

SENVISAT

Etude de faisabilité d'un Suivi
ENVironnemental par SATellite dans le
Parc naturel marin du golfe du Lion

Phase 2 : Suivi du couvert végétal avec les données satellites PLEIADES

15 novembre 2019

EID & CEFREM, 2019



Référence à utiliser pour toute citation du document

Richard, B., Gervais, M., Cros, M., Bourachot, A., Heurtefeux, H., Aleman, N., Robin, N., Raynal, O., Certain, R., 2019. Suivi du couvert végétal avec les données satellites PLEIADES. Rapport phase 2 - SENVISAT.

CONTACTS

Mathieu Gervais mgervais@eid-med.org 04.30.63.67.82	EID Méditerranée Pôle Littoral 165 avenue Paul Rimbaud 34 184 Montpellier Cedex 4
Hugues Heurtefeux hheurtefeux@eid-med.org 04.67.63.72.99	
Nicolas Aleman nicolas.aleman@univ-perp.fr 04.68.66.20.57	Laboratoire CEFREM Université de Perpignan Via Domitia, Bat U 52, av. Paul Alduy 66860 Perpignan Cedex
Nicolas Robin nicolas.robin@univ-perp.fr 04.30.19.23.00	

Le pôle Littoral de l'EID Méditerranée s'investit dans la protection et la restauration des cordons dunaires et des plages. Les connaissances scientifiques et techniques accumulées sur cet espace constituent aujourd'hui un capital mis à disposition des collectivités territoriales et des établissements publics. Ainsi nous concevons et proposons aux collectivités des actions de protection et de réhabilitation des plages et des dunes, ainsi que des modalités de gestion de la fréquentation adaptées à ces sites.

Le laboratoire CEFREM (Centre de Formation et de Recherche sur les environnements Méditerranéens) de l'Université de Perpignan Via Domitia est spécialisé notamment dans l'étude détaillée des données sédimentaires et morphologiques pour la compréhension de la dynamique littorale. Son activité de recherche est focalisée sur le thème des transferts de matière et d'énergie aux interfaces du système côtier.

Suivi du couvert végétal avec les données satellites

PLEIADES

PHASE 2 – Contrat de recherche SENVISAT

Bertrand Richard

Myriam Cross

Mathieu Gervais

Arthur Bourachot

Hugues Heurtefeux

Nicolas Aleman

Nicolas Robin

Olivier Raynal

Raphaël Certain

TABLE DES MATIERES

PREAMBULE.....	7
1. INTRODUCTION.....	7
2. SITES D'ETUDE.....	8
3. LA TELEDETECTION POUR LA CARTOGRAPHIE DE L'OCCUPATION DU SOL.....	11
4. LES DONNEES SATELLITES OPTIQUES.....	12
4.1 Les données Pléiades	12
4.2 Autres sources disponibles	13
5. LES PRINCIPALES METHODES D'EXTRACTION DE L'INFORMATION	14
5.1 Les indices radiométriques	14
5.2 La classification non-supervisée.....	15
5.3 La classification supervisée	15
5.4 L'approche orientée objet	15
6. DEFINITION DE LA METHODE	16
7. LES INDICATEURS - RESULTATS.....	17
7.1 La cartographie annuelle des milieux dunaires.....	17
La végétation	17
Les cheminements.....	18
Le trait de cote.....	19
7.2 La détection de changements avec une approche multi-dates	23
8. APPORTS METHODOLOGIQUE POUR LA FINALISATION DE LA CHAINE DE TRAITEMENT.....	27
8.1 Les indices	27
8.2 Les méthodes de classification	27
8.3 La définition de la nomenclature	28

9. CONCLUSION GENERALE	31
10. PERSPECTIVES	32
REFERENCES BIBLIOGRAPHIQUES	33
ANNEXE – METHODOLOGIE DE CLASSIFICATION BASEE SUR LA METHODE DU CEREMA (EXTRAIT DU RAPPORT DE STAGE)	35
<i>Principe général</i>	35
<i>Démarche détaillée</i>	37
<i>Protocoles de test et variations des paramètres</i>	43

PREAMBULE

Ce rapport constitue la phase 2 du contrat de recherche " Suivi ENVironnemental par SATellite" (SENVISAT) entre le Parc Naturel Marin du Golfe du Lion (PNMGL), l'Université de Perpignan Via Domitia (UPVD-CEFREM, CNRS, et équipe LMUSCA), l'Observatoire de la côte sableuse catalane (OBSCAT), et l'Entente Interdépartementale pour la Démoustication (EID- méditerranée).

La partie Occitane du littoral méditerranéen est sujette à de multiples modificateurs anthropiques, morphologiques et biologiques comme le vent ou la houle qui influent sur la santé des habitats dunaires. Le travail mené au cours de cette phase 2a a eu pour objectif de mener une étude de faisabilité sur l'utilisation de l'imagerie satellitaire pour le suivi de la végétation dunaire et de la morphologie côtière.

1. INTRODUCTION

Le littoral méditerranéen est une interface en perpétuel changement. Son fonctionnement est issu de l'interaction entre écosystèmes terrestres et marins et de la façon dont l'homme s'est approprié le territoire. L'évolution de sa géomorphologie et de sa biodiversité est donc fortement tributaire de ces interactions. Les aléas naturels et activités anthropiques sont des éléments qui le redessinent en permanence. Il est de ce fait primordial pour les collectivités d'avoir conscience des évolutions de leur territoire.

Mettre en place un suivi régulier des plages et des dunes devient un enjeu capital pour comprendre les phénomènes et définir des programmes d'actions à mener pour la préservation, la restauration et la bonne gestion des plages en tant que ressource économique mais aussi environnementale.

Le Parc naturel marin du Golfe du Lion conscient des problèmes liés à ces altérations du littoral souhaite mettre en place des appels à projet pour le suivi environnemental, ce projet SENVISAT en est une conséquence. Situé en Occitanie, le territoire du parc comprend près de 100 km de côte particulièrement touchée par les problématiques d'érosion et de submersion liés aux forçages météorologiques et aux apports fluviaux ainsi qu'aux activités anthropiques. Ces facteurs provoquent des modifications sur le milieu littoral d'un point de vue morphologique et biologique ayant des conséquences sur les habitats et les écosystèmes : recul du trait de côte, perte sédimentaire et modification des habitats dunaires notamment.

Reconnue comme étant une des approches les plus efficaces et les plus économiques en terme d'élaboration d'occupation du sol sur de larges zones (Cihlar, 2000; Manakos et Braun, 2014), la télédétection s'avère aussi être adaptée pour récupérer des informations à une échelle plus fine, notamment grâce au développement de capteurs à très haute résolution spatiale.

La mise à disposition croissante pour les acteurs publics d'images satellites tel que le Pole Theia¹ et le projet Equipex Geosud², et le développement des outils libres de traitement d'image tel que OrfeoToolbox³ (OTB) développé par le Centre national d'études spatiales (CNES) offrent un panel de possibilités qu'il était intéressant d'étudier pour son application dans le domaine de la gestion de littoraux.

Dans ce contexte, la phase 2 du projet vise à **étudier la végétation dunaire** en utilisant les **données satellites Pléiades**. Ce type de suivi permettrait alors, par rapport à un suivi terrain, d'étudier les systèmes dunaires à grande échelle et de suivre de manière plus régulière l'efficacité des méthodes et aménagements pour protéger les dunes, tout cela avec un coût réduit.

Au cours de l'année 2017 un sujet de stage a été lancé à l'EID-Méditerranée afin d'évaluer les intérêts d'une telle méthode de suivi. Ce travail a abouti à un rapport de stage qui a permis de mettre en évidence les possibilités qu'offrait l'utilisation des images Pléiades dans le logiciel de traitement d'images OrfeoToolbox et du logiciel Qgis. L'objectif principal de ce stage a été de montrer la faisabilité d'identifier à très haute résolution spatiale les principaux habitats dunaires : la dune vive (blanche), la dune fixée (grise) et la plage (Larroque et al., 2016) pour ensuite mettre en avant les changements entre deux dates afin d'avoir un aperçu concret des modifications morphologiques et floristiques. Enfin, la tâche finale a été d'automatiser les différentes étapes de la méthodologie en produisant une chaîne de traitement capable de reproduire cette démarche.

Dans la continuité de ce travail, l'EID-Méditerranée a continué au cours de l'année 2019 à perfectionner la méthode développée durant ce stage dans le but de la rendre plus efficace, plus simple, et opérationnelle.

2. SITES D'ETUDE

Le littoral du golfe du Lion tel que nous le connaissons s'est globalement formé à la fin de la transgression⁴ suivant le dernier âge glaciaire, le « Würm », 6000 ans avant notre ère. La morphologie résultante est essentiellement composée de plages à lidos⁵, où une étroite bande de sable sépare la mer des étangs, ce qui représentent le milieu caractéristique du littoral ouest-méditerranéen français (Larroque et al., 2016).

Principales indicateurs de la santé du littoral, les dunes et plages sont constituées de sédiments dont la majeure partie du stock s'est formée durant la transgression. Il ne se reconstitue que très faiblement, et les aménagements anthropiques contribuent à bloquer les apports potentiels en provenance des cours d'eau. C'est un milieu particulièrement dynamique, voire instable (Paskoff,

¹ <https://www.theia-land.fr/>

² <http://geosud.ign.fr/accueil>

³ <https://www.orfeo-toolbox.org/>

⁴ On parle de transgression marine lorsqu'une zone terrestre littorale est recouverte par la mer sur le long terme, lié à un affaissement des terres ou à une élévation progressive du niveau de la mer suite à la fonte massive des glaciers issus d'une ère glaciaire, par exemple.

⁵ Cordon littoral en position avancée à l'entrée d'une baie et pouvant isoler une lagune

1998), principalement modelé par le vent, la houle et les courants associés. Ces dynamiques sont indicatrices de l'évolution du milieu (García-Mora et al., 2000) et c'est la raison pour laquelle un suivi régulier est nécessaire pour en apprendre davantage sur l'évolution des habitats dunaires et anticiper les risques littoraux.

Le milieu dunaire présente un faciès composé de multiples habitats dont l'organisation est régulière sur le territoire, bien que de nombreuses variations et particularités distinguent les plages le long du littoral. Les principales communautés végétales du littoral méditerranéen se présentent théoriquement de la manière suivante, de la mer jusqu'à la lagune (Figure 1

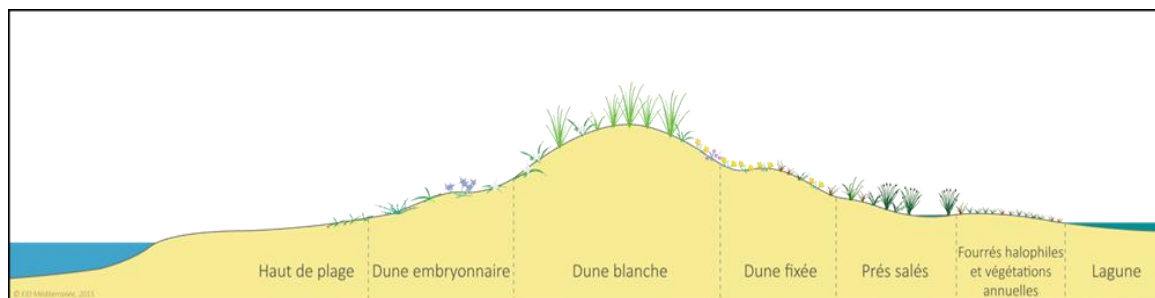


Figure 1 : Coupe transversale théorique illustrant la succession des principales communautés végétales du littoral méditerranéen – schéma EID Méditerranée

Les traitements se sont concentrés sur deux zones représentatives du littoral du PNMGL : les plages du Barcarès et de Torreilles (Figure 2). Sur l'ensemble du système littoral qui nous intéresse, on souhaite pouvoir distinguer plusieurs éléments. Dans un premier temps, il doit être possible de distinguer correctement les trois structures principales : le cortège de végétation principal, la plage, et la mer. Dans un second temps, il faut voir dans quelle mesure il est possible de différencier les types de végétation, notamment en termes de densité, et les différents types de sols. Enfin, la méthode devra également permettre d'extraire correctement les chemins de plage au travers la dune.

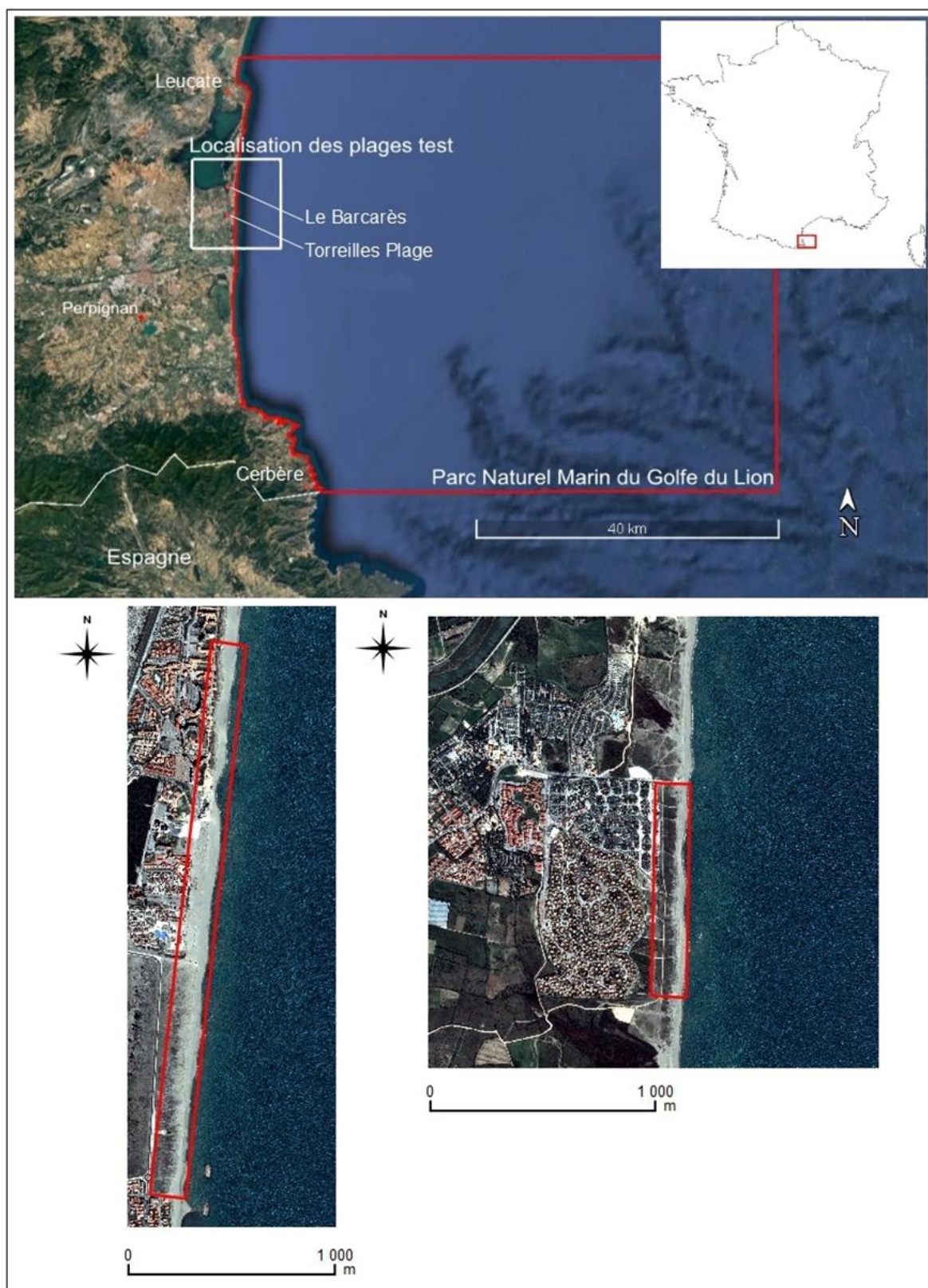


Figure 2 : Localisation des plages de test et de la zone d'étude. En haut : le Parc Naturel Marin du Golfe du Lion. A gauche : Barcarès. A droite : Torreilles

