

Table des matières

1. Introduction.....	7
1.1 La fièvre de la vallée du Rift.....	7
1.2 Le projet EDEN.....	9
1.3 Application de la télédétection et des SIG à l'épidémiologie.....	9
1.4 Problématique et objectifs du stage.....	10
2. Données.....	11
2.1 Image Quickbird.....	11
2.2 Données sérologiques.....	12
2.3 Données collectées sur le terrains.....	12
3. Matériel.....	14
3.1 Logiciel eCognition.....	14
3.1.1 Segmentation.....	15
3.1.2 Classification.....	16
3.2 Logiciel R.....	17
4. Méthode.....	18
4.1 La classification supervisée orientée objet.....	18
4.2 Méthodologie statistique.....	19
4.2.1 La régression logistique.....	19
4.2.2 Cas des donnée sérologiques.....	19
5. Cartographie de l'occupation du sol.....	19
5.1 Définitions des unités paysagères étudiées.....	19
Détails sur les unités paysagères.....	22
5.2 Choix des paramètres de segmentation.....	22
5.3 Analyse des objets générés par segmentation.....	25
5.4 Détection des Mares et du bâti.....	30
5.5 Carte d'occupation du sol.....	32
5.6 Carte détaillée de la végétation des mares.....	34
5.7 Occupation du sol à deux échelles de précision.....	35
5.8 Résultats.....	35
6. Extraction et analyse statistique des variables paysagères.....	36
6.1 Passage en classes de végétation à risque.....	36
6.2 Création des indices.....	37
6.3 Analyse statistique – Relation Incidence FVR / variables paysagères.....	39
6.4 Résultats.....	40
7. Discussion.....	41
Bibliographie.....	43
Annexes.....	44
Annexe 1 : Caractéristiques de Quickbird.....	44
Annexe 2 : Fiche « Parcelle homogène».....	45
Annexe 3 : Fiche « Transect ».....	46
Annexe 4 : Résultats de la variation du paramètre d'échelle.....	47
Annexe 5 : Texture d'Haralick.....	48
Annexe 6 : Modèle Erdas de l'indice de l'eau.....	49

Listes des figures :

Figure 1 : Répartition de la FVR dans le monde. Source : Center for Disease Control and Prevention, 2005	7
Figure 2 : Localisation de la zone d'étude. Source Arcgis data	8
Figure 3 : Représentation des couches dans le SIG. Source : www.mbeconetwork.org	10
Figure 4 : Digital Globe. [Région du Ferlo, Sénégal]. Quickbird [Image satellitaire] 04/08/2005 à 11 :49	11
Figure 5 : Zoom sur l'image multi spectrale.....	12
Figure 6 : Zoom sur l'image Panchromatique.....	12
Figure 7 : Localisation des Points GPS acquis au cours de la mission à Barkédji.....	13
Figure 8 : Interface de Segmentation eCognition.....	15
Figure 9 : Schéma de l'imbrication de différents niveaux.....	16
Figure 10 : Les attributs eCognition	16
Figure 11 : Fonction d'appartenance eCognition	17
Figure 12 : Schéma de la méthodologie générale	18
Figure 13 : Paysage de Seeno	20
Figure 14 : Paysage de Sangre.....	20
Figure 15 : Paysage de Baldiol	21
Figure 16 : Détails d'une Butte de Cuirasse	22
Figure 17 : Cassia tora.....	22
Figure 18 : Zoom de l'image Quickbird (Fusionnée Panchromatique Multispectrale).....	23
Figure 19 : Pondération des couches avant segmentation sous eCognition	23
Figure 20 : Visualisation des deux niveaux de segmentation.....	23
Figure 21 : Résultats de la segmentation avec variation des paramètres de forme /compacité.....	24
Figure 22 : Imbrication des différentes segmentations	25
Figure 23 : Analyse d'objets générés lors de la segmentation eCognition	26
Figure 24 : Profil spectrale d'objets de type « Mare » et « Sols ».....	27
Figure 25 : Signature spectrale de deux types de Mare	27
Figure 26 : Signature spectrale des objets de type « Savane ».....	28
Figure 27 : Exemple de texture sur l'image panchromatique d'un Baldiol (en haut) et d'une savane arborée claire (en bas).....	28
Figure 28 : Graphique récapitulatif des grandes classes d'objets analysés	29
Figure 29 : Capture d'écran Erdas de l'interface Breakpoint et l'image résultat après seuillage	30
Figure 30 : Interface de Recodage du Raster sous Erdas.....	30
Figure 31 : Exemple de Confusion Bâti / Eau avec l'indice de détection de l'eau	31
Figure 32 : Etapes de Création d'une couche Mare	31
Figure 33 : Hiérarchie de groupe (regroupement par thème) à gauche Hiérarchie d'héritage (regroupement par propriétés communes) à droite	32
Figure 34 : Fonction d'appartenance de la classe Végétation	32
Figure 35 : Exemple de Description de classe.....	33
Figure 36 : Méthodologie de la classification à petite échelle	34
Figure 37 : Zoom sur les deux classifications.....	35
Figure 38 : Zoom sur l'occupation du sol.....	36
Figure 39 : Passage des classes de végétation en milieu fermé/ouvert.....	36
Figure 40 : Traduction de l'occupation du sol	37
Figure 41 : Exemple d'Indice de Fermeture	37
Figure 42 : Exemple d'indice de couverture de Mare.....	38
Figure 43 : Exemple d'indice du nombre de Mare	38
Figure 44 : Résultat de l'analyse statistique.....	40
Figure 45 : Bilan - Méthodologie Générale.....	42

Listes des Tableaux :

Tableau 1 : Tableau des différents types de végétations	14
Tableau 2 : Unités paysagères étudiées	21
Tableau 3 : Résultats de la classification	35
Tableau 4 : Récapitulatif des variables explicatives (Rouge : variables explicatives Bleu : autres informations).....	39

1. Introduction

1.1 La fièvre de la vallée du Rift

La fièvre de la vallée du Rift (FVR) est une zoonose (maladie affectant les animaux et occasionnellement transmise à l'Homme) pouvant causer de graves dommages. Elle est due à un arbovirus de la famille des Bunyaviridae et du genre Phlébovirus qui affecte de nombreuses espèces animales domestiques et sauvages [a].

Le virus de la FVR fut tout d'abord isolé en 1930 [a] au Kenya. Des épidémies majeures sont survenues en 1997-1998 au Kenya et en Somalie. En 2000, le virus a été découvert en dehors du continent africain. Des cas ont été confirmés en Arabie saoudite [1] et au Yémen. Le virus s'adaptant à des écosystèmes variés, il est possible que d'autres pays soient exposés.

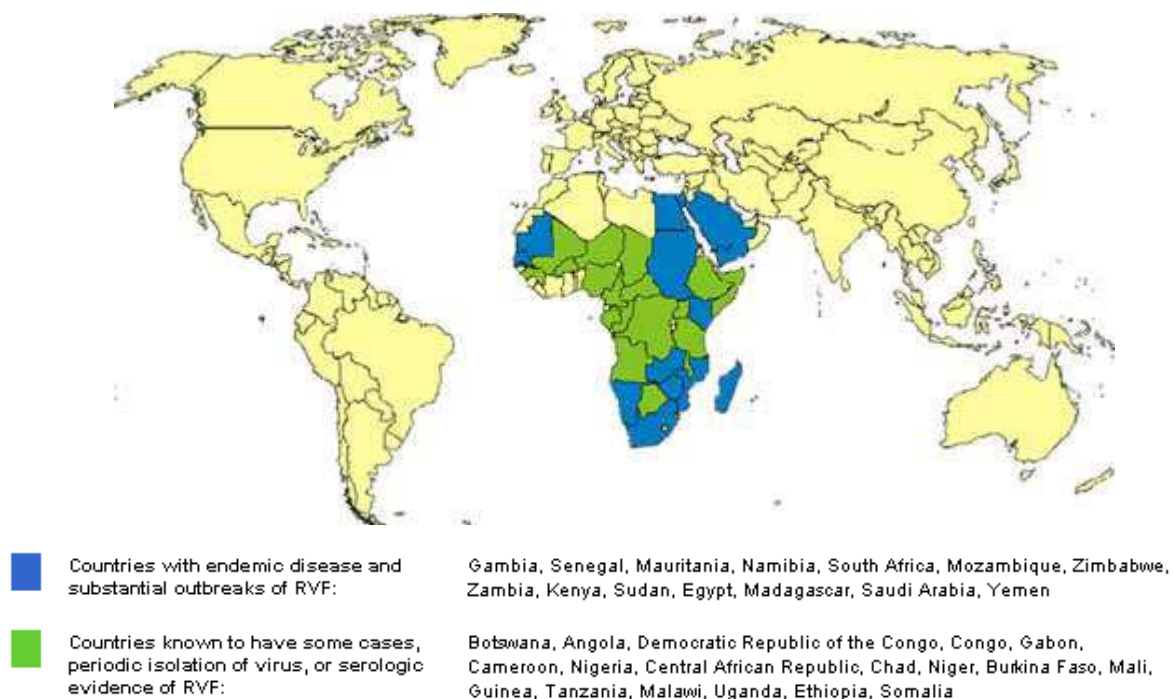


Figure 1 : Répartition de la FVR dans le monde. Source : Center for Disease Control and Prevention, 2005

Contexte au Sénégal :

La première épidémie de FVR au Sénégal a eu lieu en 1987. Depuis, des études de surveillance sérologique sont menées [2] pour prévenir une ré-émergence de la maladie.

La vaccination régulière est peu justifiable à cause de l'existence d'un pouvoir pathogène résiduel ainsi que des apparitions irrégulières d'épizooties entrecoupées de longues périodes inter épizootiques.

Il est donc plus pertinent d'étudier soit la dynamique de la maladie afin de prévoir sa ré-apparition et de prendre les mesures nécessaires, soit l'occurrence spatiale de la maladie, afin de déterminer les zones à risques et pouvoir cibler les zones pour la vaccination. Le vaccin a un pouvoir pathogène résiduel sur les femelles gestantes seulement, il est donc possible de l'utiliser sur les autres animaux sans risque. Il est en plus très immunogène, donc très protecteur.

Région du Ferlo

La région du Barkédji s'étend sur 11000 km², du centre au Nord-Est du Sénégal.

La ressource en eau y est très rare durant les trois quarts de l'année à cause du climat sahélien. Ces fortes contraintes ont conduit les populations locales à pratiquer un pastoralisme extensif.

La principale réserve en eau est constituée de mares temporaires formées lors de la saison des pluies. Ces réservoirs sont utilisés par les éleveurs de la région, ainsi que les transhumants pour alimenter les troupeaux.

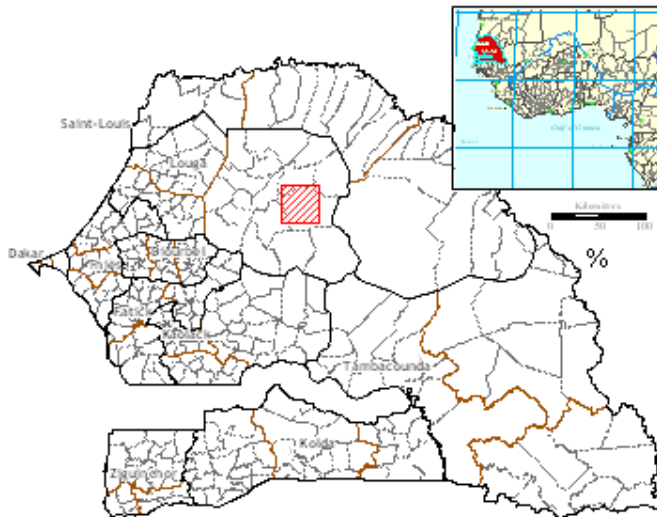


Figure 2 : Localisation de la zone d'étude. Source Arcgis data

Ces mares, constituent l'habitat privilégié des moustiques, vecteurs de la FVR composés essentiellement de *Culex* et *Aedes*. Ces mares occupées à la fois par les hôtes (bétail, hommes) et les vecteurs constituent ainsi des aires de transmission favorables à la fièvre de la vallée du rift.

Si en Afrique de l'Est, les épisodes épidémiques ont pu être corrélés avec des pluies particulièrement abondantes, permettant de définir un modèle prédictif [3] en Afrique de l'Ouest, les flambées épidémiques ont été observées lors d'années à pluviométrie normale, voire déficitaire.

Les mécanismes de transmission de la maladie en Afrique de l'Ouest restent complexes, avec de nombreux facteurs impliqués : répartition spatio-temporelle des hôtes (bétail, hommes) et des vecteurs (moustiques), trajets de transhumance des hôtes, facteurs environnementaux et climatiques.

Cycle et pathologie de la maladie

Lors de la saison des pluies, les zones de bas fond sont alimentées et se transforment en mare temporaire et zones humides [4]. Elles resteront en eau jusqu'à Janvier/Février pour les plus étendues. Les œufs immergés (ponte de la génération précédente) de l'espèce *Aedes* vont alors pouvoir éclore et donner des moustiques adultes. Les œufs d'*Aedes* se caractérisent par leur capacité de résistance à la dessiccation et peuvent attendre plusieurs saisons avant d'éclore. Les moustiques *Aedes*, s'ils sont infectés, pourront transmettre le virus à des animaux domestiques et/ou des rongeurs (réservoirs/hôtes intermédiaires de la maladie) par piqûre. D'autres espèces de moustiques, apparues en fin de saison des pluies comme ceux du genre *Culex*, pourraient servir d'amplificateur de la maladie par contamination horizontale du bétail.

La contamination à l'homme s'effectue principalement par contact avec des animaux et manipulation de produits infectés rarement par transmission directe (moustique).

La transmission transovarienne du virus par les femelles *Aedes* pourrait générer un réservoir pour les années à venir. Ce mécanisme de transmission a été démontré sur *Aedes Mcintoshii* au Kenya, mais il est fort probable que le même phénomène est lieu à Barkédji avec les *Aedes vexans*

Chez les ruminants (bovins, ovins ...), principales victimes de la maladie, elle se traduit par une augmentation importante du nombre d'avortements et une mortalité élevée des nouveaux-nés. Chez l'homme, la FVR se traduit par des poussées de fièvre, de sévères céphalées, des myalgies, des rougeurs de la conjonctive. Certains patients développent une fièvre hémorragique fatale ou des encéphalites et ceux qui survivent peuvent être atteints de maladies oculaires.

1.2 Le projet EDEN (Emerging Diseases in a changing european ENvironment):

Le Projet EDEN [b], réunissant 47 partenaires issus de 24 pays, a été retenu au titre du 6e Programme-cadre de recherche et développement de l'Union européenne.

Il a pour but de comprendre et quantifier l'impact des changements environnementaux sur le risque d'émergence, ou de réémergence de maladies humaines en Europe.

Les objectifs généraux d'EDEN sont à la fois scientifiques et techniques :

- l'enjeu scientifique porte sur la compréhension des processus impliqués dans l'introduction, l'amplification et la diffusion des ces maladies émergeant dans des espaces nouveaux et en mutation,
- le développement méthodologique et technique porte sur les outils de quantification du risque, de développement de systèmes d'alerte et de prise de décision.

Afin d'atteindre ces objectifs, le comité scientifique d'EDEN a sélectionné une série de maladies indicatrices présentant un fort potentiel d'émergence (ou de ré émergence) et de diffusion suite à la modification de notre environnement. Deux zones géographiques ont été retenues comme sites d'observation et de recherche, l'Europe et l'Afrique.

1.3 Application de la télédétection et des SIG à l'épidémiologie

L'épidémiologie étudie la distribution et les déterminants des événements de santé dans les populations. En particulier, elle s'intéresse aux relations entre l'occurrence d'une maladie et des caractéristiques environnementales.

De nombreuses pathologies sont liées plus ou moins directement à l'environnement. Ce lien est particulièrement fort pour des maladies transmises par des insectes vecteurs (moustiques, tiques...) dont la distribution spatiale est liée à des paramètres environnementaux spécifiques (température, pluviométrie, type de végétation...).

L'occupation du sol est donc un indicateur essentiel en épidémiologie [5] qui peut être cartographié à partir de photos aériennes ou d'images satellites suivant des techniques de classifications.

Ces classifications permettent d'obtenir des cartes d'unités paysagères. Des indices environnementaux dérivés de ces cartes peuvent ensuite être aisément calculés et centrés sur les zones à risque de transmission pour évaluer leur implication dans le cycle de la maladie.

D'autres informations caractérisant le milieu peuvent être extraites des images satellites comme le Normalized Difference Vegetation Index (NDVI). Cet indice de végétation met en évidence des différences de réflectances dans le proche infrarouge et le rouge. Il est sensible à l'activité chlorophyllienne.

