatisé, à l'aide notamment des avantages de l'acquisition LIDAR face à la photographie sur le milieu côtier. Les temps de calcul se sont avérés acceptables, malgré la taille des fichiers traités et produits, même si les temps d'affichage et de sélection sont encore un peu longs. Le format utilisé dans ce stage n'est par contre probablement pas pertinent, car les fichiers sont des fichiers très volumineux, nécessitant quelques compétences techniques et livrés ici avec une qualité d'information en dessous de la qualité attendus. Un MNS ou MNT accompagné de quelques données supplémentaires aurait été tout aussi efficace, bien moins volumineux et plus accessible pour des utilisateurs habitués au SIG.

Ce stage a été l'occasion de mener un vrai travail de recherche en autonomie et d'intégrer le cœur d'une chaîne de traitements de données au sein d'une structure dotée de nombreuses compétences diverses pour m'aider dans mon stage. Il a fallu non seulement produire les résultats ici présentés mais aussi diriger les axes d'études au fur et à mesure des différents résultats en collaboration avec mon maître de stage et les différents acteurs, et s'adapter aux aléas comme la vérification relativement approfondie de la qualité des données.

Bibliographie

Principes fondamentaux et tutoriels.

pythonhosted.org. <https://pythonhosted.org/laspy/tutorial.html>

Tutoriel pour l'utilisation de laspy et la gestion de fichiers .las en python

docs.qgis.org. https://docs.qgis.org/2.8/fr/docs/training_manual/forestry/basic_lidar.html

Tutoriel pour l'utilisation du plugin lastools

www.umn-cnrm.fr. Fiche de cours réalisé par Clément Mallet.

https://www.umn-cnrm.fr/ecole_lidar/IMG/pdf/Mallet-Topo_Bathy_Veget.pdf

Principe fondamental du LIDAR bathymétrique.

Rafael Grompone von Gioi , J'érémie Jakubowicz , Jean-Michel Morel , Gregory Randall. *LSD: a Line Segment Detector*. Published in Image Processing On Line on 2012-03-24.

<https://www.ipol.im/pub/art/2012/gjmr-lsd/article.pdf>

Démonstration du fonctionnement de l'algorithme LSD (Line Segement Detection)

Sébastien Labarre, *Caractérisation et modélisation de la rugosité multi-échelle des surfaces naturelles par télédétection dans le domaine solaire*, Thèse de doctorat en Physique de l'Univers présentée et soutenue publiquement le 8 novembre 2017 à l'UNIVERSITÉ PARIS DIDEROT

http://www.ipgp.fr/~jacquemoud/publications/these_Labarre_2017.pdf

Description du principe de rugosité de façon poussé et exhaustive.

Applications du LIDAR et donnée d'entraînement.

Laure Nuninger, Krištof Oštir, Ziga Kokalj, Ales Marsetic. *Lidor. Acquisition, traitement et analyse d'images LiDAR pour la modélisation des paléo-reliefs de la plaine littorale du Languedoc oriental..* 2008, pp.95. halshs-00483129.

<https://halshs.archives-ouvertes.fr/halshs-00483129/document>

Application des données LIDAR pour le suivi de variation de relief.

www.coast.noaa.gov.

<https://www.coast.noaa.gov/dataviewer/#/lidar/search/>

Site de donnée LIDAR pour le suivis de la côte américaine.

Classification

Mohamed Mahmoud Sidi Youssef, Clément Mallet, Nesrine Chehata, Arnaud Le Bris, Adrien Gressin. *Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante*. Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014, Jun 2014, France.

hal-00989026

<https://hal.archives-ouvertes.fr/hal-00989026/document>

Detection automatique des batiments par images satellites.

Wang, Sheng-Guo & Deng, Jing & Chen, M.-Z & Weatherford, M & Paugh, L. (2015). *Random Forest Classification and Automation for Wetland Identification based on DEM Derivatives*. Publié le 26 juillet 2016 sur la plateforme researchgate.net.

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/305385776_Random_Forest_Classification_and_Automation_for_Wetland_Identification_based_on_DEM_Derivatives)

[305385776_Random_Forest_Classification_and_Automation_for_Wetland_Identification_based_on_DEM_Derivatives](https://www.researchgate.net/publication/305385776_Random_Forest_Classification_and_Automation_for_Wetland_Identification_based_on_DEM_Derivatives)

Application d'indice géométrique à l'archéologie.

Arthur Le Guennec Dimitri Lague Sébastien Lefèvre Thomas Corpetti. *Classification de données LiDAR bi-spectral topo-bathymétriques par une approche multi-échelle : Application en milieu fluvial*.