3. LA TELEDETECTION POUR LA CARTOGRAPHIE DE L'OCCUPATION DU SOL

Le principe de la télédétection optique (dite « passive » à l'inverse du radar) est d'utiliser le rayonnement solaire comme source d'information. En entrant en contact avec un objet au sol, ce dernier va absorber une part de l'énergie reçue et réémettre le reste. Le capteur du satellite est chargé de récupérer l'information émise par l'objet au sol et de la traduire sous la forme d'une image.

Chaque objet émet donc différemment dans chaque bande, et la génération d'une image est issue de la traduction de la réponse de l'objet dans ces bandes en valeurs par pixel. En associant sur l'image une couleur à la bande correspondante. La manipulation de cette information spectrale permet ensuite de mettre en valeur des phénomènes ou d'extraire des éléments en particulier.

La télédétection, et plus particulièrement l'usage de l'imagerie optique, est une discipline plébiscitée pour la cartographie de l'occupation du sol et de la végétation spécifiquement. De nombreuses applications existent, et le suivi des territoires pour évaluer la santé de ses milieux naturels est une catégorie à part entière des multiples champs de recherche en la matière. En fonction des besoins exprimés, il est possible de travailler à des échelles et précisions variables, sur des espèces précises ou sur de larges zones avec des densités plus ou moins importantes. De ce fait, le développement de la télédétection à très haute résolution spatiale a permis une avancée considérable pour l'exploitation de zones hétérogènes ou pour le travail de précision puisque cela permet notamment de se mettre à une échelle supérieure à celle de l'objet étudié en espérant notamment extraire des informations inatteignables jusqu'alors, tels que les habitats dunaires.

4. LES DONNEES SATELLITES OPTIQUES

4.1 Les données Pléiades

Afin de mettre en place ce suivi il a été décidé de se porter sur les données issues du programme Pléiades. Ces données présentent l'avantage d'être ce qui est le plus précis en ce qui concerne l'imagerie satellite civile européenne. Ces images à très haute résolution spatiale (THRS) offre une résolution de 50 cm dans les canaux Bleu, Vert, Rouge et Proche Infra-Rouge (PIR). La végétation ayant une réponse très forte dans l'infra-rouge (IR) la présence de la bande PIR contient une information capitale en ce qui concerne l'étude de la végétation, permettant de la distinguer des autres éléments de l'image mais permettrait également de distinguer les différents types de végétation caractéristiques des différents habitats dunaires. Ces images codées sur 16 bits sont calibrées optiquement (niveau *top-of-atmosphere*) et géométriquement, apportant précision géographique et éliminant les effets liés à la position du capteur, au relief, mais pas des effets atmosphériques.

Les acquisitions ont été faites durant le printemps (Tableau 1), période plus propice à la vivacité de la végétation qui se situe entre les grandes houles hivernales qui perturbent la morphologie côtière et les sécheresses estivales qui endommagent grandement les habitats végétaux.

Tableau 1. Récapitulatif des images Pléiades acquises par le Parc Naturel Marin du Golfe du Lion via les acquisitions à la demande de Geosud

Prise de vue	Référence
31.03.2015	ORTHO-SATr_1-0 LITTO-MED-66_2015033139184457 ORTHO-SATr_1-0 LITTO-MED-66_2015033139189628
29.04.2017	ORT_2017042939393822_LA93 ORT_2017042939431069_LA93
07.05.2017	ORT_2017050738742331_LA93

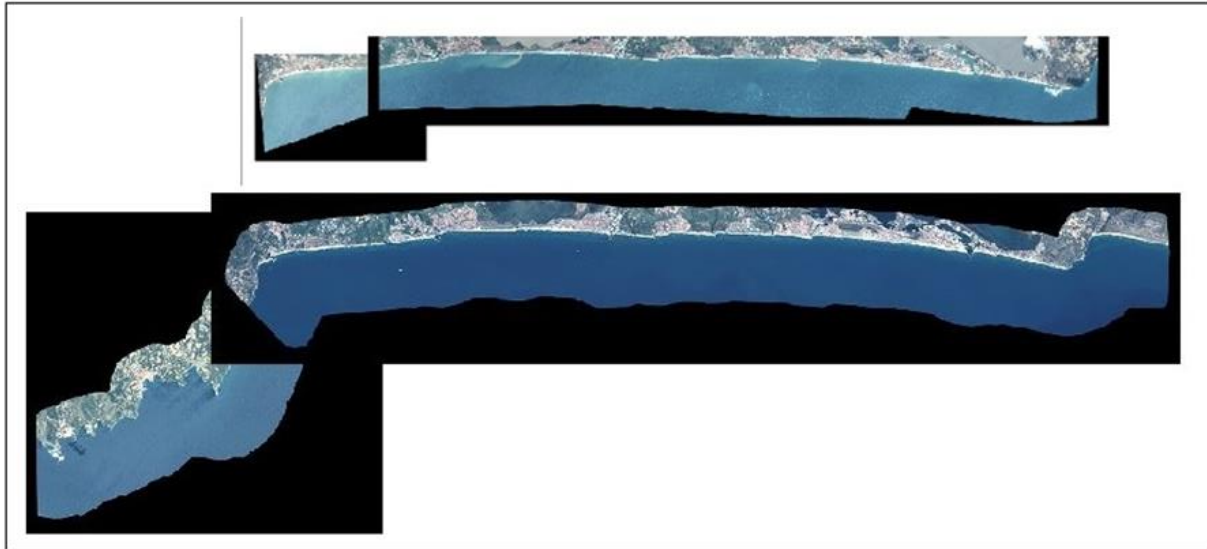


Figure 3 : Sets d'images Pléiades mis à disposition par l'Equipex GEOSUD. En haut: dalles du 31/03/2015 ne comprenant que la côte sableuse de Leucate à Argelès. En bas: dalles du 29/04 2017 et 07/05/2017 de l'ensemble du PNM, de Cerbères à Leucate.

4.2 Autres sources disponibles

Il est à noter que la possibilité d'utiliser des images Spot 6/7 avait été envisagée cependant les traitements présentés dans ce rapport sont effectués sur les images Pléiades disposant de la meilleure résolution spatiale, Spot 6/7 disposant seulement d'une résolution de 1,5 m (Figure 4). Elles sont fournies pré-fusionnées, ce qui épargne une étape de traitement. Les méthodes présentées dans ce rapport pourraient s'appliquer de la même manière sur ces images car seul un *pansharpening* est nécessaire afin de d'obtenir une résolution à 1,5 m dans le multi-spectral. L'avantage de Spot est qu'il y a beaucoup plus d'images gratuites disponibles sur la plateforme Equipex Geosud car chaque année une couverture nationale est réalisée.

Une approche basée sur une analyse de série temporelle d'images Sentinel-2 avait également été envisagée pour apporter une information supplémentaire en prenant en compte les cycles végétatifs de plantes. Cependant bien que gratuites et facilement accessibles (voir le tutoriel de téléchargement des images Sentinel-2, Annexe 2, en lien avec la phase 3 du projet) la résolution de 10 m n'aurait pas été suffisante pour identifier avec autant de finesse les morphologies dunaires.

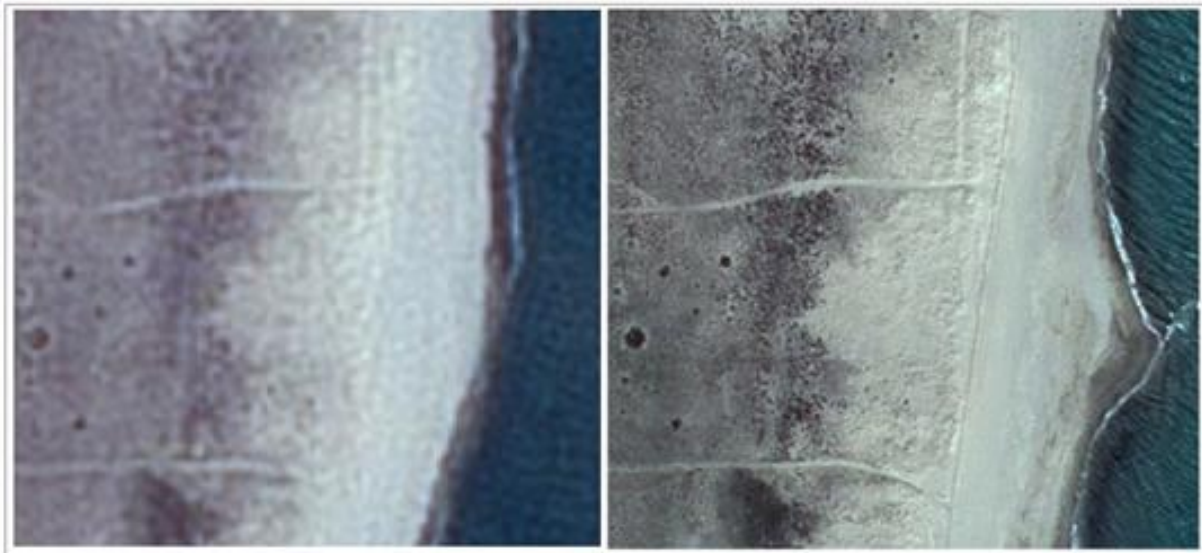


Figure 4 : Comparaison des deux acquisitions sur la plage du Barcarès (Occitanie). A gauche un échantillon SPOT-7 dont la résolution spatiale après fusion est de 1.5 m. A droite un échantillon Pléiades de résolution 0.5 m après fusion.

5. LES PRINCIPALES METHODES D'EXTRACTION DE L'INFORMATION

Pour extraire des éléments d'une image satellite optique, il est possible d'utiliser des méthodes de classification (Bonn et Rochon, 1992; Collet, 2001) ainsi que le calcul de différents indices radiométriques. On peut distinguer sur une image des objets de différentes natures, telles que la végétation, les surfaces artificialisées, les ombres, ou l'eau par exemple. Les objets de même nature possèdent des propriétés spectrales similaires, ce qui se traduit sur une image optique par des couleurs similaires, puisque les signatures spectrales des objets sont représentées numériquement par des valeurs de pixels.

Une classification consiste à effectuer cette distinction de manière automatique à l'aide d'un algorithme de classification qui va classer les pixels en fonction de leurs valeurs numériques. Le principe général étant d'identifier l'occupation du sol, et non pas l'utilisation du sol : un parking et le toit d'un immeuble auront une signature spectrale quasi-identique bien que ce soit des éléments urbains totalement différents.

5.1 Les indices radiométriques

Les indices radiométriques en particulier ont été développés pour répondre à ce besoin d'observation du territoire et d'extraction à grande échelle de la végétation pour constituer de vastes bases de données d'occupation du sol.

Le NDVI « Normalized difference vegetation index » (Rouse Jr et al., 1974) est le plus connu : il s'agit d'un indicateur permettant d'établir un rapport entre la bande spectrale du rouge et du proche infrarouge. Compte tenu de la réflectance élevée de la végétation dans le proche-

infrarouge, l'indice permet de mettre en valeur cette différence et de percevoir les variations dans la couverture végétale ou même de différencier les espèces entre elles (Jackson et Huete, 1991).

Ce type d'indice s'avère particulièrement efficace pour détecter la végétation littorale ou urbaine lorsqu'il est couplé à des méthodes de discrétisation précises (Crombette, 2016) ou à des analyses in-situ de la biomasse pour paramétrer le choix des capteurs hyperspectraux (Bargain, 2012).

Il existe néanmoins d'autres indices radiométriques, tous plus ou moins dérivés du NDVI et de la droite des sols (Baret et al., 1993) qui cherchent à s'affranchir des effets causés par l'atmosphère ou par le sol nu (Bausch, 1993; Elvidge et Chen, 1995; Elvidge et Lyon, 1985; Escadafal et Huete, 1991). On parle notamment du SAVI « Soil-adjusted vegetation index » (Huete, 1988), du MSAVI « Modified soil-adjusted vegetation index » (Qi et al., 1994) ou du TSARVI « Transformed soil atmospherically resistant vegetation index » innovant dans son utilisation de la bande bleue (Bannari et al., 1996). Pour une comparaison de l'efficacité de ces indices, se reporter sur les travaux de Bargain (2012) et Bannari et al. (1995, 1998).