Toutes ces informations géoréférencées peuvent être collectées, stockées, analysées dans un Système d'Information Géographique (SIG). Le SIG permet de structurer et de manipuler des objets géographiques organisés en couches d'informations distinctes selon leur thème spécifique (occupation du sol, habitat, réseau hydrographique...). Dotés d'outils puissants dédiés à l'analyse spatiale (requêtes, interrogation...), il permet d'étudier les interfaces spatiale et temporelle et plus spécifiquement les relations Hôte/Vecteur/Environnement.

L'arrivée des images de télédétection à Très Haute Résolution Spatiale (THRS) impose actuellement de reconsidérer les méthodes de traitement numérique. En effet, ces nouveaux produits à haute précision permettent de discriminer une nouvelle gamme d'objets au sol jusque-là encore non observable, et pour laquelle les méthodes de classification sont encore à développer.

Il est possible d'atteindre une information très riche, nécessaire pour analyser la complexité de la réalité. La THRS ouvre ainsi de nouvelles perspectives de recherche permettant de travailler à l'échelle de l'habitat du moustique (la Mare) et apporte une nouvelle information pour comprendre et décrire les phénomènes de propagation d'une maladie.

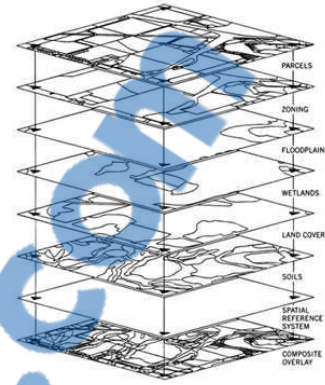


Figure 3 : Représentation des couches dans le SIG. Source : www.mbeconetwork.org

1.4 Problématique et objectifs du stage

Certaines études ont montré que certaines mares/lieux de repos et de reproduction des moustiques vecteurs étaient à risque pour la transmission de la FVR [6]. D'autres études ont le plus souvent été orientées sur la dynamique saisonnière des mares (vidage/remplissage) [7] pour expliquer la « production » de vecteurs de FVR. Il est également possible que les facteurs écologiques et plus spécifiquement les biotopes associés aux mares entrent en jeu pour expliquer la présence, la densité et l'absence des moustiques. En partant de l'hypothèse que ces mares constituent une zone de production des vecteurs et les parcs à bétail celle de contact favorable pour la transmission entre vecteurs et réservoirs de la FVR, nous allons étudier les relations statistiques et spatiales entre des variables paysagères synthétisant l'information relative au pourtour des mares et les données sérologiques moutons qui ont été collectées dans le cadre de deux projets antérieurs à EDEN, Corus (Coopération pour la recherche universitaire et scientifique du ministère des affaires étrangères) et ACI Ecologie Quantitative (Action concertée incitative supportés par le ministère et le ministère français pour la Recherche).

La première partie du stage sera consacrée à la réalisation de la carte de la végétation de la zone de Barkédji ainsi qu'à la détection des mares à partir de l'image Quickbird très haute résolution

La deuxième partie consistera à mettre en relation des variables paysagères dérivées de la carte de végétation avec des données sérologiques collectées sur des moutons en 2003. L'objectif étant de trouver un lien statistique entre l'incidence et une ou plusieurs caractéristiques « écologiques » des mares et à plus long terme, de caractériser celles étant potentiellement à risque pour la transmission de la maladie.

2. Données

2.1 Image Quickbird

Figure 4

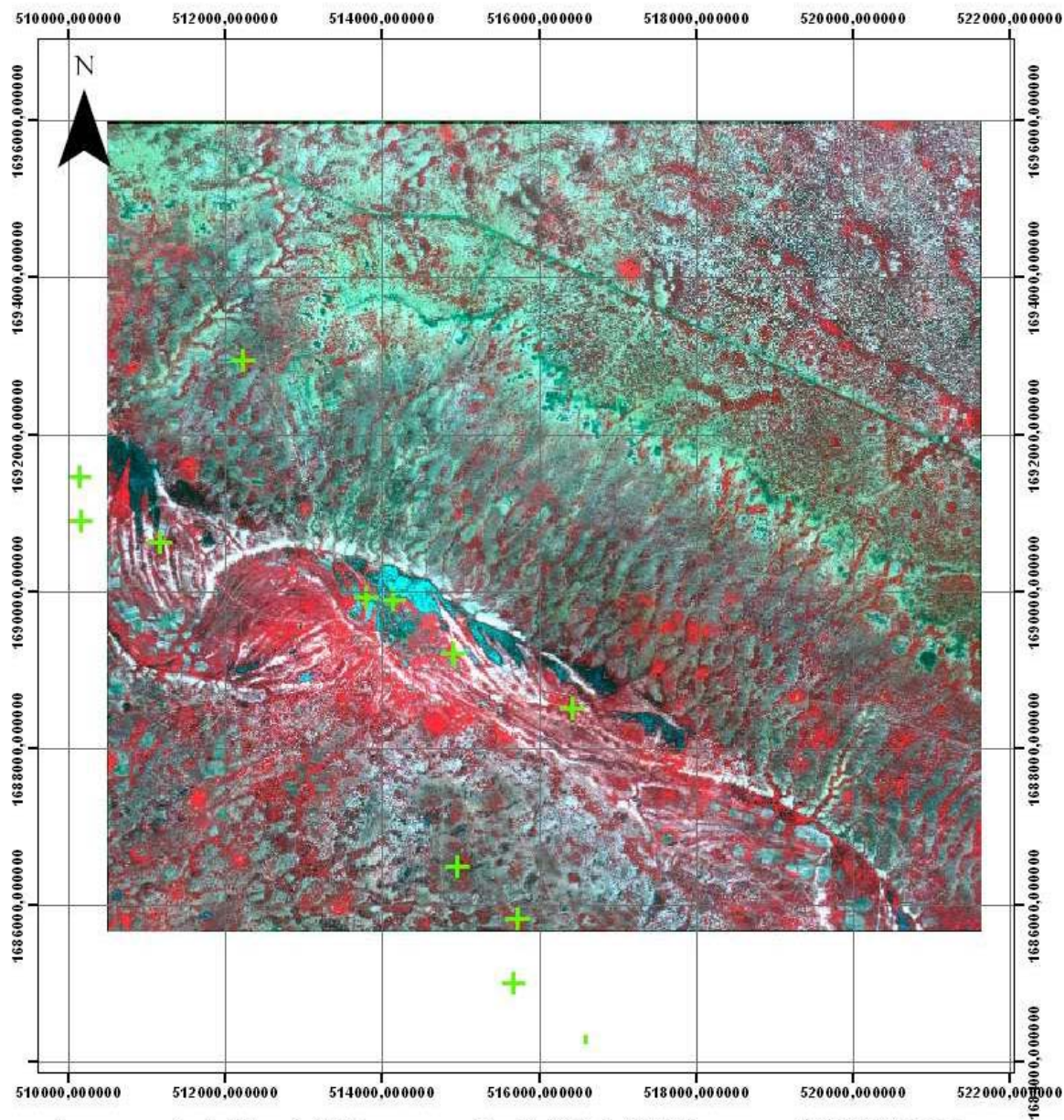


Image acquise le 04 août 2005

Affichage des bandes 4 / 3 / 2

Système de projection: WGS_1984_UTM_Zone_28N

Figure 5 : Digital Globe. [Région du Ferlo, Sénégal]. Quickbird [Image satellitaire] 04/08/2005 à 11 :49

L'affichage dite en « fausse couleur », est une composition colorée de type 4 (PIR), 3 (R), 2 (Vert) où le rouge correspond à de la végétation, le bleu/vert à l'eau et le vert à des types de sol nu.

Caractéristiques de Quickbird en annexe 1

Une image Quickbird (Figure 5) a été acquise le 04 août 2005 à 2.44 m de résolution spatiale pour la multispectrale (XS) et 0.61 m pour la panchromatique. Couvrant un zone de 13 x 13 km centrée sur le village de Barkédji, 4 bandes spectrales compose l'image multispectrale : le Bleu, Vert, Rouge et Proche infrarouge. La date d'acquisition a été volontairement choisie au mois d'août, correspondant au pic de la saison des pluies, afin de cartographier le plus grand nombre de mares temporaires qui disparaissent assez rapidement avec la montée des températures.

Système de projection: WGS_1984_UTM_Zone_28N

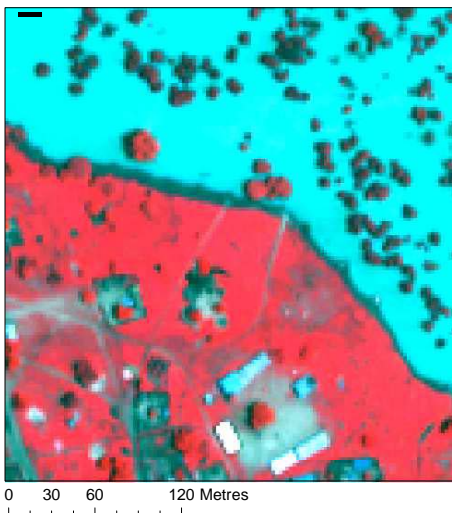


Figure 6 : Zoom sur l'image multi spectrale

Système de projection: WGS_1984_UTM_Zone_28N



Figure 7 : Zoom sur l'image Panchromatique

L'image a été acquise avec un niveau de traitement « Standard » signifiant que des corrections radiométriques et géométriques ainsi que des corrections des distorsions dues aux détecteurs ont été préalablement appliquées.

2.2 Données sérologiques

Il s'agit de données sérologiques prélevées sur des ovins dans la zone de Barkédji. On dispose de coordonnées géographiques enregistrées au GPS (Global Positioning System) de chaque lieu (campements) où ont été prélevés les échantillons pour les tests sérologiques. Ces campements se répartissent autour de mares (accès aux ressources plus facile). Chaque campement est en quelques sortes liés à une seule mare tant quelle peut fournir de l'eau au troupeau. L'échantillonnage a été réalisé une première fois en Août et la seconde de mi novembre à mi décembre 2003 [6].

L'image Quickbird étant plus petite que la zone d'étude couverte par le jeu de donnée, nous avons réduit nos échantillons à 8 points sérologiques répartis autour de 4 mares (Barkedji, Kangeljji, Furdu et Niaka) contre 15 points sérologiques et 7 mares à l'origine (Figure 4).

2.3 Données collectées sur le terrains

Pour chaque mare visitée, une fiche de terrain a été remplie, dédiée à la description phénologique des mares. Ces relevés serviront de parcelles d'entraînement pour le traitement numérique sous eCognition de l'image satellitaires Quickbird

La seconde priorité a été d'orienter les observations sur les points de piégeages des équipes de l'Institut Pasteur de Dakar. Des relevés de végétation précis sur les 80 lieux de piégeage de Pasteur

ont été effectués dans la perspective d'une étude future sur la relation spatiale entre milieu naturel et densité/présence/absence de moustique.

Nous disposons ainsi de deux types de fiches celles associées à la description phénologique des mares et celles utilisées pour la description de la végétation en dehors du pourtour immédiat des mares (Annexe 2 et 3). Un GPS a permis d'enregistrer la position géographique des points de collecte. Pour les fiches « parcelles homogènes », au moins 3 points étaient enregistrés afin de délimiter une zone « homogène » d'une superficie au moins égale à 60*60 m.

A la fin de la journée, les points GPS relevés sur le terrain étaient intégrés puis visualisés sous SIG. Cette opération permettait d'effectuer le bilan de la journée et de repérer les zones oubliées ou sous-exploitées. En fonction du bilan, de nouveaux points étaient alors choisis puis transférés dans le GPS. La figure 8 montre les sites où ont été réalisés des relevés de végétation :

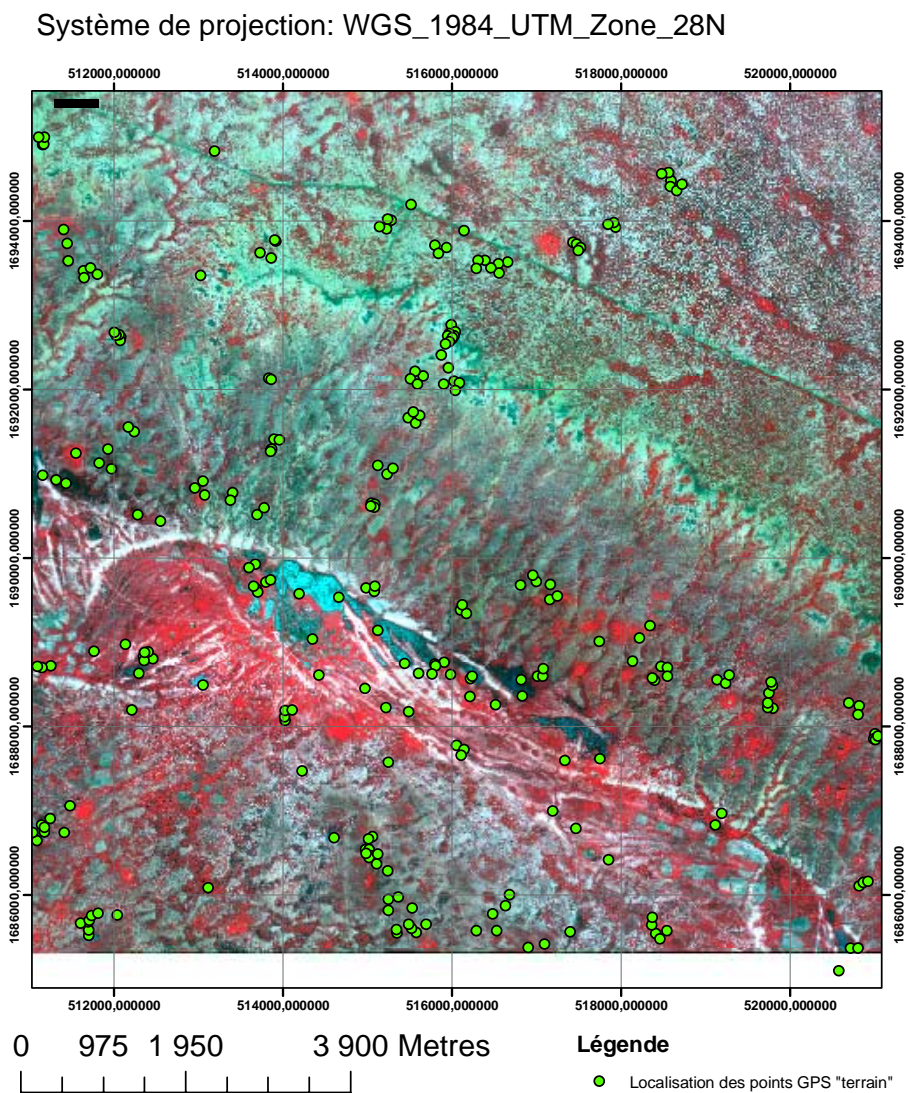


Figure 8 : Localisation des Points GPS acquis au cours de la mission à Barkédji

Au total, 293 points GPS ont été collectés, comprenant plus de 50 mares et dépressions humides, plus de 40 parcelles homogènes (savane herbeuse, savane arborée, sangre, baldiol...) et enfin une dizaine de zones de culture (essentiellement mil et quelques parcelles d'arachide).

Les différents types de végétation rencontrés ont été relevés dans trois dénominations : latine, peul et wolof, dans la perspective d'une carte qui pourra être utilisable pour une communication avec les populations locales (Tableau 1).

Latin	Peul	Wolof
Ligneux		
Acacia nilotica	gawdi	goniakié
Acacia raddiana	tchiluki ou tchili	sêng
Acacia seyal	bulbi (pl : bulbé)	surur
Adansonia digitata		
Anogeissus leocarpus	kodioli	guédiane
Balanites aegyptiaca	mutchétéki (pl : mutchétéké)	soump
Capparis corymbosa	gumi balevi	-
Combretum glutinosum	dôkki (pl : doke)	ratt
Grewia bicolor	kelli (pl : kelle)	kelle
Guiera senegalensis	gelôki	n'guer
Myrtagyna inermis	koyli (pl : koylé)	kross
Pterocarpus lucens	tchagi (pl : tchagé)	bey bey
Ziziphus mauritania	diabi (pl : diabé)	dèm
Herbacés		
Brachiaria *...	paggini	-
Cassia tora	ulo	n'dour
Chloris prierii	tchelbi	gén u dor
Cyperus *...	baké bolyé	noup noup
Gisekia pharnaceoides	takel pôli	-
Kyllinga welwitchii	-	-
Oryza banthii	moro véndu	tieb
Oryza beviligulata	moro véndu	tieb
Sphaeranthus senegalensis	débté	lugut
Zornia glochidiata	déggo	ndéger...

Tableau 1 : Tableau des différents types de végétations

3. Matériel

3.1 Logiciel eCognition

Une société privée, Definiens, a développé un logiciel d'analyse d'image nommé « eCognition ». A l'origine, le logiciel était adapté à l'imagerie médicale puis s'est généralisé à l'étude d'images très haute résolution (photo aérienne, image satellite...).