Congrès RFIAP 2018 et CFPT 2018, Cité Descartes (Marne-la-Vallée), organisé par les associations AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes), SFPT (Société Française de Photogrammétrie et de Télédétection) et le chapitre français de la société GRSS de l'IEEE

[https://rfiap2018.ign.fr/sites/default/files/ARTICLES/CFPT2018/Oraux/](https://rfiap2018.ign.fr/sites/default/files/ARTICLES/CFPT2018/Oraux/CFPT2018_paper_leguennec.pdf)

[CFPT2018_paper_leguennec.pdf](https://rfiap2018.ign.fr/sites/default/files/ARTICLES/CFPT2018/Oraux/CFPT2018_paper_leguennec.pdf)

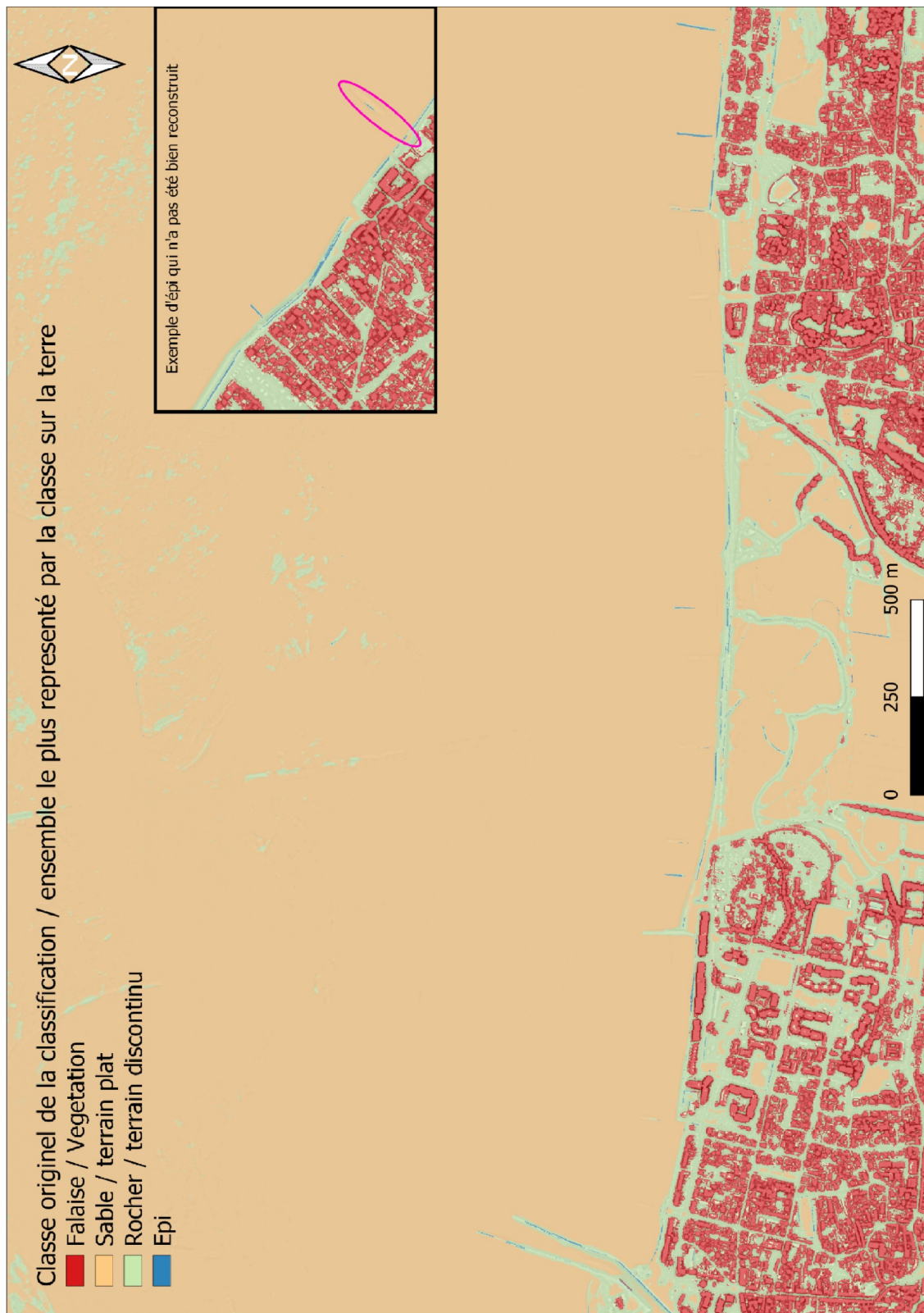
Une des méthodes fonctionnelles vu en phase de pré-recherche.

Annexes

La classification a été réalisée en utilisant uniquement des éléments de l'estran ou des fonds marins. Cependant les classes produites restent intéressantes sur la terre pour la visualisation et pour se repérer, elles ne sont donc pas masquées.



Annexe 1: Classification dans la zone de Granville



Annexe 2 : Classification sur le Plateau du Calvados



Annexe 3 : Classification sur la zone à l'ouest de Dieppe