5.2 La classification non-supervisée

On distingue deux approches de classification : la première est la classification non supervisée, où on laisse l'algorithme analyser les signatures spectrales des pixels de l'image puis les ranger dans des catégories naturelles, les pixels devant être le plus homogènes possibles au sein d'une catégorie. On retrouve souvent la classification non-supervisée dans la littérature scientifique puisqu'elle permet d'éviter le biais d'interprétation en automatisant au maximum les traitements. Pour la détection des faciès géomorphologiques littoraux, on retrouve par exemple des travaux sur FORMOSAT-2 (Mallet et al., 2007) avec notamment l'utilisation de fonctions d'optimisation des résultats (Lafon et al., 2008). La force de la classification non-supervisée réside dans sa rapidité, mais malgré ses performances dans l'imagerie à très haute résolution spatiale de type Pléiades (Sublime et al., 2017) les publications indiquées ci-dessus tendent à montrer que la détection en milieu littoral se passe difficilement de l'intervention humaine.

5.3 La classification supervisée

La seconde approche est la classification supervisée, où l'interprète associe lui-même des zones de l'image à des catégories sous forme d'échantillons de données vectorielles. L'algorithme va alors se charger de répartir les pixels dans les classes en fonction des réponses spectrales des échantillons. Plusieurs algorithmes de ce type existent. Les plus utilisées sont *Support Vector Machine* et *Random Forest*.

5.4 L'approche orientée objet

L'approche orientée objet est une méthode dérivée de la classification précédente. Il s'agit pour l'algorithme de ne traiter non pas chaque pixel de manière isolée, mais de les segmenter sous forme de groupes de pixels afin de les traiter dans leur contexte ou leur voisinage à l'aide de critères de texture, de forme et des caractéristiques spectrales. On peut dès lors traiter de manière bien plus complète l'information à très haute résolution spatiale et s'affranchir des

limites de l'approche « pixel » lorsque des objets sont trop détaillés pour n'être décrits que par un ensemble de pixels homogènes.

Une fois que l'on a procédé à l'étape de segmentation de l'image, qui peut prendre beaucoup de temps en fonction de la taille de l'image, il faut lancer une étape de classification supervisée afin de classer les segments générés. La méthode demande cependant une quantité relativement importante d'échantillons (Rougier et al., 2016). On retrouve par ailleurs une bonne partie des travaux sur l'approche orientée objet pour la classification du milieu urbain, notamment pour faire ressortir les éléments de végétation (Rougier et Puissant, 2014), ainsi que pour la cartographie de l'occupation du sol. On ne trouve cependant que peu d'informations sur l'approche orientée objet dans les milieux littoraux.

6. DEFINITION DE LA METHODE

La méthode qui a été définie lors du stage s'est basée sur une méthodologie (Bouffier et al., 2015) développée par le centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (CEREMA) dans le but d'établir une cartographie de l'occupation du sol du département de la Drôme. Cette approche utilise l'indice de végétation par différence normalisé (NDVI) qui est un rapport entre les bandes spectrales rouge (R) et PIR selon : $NDVI = (PIR - R) / (PIR + R)$. Les indices de textures d'Haralick sont également utilisés (Figure 5 - a). L'ensemble de l'information contenue dans les différentes bandes du satellite plus l'information contenue dans le canal NDVI et dans les bandes de descripteurs texturaux va être ensuite analysé par un algorithme de classification supervisée *Support Vector Machines* (SVM) (Figure 5 - b). Cette méthode a pour but d'imiter l'approche objet sans à avoir à effectuer une segmentation de l'image permettant ainsi de gagner en vitesse de calcul. Afin d'effectuer cette simulation il est nécessaire d'effectuer des pré-traitements sur le jeu de données d'apprentissage. Cette méthode consiste à utiliser une pré-classification des données d'entraînement via la méthode de classification non-supervisée *Kmeans* qui va diviser chaque macro-classe (eau, sable, végétation...) en plusieurs sous-classes (Figure 5 - c et d). Ces classes n'ont pas de sens thématique mais sont seulement des classes radiométriques qui vont faciliter la « vraie » classification. Un fois la classification terminée l'opérateur doit alors fusionner les sous classes obtenues en fonction des classes qu'il souhaite obtenir.

La méthodologie utilisée ici est une approche pixelaire, car c'est cet élément élémentaire de l'image qui va être classifié. L'autre approche proposée aurait été l'approche objet, ou l'on aurait classifié des ensembles de pixels homogènes au niveau de leur forme et de leur radiométrie, nécessitant de passer par un processus de segmentation de l'image. Cependant bien qu'OTB dispose de ces outils c'est une approche complexe pour quelqu'un n'ayant de formation en télédétection. Les temps de traitements, assez courts pour l'approche pixelaire, aurait été fortement allongés surtout pour traiter des images complètes en utilisant une machine classique.

Pour une description détaillée de la méthode se reporter à l'ANNEXE – Méthodologie de classification basée sur la méthode du CEREMA (extrait du rapport de stage) à la fin de ce document.

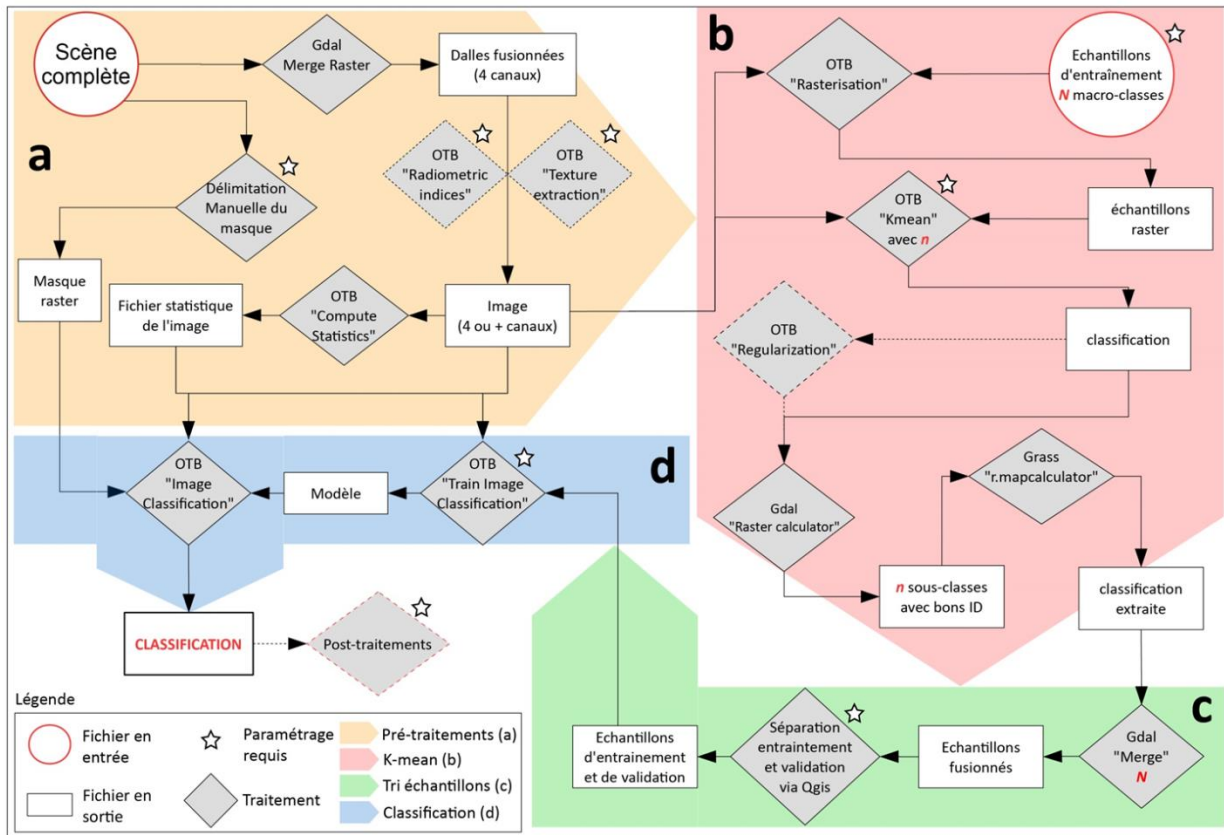


Figure 5 : Représentation schématique complète de la méthodologie de classification.

7. LES INDICATEURS - Résultats

7.1 La cartographie annuelle des milieux dunaires

Cette première méthode a été appliquée sur les plages du Barcarès en 2015 (Figure 7) et 2017 (Figure 8) et de Torreilles en 2015 (Figure 9) et 2017 (Figure 10). **Le bilan global apparaît satisfaisant avec une séparation très nette des zones végétalisées du reste des autres secteurs de la plage.** Il est cependant difficile d'identifier comme prévu les habitats dunaires de façon précise, bien que ce fût un objectif volontairement ambitieux car les différents habitats étant particulièrement difficiles à interpréter même à l'œil nu. Il est également possible de distinguer la plage de la dune, et d'extraire partiellement la zone transitoire où se mêlent le sable et la végétation.

La végétation

Il est possible de distinguer trois états de conservation de la dune (Figure 7, Figure 8, Figure 9, Figure 10) avec les différentes densités de végétation, pouvant potentiellement correspondre à

des types d'habitats, permettant quand même de fournir des informations précieuses sur l'organisation de la morphologie dunaire.

Quantifier l'état des milieux dunaires est complexe car cela dépend beaucoup du site et donc des connaissances terrain. La classification d'images offre ainsi l'avantage d'**avoir une vision globale d'un site et des superficies recouvertes par les différents milieux**. Les données obtenues donnent alors la possibilité de calculer des taux de végétalisation homogène sur un secteur car correspondant à une même réalité radiométrique. Un indicateur de densité de végétalisation globale avec tout type végétation confondue peut donc être obtenu en fusionnant les classes de végétation.

On peut également calculer un taux de recouvrement prenant en compte les différents états de conservation de la végétation dunaire, avec cependant une marge d'erreur plus importante. Une fois la classification effectuée on peut imaginer découper la zone en plusieurs secteurs pour calculer des surfaces de recouvrements qui décriront plus en finesse certains sites spécifiques.

Ces indicateurs de surface de recouvrement permettraient sur plusieurs années de caractériser les évolutions de milieux dunaires et de les mettre en rapport avec les facteurs climatiques et anthropiques qui influencent ou non la vivacité des milieux dunaires tel que les précipitations, les tempêtes, et le niveau de protection (ex. : mise en place de nouvelles ganivelles...) afin de mieux cerner les dynamiques de ce milieu sur le long terme. La télédétection permettrait également un suivi des chantiers de végétalisation des dunes, à partir de l'état post-chantier jusqu'au suivi complet des différentes étapes de végétalisation. Il serait dès lors possible d'analyser l'efficacité des différentes mesures mises en place et même de comparer des sites entre eux.

Les cheminements

Sur les différentes cartographies réalisées on voit ressortir les cheminements de plage. Cependant tous ne se sont pas forcément visibles après classification car la résolution des images est limitée à 0,5 m. Les chemins identifiables sont ceux qui ont au moins 1 pixel d'information radiométrique pure (sable), c'est-à-dire, ceux d'au moins 1,5 m de largeur. Pour les cheminements plus étroits les résultats seront plus aléatoires car représentés par une réponse spectrale mixte (sable et végétation). On les retrouvera ainsi souvent dans la classe végétation épars (Figure 7).

Certains cheminements en état de fermeture avancée sont parfois recouverts entièrement par de la végétation et se seront pas identifiables comme le montre la photographie ci-dessous (Figure 6). Devant ce constat il est impossible de créer un indicateur assez fiable pour décrire cet élément. Pour parvenir à une bonne détection de cet élément une approche objet serait à envisager afin de prendre le chemin comme une entité complète et non comme un ensemble de pixels isolés. De cette manière on pourrait extraire les caractéristiques de la forme des objets ce qui permettrait d'obtenir un descripteur intéressant pour la classification, vu l'élongation particulière des cheminements permettant peut-être d'éviter ce problème de « contamination » radiométrique par la végétation.



Figure 6 : Un chemin de plage étroit partiellement refermé par la végétation

Le trait de cote

La détection du trait de côté est également possible avec une méthode de classification. Cependant les résultats sont mitigés car la zone de sable mouillé ainsi que les effets du jet de rive (battement des vagues sur la plage) entraînent des confusions entre les classes, n'offrant pas une limite nette et constante. De plus la classification est un processus qui est difficilement paramétrable, ainsi pour détecter une limite aussi fine il est plus facile d'utiliser un seuillage sur la bande NDVI (voir rapport de Phase 1). Cette technique permet de faire varier facilement le seuil qui va déterminer la limite entre le sable et l'eau. Toutefois la limite obtenue par classification reste utilisable et permet une différenciation correcte de l'eau et du sable pour un rendu cartographique global du site en permettant d'obtenir un polygone qui correspond à la surface de la plage.