Son atout principal étant la description des surfaces par une approche orientée objet. C'est-à-dire que l'on travaille sur des objets radiométriquement « homogènes » et non pas sur des pixels. C'est pour cette raison qu'il est employé pour la très haute résolution, car il évite l'effet « poivre et sel » (Apparition de points correspondant à des objets trop petits pour impressionner plus qu'un pixel de l'image).

Il permet de développer des méthodes hybrides couplant segmentation d'images, analyse de texture, reconnaissance de formes, classification par plus proche voisin ou par logique floue...

Il fournit un outil complet pour le passage d'une information raster à une couche vectorielle d'objets ainsi que pour la classification des objets créés. Je présenterai ici les étapes spécifiques à l'approche orientée objet, nécessaires au traitement des images satellites :

3.1.1 Segmentation

Pour passer une information du format raster au format vecteur, il est nécessaire de segmenter l'image. Cette Première étape est une particularité de la classification orientée objet.

La segmentation multi-résolution permet de générer des objets de tailles variables (structures fines ou grossières) selon le choix de l'utilisateur. Elle permet d'adapter facilement l'extraction de primitives d'objets (Donner ici définition de primitives d'objet).

Deux modes de segmentation sont possibles :

- La procédure de segmentation dite Normale (Multiresolution Segmentation) est basée sur une technique ascendante de croissance de région commençant par un objet constitué d'un seul pixel de l'image. Durant le processus les objets sont fusionnés par paires pour créer des objets de taille supérieure tout en réduisant au minimum l'hétérogénéité. A chaque itération sont fusionnés les deux objets adjacents de l'image ayant une fois réunit la plus faible croissance d'hétérogénéité. Si la plus faible croissance d'hétérogénéité excède le seuil défini par le paramètre d'échelle, le processus s'arrête.
- Le mode de segmentation « Différence Spectrale » fonctionne selon le même procédé en prenant en compte uniquement des valeurs spectrales de l'image

L'utilisateur joue sur des paramètres d'échelle (Figure 9), de forme et de compacité pour créer les éléments porteurs de l'information :

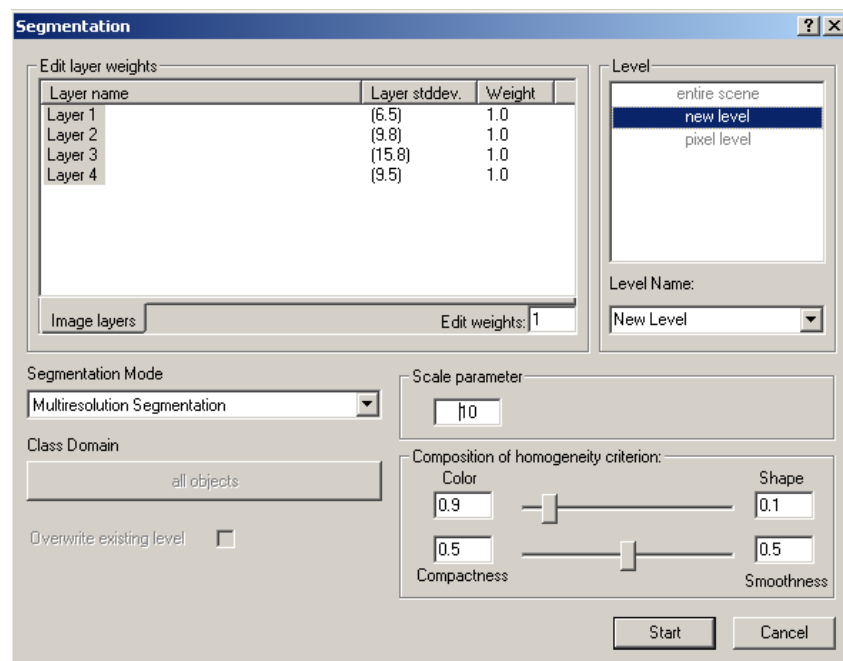


Figure 9 : Interface de Segmentation eCognition

Le Paramètre de Forme / Couleur et de compacité / Lissé font partis des critères d'Homogénéité. Le critère de couleur définit le pourcentage de l'homogénéité globale issu des valeurs spectrales des canaux de l'image par opposition aux critères d'homogénéité de forme. Ce dernier est à employer dans le cas d'image radar (données très texturées) afin d'éviter une forme trop « fractale » des objets. En plus de l'information spectrale, eCognition permet d'optimiser selon la forme des objets à bordures compactes (rapport entre la longueur de la bordure et la racine carrée du nombre de pixel formant l'objet) ou lisses (rapport entre sa longueur et la longueur du plus petit côté).

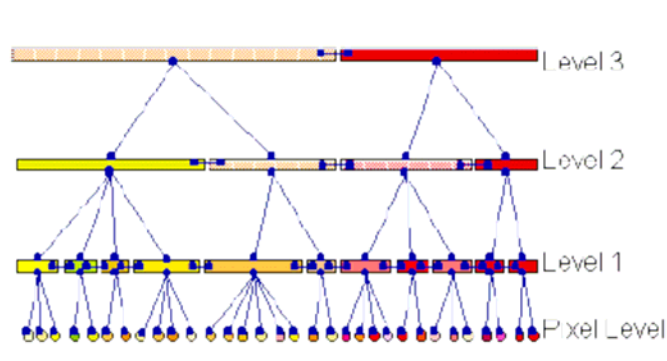


Figure 10 : Schéma de l'imbrication de différents niveaux

Il est possible de pondérer l'intervention des couches en entrée. Le système de Niveau permet de créer des segmentations imbriquées (Figure 10) de plus grande ou plus petite taille. eCognition permet aussi d'intégrer des couches thématiques (vectorielle) pour affiner la segmentation.

3.1.2 Classification

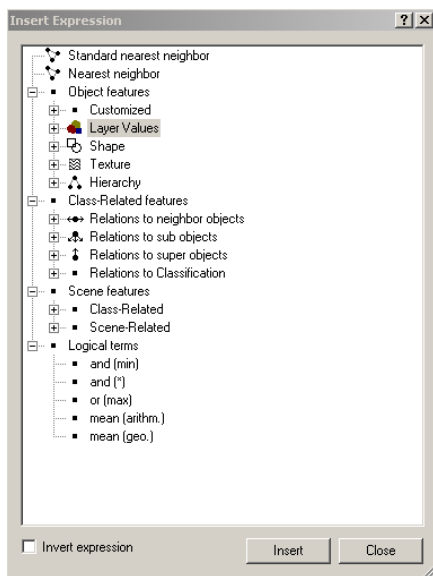


Figure 11 : Les attributs eCognition

eCognition propose deux techniques de classification supervisée : l'approche par plus proche voisin et les fonctions d'appartenance. Les deux approches peuvent être combinées pour constituer la stratégie la plus appropriée pour chaque classe.

La classification est conduite en utilisant la logique floue. La classification floue assigne non seulement un objet d'image à une classe, mais elle fournit en plus le degré d'affectation à toutes les classes considérées.

Le classification au plus proche voisin (Nearest Neighbor algorithm) exige des échantillons pour chaque classe, ainsi qu'un espace des attributs (Figure 11) qui peut être constitué par n'importe quelle combinaison d'attributs (Valeurs des canaux, Forme, texture...). Chaque objet de l'image est ensuite comparé aux échantillons et rangé dans la classe ayant l'échantillon le plus proche de l'objet selon les attributs retenus. Le Standard Nearest Neighbor applique les mêmes attributs simultanément à toutes les classes.

Les fonctions d'appartenance (Figure 12) constituent une méthode simple pour traduire une gamme arbitraire d'attributs en une gamme uniforme comprise entre [0...1]. Une valeur d'appartenance entre 0 et 1 est renvoyée par la fonction d'appartenance après évaluation de chacun des attributs formant une description de classe. Ces valeurs sont combinées en une valeur globale d'affectation de classes au moyen d'opérateurs logiques.

L'édition graphique est limitée aux fonctions d'appartenance à une dimension, leur utilisation est recommandée lorsque une classe peut être séparée des autres par un nombre limité de critère.

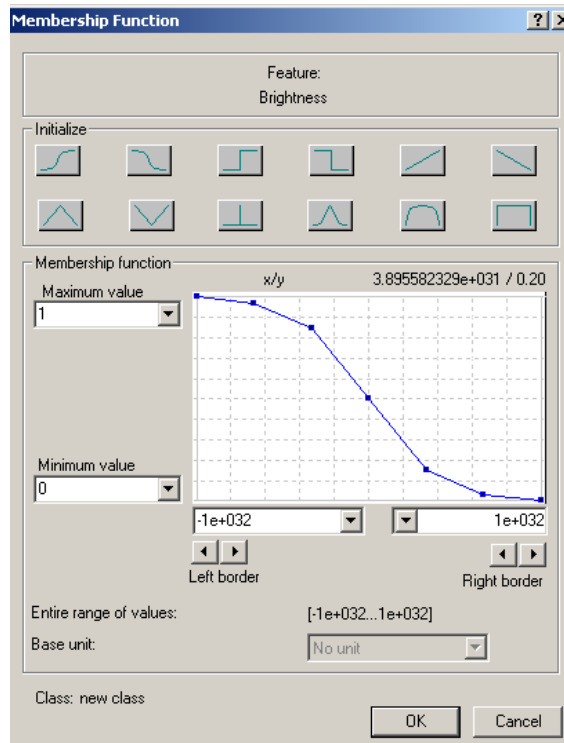


Figure 12 : Fonction d'appartenance eCognition

La classification orientée objet permet la création de différents niveaux qui sont donc liés les uns aux autres, on peut ainsi définir des relations de classification entre objets de même niveau mais aussi entre objets de niveau inférieur ou supérieur.

Le langage orienté objet permet aussi de différencier la hiérarchie d'héritage et de groupe. La première notion concerne la transmission des descriptions de classes vers des classes filles et la deuxième de regrouper sémantiquement des classes aux descriptions d'attributs variés.

eCognition cumule donc de nombreuses méthodes pour traiter avec facilité les images très haute résolution spatiale.

3.2 Logiciel R

Le langage R, parfois décrit comme « GNU S », est un langage de programmation et un environnement mathématique utilisés pour l'analyse statistique. R dispose de nombreuses fonctions graphiques. Les sites miroirs [c] proposent des centaines de bibliothèques sous licence libre. **Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.**

4. Méthode

4.1 La classification supervisée orientée objet

La première étape de la classification orientée objet est la segmentation. Elle conditionnera la création des objets sur lesquels la classification sera effectuée.

Les points GPS récoltés sur le terrain vont servir à vectoriser sous le SIG les parcelles dites terrain ou d'apprentissage. Elles seront ensuite intégrées dans eCognition et serviront à la constitution de l'information de Référence. En s'appuyant sur cette référence, les algorithmes de classification réalisent l'occupation du sol qui devra être contrôlée par les parcelles « vérité terrain ».

La couche d'occupation du sol est obtenue au format shape (.shp) directement intégrable sous le SIG. Enfin, une matrice de confusion sera réalisée, permettant d'évaluer de la qualité de la classification.

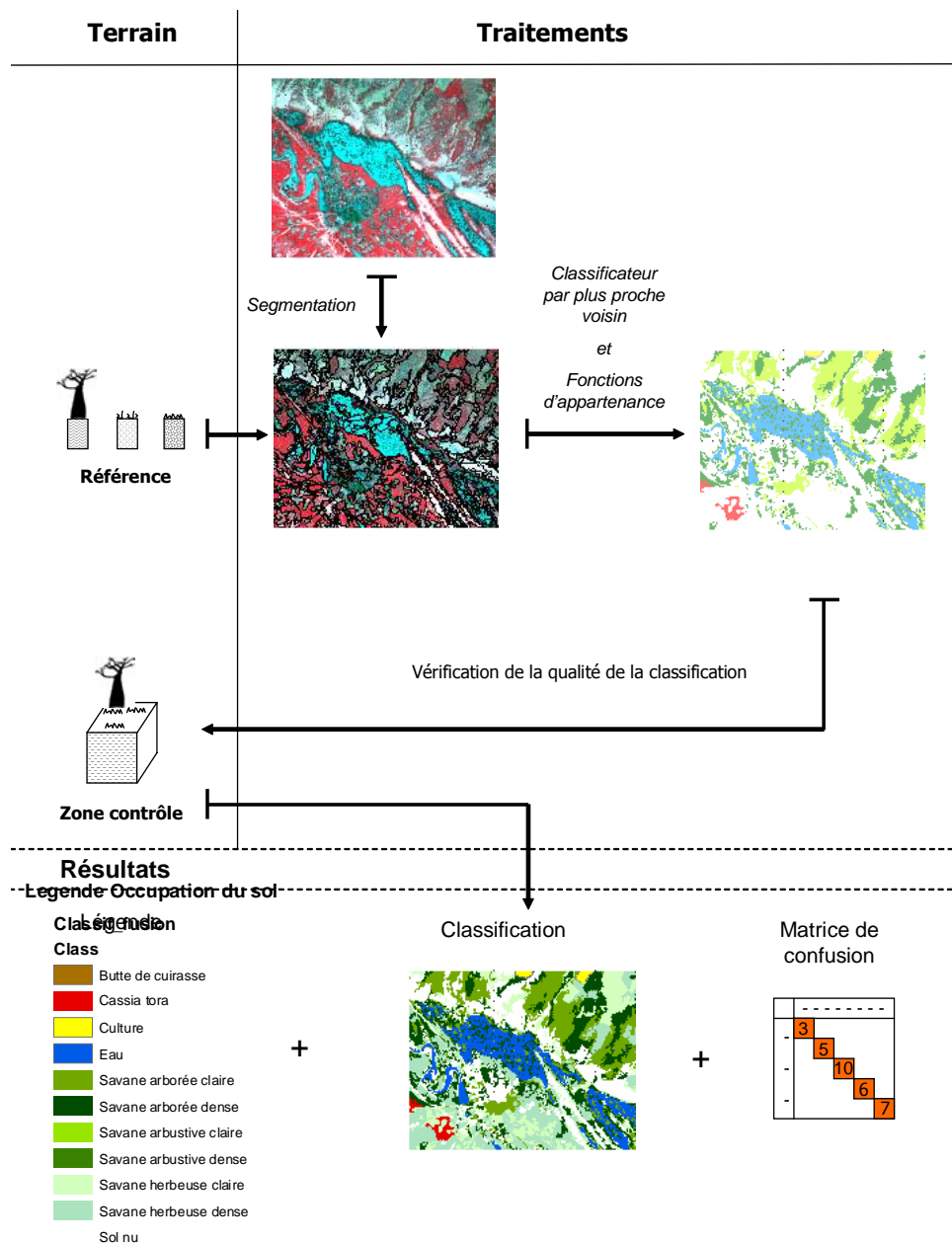


Figure 13 : Méthodologie de la classification orientée objet

4.2 Méthodologie statistique

4.2.1 La régression logistique

Pour analyser les données sérologiques, un modèle basé sur la régression logistique est choisi. C'est une technique statistique qui a pour objectif, à partir d'un fichier d'observations, de produire un modèle permettant de prédire les valeurs prises par une variable catégorielle, le plus souvent binaire, à partir d'une série de variables explicatives continues et/ou binaires.

En tant que méthode de prédiction pour variable catégorielle, la régression logistique est tout à fait comparable aux techniques supervisées proposées en apprentissage automatique (arbre de décision, réseaux de neurones, etc.), ou encore l'Analyse Discriminante prédictive en statistique exploratoire. Il est notamment possible de les mettre en concurrence pour choisir le modèle le plus adapté pour un problème de prédiction à résoudre.

4.2.2 Cas des données sérologiques

Les données sérologiques ont donc été analysées en utilisant un modèle mixte de régression logistique. L'incidence représente dans le modèle la réponse, agrégée au niveau du campement, car les échantillons ont été prélevés dans les parcs de nuit et non pas par mare. La mare d'origine (c'est à dire la mare la plus proche du lieu de prélèvement de la sérologie) a été incluse dans le modèle comme un effet aléatoire, permettant de prendre en compte de variables liées aux mares, telles que la turbidité de l'eau ou la composition végétale qui ne figurent pas dans le modèle en tant que variable explicative.

Les autres facteurs testés (Cf. **6.3**) seront inclus dans la partie fixe du modèle, non lié au hasard. Cette méthode permet d'estimer à la fois la moyenne du niveau de population (incidence générale) et les moyennes liées aux mares.

5. Cartographie de l'occupation du sol

Pour réaliser l'occupation du sol, plusieurs étapes sont nécessaires. Il faut définir les unités paysagères qui serviront de repère pour la segmentation. L'analyse des objets créés par segmentation permettra de définir les règles pour la classification. La détection des mares s'effectuera en parallèle.

5.1 Définitions des unités paysagères étudiées

La zone est en majorité composée des paysages typiques du Sahel, décrits ci-après et dénommés par leur nom vernaculaire *Fulani* (Raynaut, Gregoire et al. 1997) :

Le *seeno* (Figure 14) est une steppe arbustive sur dune et inter-dune (substrat sableux), dominée par les espèces ligneuses soudaniennes *Balanites aegyptiaca*, *Guiera senegalensis* et *Combretum glutinosum*. Cette formation est propice à l'agriculture et offre une productivité fourragère élevée, grâce aux bonnes propriétés hydriques des sols sableux. Elle est cependant sensible à l'érosion éolienne et au surpâturage.