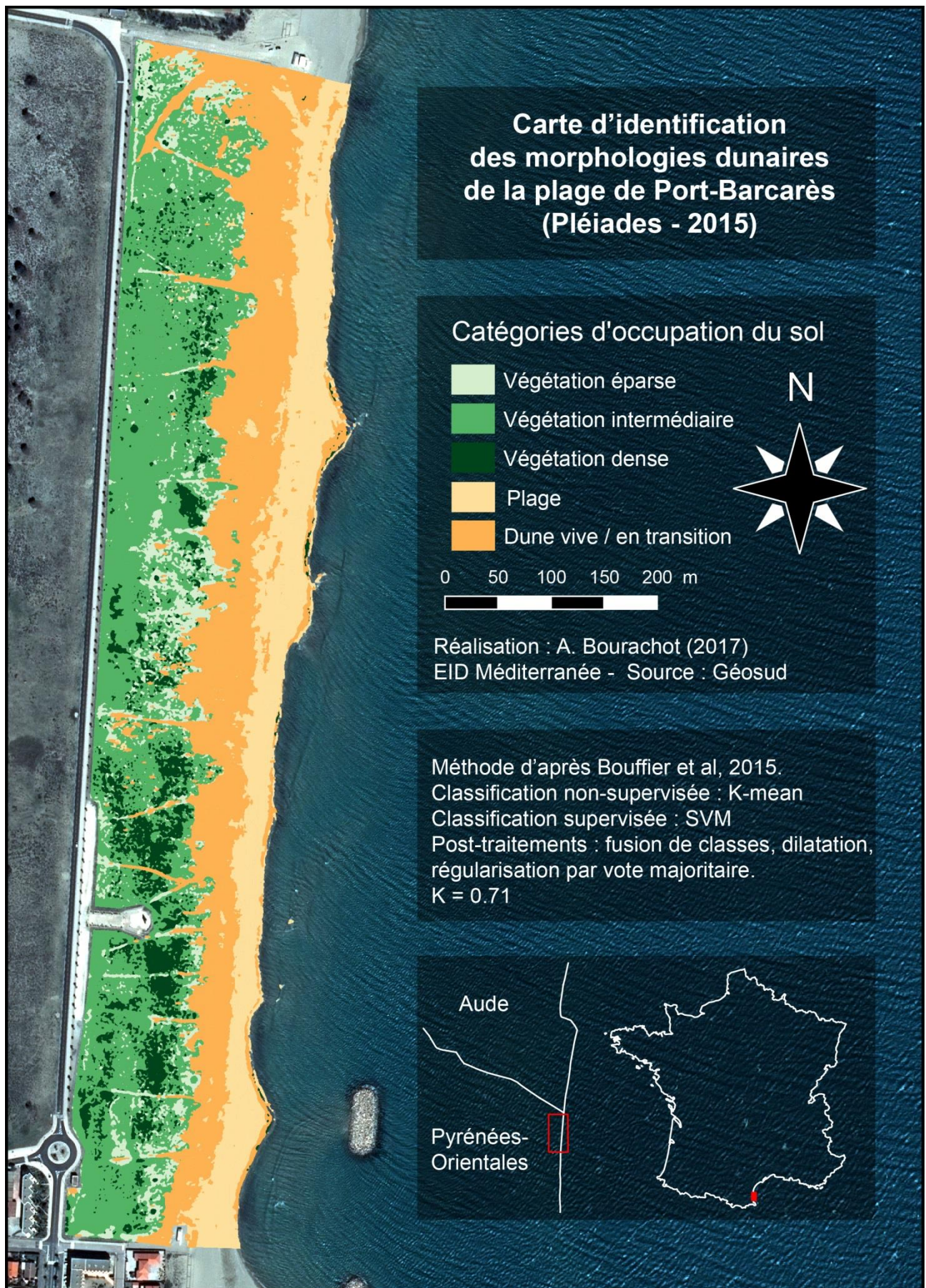


Figure 7 : Cartographie d'occupation du sol de la plage du Barcarès en 2015, avec sous-classes fusionnées et post-traitements morphologiques.

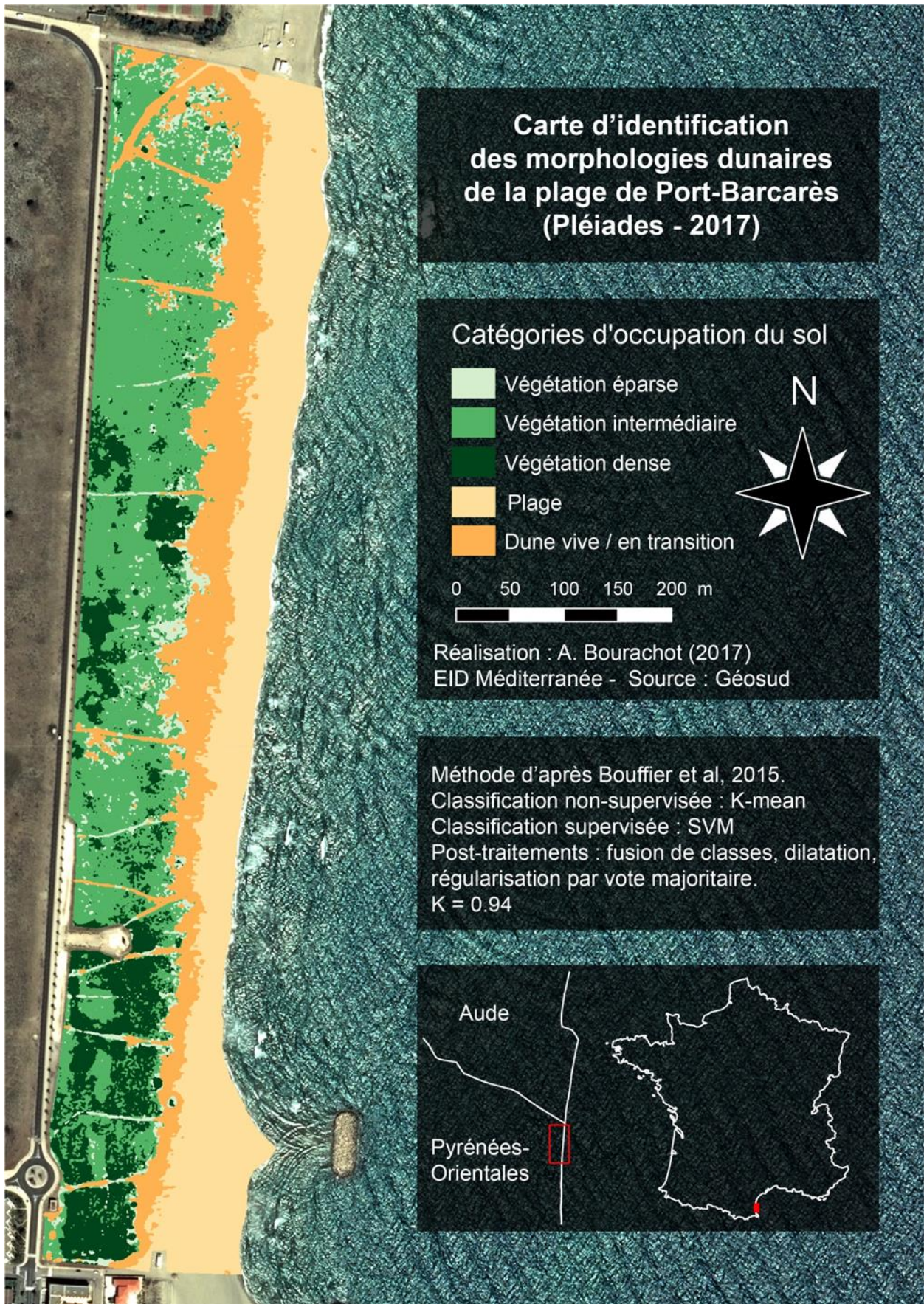


Figure 8 : Cartographie d'occupation du sol de la plage du Barcarès en avril 2017, avec sous-classes fusionnées et post-traitements morphologiques.

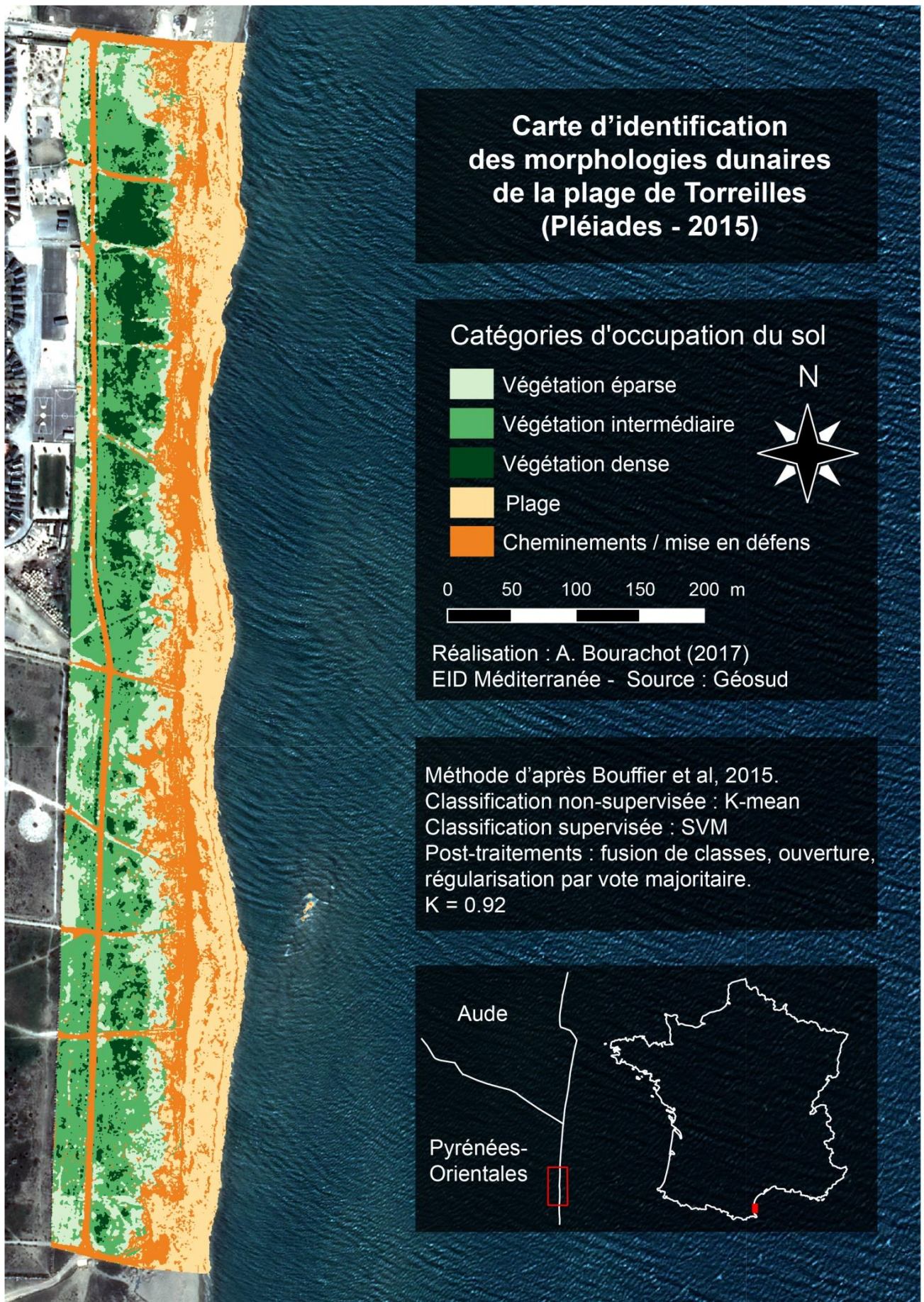


Figure 9 : Cartographie d'occupation du sol de la plage du Torreilles en 2015, avec sous-classes fusionnées et post-traitements morphologiques.

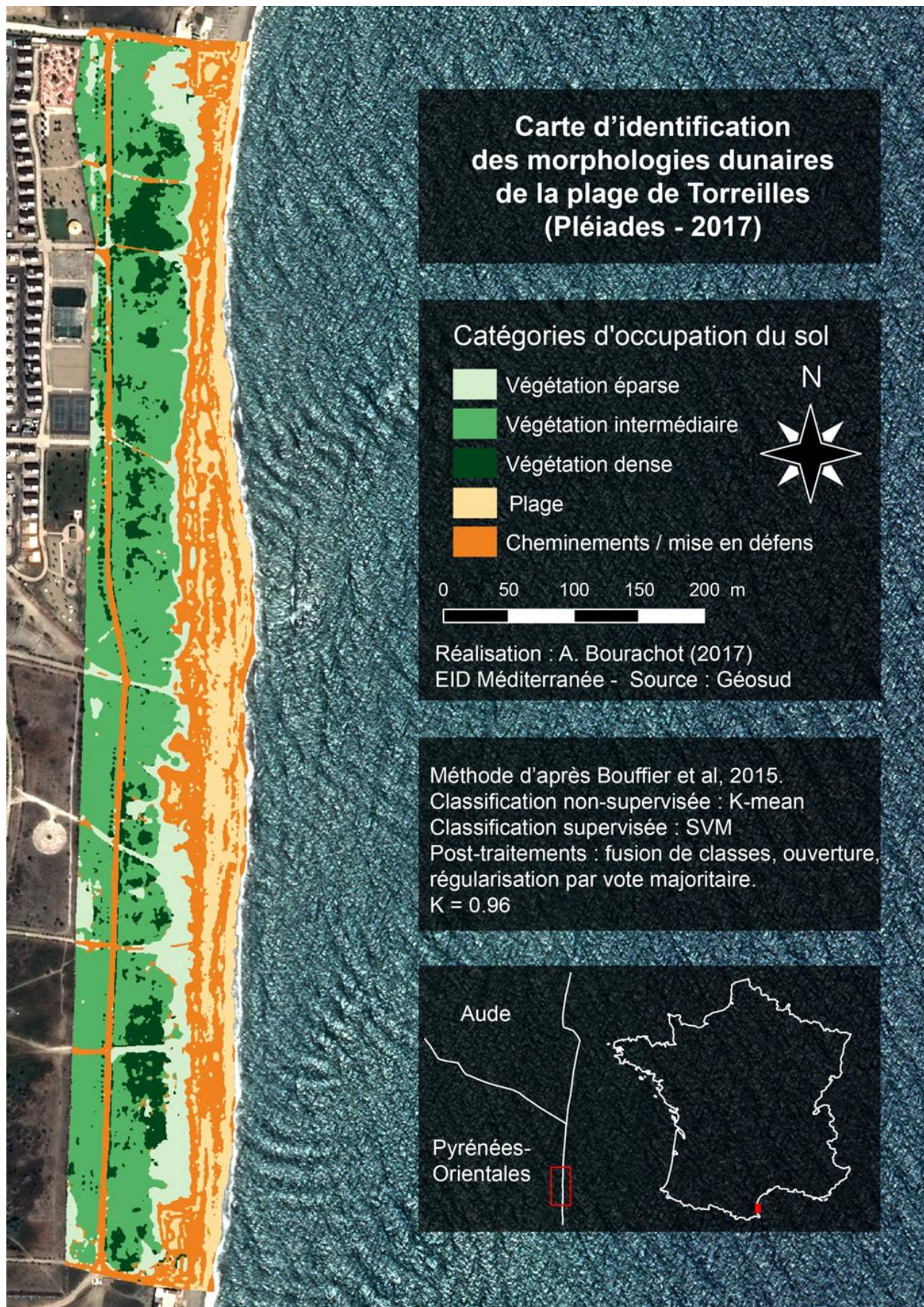


Figure 10 : Cartographie d'occupation du sol de la plage de Torreilles en 2017, avec sous-classes fusionnées et post-traitements morphologiques

7.2 La détection de changements avec une approche multi-dates

Pour obtenir des résultats précis sur l'évolution diachronique, il est nécessaire que les images soient comparables entre elles. Hors nous disposons ici de données de calibrées niveau *Top-Of-Atmosphere*, c'est-à-dire non corrigées des effets de l'atmosphère (niveau *Top-Of-Canopy*). L'atmosphère ayant une composition chimique qui va varier entre les dates étudiées, des différences de valeurs pourraient s'observer sur un même pixel qui représenterait un même objet sur le terrain qui n'aurait pas évolué d'une image à l'autre. On ne pourrait pas, par exemple, passer par un simple seuillage de NDVI car les valeurs pourraient être surestimées sur une image par rapport à l'autre donnant l'impression d'une expansion de la végétation.