Figure 14 : Paysage de Seeno

Le *sangre* (Figure 15) est une steppe arbustive associée aux affleurements de cuirasse latéritique ; *Pterocarpus lucens*, *Sterculia setigera*, *Commiphora africana*, *Grewia bicolor* sont les espèces dominantes. Ses faibles réserves hydriques le rendent impropres à l'agriculture, mais pas au pâturage.



Figure 15 : Paysage de Sangre

Le *baldiol* (Figure 16) est une steppe arborée poussant dans les dépressions hydromorphes, à substrat argilo-sableux et caractérisée par *Acacia seyal*, *Adansonia digitata*, *Balanites aegyptiaca*, *Boscia senegalensis*. La pression pastorale y est importante en saison des pluies. Dans le Ferlo, la plupart des Baldiols correspondent aux anciens bras du fleuve Sénégal.



Figure 16 : Paysage de Baldiol

Ces 3 unités écologiques ne sont pas directement détectables par télédétection, car elles expriment une unité prenant en compte à la fois le type de sol et la végétation associée. Seule la végétation et les éléments présents au sol pourront être cartographiés. On analysera donc des savanes arborées, arbustives ou herbeuses de densités variables. Les unités du type seeno, sangre et baldiol pourront être définies en tant que tel, que si l'on recoupe la couche occupation du sol avec une couche pédologique suffisamment précise.

Les unités paysagères étudiées sont présentées dans le tableau 2, ci-dessous :

Butte de cuirasse
 Cassia Tora
 Culture
 Eau
 Savane arborée claire
 Savane arborée dense
 Savane arbustive claire
 Savane arbustive dense
 Savane herbeuse claire
 Savane herbeuse dense
 Sol nu

Tableau 2 : Unités paysagères étudiées

Ces unités ont ensuite été regroupées en 3 thèmes : les sols nus / la végétation / l'eau. Au niveau Entomologique, nous avons traduit ces paysages en indices d'ouverture et de Fermeture des milieux.

Détails sur les unités paysagères

La butte de cuirasse (Figure 17) correspond à un affleurement de la couche de latérite.



Figure 17 : Détails d'une Butte de Cuirasse

Les cultures sont considérées comme une sorte de sol nu, car à la date d'acquisition de l'image satellite (août), les cultures ne sont pas encore semées.

Cassia tora (Figure 18) correspond à une plante ou petit arbuste à feuilles pennées, cespiteux (désigne une plante poussant en touffe dense). On la trouve souvent à l'emplacement des anciens campements. Les touffes denses sont facilement détectables sur l'image satellite.



Figure 18 : Cassia tora

5.2 Choix des paramètres de segmentation

En entrée, deux types d'images sont disponibles. L'image Brute multi-canal ou l'image Fusionnée dites P+XS. La fusion issue de la fusion entre l'image panchromatique 0.6m et l'image multispectrale 2.4m (Figure 19) permet d'obtenir une image multi-canal de résolution plus fine (0,6m), mais en contrepartie transforme l'information spectrale d'origine. Nous avons testé la segmentation sur l'image P+XS mais le résultat à été décevant, générant des objets moins proches de la réalité, plus bruités.

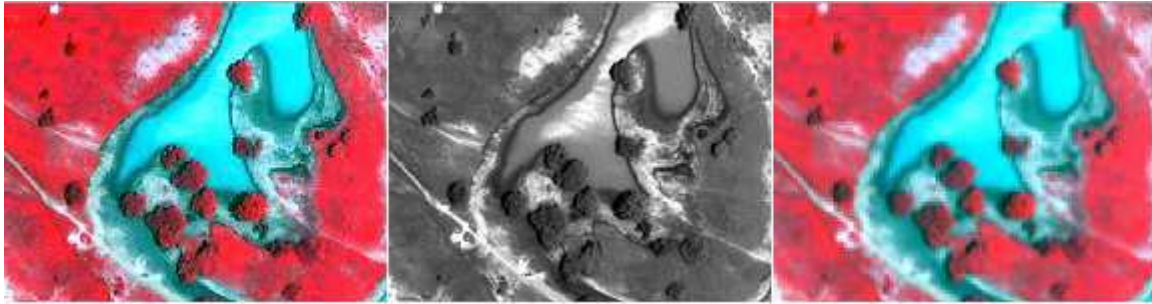


Figure 19 : Zoom de l'image Quickbird (Fusionnée | Panchromatique | Multispectrale)

On travaillera donc uniquement sur l'image multi canal brute pour la segmentation et non sur l'image fusionnée.

Le mode de Segmentation choisi sera Multi résolution, afin de pouvoir jouer sur les paramètres de Forme et de Couleur.

Layer name	Layer stddev.	Weight
Layer 1	(6.5)	1.0
Layer 2	(9.8)	1.0
Layer 3	(15.8)	1.0
Layer 4	(9.5)	1.0
Layer 5	(7.5)	0.0

Image layers Edit weights: 0

Figure 20 : Pondération des couches avant segmentation sous eCognition

Une pondération de 1 sur les 4 couches de l'image multi spectrale est appliquée (Figure 20). Une pondération de 0 sur l'image panchromatique (Layer 5) permet de ne pas la prendre en compte dans la segmentation. De la même manière que pour l'image fusionnée, la résolution plus fine de l'image panchromatique générant du bruit lors de la segmentation n'a pas été prise en compte.

Le premier paramètre à déterminer est le paramètre d'échelle ou « Scale Parameters ». La classification va se faire sur deux niveaux (Figure 21). Le premier focalisé sur les mares afin de pouvoir décrire la végétation couvrante de la mare ainsi que le pourtour. Le Paramètre d'échelle choisi est celui conservant l'individualité des arbres de taille importante (Baobab) soit un paramètre de **50**.

Un deuxième niveau de classification concerne l'occupation du sol. Le paramètre d'échelle générant des objets de l'ordre d'une parcelle a été choisi soit un paramètre de **200**.

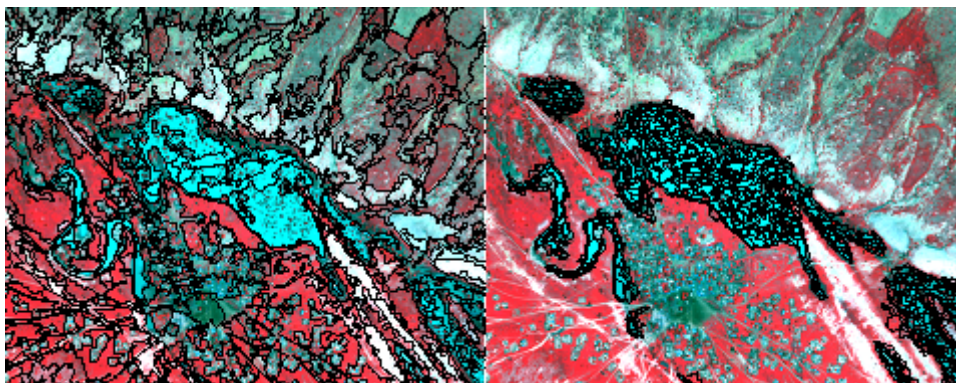


Figure 21 : Visualisation des deux niveaux de segmentation

Il reste à choisir le paramètre de forme, ainsi que celui de compacité (Figure 22).

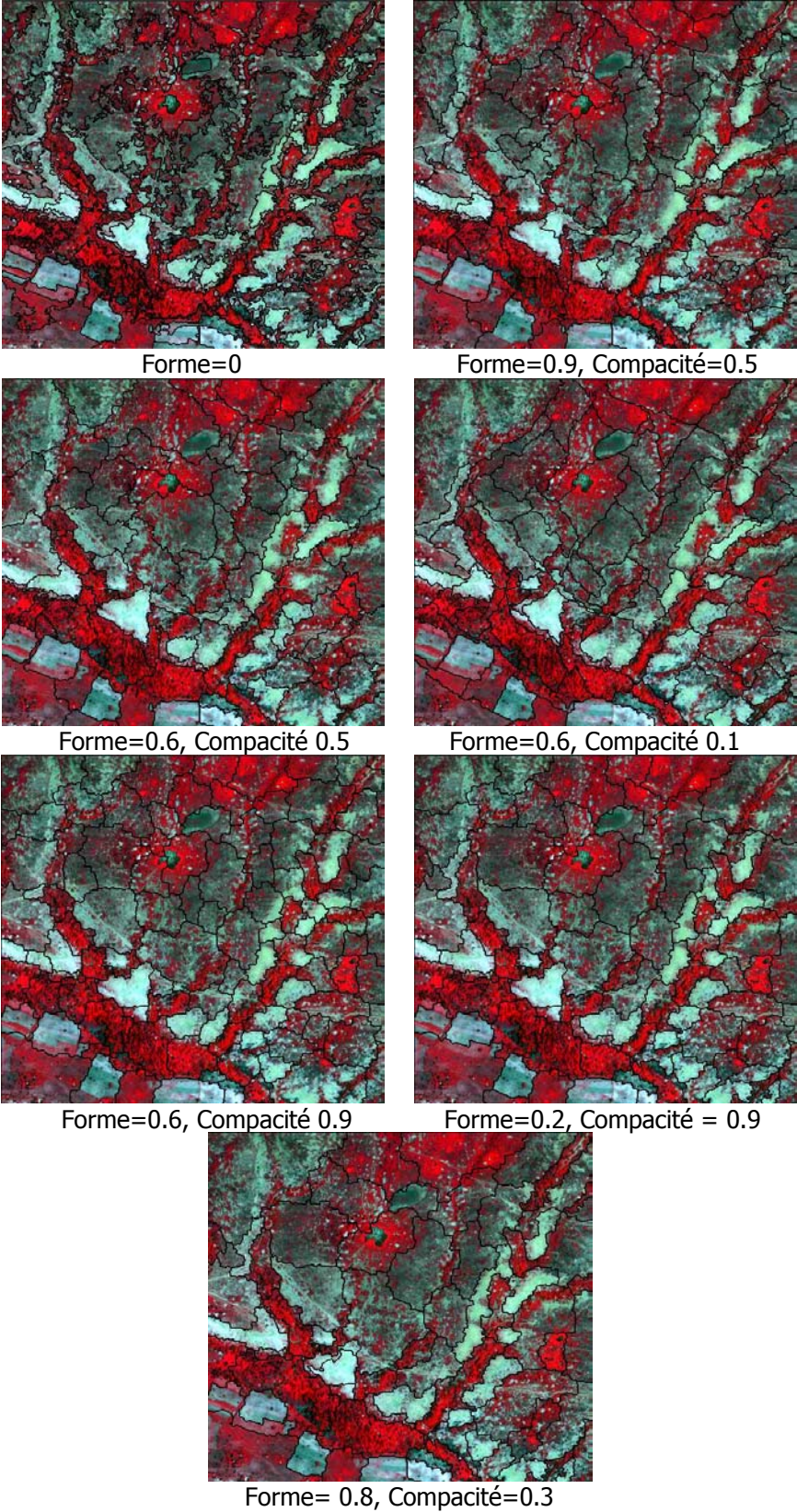


Figure 22 : Résultats de la segmentation avec variation des paramètres de forme /compacité

Les paramètres suivants : $\text{Forme}=0.8$, $\text{Compacité}=0.3$ ont été sélectionnés, car ils conservent au mieux la forme des Baldiols, ces dépressions humides à végétation dense correspondant souvent aux anciens bras du fleuve Sénégal, tout en fournissant des objets de tailles suffisantes pour les autres classes d'objets.

Le baldiol est une des plus intéressantes unités paysagères à étudier car certainement la plus propice à la présence de moustiques.

Les deux niveaux de segmentation choisis sont imbriqués l'un dans l'autre (Figure 23). eCognition permet de générer un premier niveau, avec un paramètre d'échelle de 50, puis de re-segmenter en fonction de nouveaux paramètres tout en conservant les bordures des objets de niveau inférieur.

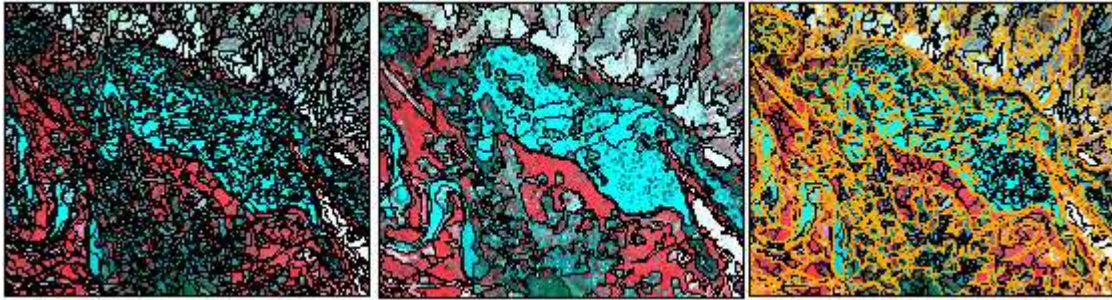
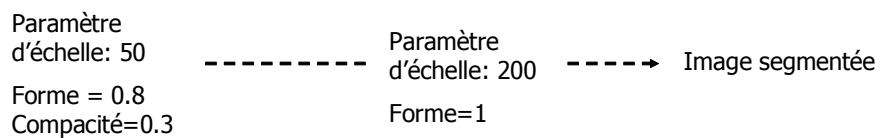


Figure 23 : Imbrication des différentes segmentations

Le schéma suivant sera donc retenu pour la segmentation de l'image Quickbird :



5.3 Analyse des objets générés par segmentation

Afin de contrôler les objets générés lors de la segmentation (Figure 24), d'analyser leur homogénéité intra et inter objets, nous avons extrait des objets de chaque classe d'occupation du sol pour étudier leurs signatures spectrales respectives. Cette opération a été effectuée aux deux niveaux, (paramètres d'échelle de 200 et de 50) de segmentation retenus.

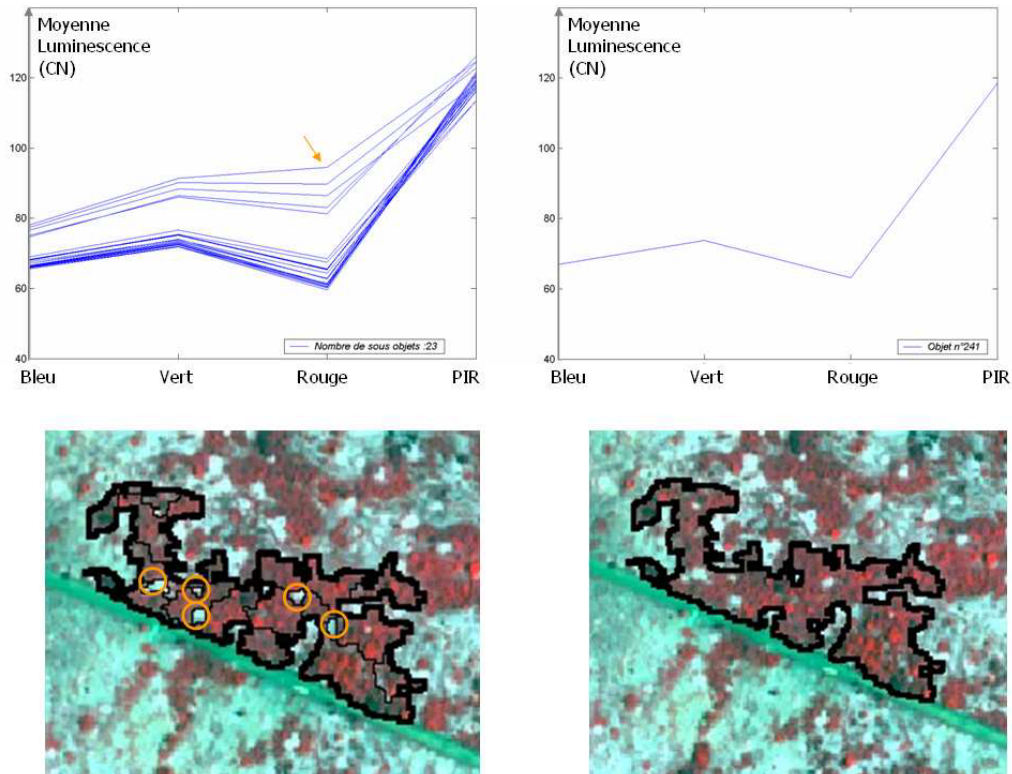


Figure 24 : Analyse d'objets générés lors de la segmentation eCognition

Nous observons une certaine hétérogénéité des objets segmentés avec un paramètre d'échelle de 200. Le problème ne se posera pas pour des objets comme les savanes, mais pour des objets plus petits comme les mares, il ne sera pas adapté. eCognition effectuant une moyenne spectrale sur l'objet, si le paramètre d'échelle est important (comme 200), il est possible que certaines mares ne soient pas détectées. Nous avons donc appliqué un paramètre d'échelle de 50 afin que toutes les mares soient prises en compte.

Pour une question de temps de calcul, la classification n'est pas réalisable sur l'image entière segmentée à un paramètre d'échelle de 50 (représentant 1 millions de polygones contre 100 000 à un paramètre d'échelle de 200). Nous avons donc été contraint d'effectuer en plus de l'occupation du sol, un traitement spécifique pour la détection des mares de la zone sur laquelle nous avons ensuite appliquée une classification avec un paramètre de d'échelle de 50 pour la segmentation.

L'analyse inter objet nous donne des informations sur les confusions possibles lors de la classification. Les profils spectraux ne sont issus que d'un nombre limité d'objets. Le problème de la représentativité peut ainsi se poser, mais il s'agit uniquement d'évaluer grossièrement les classes aux caractéristiques spectrales proches.

La plupart des sols nus, dont les cultures que nous avons considérées comme tel car encore non semés, sont spectralement proches. De plus, quelques confusions peuvent apparaître entre l'eau de certaines mares comme le montre la figure 25.

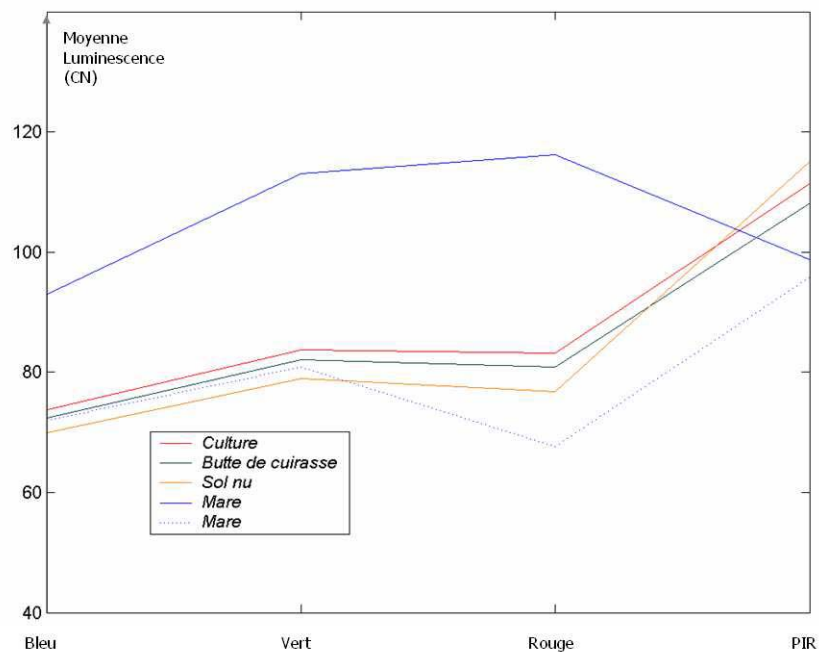


Figure 25 : Profil spectrale d'objets de type « Mare » et « Sols »

Concernant les mares, on observe deux signatures spectrales différentes (Figure 26). Elles peuvent s'expliquer par des différences de turbidité de l'eau, de profondeur de la lame d'eau ou encore par la présence de dépôts organiques.

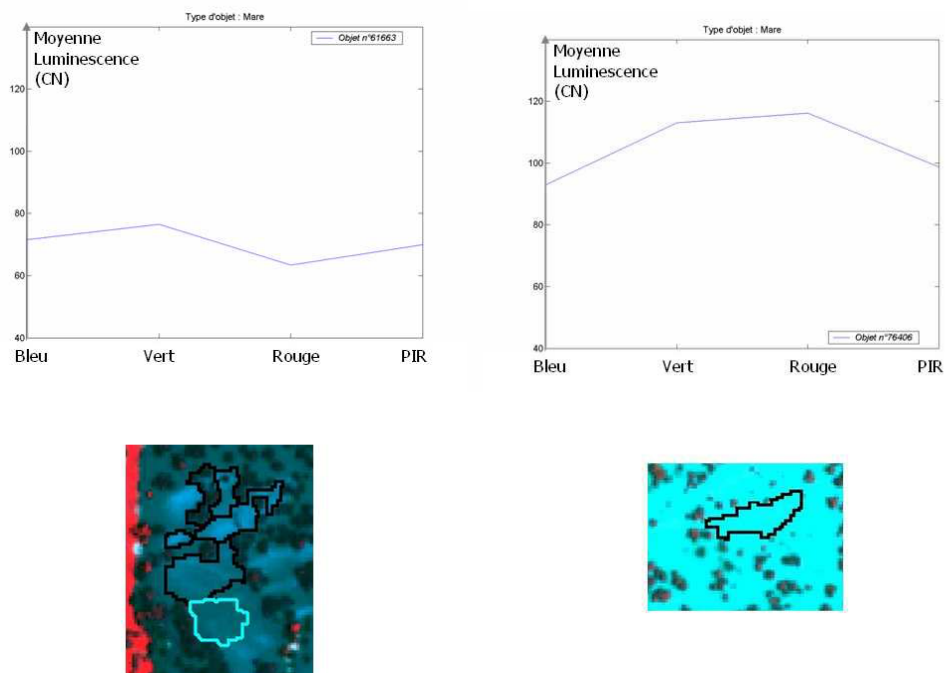


Figure 26 : Signature spectrale de deux types de Mare

Concernant les différents types de savanes il semble difficile de les différencier par type, du fait d'une signature spectrale très proche (Figure 27). Pour ce qui est des densités, et plus spécifiquement les densités de végétation herbacées, les résultats sont difficilement interprétables et comparables, car nos relevés de végétations effectués sur le terrain ne correspondent pas à la date d'acquisition de l'image Quickbird.

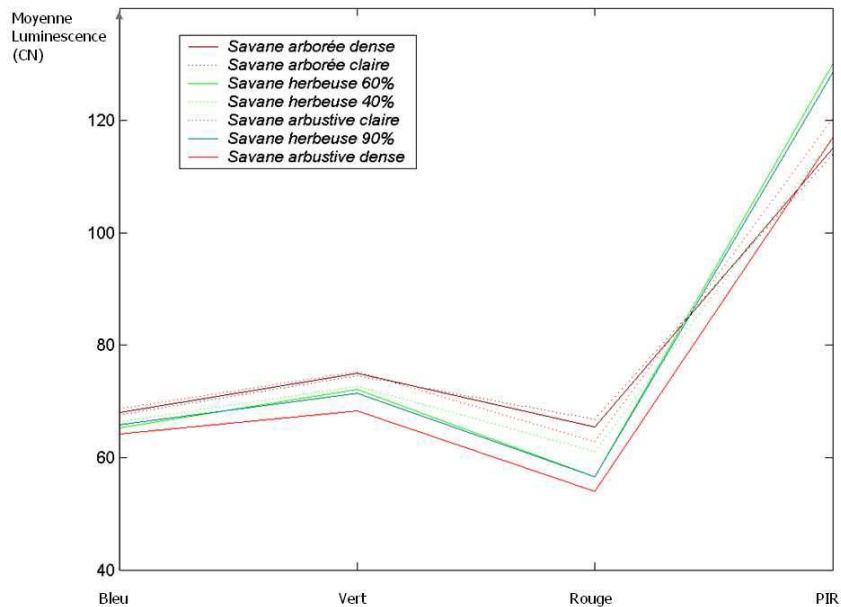


Figure 27 : Signature spectrale des objets de type « Savane »

Ce constat, nous dirige plutôt vers une règle prenant en compte le NDVI (Indice de végétation normalisée) et un seuil à partir duquel la végétation sera considérée comme dense ou claire.

Pour les différents types de savanes arborée/arbustive/herbeuse, la réponse spectrale ne semble pas suffisamment discriminante. eCognition permet aussi de travailler sur les textures de l'image Panchromatique (Figure 28) à l'aide de différents indices d'Haralick (Annexe 5)

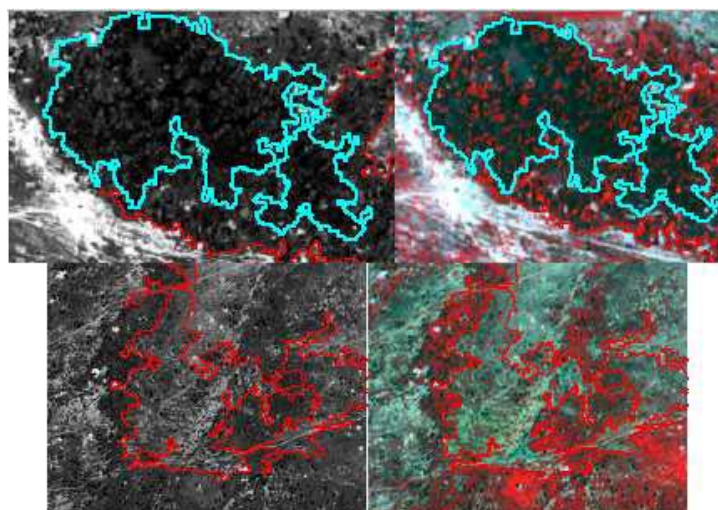


Figure 28 : Exemple de texture sur l'image panchromatique d'un Baldiol (en haut) et d'une savane arborée claire (en bas)

La figure 29, regroupant l'ensemble des signatures spectrales des objets que l'on veut cartographier, permet d'évaluer les confusions prévisibles, que ce soit sur la nature de la savane (arborée/arbustive/herbeuse) que sur sa densité. Suite à cette pré-analyse, les moyennes spectrales et les écarts types des objets dans les canaux 2 (0,52 - 0,60 μm), 3 (0,63 - 0,69 μm) et 4 (0,76 - 0,90 μm), les textures d'Haralick et le NDVI ont été sélectionnés comme critères discriminants pour la classification. Le canal 1 (Bleu) ne semble pas contenir une information suffisamment variée pour les différentes classes. L'eau devra être détectée selon d'autres critères (Cf. 5.4).

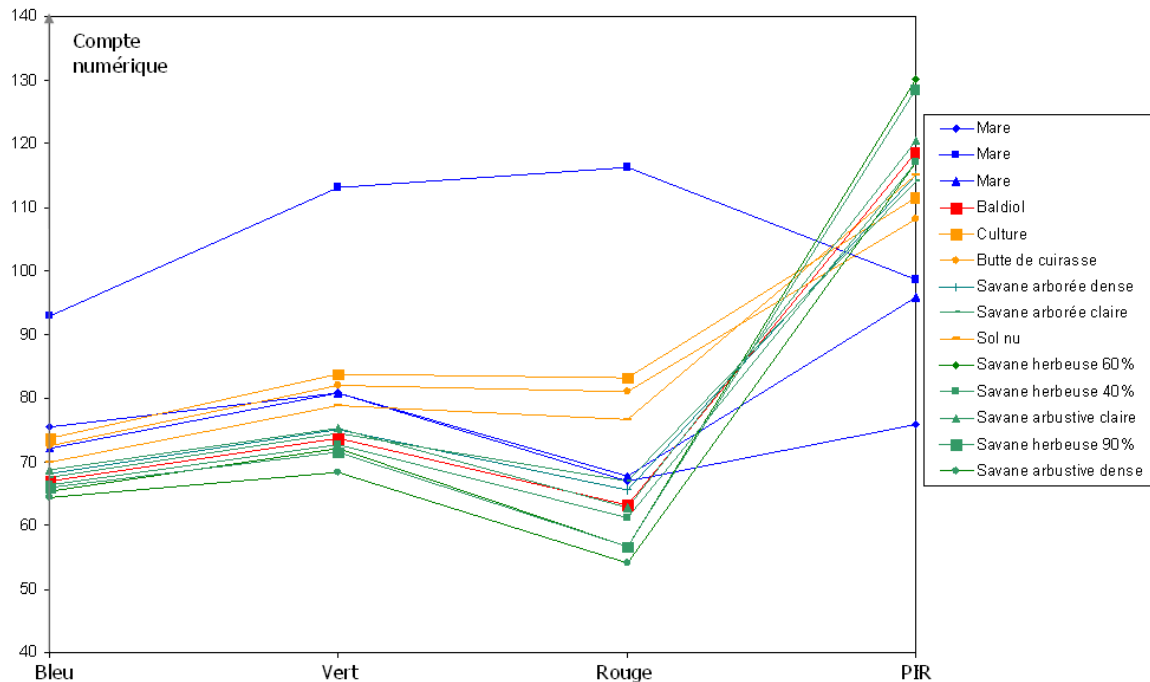


Figure 29 : Graphique récapitulatif des grandes classes d'objets analysés

L'analyse intra et inter objets permet la mise en place de la méthodologie d'extraction des différentes informations, ainsi que de d'identifier les premières pistes pour le choix des règles de classification supervisée.

5.4 Détection des Mares et du bâti

Afin de détecter les Mares de manière précise, on crée un indice utilisant l'information contenue dans les canaux 2 (Vert) et 4 (Proche infra-rouge) de l'image Quickbird :

$$[V - PIR] / [V + PIR]$$

Sous Erdas, grâce au Model maker (Annexe 6), on réalise une image raster de l'indice puis un seuillage (Breakpoint) permettant une meilleure visualisation avant traitement (Figure 30). Il s'agit d'une étape de vérification.

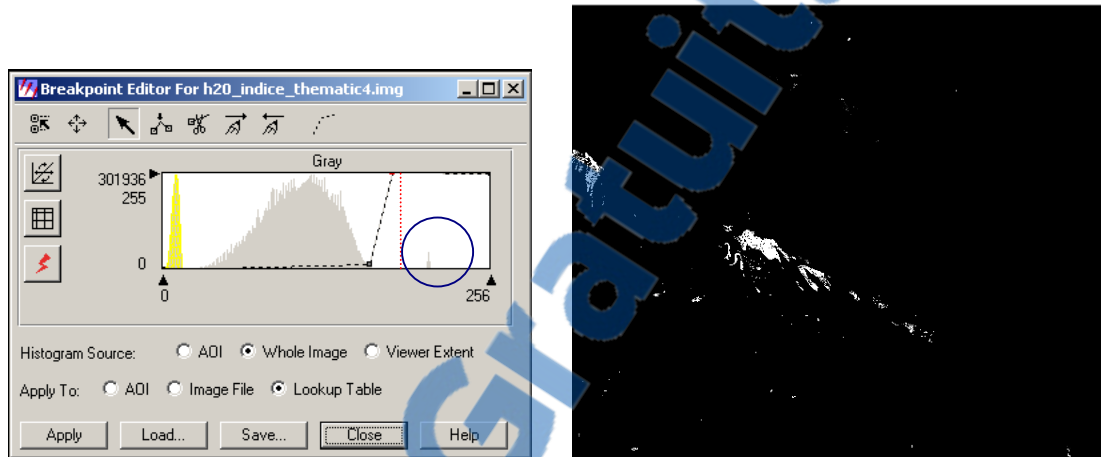


Figure 30 : Capture d'écran Erdas de l'interface Breakpoint et l'image résultat après seuillage

Après vérification des valeurs de l'histogramme correspondant à l'eau (ici les valeurs vers 256), l'image a été codées en binaire (Figure 31) : les valeurs de pixels correspondant à l'eau en 1 et le reste en 0. On obtient ainsi un Raster codé 0 ou 1 qui permet de détecter pixel à pixel l'eau.

Figure 31 shows the 'Recode: h20_index_thematic.img' window. It contains a table with columns 'Row', 'Old Value', 'New Value', and 'Histogram'. The 'New Value' column is currently set to 1 for all rows. Below the table is a 'New Value' input field set to 1 and a 'Change Selected Rows' button. At the bottom are 'Apply', 'Close', and 'Help' buttons.

Row	Old Value	New Value	Histogram
0	0	1	3
1	1	1	7
2	2	1	10
3	3	1	34
4	4	1	28
5	5	1	41
6	6	1	66
7	7	1	167
8	8	1	114
9	9	1	328
10	10	1	403
11	11	1	467
12	12	1	1005
13	13	1	917
14	14	1	1583

Figure 31 : Interface de Recodage du Raster sous Erdas

Le résultat n'est pas parfait, quelques confusions sont observables (Figure 32) entre l'eau et le bâti. Un nettoyage est nécessaire pour améliorer la détection des mares.