L'avantage de passer par la classification est que l'on s'affranchit des paramètres atmosphériques. La classification est un processus au cours duquel un modèle d'entraînement propre à l'image est généré. Ce modèle sert à fixer des limites radiométriques entre les classes afin de les séparer, elles sont donc propres à l'image car elles dépendent des données d'entraînement fournies par l'utilisateur. Classifier agit alors comme un filtre qui va normaliser les images entre elles lors du processus de classification.

Pour illustrer cette possibilité nous allons prendre en exemple le calcul de surface de plage

Tableau 2). Bien qu'une démarche similaire puisse être appliquée afin d'étudier la végétation. Ce tableau montre qu'il est possible d'estimer un pourcentage de changement en termes de surface de plage. Cependant il faut prêter attention à plusieurs biais liés à ce calcul tel que la répartition des sous-classes sol nu et de végétation entre 2 dates. Afin de compenser ce biais on peut effectuer le calcul en se basant sur un masque qui prend la totalité du trait de côte mais qui s'arrête vers le milieu de la plage, pour que la partie dunaire n'influence pas le résultat (Figure 11). Ce même masque étant utilisé pour les deux dates, la différence de trait de côte entre les deux n'indiquera donc que la surface gagnée ou perdue par la plage durant cette période.

Tableau 2. Evolution des surfaces des plages de Barcarès et Torreilles entre 2015 et 2017. Le code couleur indique une information fiable (vert), moyennement fiable (orange), pas fiable (rouge) et informative (italique).

		Surface en 2015 en m ²	Surface en 2017 en m ²	Evolution de la largeur de plage en m ²	Evolution théorique de la largeur de plage en %
Le Barcarès	Masque	56 797	50 449	-6348	- 11,2
	Plage et dune	93 352	83 375	-9977	-10,6
Torreilles	Masque	37293	32385	-4908	-13.2
	Plage et dune	55881	42381	-13500	-24.1



Figure 11 : Exemple d'utilisation du masque de trait de côte pour éviter le biais de classification sur la dune.

8. APPORTS METHODOLOGIQUE POUR LA FINALISATION DE LA CHAINE DE TRAITEMENT

8.1 Les indices

Le travail réalisé initialement souligne l'importance de l'utilisation des indices spectraux et en l'occurrence celui du NDVI comme un descripteur pertinent des différentes classes de végétation, mais également celles pour l'eau et le sable. L'ajout ou non d'indice radiométrique ou d'indice de texture peut alors grandement modifier le résultat de la classification, notamment pour la végétation. Les essais effectués avec et sans le NDVI en tant que néocanal démontrent clairement une amélioration visuelle et statistique des résultats, notamment au niveau de la limite avec le rivage, où le sable humide semble souvent se confondre avec la végétation. Cet apport du NDVI a notamment été utile lors de la Phase 1 du projet dans la détection du trait de côte.

Avec les essais effectués pour la détection des bois flottés lors de la Phase 4 il est apparu que les indices de textures ne peuvent pas s'appliquer à la détection de la végétation dunaire car ils n'apportent pas d'information suffisamment discriminante pour caractériser les classes mais auraient plutôt tendance à bruiser l'information. Ces indices peuvent apporter une information intéressante lorsqu'il s'agit de structures qui se répètent à une certaine fréquence spatiale. Ils ne seront alors pas inclus dans la chaîne de traitement finale.

8.2 Les méthodes de classification

L'approche de classification supervisée implique les paramétrages à effectuer lors de l'entraînement. L'utilisation de l'algorithme SVM, et les classes ainsi générés dépendent fortement de la zone d'étude et des divers paramétrages complexes. Pour éviter cette variabilité liée à l'utilisation de cet algorithme, nous avons choisi de recommander plutôt l'utilisation de *Random Forest*, qui nécessite des paramétrages plus simples, qui peuvent rester fixes. En plus de simplifier la classification, la donnée produite sera également plus homogène.

Effectuer une pré-classification non-supervisée *Kmeans* des données d'entraînement offre une possibilité intéressante, en l'absence de données terrain, pour permettre d'obtenir plus de classes de manière automatique que l'on regroupe après selon la connaissance experte en fin de processus. Cependant le nombre de sous classes reste très important à déterminer car il peut varier beaucoup en fonction de la zone d'étude et de l'image. En moyenne pour classer la mer, 2 sous-classes suffisent, tandis que la végétation et le sol nu peuvent être catégorisés correctement

avec 4 à 6 sous-classes au minimum. Cette méthode d'échantillonnage rend donc la méthode difficilement automatisable pour être reproductible d'une zone à l'autre, et d'une date à l'autre. D'autre part, même en automatisant ce qui est possible, elle nécessite un certain temps pour sa mise en œuvre car elle comprend beaucoup d'étapes nécessitant de faire des choix arbitraires, notamment en fusionnant les classes.

Ainsi il est fortement recommandé de constituer une base d'apprentissage solide en allant soit réaliser des vérités terrain, soit en réalisant une photo-interprétation dans le but de fournir à l'algorithme des classes qui correspondent à la réalité radiométrique des classes que l'on souhaite obtenir. En procédant ainsi il n'est plus nécessaire forcément de partitionner le jeu d'entraînement avec la méthode des *Kmeans* et ainsi on s'affranchit de la partie traitement la plus complexe.

8.3 La définition de la nomenclature

Les particularités de certains environnements dunaires du littoral occitan comme ceux étudiés ici rendent l'automatisation complexe notamment en ce qui concerne la définition des classes à l'échelle régionale :

En effet, contrairement aux dunes que l'on peut trouver sur la plage des Orpellières ou à la pointe de l'Espiguette, les côtes du Parc naturel marin du Golfe du Lion sont souvent caractérisées par des morphologies très dégradées, avec peu de relief et une végétation éparse. Il est ainsi difficile d'adapter la méthodologie à toutes les morphologies dunaires possibles. Pour faire face à ce problème il est conseillé de cartographier les espaces en questions de manière indépendante en prenant en compte les spécificités de chaque site. Une bonne définition de la nomenclature en amont permettrait alors de relever sur le terrain les échantillons qui serviront à l'entraînement de l'algorithme classifieur afin de mettre en avant les spécificités de chaque site. De cette manière, les données que l'on va rentrer dans l'algorithme de classification correspondront exactement aux classes que l'on a définies dans la nomenclature.

Pour cela il serait nécessaire, grâce à la connaissance terrain, d'adapter la nomenclature en fonction des secteurs. Si la classification n'apparaît pas satisfaisante il ne faut pas hésiter à relancer le processus en ajoutant/supprimant des classes ou bien en divisant certaines classes en plusieurs sous-classes à partir de la base de donnée d'entraînement pour obtenir un compromis entre les détails que l'on souhaite observer et la précision qu'il est possible d'atteindre.

Concernant les différents types de végétation, éléments les plus complexes à identifier en raison des signatures spectrales proches de plusieurs espèces dunaires de la région (ex. : chiendent et oyats), cette étude a montré qu'avec la résolution de Pléiades il n'est pas possible de cibler une précision au niveau des espèces. Les espèces devront alors être regroupées par classe de manière à bien représenter les éléments morphologiques déterminants d'un site d'étude. On pourra néanmoins facilement définir de manière générale 3 classes : végétation dense, intermédiaire et éparse ; ou bien séparer une végétation d'arrière-dune très sombre (dune grise) de la végétation très verte du devant de la dune.

Il est également nécessaire de garder des proportions relativement similaires entre chaque classe pour éviter que certaines classes soient sur/sous-représentées afin de ne pas biaiser la classification. En effet, plus une classe contient un grand nombre d'échantillons, plus cette classe sera représentée dans le résultat final. On pourrait par exemple éliminer la classe végétation dense sur un site ou cette classe serait sous-représentée.

Une grande partie des bons résultats est issue du soin que l'on va apporter à la base de données, pour cela il faudra qu'elle soit correctement numérisée. Il faudra dans la mesure du possible éviter au maximum de mélanger les informations radiométriques quitte à dessiner des objets légèrement plus petits que ceux qui sont visibles sur les images, afin d'éviter le bord des objets qui présentent une information radiométrique mixte (ex : sable + végétation). Il faut veiller également à créer un nombre d'échantillons assez important, afin de représenter toute la diversité qui la constitue.



Figure 12 : Cartographie de la végétation dunaire de la plage de Port-Barcarès en septembre 2017 obtenue en utilisant uniquement la chaîne de traitement et une base de données d'entraînement sans pré-classification non-supervisée *Kmeans*

9. CONCLUSION GENERALE

Grace à cette étude, il a été possible de montrer l'intérêt de l'utilisation de l'imagerie satellite dans un objectif de cartographie des milieux littoraux du Parc Naturel Marin du Golfe du Lion. La très haute résolution spatiale de Pléiades couplée à la boîte à outils de traitement d'images libre *OrfeoToolbox* a permis d'obtenir une cartographie des trois structures morphologiques des systèmes plages-dune, en séparant la végétation du sable nu de la plage et en révélant le trait de côte et la majorité des cheminements.

Grace aux faibles coûts de mise en pratique comparés aux levés terrain, il est maintenant envisageable de mener un suivi régulier de ces espaces avec une fréquence annuelle permettant de produire des cartographies servant à calculer des indicateurs de surfaces de recouvrements afin d'évaluer l'état des milieux dunaires dans le temps.

Elaborer une chaîne de traitement automatisée, répliquable sur différents sites présentant des morphologies dunaires différentes a constitué le deuxième enjeu de cette étude de faisabilité. L'adaptation au territoire est d'ailleurs un des enjeux majeurs de la télédétection depuis ces origines. Les méthodes développées résultent de la recherche d'un compromis entre robustesse et facilité de mise en œuvre pour un novice en télédétection afin que celle-ci soit utile dans un contexte opérationnel qui est aussi celui des actions du Parc marin.

Afin de mettre en œuvre une classification des milieux dunaires deux possibilités sont donc envisagées :

S'il n'est pas possible d'obtenir des données terrain, il est conseillé d'utiliser la démarche initiale développée pendant le stage d'A. Bourachot (ANNEXE – Méthodologie de classification basée sur la méthode du CEREMA (extrait du rapport de stage)), utilisant la pré-classification *Kmeans* pour permettre de générer une base d'apprentissage plus diversifiée que ce qu'il est possible d'obtenir par simple photo-interprétation. Puis de réaliser une classification avec la chaîne de traitement finalisée utilisant *Random Forest* sur le NDVI et les bandes initiales de l'image (voir la méthode Annexe 3 et 4).

Dans un deuxième cas, lorsqu'on dispose de données terrain fiables, il est conseillé d'utiliser la chaîne de traitement finalisée directement. Le plus gros du travail à effectuer restant la création d'une base de données permettant de fournir des échantillons les plus qualitatifs possible à l'algorithme classifieur (dans ce cas, il faut entourer par photo-interprétation les espaces que nous savons être de la dune vive, de la dune éparse, de la dune grise, les espaces de sable nu, ainsi que l'eau ; voir la méthode Annexe 3 et 4).

Les résultats dépendront aussi bien, si ce n'est plus, de la bonne définition des classes et des données d'entraînement que de la méthode en elle-même. Par définition un algorithme n'est qu'« une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre une classe

de problèmes »⁶. Ici, les règles qu'utilisera l'algorithme seront fixées par le jeu de données d'entraînement que va construire l'utilisateur. Réaliser automatiquement un tel travail en amont reste trop complexe aujourd'hui pour un ordinateur classique. Pour construire une telle base de données, une connaissance de la zone géographique est alors essentielle afin d'identifier les spécificités des classes que l'on souhaite identifier afin d'exploiter au mieux l'information que fournissent les images satellites.

10. PERSPECTIVES

La résolution de Pléiades ne permet pas d'identifier directement les espèces végétales dunaires. Cependant à l'avenir d'autres sources de données satellites vont pouvoir également être utilisées, on pourrait penser notamment au programme Pléiades NEO qui va prendre la suite de Pléiades et dont le lancement se déroulera entre 2020 et 2022. Cette constellation de 4 satellites offrira des images à 30 cm avec 6 longueurs d'onde allant du bleu au proche infra-rouge, aura un temps de revisite de 1 jour ainsi qu'une fauchée de 14 km. De telles données pourront apporter ainsi plus de précision sur les classes, cependant cette augmentation de résolution ne permettra peut-être pas encore d'identifier les espèces. Pour cela l'imagerie par drone infra-rouge est plus adaptée à l'échelle spatiale étudiée en raison d'une résolution spatiale inédite de l'ordre du centimètre, voire du millimètre. Ce type de capteur oblige par contre à se restreindre à une étude sur des surfaces bien moindres (quelques km²).

⁶ wikipedia.org

REFERENCES BIBLIOGRAPHIQUES

Bannari, A., He, D.-C., Morin, D. et Anys, H. (1998) Analyse de l'apport de deux indices de végétation à la classification dans les milieux hétérogènes. *Canadian Journal of Remote Sensing*, vol. 24, n°3, p. 233-239.

Bannari, A., Huete, A. R., Morin, D. et Zagolski, F. (1996) Effets de la couleur et de la brillance du sol sur les indices de végétation. *International Journal of Remote Sensing*, vol. 17, n°10, p. 1885-1906.

Bannari, A., Morin, D., Bonn, F. et Huete, A. R. (1995) A review of vegetation indices. *Remote Sensing Reviews*, vol. 13, n°1-2, p. 95-120.

Baret, F., Jacquemoud, S. et Hanocq, J. F. (1993) The soil line concept in remote sensing. *Remote Sensing Reviews*, vol. 7, n°1, p. 65-82.

Bargain, A. (2012) Etude de la structure et de la dynamique des herbiers de *Zostera noltii* par télédétection multi et hyperspectrale. Thèse de doctorat, Université de Nantes.