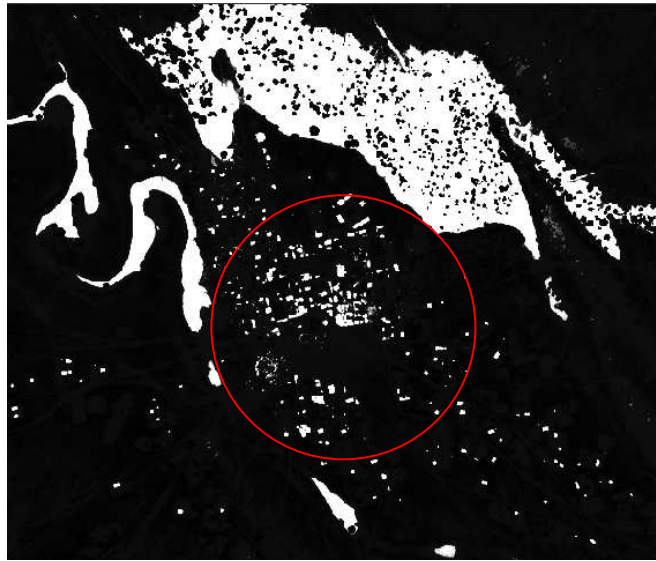


Figure 32 : Exemple de Confusion Bâti / Eau avec l'indice de détection de l'eau

La zone comportant peu d'habitations, un nettoyage manuel sous Erdas est suffisant. Une « Area of Interest » (AOI) doit être créée sur les bâtiments puis l'outil « fill » permet de recoder la zone sélectionnée en 0. Une fois la couche mare réalisée en Raster, il faut effectuer un « Zonal attributes » sous Erdas. Il s'agit de mettre en entrée d'algorithme la couche Raster effectuée ainsi que la couche vectorielle segmentée à 50. L'algorithme fournit en sortie, pour chaque polygone de la couche vectorielle, un résumé de l'information raster. C'est-à-dire que la couche vectorielle va récupérer les données du raster dans un de ces champs. Les « données » peuvent être le minimum, le maximum, la moyenne ou l'écart type de l'objet.

Sous le SIG, une requête de sélection des objets possédant plus de 20% de pixel « eau » permettra de les considérer comme faisant partie de la couche Mare.

Ici, le seuil est relativement bas (20%), pour obtenir une couche couvrant intégralement les Mares ainsi que leurs bordures (Figure 33).

Cette couche Mare peut servir d'AOI sous Erdas. Si la couche de détection de l'eau est recrée et si l'outil « fill » est appliqué sur l'AOI « mare », une nouvelle information sera sélectionnée, celle du bâti.

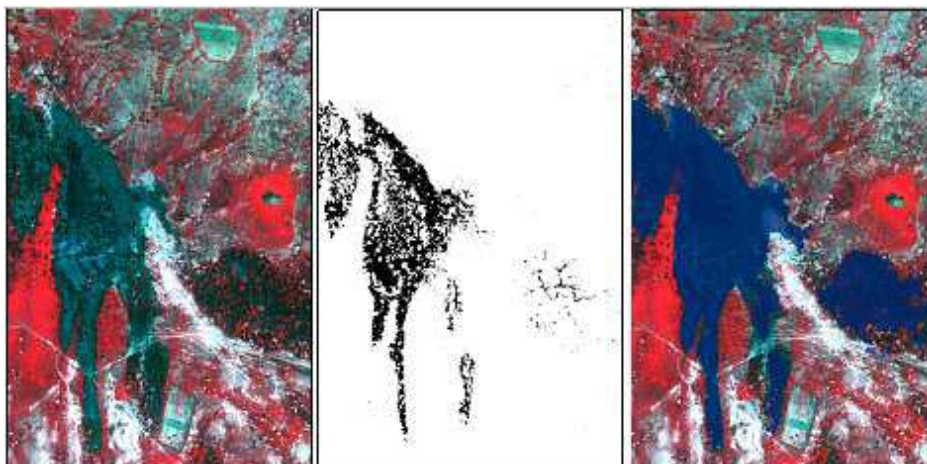


Figure 33 : Etapes de Création d'une couche Mare

A partir de cette couche Mare, un buffer de rayon 100m permettant de sélectionner des polygones dans la couche segmentée à 50 a été calculé. Cette opération a été nécessaire afin de réduire le nombre de polygones à classer à cette échelle sous eCognition, car l'opération demande des temps de calcul très longs (~Plus d'une semaine).

5.5 Carte d'occupation du sol

Pour réaliser l'occupation du sol, une règle de classification sous eCognition a été créée (Figure 34) basée sur les hypothèses issues de l'analyse intra et inter objets (Cf. 5.3).

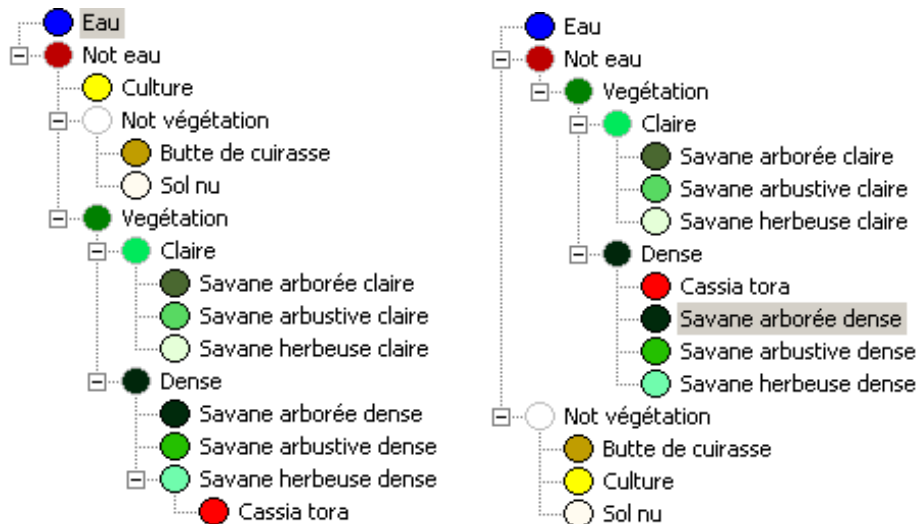


Figure 34 : Hiérarchie de groupe (regroupement par thème) à gauche | Hiérarchie d'héritage (regroupement par propriétés communes) à droite

Au final 11 classes pour la classification ont été obtenues. La classe Eau a été définie par une fonction d'appartenance basée sur l'indice de détection d'eau. Il s'agit d'une détection sous estimée des mares. La classe végétation (Figure 35) quant à elle, est issue d'une fonction d'appartenance liée au NDVI (strictement supérieur à 0.19).

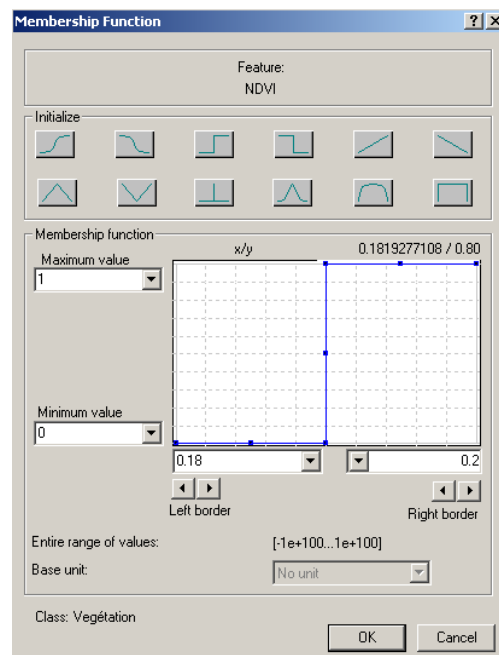


Figure 35 : Fonction d'appartenance de la classe Végétation



« Dense » et « Claire » sont des sous classes de Végétation dans la hiérarchie d'héritage, c'est-à-dire qu'elles héritent des attributs de Végétation (NDVI >0.19). Elles possèdent en plus leurs propres fonctions d'appartenance.

NDVI Claire : 0.19->0.27

NDVI Dense : 0.27->Max

Toutes les classes « filles » de type Savane héritent directement ces valeurs de NDVI.

« Dense » et « Claire » sont aussi des classes filles pour la Hiérarchie de groupe, c'est-à-dire que ces deux classes sont regroupées car elles ont une sémantique commune.

En plus de ces fonctions d'appartenance, les classes de sols nus (Butte de cuirasse, Culture et Sol nu) et de végétation ont été combinés à l'algorithme de classification au « Plus proche voisin » par les attributs de moyenne spectrale, d'écart type sur les canaux 2, 3, 4 et de texture d'Haralick (Dissymétrie, Entropie, Homogénéité dans toutes les directions) sur le canal panchromatique.

Pour réaliser cette classification supervisée au « plus proche voisin », 50 parcelles terrain ont été exploitées comme parcelles d'entraînement.

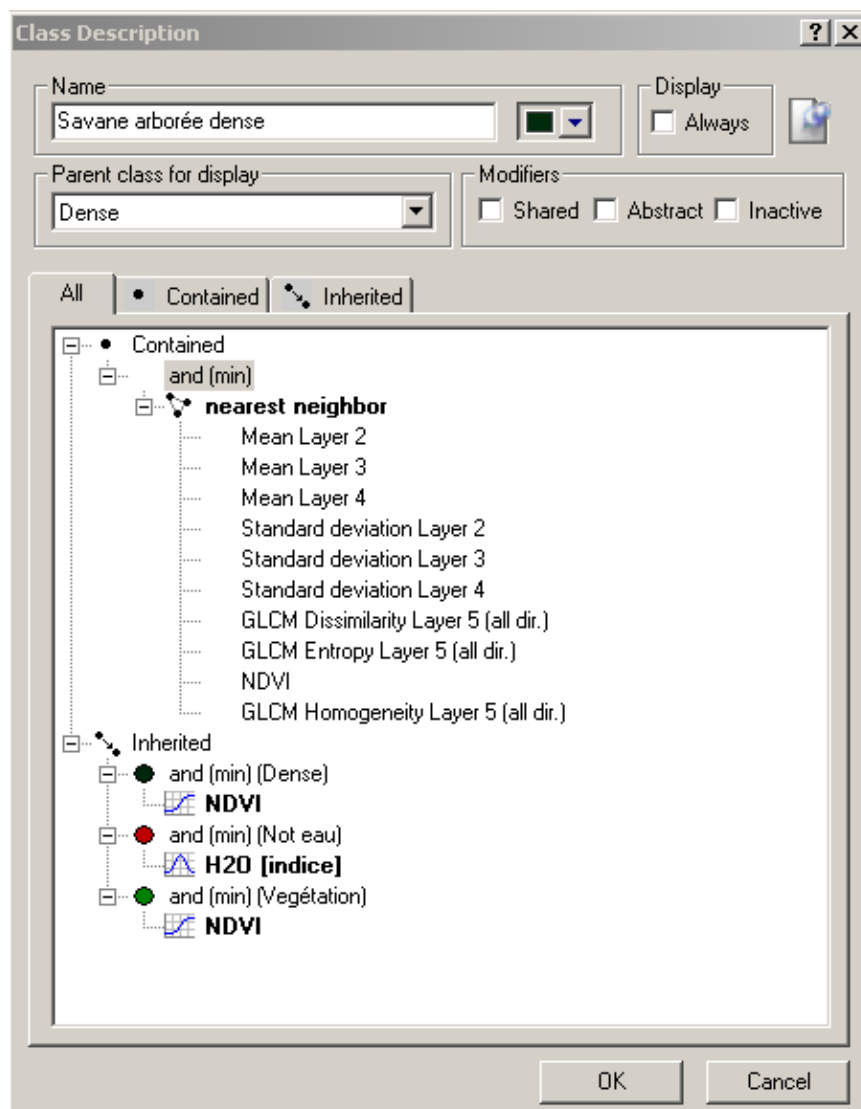


Figure 36 : Exemple de Description de classe

5.6 Carte détaillée de la végétation des mares

La classification effectuée sur la segmentation de paramètre d'échelle 50 a deux buts :

- Fournir une occupation du sol plus détaillée aux abords immédiats de la mare
- Permettre la description de la végétation de surface de la mare

Les mêmes règles de classification que l'occupation du sol général sont utilisées. Néanmoins, pour réaliser une classification à une échelle plus fine centrée autour des mares (Cf. Détection Mare) plusieurs étapes supplémentaires de traitements numériques ont été nécessaires pour ne travailler que sur les zones situées dans un périmètre de 100 m autour des mares (Figure 37).

L'extraction des zones d'intérêt s'est faite avec le Logiciel Erdas Imagine par l'utilisation de l'outil AOI qui permet par codage en 0 d'éliminer les zones exclues de l'étude (Il ne reste plus que la zone de 100m autour des mares).

Sous eCognition, le raster réduit ainsi que la segmentation à 50 sont introduits en entrée. Il suffit de lancer la segmentation en pondérant la « thematic layer » de 1 et les autres couches à 0 pour obtenir une segmentation identique sur la nouvelle image que celle employée pour l'occupation du sol. Ces étapes sont nécessaires pour travailler sur une zone réduite et rester homogène avec la classification de l'occupation du sol.

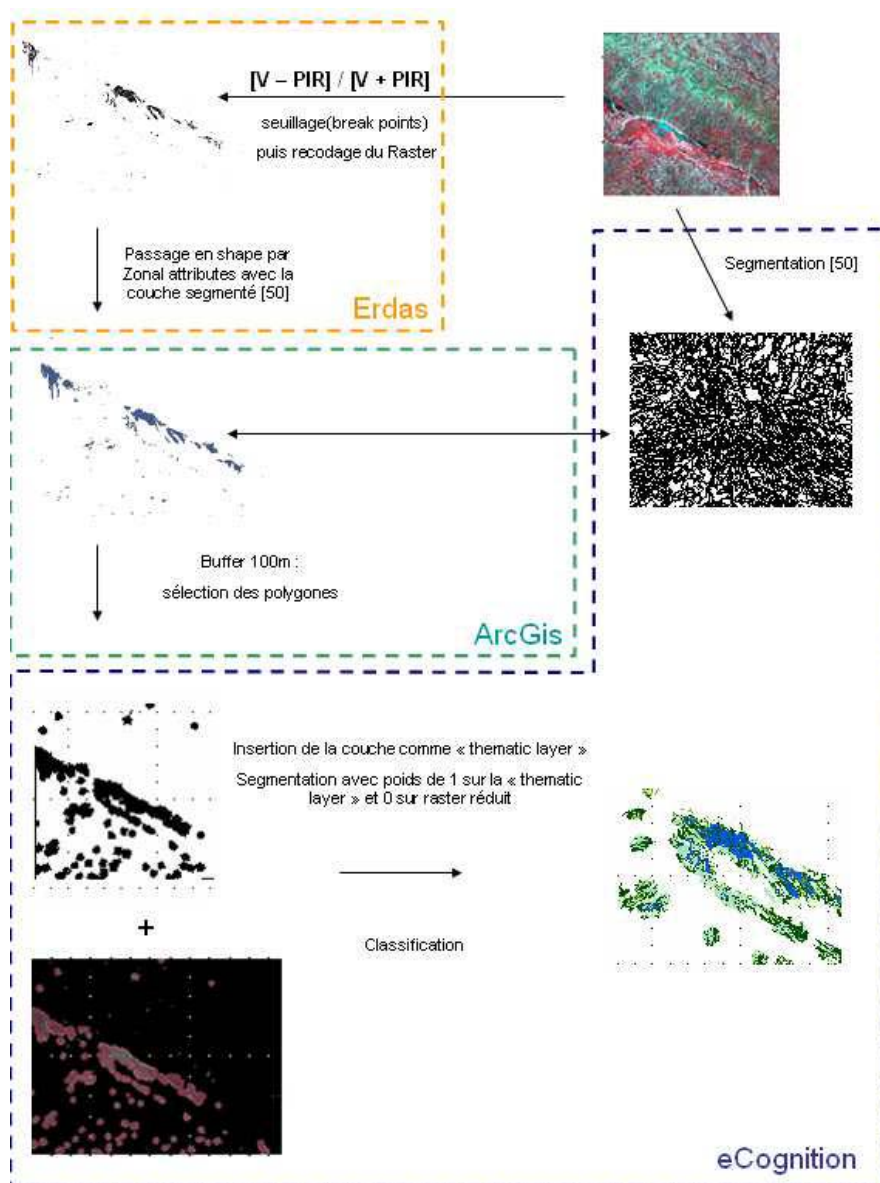


Figure 37 : Méthodologie de la classification à petite échelle

5.7 Occupation du sol à deux échelles de précision

Après avoir effectué les deux classifications (celle à 200 et la sélection à 50 – Figure 38), il est nécessaire de fusionner les deux couches afin d’obtenir une unique couche décrivant à la fois l’occupation du sol et les mares. Sous le SIG, les deux couches sont intersectées afin d’obtenir la couche finale.