Bausch, W. C. (1993) Soil background effects on reflectance-based crop coefficients for corn. *Remote Sensing of Environment*, vol. 46, n°2, p. 213-222.

Bonn, F. et Rochon, G. (1992) Précis de télédétection. Volume 1 : Principes et méthodes (Vol. 1). Presses de l'université du Québec.

Bouffier, J., Goarant, T., Hebrard, D., Mayis, A. et Rucquoi, S. (2015) Mise en place d'une couche d'occupation des sols à partir d'images satellites (p. 105). Cerema - Direction territoriale Sud-Ouest.

Cihlar, J. (2000) Land cover mapping of large areas from satellites: status and research priorities. *International journal of remote sensing*, vol. 21, n°6-7, p. 1093-1114.

Collet, C. (2001) Précis de Télédétection. Volume 3 : Traitements Numériques d'Images de Télédétection. PUQ.

Elvidge, C. D. et Chen, Z. (1995) Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote sensing of environment*, vol. 54, n°1, p. 38-48.

Elvidge, C. D. et Lyon, R. J. (1985) Influence of rock-soil spectral variation on the assessment of green biomass. *Remote Sensing of Environment*, vol. 17, n°3, p. 265-279.

Escadafal, R. et Huete, A. (1991) Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. *Comptes Rendus de l'Académie des Sciences.Série 2 : Mécanique...*, vol. 312, p. 1385-1391.

Lafon, V., Dehouck, A., Robinet, A., Marieu, V., Curti, C. et Durou, N. (2014) Détection d'indicateurs géomorphologiques littoraux par télédétection à très haute résolution spatiale, p. 601-610. Editions Paralia.

Lafon, V., Marieu, V., Butel, R., Dehouck, A., Froidefond, J.-M. et Trut, G. (2008) Cartographie des faciès bio-sédimentaires du bassin d'Arcachon à partir de l'imagerie FORMOSAT-2, p. 563-572. Editions Paralia.

Larroque, B., Favennec, J., Andrieu, F. et Rouveyrol, P. (2016) Guide de la flore du littoral sableux méditerranéen (Sud Ouest.). Sud Ouest.

Mallet, C., Lafon, V. et Desprats, J. F. (2007) Cartographie des faciès géomorphologiques du littoral aquitain à partir de données FORMOSAT-2 (No. RP-56101-FR) (p. 70). BRGM.
Manakos, I. et Braun, M. (2014) Land use and land cover mapping in Europe: Practices & trends (Springer.).

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H. et Sorooshian, S. (1994) A modified soil adjusted vegetation index. Remote sensing of environment, vol. 48, n°2, p. 119-126.

Rougier, S. et Puissant, A. (2014) Improvements of urban vegetation segmentation and classification using multi-temporal pleiades images. 5th International Conference on Geographic Object-Based Image Analysis, vol. 6.

Rougier, S., Puissant, A., Stumpf, A. et Lachiche, N. (2016) Comparison of sampling strategies for object-based classification of urban vegetation from Very High Resolution satellite images. International Journal of Applied Earth Observation and Geoinformation, vol. 51, p. 60-73.

Rouse Jr, J., Haas, R. H., Schell, J. A. et Deering, D. W. (1974) Monitoring vegetation systems in the Great Plains with ERTS.

Sublime, J., Troya-Galvis, A. et Puissant, A. (2017) Multi-Scale Analysis of Very High Resolution Satellite Images Using Unsupervised Techniques. Remote Sensing, vol. 9, n°6, p. 495.

ANNEXE – Méthodologie de classification basée sur la méthode du CEREMA (extrait du rapport de stage)

Principe général

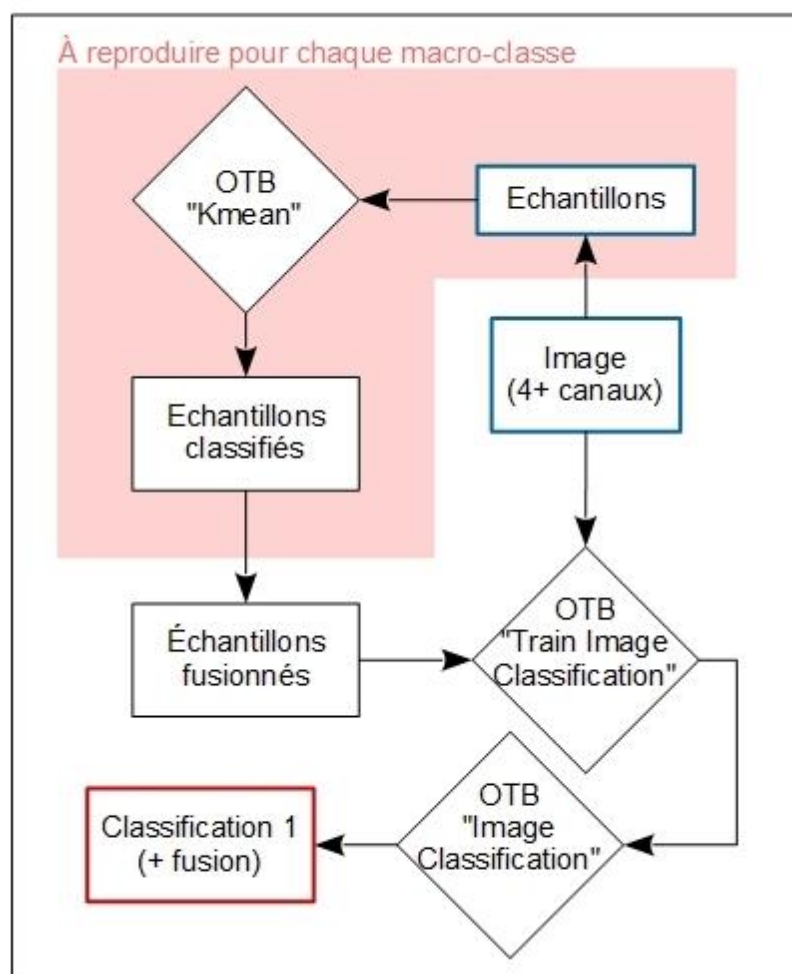


Figure 13. Schéma résumé de la méthodologie du CEREMA adaptée à notre problématique. La partie rouge désigne l'exécution par lot du modèle de traitement Qgis afin de traiter l'ensemble des échantillons (un par macro-classe telle que la végétation, le sol nu ou l'eau) en une seule fois.

La méthodologie utilisée par le CEREMA a pour but d'établir une cartographie de l'occupation du sol du département de la Drôme en conservant une résolution spatiale convenable mais en réduisant au maximum les erreurs causées par les confusions liées à la classification. Il s'agit en quelques sortes d'imiter le mode de fonctionnement de l'approche orientée objet sans pour autant passer par la segmentation de toute l'image. L'approche orientée objet a pour avantage d'utiliser les caractéristiques de spectre, de

texture et de forme pour distinguer les objets d'intérêt. La méthode que nous appliquons perd l'utilisation du critère de forme, mais gagne considérablement en vitesse de calcul. Pour imiter le fonctionnement de l'approche orientée objet, sont combinées une classification non supervisée pour distinguer les échantillons homogènes et les segmenter puis une classification supervisée afin de classer toute l'image (Figure 14). La majeure partie de la méthodologie concerne donc le prétraitement des échantillons d'entraînement et de validation, avec un fichier par macro-classe. Par conséquent, cette partie est exécutée par lot sous Qgis. Il est possible de résumer rapidement la méthodologie par ces quelques étapes :

1. L'interpréteur détermine les classes « de base » dites « macro-classes » qu'il veut distinguer (Végétation, sol nu et eau par exemple).
2. Tracé manuel des sets d'entraînements sous Qgis, un seul fichier par macro-classe.
3. Exécution du modèle de traitement, qui effectuera une classification *k-mean* sur chaque échantillon à partir de l'image cible (Figure 15). Il va diviser chaque macro-classe en plusieurs « sous-classes ».
4. Fusion de tous les sets, que l'on divise en un set d'entraînement et un set de validation.
5. Lancement de la classification supervisée.
6. Fusion éventuelle des sous-classes.

Lorsque l'on ne dispose pas de base de données pour lancer une classification supervisée, il faut s'en construire une manuellement via photo-interprétation en créant des échantillons d'entraînement et de validation sous forme de polygones sur le logiciel Qgis (**Erreur ! Source du renvoi introuvable.**).

L'étape suivante serait normalement d'utiliser ces échantillons directement pour lancer la classification, mais la méthode employée nécessite d'ajouter une étape supplémentaire pour faciliter le travail de l'algorithme : chaque échantillon d'entraînement et de validation, qui représente un petit groupe de pixels homogènes dans leur nature mais hétérogènes dans leur réponse spectrale, sera soumis à une classification

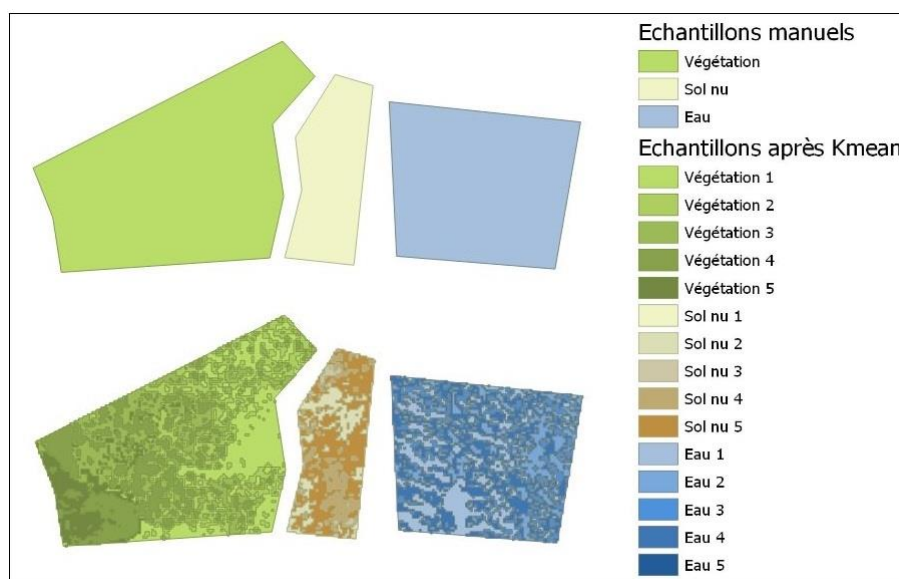


Figure 14. Exemple de set d'échantillons d'entraînement/validation avant et après l'étape de classification non-supervisée du K-mean. Paramétrage de 5 sous-classes par macro-classe, 10 000 pixels, 0.95 taux de convergence, 5000 itérations.

non-supervisée de type *K-mean* (MacQueen, 1967). L'algorithme du *K-mean* va, en fonction du nombre de

sous-classes* demandées par l'interpréteur, diviser chaque échantillon en segments de taille inférieure appartenant chacun à une sous-classe radiométrique précise.

Ces « sous-classes » n'ont aucune signification thématique apportée par l'interpréteur, il s'agit uniquement de classes « radiométriques » dans le sens où l'algorithme a simplement découpé à l'intérieur de chaque échantillon des petites zones dont il a rassemblé les plus semblables, formant ainsi de petits groupes : les « sous-classes » (Figure 15).

Cela permet en bout de traitement de multiplier le nombre de classe automatiquement, sans avoir à créer 20 sets d'échantillons différents pour être sûr de distinguer les zones les plus difficiles à interpréter, tout en distinguant efficacement les changements texturaux de la zone étudiée. Une fois les échantillons découpés ainsi, de la même manière que le ferait une segmentation ou presque, on lance une classification supervisée dont le modèle choisi doit varier en fonction de la zone d'étude et des échantillons récupérés. Les deux méthodes principales sont *SVM* (Cortes et Vapnik, 1995) et *Random Forest* (Breiman, 2001; Ho, 1995).

La méthodologie menée par le CEREMA se termine en fusionnant sans distinction l'ensemble des sous-classes générées pour chaque macro-classe, puisque l'objectif n'est pas d'avoir une grande précision. Dans notre cas, des post-traitements sont nécessaires pour rendre la classification utilisable. Le principal traitement à effectuer est une fusion des sous-classes de façon experte : on veut conserver une certaine richesse dans la nomenclature en distinguant différents cortèges de végétation ou de type de dunes, sans pour autant devoir cartographier un trop grand nombre de sous-classes. On tâchera donc de les fusionner en fonction de critères déterminés par l'interpréteur, en se servant notamment d'opérateurs morphologiques ou de vote majoritaire pour faire ressortir des éléments particuliers (tels que les cheminements).

L'ensemble du processus nécessite davantage de manipulations « expertes » que l'objectif initial, mais il permet d'avoir la main à plusieurs étapes du traitement sans pour autant mobiliser l'interpréteur en permanence. Le biais lié à l'utilisateur peut néanmoins devenir un avantage lorsque les résultats de la classification doivent se plier à son interprétation, particulièrement dans notre cas où le littoral souvent dégradé permet difficilement de distinguer à l'œil nu les différentes morphologies dunaires. Un algorithme entièrement automatisé ne saurait s'adapter aux différentes formes et textures que peuvent former le système dune-plage et la nécessité de créer des échantillons permet ainsi de faire coïncider les résultats attendus avec l'interprétation initiale du spécialiste.

Démarche détaillée

Des prétraitements sont nécessaires (Figure 16 partie a) pour lancer la classification entière. Dans un premier temps, compte tenu du fait que les scènes Pléiades sont divisées en dalles d'1 km de côté, il faut au préalable fusionner les dalles si la zone d'étude s'étend sur plusieurs d'entre elles à l'aide de Qgis. Dans le même temps, si l'on cherche à travailler sur une zone très précise tout en évitant les perturbations de la classification liées à des surfaces de nature indésirables (surfaces artificielles, ombres), en ne sélectionnant que les dunes, la plage, et l'eau par exemple, il est conseillé de réaliser un masque au format vecteur, puis de le transposer en raster. De cette façon, la classification finale ne se fera que sur la zone désirée, ce qui limite le nombre de macro-classes, et par la même occasion les erreurs possibles de la classification et son temps de traitement.