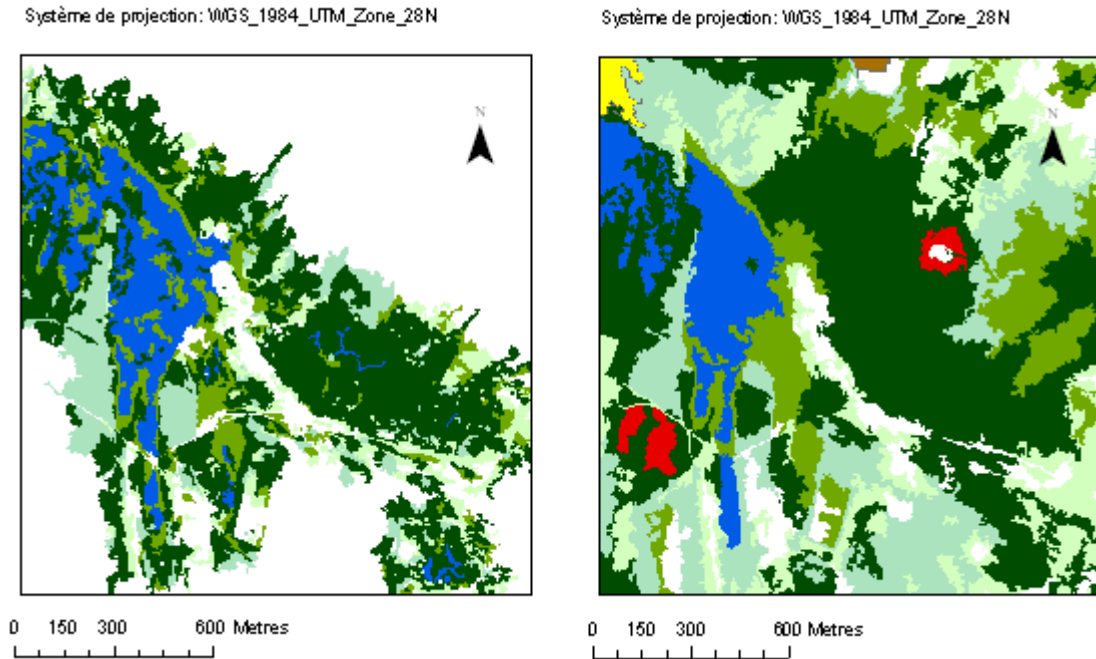


Figure 38 : Zoom sur les deux classifications

5.8 Résultats

60 parcelles « vérité terrain » sont disponibles pour vérifier la qualité de la classification.

	Savane arborée claire	Savane herbeuse dense	Cassia tora	Culture	Sol nu	Savane arbustive claire	Butte de cuirasse	Savane arbustive dense	Savane arborée dense	Savane herbeuse claire
Producer	0,65	1,00	0,44	0,54	0,98	0,50	0,80	0,99	0,89	0,99
User	0,88	0,31	1,00	1,00	0,44	0,99	1,00	1,00	0,87	0,77
Overall Accuracy	0,78									
KIA	0,75									

Tableau 3 : Résultats de la classification

La classe Savane arborée claire présente des confusions avec la classe Sol nu (18%) et Savane arborée dense (10%). La confusion avec le sol nu était prévisible, car le passage à des objets a toujours tendance à masquer les éléments minoritaires. ECognition effectue une moyenne spectrale sur l’objet, la différence entre une surface de sol parfaitement nue et une surface de sol nu avec un nombre réduit d’arbres n’est pas évidente.

La classe Savane herbeuse dense présente un bon résultat de reconnaissance. Les critères de NDVI et de Texture sont pertinents pour ces objets.

La classe Cassia tora semble ne pas fournir un bon taux de reconnaissance. En fait la seule confusion est entre la classe Cassia tora et Savane herbeuse Dense. Le Cassia tora étant un sous ensemble de la classe Savane herbeuse Dense, on peut considérer que le pourcentage de reconnaissance est sous-estimé.

La classe culture a 33% de confusion avec la classe Sol nu et 12% avec la classe Savane arborée claire. L'objectif final étant un travail en terme d'ouverture et de fermeture des milieux, ces confusions sont acceptables.

Il en est de même pour la classe Savane arbustive claire (Savane herbeuse claire=42%, Sol nu=7%). Les autres classes présentent des pourcentages de reconnaissance acceptables. Le pourcentage de reconnaissance global est donc de 78%. Les résultats de la classification peuvent être considérés comme acceptable.

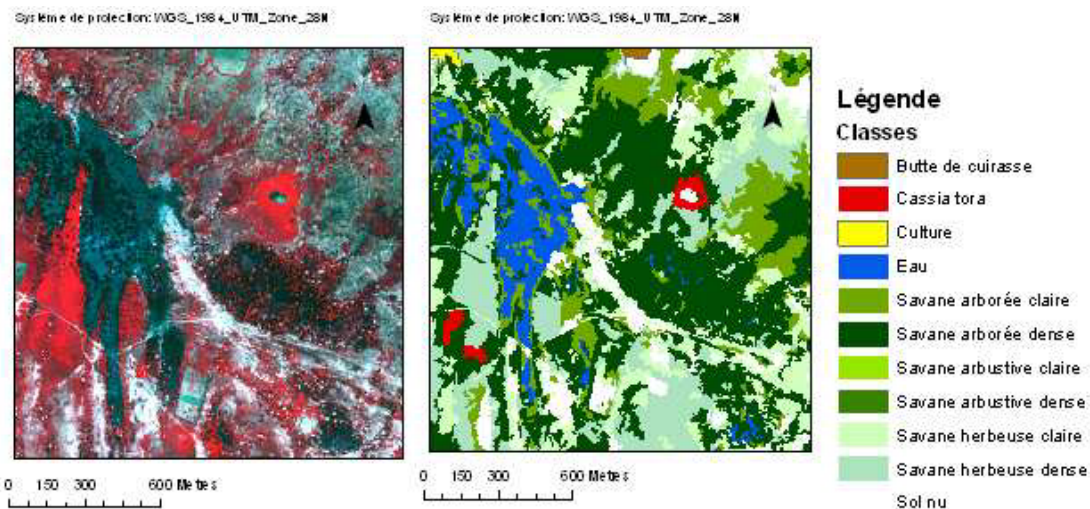


Figure 39 : Zoom sur l'occupation du sol

6. Extraction et analyse statistique des variables paysagères

6.1 Passage en classes de végétation à risque

Grâce à la connaissance terrain et de la biologie du vecteur [8], les 11 classes de la classification vont être traduites en 3 classes sur le thème de l'ouverture/ Fermeture du milieu (Figure 40). En effet, si l'on se concentre sur la vision « épidémiologie », sur le comportement du moustique, les zones propices au déplacement des moustiques seront plutôt les couverts végétaux fermés. Ce sont des milieux ombragés dans lesquels la probabilité de présence de moustique sera plus forte que dans un milieu ouvert type sol nu.

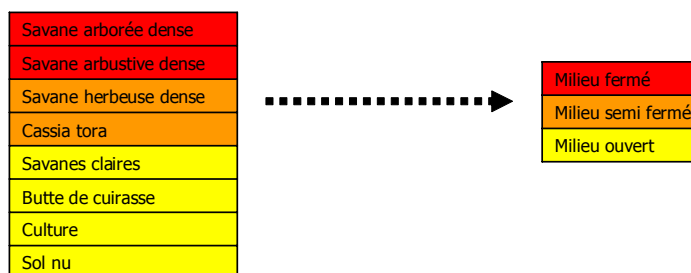


Figure 40 : Passage des classes de végétation en milieu fermé/ouvert

Afin de transformer les classes en indice de risque, on ajoute à la table liée à la couche occupation du sol, un champ. Une requête sous le SIG permet de sélectionner les classes et de remplir automatiquement le champ. On retrouve l'occupation du sol codé comme suit : 1= Milieu fermé ; 2= Milieu semi fermé ; 3= Milieu ouvert (Figure 41).



Figure 41 : Traduction de l'occupation du sol

6.2 Création des indices

Après cette synthèse de l'information contenue dans l'occupation du sol, on cherche des indices pour caractériser les mares. Trois indices ont été définis préalablement : l'indice de fermeture, l'indice du taux de couverture végétale dans la mare, et l'indice de densité des mares qui sont présentés ci-dessous.

- L'indice de fermeture :

Il a pour but de caractériser l'environnement immédiat autour des mares (Figure 42). Pour cela une mesure de la fermeture du milieu à trois différentes échelles (100m / 500m / 1km) a été faite à l'aide de buffer.

Indice de Fermeture = % Aire de Végétation fermée / [% Aire de Végétation semi fermé & ouverte]

Une intersection a été effectuée entre les buffers et la couche occupation du sol.

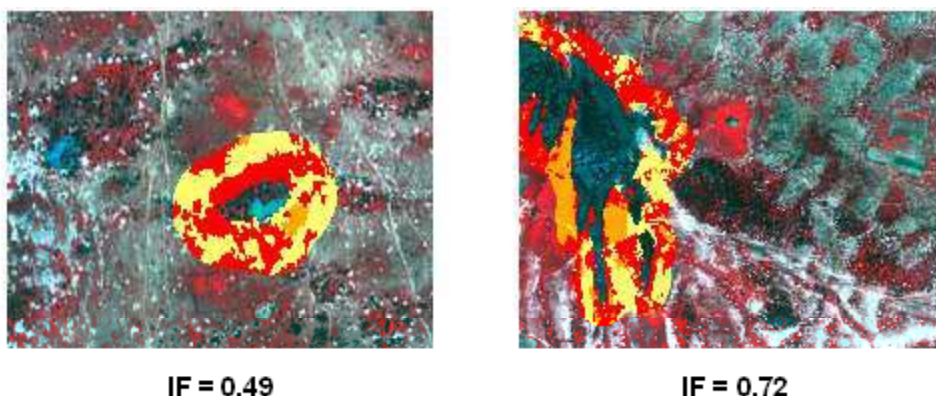


Figure 42 : Exemple d'Indice de Fermeture

- L'indice de couverture de la mare :

Un autre objectif est d'analyser la végétation couvrante de la mare (Figure 43). L'hypothèse liée à cet indice est que les mares dont la couverture végétale est plus importante, auront plus de probabilités d'être des habitats favorables aux moustiques.

Indice de couverture de la mare = Aire des classes "Végétation" / Aire totale de la Mare

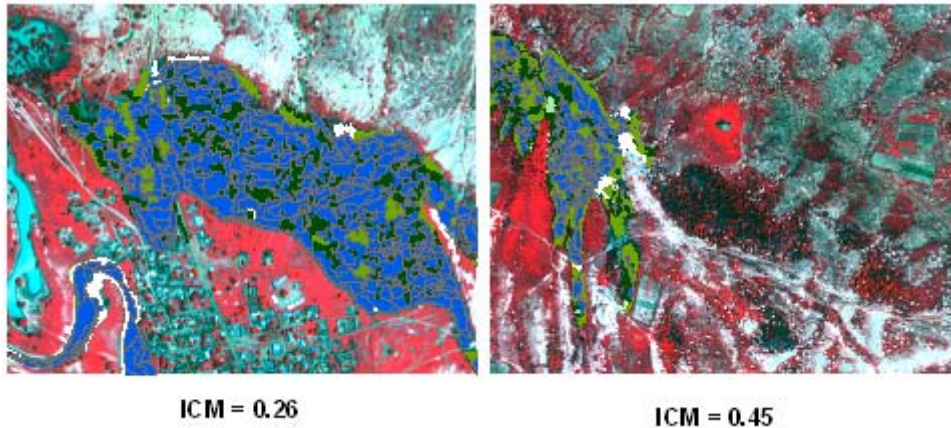


Figure 43 : Exemple d'indice de couverture de mare

- L'indice de densité des mares :

Le dernier Indice est basé sur le buffer de 1 km, il consiste à compter le nombre de mare autour de la mare étudiée. Chevalier et al. 2003 [6] ont mis en évidence un effet « petite mare » influençant la prévalence chez les ovins. C'est-à-dire, plus la mare est petite plus la prévalence est importante. Cet indice prend donc en compte la taille de chaque mare.

Indice de nombre de mares = [1/Surface de la mare 1 + + 1/ Surface de la mare i]

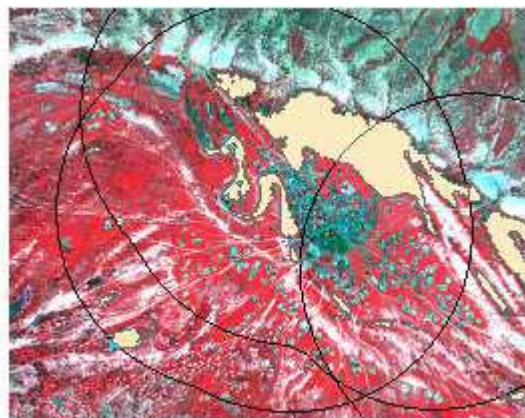


Figure 44 : Exemple d'indice du nombre de mares

Pour obtenir l'aire totale de la mare et pas l'aire intersectée par le buffer, il est nécessaire d'effectuer le calcul d'aire avant l'intersection.

6.3 Analyse statistique – Relation Incidence FVR / variables paysagères

Cinq facteurs et leurs interactions ont donc été testées dans la partie fixe du modèle : 1) L'indice de végétation 2) L'indice de couverture des Mares 3) L'indice de nombre de mares. On ajoute aussi les deux effets trouvés par *Chevalier et al. 2003* [6] qui sont la localisation de la mare (à l'intérieur ou à l'extérieur de l'ancien lit de la rivière Ferlo) et la taille de la mare exprimée en hectares (ha) et centrée sur la surface de la plus petite mare.

Les variables explicatives sont détaillées dans le tableau 4.

Mare	Campement	INM	ICM	IF 100m	IF 500m	IF 1000m	S	Taille	M	Denom	I
Barkedji	BK1	0,004610	0,26	0,43	0,38	0,44	33,62	20	0	20	0
	BK2	0,004610	0,26	0,43	0,38	0,44	33,62	50	28	36	0
	BK3	0,004610	0,26	0,43	0,38	0,44	33,62	30	15	22.5	0
Furdu	FUR	0,009296	0,72	0,49	0,54	0,43	1,00	40	1	39.5	1
Kangaledji	KAN	0,009067	0,52	0,76	0,76	0,68	9,81	86	34	69	14
Niaka	BEL	0,001068	0,45	0,72	0,67	0,54	34,74	17	1	16.5	0
	NIA	0,001068	0,45	0,72	0,67	0,54	34,74	37	10	32	4
	NIK	0,001068	0,45	0,72	0,67	0,54	34,74	20	18	11	1

INM : Indice du Nombre de Mare
 ICM : Indice de Couverture de mare
 IF : Indice de Fermeture
 S : Surface
 Taille : Taille du troupeau d'ovins
 M : Nombre d'animaux manquant
 Denom : Nombre d'animaux restants
 I : Incidence

Tableau 4 : Récapitulatif des variables explicatives (Rouge : variables explicatives | Bleu : autres informations)

Le travail de *Chevalier et al* permet de connaître à priori un modèle plausible, celui liant la surface de la mare et la position dans/dehors du lit de la rivière.

On va donc tester tous les modèles possibles avec ces 5 effets principaux et leurs interactions. Pour éviter le problème de la multiple comparaison de modèles, le critère d'information Bayésien est utilisé pour sélectionner le meilleur modèle.

$$BIC = -2 \log (ML) + k \times \log (n),$$

Les paramètres de l'équation sont : ML le maximum de vraisemblance, k le nombre de paramètres dans le modèle et n le nombre d'observations. Pour ce critère d'information, le modèle le plus adapté est celui ayant la valeur la plus faible. Un modèle témoin est calculé pour fournir la valeur de BIC seuil, au-delà duquel les modèles sont significatifs.

Le software R a été utilisé pour l'analyse des données [(C) R Foundation, <http://www.r-project.org>].

6.4 Résultats

Tous les modèles n'ont pu être testés car la variance estimée du facteur aléatoire est nulle. Ceci peut être dû à la réduction du nombre de données ou alors à un choix du type de modèle qui n'est plus pertinent avec les nouveaux facteurs.

On teste alors le jeu de données avec un Modèle linéaire général plus simple. Mais là encore, l'analyse statistique n'est pas possible à cause d'un problème de sur dispersion des données. En effet, il faut plus de données pour pouvoir arriver à une conclusion. Un résultat est exploitable qui est celui présenté à la figure 45.

	BIC
Indice de Fermeture 1000m	13.02080
1	19.61590
Surface	19.82849
Indice de couverture	21.08509
Dedans/Dehors Ferlo	21.38484
Indice du nombre de mares	21.59545

~ Donnée d'incidence =

Figure 45 : Résultat de l'analyse statistique

Le 1 indique le modèle témoin servant à déterminer la valeur du BIC significative.

Le modèle mettant en relation l'indice de Fermeture à 1000m et l'incidence semble donc le plus significatif (les IF à 100 et 500m n'ont pu être testés pour cause de variance estimée nulle du facteur aléatoire).