Dans un second temps, toujours dans le cadre des prétraitements, il faut garder à l'esprit que la classification non supervisée *k-means* se sert essentiellement des valeurs radiométriques des pixels de

l'image, pour les bandes du rouge, du vert, du bleu et du proche infrarouge. Un des moyens disponibles pour tenter d'influencer les résultats, autre que de modifier les paramètres, est de rajouter des néocanaux à l'image initiale. Il s'agit simplement de couches d'informations complémentaires qui s'ajoutent à l'image en tant que bande, et qui permettent potentiellement de rendre la discrimination des pixels plus fiable, puisque l'information est plus complète. Il est par exemple possible de rajouter un indice radiométrique, comme le NDVI, ou des indices de texture, comme la corrélation d'Haralick (Haralick *et al.*, 1973). Il est déconseillé d'en ajouter un trop grand nombre au risque de bruite l'image plutôt que d'améliorer les résultats. Pour les différents tests, il faut s'adapter à la zone d'étude et aux paramètres utilisés, l'idéal est donc de se contenter d'un indice radiométrique (NDVI ou plus complet si disponible, comme les déclinaisons du SAVI (Huete, 1988)) et d'un ou deux indices de texture, comme la corrélation d'Haralick ou l'IC2 qui sont conseillés par (Bouffier *et al.*, 2015)). Néanmoins, d'autres indices de texture connus peuvent tout à fait convenir, tels que l'extraction de gradient qui possède son outil propre sous l'OTB, et la variance. Cette dernière est réalisable via le même traitement que l'ensemble des indices d'Haralick, tous sont extraits lorsque l'outil est lancé, donc il est possible de sélectionner ceux qui nous intéressent uniquement.

Une fois que l'image est découpée convenablement, le masque crée et les néocanaux concaténés, il faut créer les échantillons d'entraînement et d'apprentissage. En règle générale, il faut suivre un certain nombre de règles pour créer correctement les polygones qui serviront d'échantillons, que l'on peut retrouver en application dans le même domaine de recherche (Bainbridge et Reichelt, 1988; Chauvaud *et al.*, 2001; Gambarova *et al.*, 2010). Difficile de se servir dans ce contexte de l'unité minimale de collecte, puisque la résolution spatiale sous Pléiades est de 50 cm et que l'on cherche à obtenir la meilleure précision possible, notamment pour l'extraction des cheminements, qui traversent les dunes et qui peuvent être particulièrement étroits. Il est conseillé cependant de créer des polygones de taille suffisamment importante pour être pris en compte par l'algorithme, car des échantillons trop petits bloquent le traitement de l'OTB (on peut néanmoins fixer un intervalle compris entre 10 m² et 2000 m²). Enfin, toujours dans l'esprit de trouver un compromis entre rapidité et robustesse, l'interpréteur peut se contenter de ne créer qu'un seul fichier vecteur par macro-classe, puisque la distinction entre les pixels d'entraînement et de validation se fera en fin de traitement, juste avant la classification. En revanche, il est conseillé de respecter la proportionnalité des macro-classes, en créant un nombre de polygones qui correspond à l'importance de la répartition de l'objet représenté sur l'image, si possible sur l'ensemble de la zone pour correctement traduire l'ensemble du spectre qui peut caractériser un type d'objet.

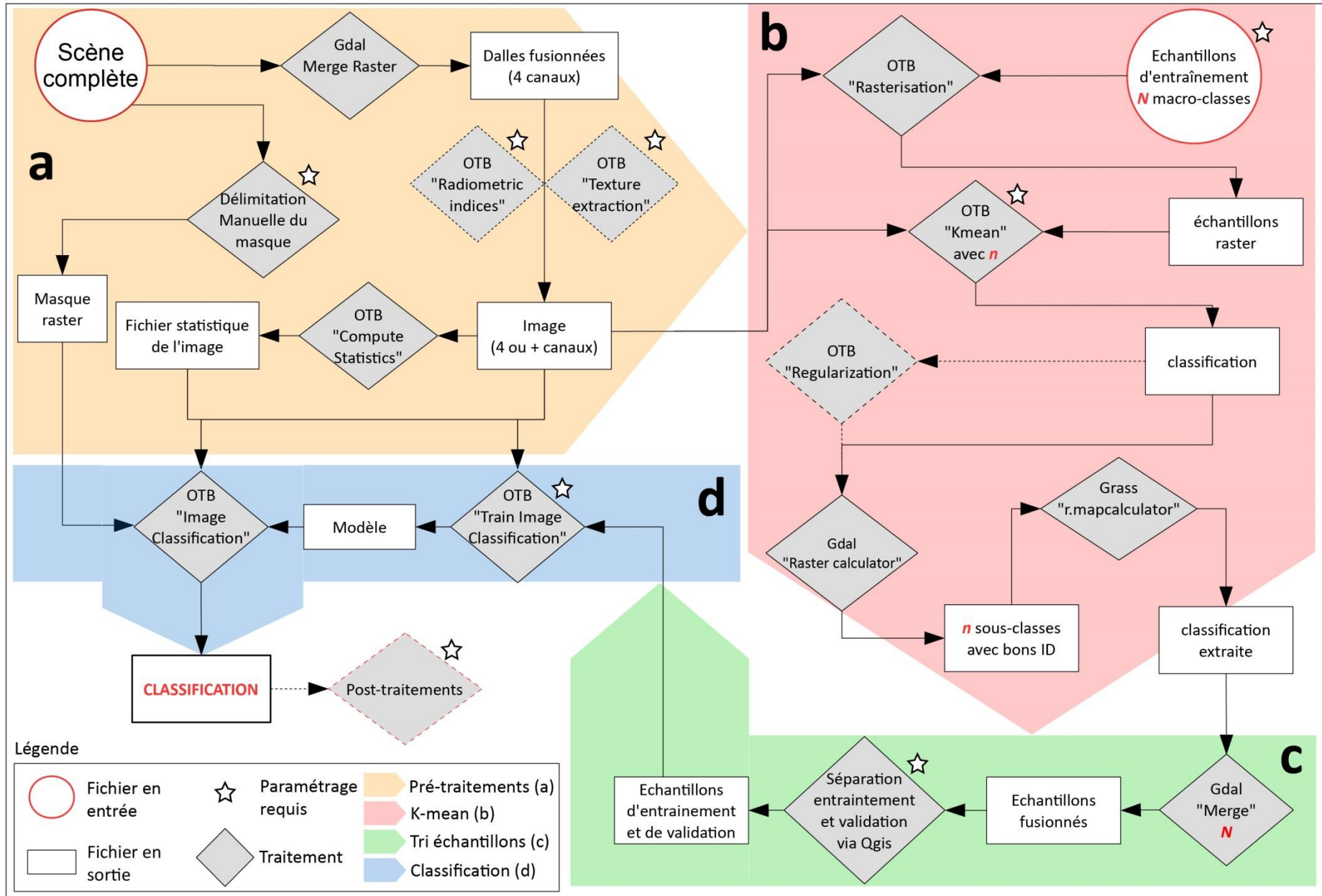


Figure 15. Représentation schématique complète de la méthodologie de classification supervisée.

L'objectif de la seconde partie (Figure 16 partie **b**), (le contenu de ce cadre est à exécuter par lot en fonction du nombre **N** de macro-classe) est d'utiliser un algorithme de classification non-supervisée pour découper tous les échantillons en segments plus petits, distribués dans des sous-classes qui sont fonctions des caractéristiques radiométriques des pixels. Par exemple, les échantillons de la classe « Végétation » sont divisés en segments rangés dans **n** sous-classes qui se distinguent par un identifiant unique, **n** étant paramétré par l'utilisateur lorsqu'il lance l'algorithme *k-mean*. Ces sous-classes n'ont pas de signification thématique, les identifiants servent juste à différencier correctement les sous-classes d'une même macro-classe entre elles dans la table attributaire.

Concrètement pour procéder, il est nécessaire de passer les sets d'échantillon au format raster, de façon à créer des masques. L'algorithme *k-mean* de l'OTB classe toujours l'ensemble de l'image qui lui est fournie, donc pour restreindre le calcul aux zones qui nous intéressent, il faut paramétrer un masque au format raster. Visuellement, il *semble* que l'algorithme classe l'ensemble de l'image car le fichier en sortie le présente comme tel, mais comme l'indique le manuel utilisateur de l'OTB, seules les zones sélectionnées par le masque indiqué en paramètres sont utilisées pour estimer les

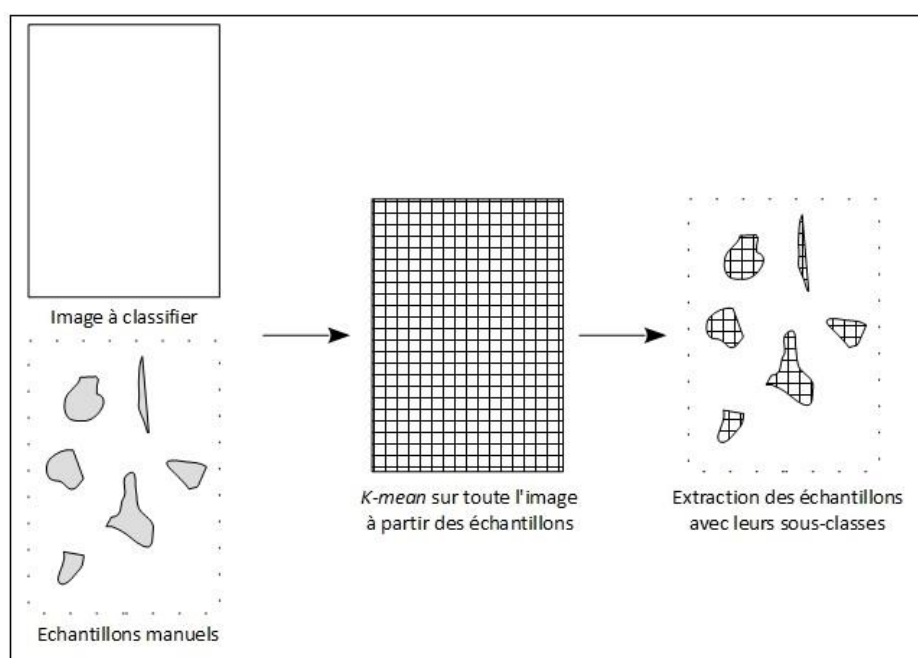


Figure 16. Schématisation du traitement qui consiste à segmenter partiellement l'image à l'aide d'échantillons d'entraînement/validation. Via l'utilisation du modèle Qgis, ces fichiers intermédiaires ne sont pas visibles.

modes statistiques (et sont donc correctement classifiées) (Figure 17).

Une fois la rasterisation terminée, il faut lancer l'algorithme du *k-mean*. On peut faire varier la taille du set d'entraînement, ce qui se voit nécessaire lorsque l'on souhaite un grand nombre de sous-classes et que la valeur par défaut est trop faible. La valeur suivante est celle permettant de paramétrer le nombre de sous-classes. Il est ensuite possible de paramétrer le nombre maximum

d'itérations que va utiliser l'algorithme avant d'arrêter le calcul sur une classification finale, et le taux de convergence qui désigne le pourcentage de pixel qui reste dans une classe entre chaque itération.

Le résultat obtenu après une classification non supervisée est souvent très bruité, avec de nombreux pixels isolés qui peuvent nuire à la classification globale. Une étape optionnelle consiste donc à lancer un traitement de régularisation par vote majoritaire, qui en fonction du rayon indiqué en paramètre va lisser le résultat de classification pour homogénéiser l'image et faciliter le travail de la classification supervisée.

Une fois ces prétraitements réalisés, on récupère le résultat de la classification non-supervisée tel que l'illustrent la Figure 15 et la Figure 17 (ce qui concrètement se traduit par autant de fichier raster que de macro-classes choisies). Pour rappel, l'ensemble de l'image est représenté comme étant classifié, or seules les zones de nos échantillons le sont, on tâchera donc de les extraire à l'aide d'un algorithme GRASS. Il s'agit de : *r.mapcalculator*, avec en entrée l'image classifiée issue de la classification (ou de la régularisation optionnelle), et les masques au format raster des échantillons qui ont déjà été utilisés pour le *k-mean*. Une fois les échantillons classifiés extraits, il suffit de les passer au format vecteur et d'adapter les couches pour la fusion et la classification supervisée.

Avant de procéder à la classification, quelques légers prétraitements manuels sont nécessaires. Il s'agit notamment de supprimer le polygone résiduel obtenu après vectorisation (le polygone qui entoure les échantillons quadrillés sur la dernière étape de la Figure 17), et de modifier les identifiants des sous-classes pour pouvoir toujours les distinguer après avoir fusionné les couches vecteur (Figure 18). Il suffit ensuite de fusionner les sets vectoriels entre eux.

L'étape suivante consiste à sélectionner les segments qui serviront à l'entraînement, et ceux qui serviront à la validation. Plusieurs méthodes sont disponibles (Figure 19).

Dans notre cas (Figure 16 partie c), il suffit de lancer l'algorithme Qgis « sélection aléatoire depuis des sous-ensembles » en choisissant un pourcentage (60% par exemple) puis d'enregistrer la sélection résultante dans un seul fichier qui sera notre set d'entraînement. Il suffit d'inverser la sélection et d'enregistrer à nouveau le résultat dans un second fichier pour obtenir le set de validation. Cela permet de choisir le ratio désiré entre les deux sets, tout en évitant le biais lié au choix de l'opérateur pour la sélection des échantillons.

L'algorithme de classification supervisée de l'OTB propose une fonction qui ne demande en entrée qu'un seul fichier vecteur et qui se charge de diviser en deux les échantillons selon un ratio donné. C'est une alternative qui a l'avantage d'être rapide, mais qui introduit un biais lié à la proximité des pixels potentiellement choisis par le logiciel. Les pixels sont choisis aléatoirement au sein des polygones du set d'échantillons, il risque donc de sélectionner des pixels d'entraînement et de validation trop proches, améliorant artificiellement l'indice de performance statistique utilisé (Kappa). L'autre possibilité encouragée par l'équipe OTB est de créer deux échantillons distincts, l'un

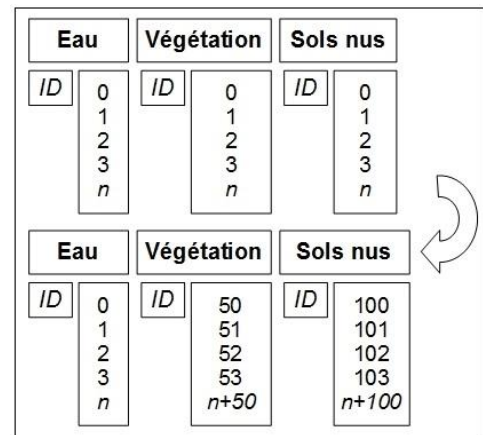


Figure 17. Exemple de modification des identifiants des sous-classes. Cela permet de fusionner les fichiers vecteurs des macro-classes sans fusionner aussi les sous-classes.

pour l'entraînement et l'autre pour la validation. L'avantage est de dissocier les pixels des deux sets mais induit un important biais opérateur.

L'alternative utilisée ici consiste donc en un simple compromis entre biais opérateur et proximité des pixels en faisant la distinction entraînement / validation après l'étape des *K-means* plutôt qu'avant. Ainsi, en sélectionnant au hasard des pixels parmi des segments distincts plutôt que sur l'ensemble du set, on limite l'influence de l'interpréteur tout en diminuant la probabilité qu'un échantillon d'entraînement se retrouve trop proche d'un échantillon de validation.

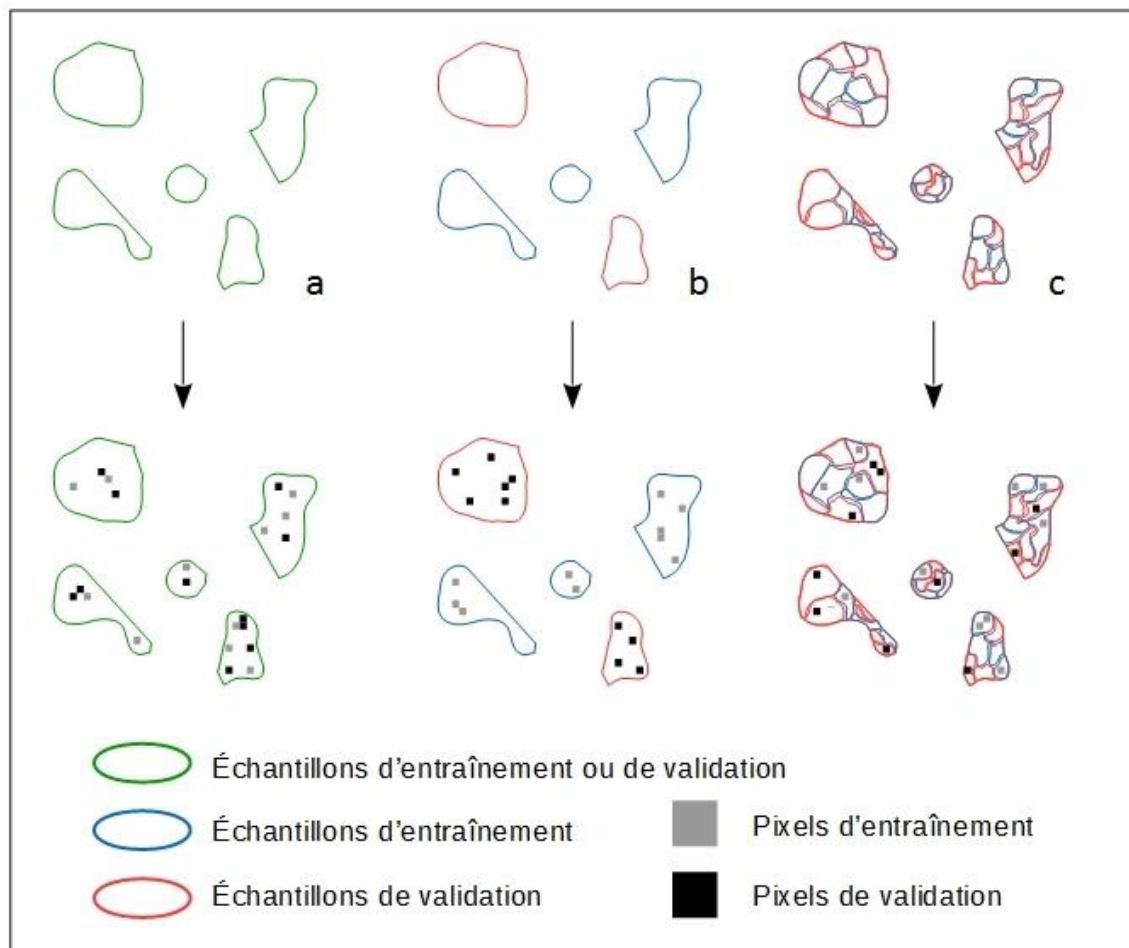


Figure 18. Alternatives de sélection d'échantillons disponibles. (a) méthode par défaut fournie par l'algorithme OTB. (b) alternative conseillée par l'équipe en charge du développement de l'OTB. (c) compromis choisi pour cette méthodologie.

La dernière étape obligatoire (Figure 16 d) et de lancer la classification supervisée à l'aide de l'OTB. Il est conseillé ici de se servir de l'application *standalone* plutôt que des traitements Qgis issus de l'OSGEO car ces derniers ne sont pas forcément à jour. Deux algorithmes particuliers sont intéressants dans notre cas : *Support Vector Machine (SVM)* et *Random Forests*. D'après Bouffier *et al.*, 2015, il est conseillé d'utiliser le SVM car celui-ci est plus long en traitement mais donne des résultats légèrement meilleurs et est plus simple à paramétrer. En effet, les seuls éléments à indiquer sont liés à la taille des échantillons pris par la classification et doivent être précisés quel que soit l'algorithme choisi. L'algorithme *Random Forest* est plus rapide et potentiellement plus malléable car possède davantage de paramètres, mais il induit des tests supplémentaires pour adapter et optimiser

les paramètres à l'objet d'étude et à la zone. Il est possible de fournir un fichier xml à calculer au préalable contenant les moyennes et la variance de chaque bande de l'image pour aider à la classification.

Une fois la classification effectuée, la majeure partie du travail est terminée et l'objectif est globalement atteint. On peut cependant vouloir une qualité supplémentaire en utilisant des opérateurs morphologiques, tels que la dilatation ou l'érosion (Figure 20), de façon à faire ressortir certains éléments de l'image tels que les cheminements, ou atténuer des sous-classes non-désirées et éliminer les pixels isolés. Cela se combine très bien avec une étape de fusion des sous-classes, consistant simplement par interprétation à fusionner les sous-classes qui semblent les plus proches pour pouvoir créer une nomenclature. Les manipulations morphologiques sont à adapter au cas par

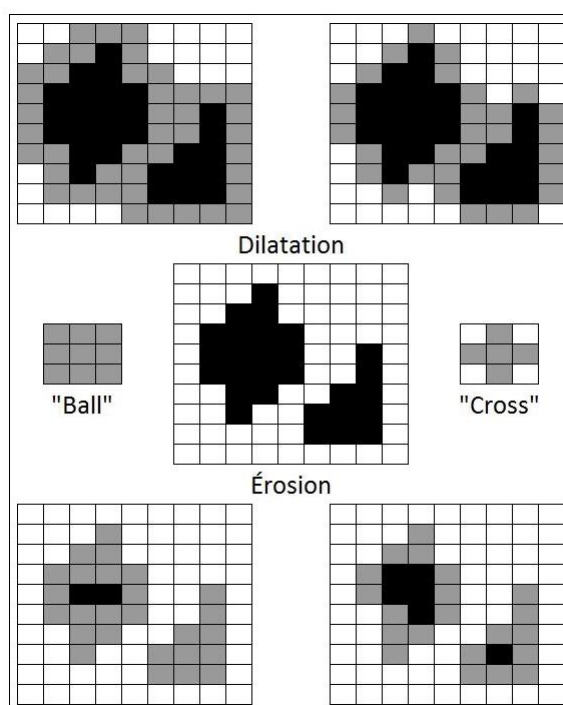


Figure 19. D'après Collet et al., 2001. Filtrages morphologiques élémentaires: Dilatation et érosion. La fenêtre de traitement est paramétrable sous OTB sous les noms de "Ball" et "Cross" et induisent des modifications différentes.

cas, et sont détaillées dans le chapitre sur les résultats.

Protocoles de test et variations des paramètres

La méthodologie a pour intérêt principal de compenser les limites de la classification supervisée en termes de combinaison de classe et détection de zones floues à l'aide de la classification non supervisée qui permet de générer autant de sous-classes que nécessaire. C'est un compromis qui induit un inconvénient notable : le choix des paramètres. En multipliant les étapes, l'opérateur multiplie nécessairement la quantité de paramètres à préciser pour optimiser la classification et potentiellement améliorer les résultats. Comme indiqué dans la méthodologie menée par le

CEREMA, il serait possible de tester toutes les combinaisons en lançant un programme qui simulerait des combinaisons aléatoires de paramètres et finalement trouver le set optimal, mais cela prendrait un temps considérable se comptant en semaines. Plutôt que de laisser la totalité des paramètres par défaut (les algorithmes de l'OTB fonctionnent correctement pour la plupart même lorsque les valeurs ne sont pas personnalisées), nous fixons plusieurs valeurs rapidement en fonction des besoins ou de différences flagrantes dans les résultats (le nombre de sous-classes par exemple, apparaît vite comme étant suffisant à 5, mais trop important à 20), puis nous faisons brièvement varier les autres paramètres majeurs en fonction notamment de la zone d'étude et des erreurs ponctuelles à éliminer. Sur la Figure 16, chaque étape de traitement dont les paramètres peuvent varier, ainsi que les paliers qui nécessitent une intervention manuelle (Introduisant le biais de l'interprète) sont indiqués par des étoiles.

Dans le cas d'un littoral très emprunté, avec de nombreux chemins de sable qui passent à travers les dunes mais qui sont difficiles à extraire à l'aide du NDVI ou de la classification simple, on peut ajouter des indices de texture dont les paramètres seront spécifiquement réglés pour tenter de faire ressortir au mieux les structures linéaires des chemins. Certaines valeurs sont laissées par défaut, telles que les X et Y de décalage de la fenêtre de traitement, d'autres sont fixées assez rapidement, tels que les rayons de la fenêtre en fonction de la taille des objets visés, et d'autres enfin sont paramétrés de manière empirique via tests successifs pour déterminer le résultat le plus satisfaisant. C'est le cas des valeurs minimum et maximum de l'image ou du nombre de boîte de l'histogramme (*Histogram number of bin*). Pour notre image de la zone du Barcarès en 2015 par exemple, la valeur maximum (*Image Maximum*) par défaut est réglée à 255 mais elle ne donne aucun résultat, il est donc nécessaire de faire varier progressivement cette valeur pour atteindre une visualisation satisfaisante : 1000 semblant a priori être un bon compromis sachant que l'indice de texture ne doit pas brouiller l'information sur le reste de la dune à seule fin de mieux faire ressortir les chemins.

Toujours pour lancer une classification cohérente par rapport à la zone étudiée, on peut par exemple se passer de NDVI lorsque l'image a un fort bruit lié à une importante réflectance du sol, avec très peu de végétation, comme pour la plage de Sainte-Marie. Le néocanal NDVI ne suffirait pas à distinguer correctement les maigres cortèges de végétation du reste de la dune, et risquerait de constituer au contraire un critère d'homogénéité pour l'algorithme.

Enfin, il convient également de régler les paramètres en fonction des résultats attendus et obtenus lors des tests, ou en correspondance avec le capteur utilisé. Pour le *K-mean* par exemple, il est nécessaire de paramétrer la taille des échantillons en nombre de pixel pour la classification. La valeur par défaut est 100, mais ne correspond pas à la taille des sets d'échantillon ou de l'image que nous fournissons. Chaque dalle Pléiades a des dimensions de 1 * 1 km, et une résolution spatiale de 0.5 m. Les polygones d'entraînement, malgré leur taille modeste liée à l'étroitesse de la dalle, recouvrent donc une quantité plus importante de pixel, neuf fois plus que pour une image SPOT de la même zone par exemple (Un pixel SPOT d'1.5 m de côté peut contenir 9 pixel Pléiades de 0.5 m de côté). Ainsi avec une valeur de 100 ou de 1000, le processus ne fonctionne pas et s'interrompt purement et simplement ou, lorsqu'il se termine, ne fournit pas pour certaines macro-classes autant de sous-classes que demandé dans les paramètres. Une valeur de 5000, voire de 10 000 est donc utilisée par sécurité pour se garantir une certaine marge. De la même manière le nombre de sous-classe dépend de la zone et du compromis entre précision et identification. Il est possible de monter jusqu'à 20 sous-classes de façon à distinguer au mieux les objets de même nature qui possèdent des

différences très peu marquées, comme une surface sableuse sur la plage régulièrement piétinée, retournée, voir humidifiée par la mer et une surface sableuse ayant très légèrement plus de relief car se situant au niveau de la dune vive et possédant quelques traces de végétation éparse. Néanmoins, l'objectif étant ici d'extraire les différentes morphologies dunaires, on ne veut pas nécessairement une trop grande diversité au sein d'objets de même nature, et l'on peut se contenter de 2 sous-classes pour la mer, 4 ou 5 sous-classes pour la végétation de façon à distinguer des cortèges plus ou moins denses, et 3 sous-classes de sol nu pour séparer la plage de la dune et du sable humide.

Deux autres paramètres importants sont à régler pour le *K-mean*, il s'agit du taux de convergence et du nombre d'itérations. Ce dernier régit le nombre de fois que l'algorithme va effectuer le calcul sur un mode avant d'en changer, et le taux de convergence indique la quantité de pixel qui changera de classe à chaque changement d'itération. Multiplier les itérations en laissant un taux de convergence bas peut être contreproductif, car la classification est sans cesse entièrement remodelée. En revanche, augmenter également le taux de convergence permet de garder une cohérence dans la classification tout en maintenant un important nombre d'itération sans pour autant le multiplier au risque d'accroître considérablement les temps de calcul.



EID Méditerranée

165, avenue Paul-Rimbaud
34184 Montpellier Cédex 4
Tél. 04 67 63 67 63



**Parc Naturel Marin
du Golfe du Lion**

2 impasse Charlemagne
66700 Argelès-sur-Mer
Tél : 04 68 68 40 20



LMUSCA

CEFREM – UPVD
52 avenue Paul Alduy
66860 Perpignan Cedex 9
Tél : 04 68 66 20 57



OBSCAT

Agence d'urbanisme catalane (AURCA)
19 Espace Méditerranée
66000 Perpignan
Tél : 04.68.87.75.52