En analysant les résumés statistiques fournis par R (Figure 46), le modèle avec l'indice de Fermeture est pertinent ($p < 0.05$).

```
> summary(fm06)
Generalized linear mixed model fit using Laplace
Formula: Incidence/denom ~ IF_1000m_1 + (1 | Pond)
Data: Data2
Family: binomial(logit link)
AIC   BIC logLik deviance
12.78 13.02 -3.391   6.782
Random effects:
Groups Name          Variance Std.Dev.
Pond      (Intercept) 0.036190 0.19024
number of obs: 8, groups: Pond, 4

Estimated scale (compare to 1 ) 0.7118505

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.5944    4.3701  -2.882  0.00395 **
IF_1000m_1   0.2876    0.1174   2.450  0.01428 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr)
IF_1000m_1 -0.998
```

Figure 46 : Résumé statistique de R pour le modèle avec l'IF 100m

7. Discussion

Concernant la carte de végétation, la matrice de confusion indique une classification satisfaisante. De plus, les confusions s'effectuent entre les classes d'occupation du sol présentant un certain degré d'ouverture. Par exemple, la confusion entre savane arborée claire et sol nu est acceptable car ces deux unités seront par la suite agrégées dans la classe milieu ouvert.

La très haute résolution spatiale couplée à un langage orienté objet permet donc de traiter et de manipuler une information plus complète, pouvant apporter une nouvelle source de données pour comprendre les interactions moustiques / environnement.

Notamment, c'est cette riche information de l'image Quickbird qui a permis de créer les trois indices. La création de ces indices n'est qu'une des étapes pour caractériser la végétation à risque. Il reste à confirmer une relation variable paysagère et incidence sérologique.

La richesse de l'information nécessite des traitements plus longs. Sous la configuration suivante : Pentium 4, CPU 2.8GHz et 1 giga de ram ; une classification sous eCognition prend 1 nuit de calcul avec la classification à un paramètre d'échelle de 200. A l'échelle la plus fine, nous avons obtenu seulement 1/3 de l'image au bout de 5 jours de calcul.

Cette contrainte du logiciel a beaucoup limité les expérimentations visant à déterminer l'apport de l'information très haute résolution spatiale.

L'analyse statistique a été limitée par le manque de données sérologiques spatialisées disponibles sur la zone couverte par l'image Quickbird. Néanmoins, elle met en évidence une potentielle relation entre la fermeture du milieu et l'incidence sérologique qui reste donc à confirmer.

Plusieurs étapes statistiques seront à effectuer pour valider le modèle. La première sera de vérifier que la pente du modèle varie bien dans le sens attendu : plus le milieu est fermé, plus l'incidence sera forte. Cette relation traduirait bien l'influence d'un type de végétation à risque (les savanes arborées et arbustives denses) sur l'incidence.

La deuxième étape pourra être de valider avec un autre jeu de données sérologiques plus récent cette relation.

Il sera aussi nécessaire de comparer les profils temporels de précipitations 2003/2005 (données sérologiques / date d'acquisition de l'image) pour vérifier s'il y a similitude. Un profil totalement différent sur les deux années (exemple : majorité des précipitations décalée) devra faire relativiser le résultat.

Le travail sur le jeu de donnée sérologique a été un choix par défaut car les données de densité de moustiques collectées par l'institut Pasteur de Dakar n'étaient pas encore disponibles.

Il semble donc important d'acquérir ce jeu de données réparti de façon homogène sur l'ensemble de la zone d'étude pour tester une relation entre les unités écologiques à risque de transmission de la FVR et la présence ou la probabilité de présence des moustiques.

La nature des données d'incidence sérologique peut être discutée. En effet, lier la végétation à des incidences est plus délicate à interpréter, car le processus de contamination des animaux et des hommes met en jeu d'autres paramètres comme les densités du bétail. Les animaux étant la nuit dans des parcs à proximité des mares et les vecteurs ayant un déplacement nocturne, l'incidence reste tout de même un indicateur robuste de l'état sanitaire de la mare.

La méthodologie visant à étudier l'impact de variables environnementales sur la présence / le déplacement de moustiques est mise en place (Figure 47) et reste à être testé dès que les données de densité des moustiques seront disponibles ou encore sur un autre jeu de données sérologiques plus récent pour confirmer ces premiers résultats statistiques.

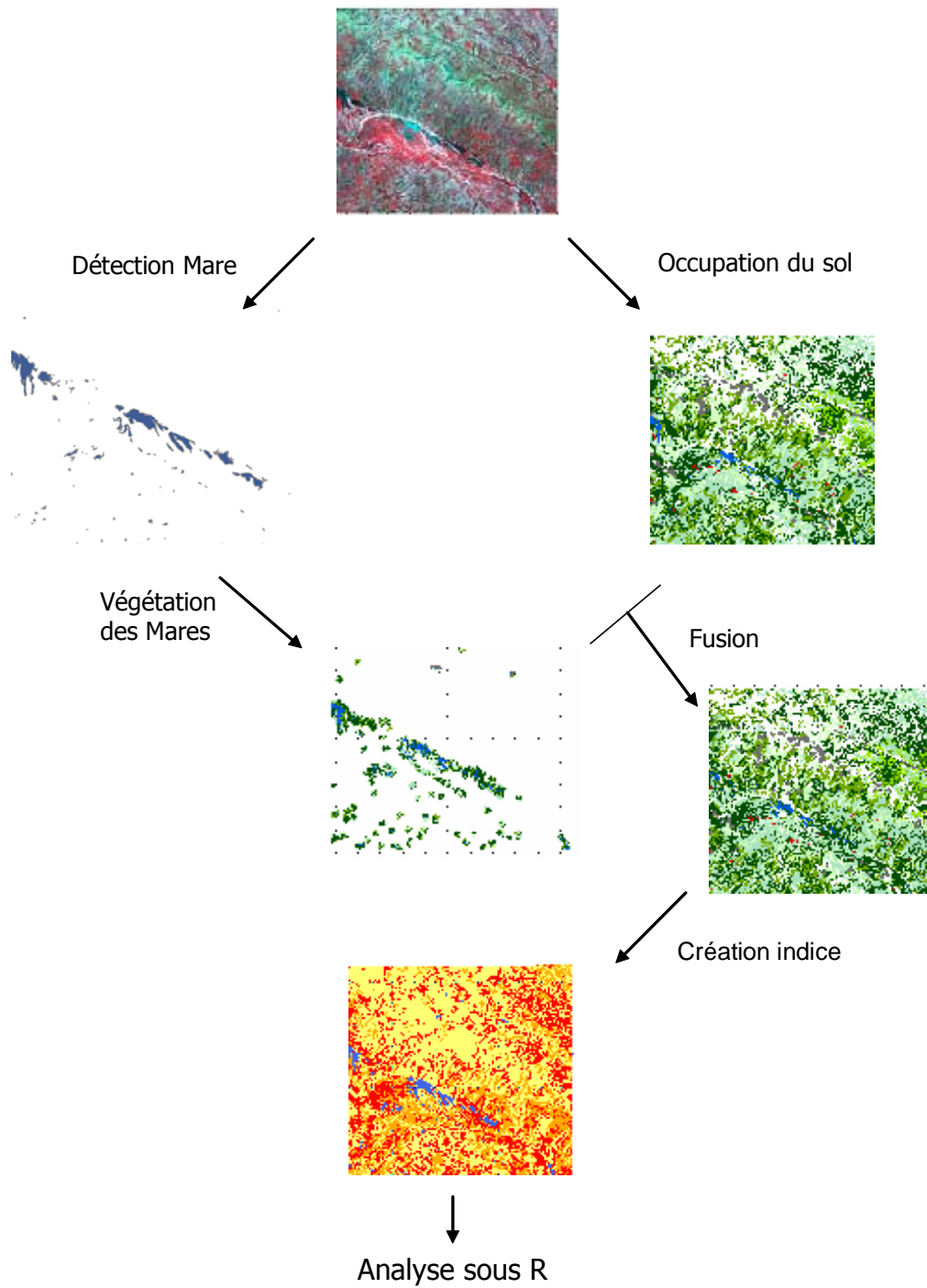


Figure 47 : Bilan - Méthodologie Générale

Bibliographie

- [1] **Jupp P.G, Kemp A, Grobbelaar A, Leman P, Burt F.J, Alahmed A.M, Mujalli D.AL , Al Khamees M and Swanepoel R, 2002.** The 2000 epidemic of Rift Valley fever in Saudi Arabia: mosquito vector studies. *Medical and Veterinary Entomology* 16, 245-252.
- [2] **Thonnon J, Picquet M, Thiongane Y, Lo M, Sylla R and Vercruyse J, 1999.** Rift valley fever surveillance in the lower Senegal river basin: update 10 years after the epidemic. *Tropical Medicine and International Health* volume 4 No8, 580-585.
- [3] **Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ, 1999.** Climate and satellite indicators to forecast RVF epidemics in Kenya. *Science* 285: 397-400
- [4] **Pin Diop R, 2005.** Spatialisation du risque de transmission de la Fièvre de la Vallée du Rift en milieu agro-pastoral sahélien du Sénégal septentrional. Rhèse de doctorat, Université d'Orléans.
- [5] **TRAN A , 2004.** Télédétection et Épidémiologie : Modélisation de la dynamique de populations d'insectes et application au contrôle de maladies à transmission vectorielle. Thèses de doctorat, Université Louis Pasteur.
- [6] **Chevalier V, Mondet B, Diaite A, Lancelot R, Fall A.G and Ponçon N, 2004.** Exposure of sheep to mosquito bites: possible consequence for the transmission risk of Rift Valley Fever in Senegal. *Medical and Veterinary Entomology* 18, 247-255.
- [7] **França L.Riou, 2002.** Modélisation de la dynamique d'arthropodes vecteurs de la Fièvre de la vallée du Rift en fonction de processus environnementaux dans le Nord du Sénégal.
- [8] **Ba Yamar, DIALLO Diawo, MOUHAMED FADEL KEBE Cheikh, DIA Ibrahima and DIALLO Mawlouth, 2005.** Aspects of bioecology of two Rift Valley Fever Virus vectors in Senegal (West Africa): *Aedes vexans* and *Culex poicilipes* (Diptera: Culicidae). *J. med. Entomol.* vol. 42, n°5, pp. 739-750
- [a] WHO Media centre « OMS | Fièvre de la vallée du rift » [En ligne]. <http://www.who.int> (Page consultée le 15/01/2007)
- [b] The European Commission community Research "EDEN (Emerging Diseases in a changing European eNvironment)" [En ligne]. <http://www.eden-fp6project.net/> (Page consultée le 16/01/2007)
- [c] The Statistics Department of the University of Auckland "The R project for statistical computing" [En ligne]. <http://www.r-project.org/> (Page consulté le 17/01/2007)

Annexes

Annexe 1 : Caractéristiques de Quickbird

QUICKBIRD

Sensor-system	Spectral resolution (μm)	Spatial resolution (m)	Scan-width (km)	Revisit period	Orbit	Operation periode
	Panchromatic: 0,45 - 0,90	0.7	16.5	1 to 3.5 days	450 km, near polar, sun- synchronous	18/10/2001 -
	channel 1: b 0,45 - 0,52	2.8				
	channel 2: v 0,52 - 0,60					
	channel 3: r 0,63 - 0,69					
	channel 4: Nir 0,76 - 0,90					

Annexe 2 : Fiche « Parcelle homogène »

Nom :

Relevés de Végétation

Zone :

Date :

Heure :

N° Photo	Orientation	N°

Points GPS		
N°	X (m) Long	Y (m) Lat

Description graphique

Topographie

Dune	
Inter dune	
Terrain plat	
Bas-fond	
Haut de pente	
Replat	
Bas de pente	
mi-pente	
Butte de cuirasse	
autres	

Herbacée	
Arbustive	
Arborée	
Boisée	
Claire	
Dense	

Type de formation :

1 ère espèce arbustive dominante.....

2eme espèce arbustive dominante.....

Formation ligneuse : Ligneux hauts (>2m) Ligneux bas (<2m)

Annexe 3 : Fiche « Transect »

Nom :

Transect n°.....

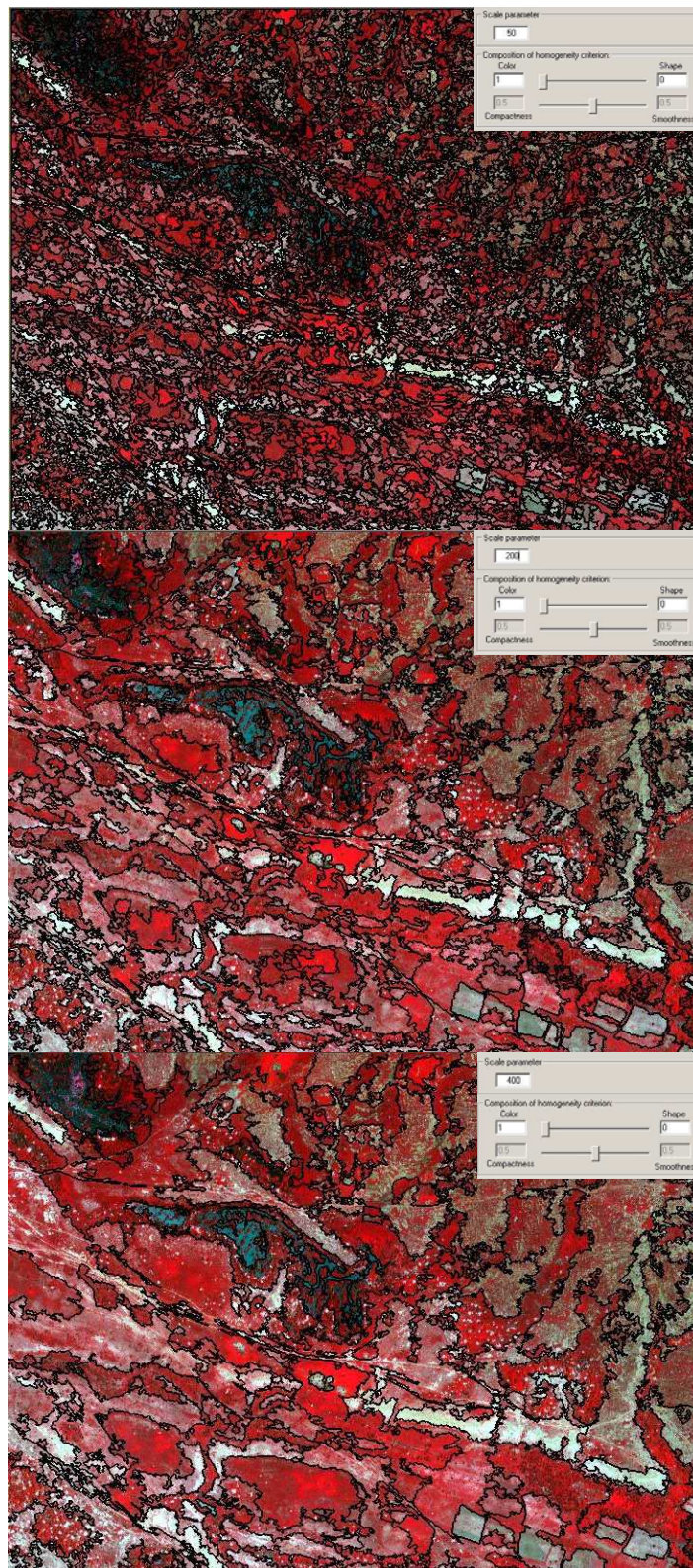
Zone : Date : Heure : Nom de la mare :

Points GPS			N° Photo	Description	H	A u	A o	B	C
N°	X Lon	Y Lat							

Eau turbide	Mare temporaire	Végétation couvrante	
Eau claire	Mare permanente		

H : Herbacée B : Bois
 Au : Arbustive C : Clair
 Ao : Arborée D : Dent

Annexe 4 : Résultats de la variation du paramètre d'échelle



Résultats de segmentation à différents paramètres d'échelle

Annexe 5 : Texture d'Haralick

Les statistiques sont les suivantes :

• **Second moment angulaire (ASM)**

mesure l'homogénéité ; une forte valeur caractérise une texture homogène.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}^2$$

• **Contraste (CON)**

un forte valeur traduit de grandes variations locales (frontières).

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 * P_{ij}$$

• **Corrélation (COR)**

coefficient de corrélation de la distribution des tons de gris.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu)*(j-\mu) * P_{ij} / \sigma^2$$

• **Variance (VAR)**

mesure l'étalement des valeurs de la matrice normalisée autour de la diagonale principale. Une forte valeur caractérise une texture fine (poivre et sel).

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu)^2 * P_{ij}$$

• **Covariance (COV)**

Proche de la variance. Une forte valeur caractérise une texture fine.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu)_x * (j-\mu)_y * P_{ij}$$

• **Probabilité maximum (MAX)**

$$\max_{ij} P_{ij}$$

• **Moment de la différence inverse (IDM)**

Une forte valeur indique que les éléments texturaux sont de grande taille.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} / (1 + (i-j))^2$$

• **Entropie (ENT)**

mesure le degré d'organisation de la texture (varie de façon opposée au second moment angulaire).

$$-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * \log P_{ij}$$

• **Dissimilarité (DIS)**

mesure l'homogénéité ; une faible valeur caractérise une texture homogène.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * |i-j|$$

• **Homogénéité (HOM)**

mesure l'homogénéité ; une forte valeur caractérise une texture homogène.

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} / (1 + |i-j|)$$

avec

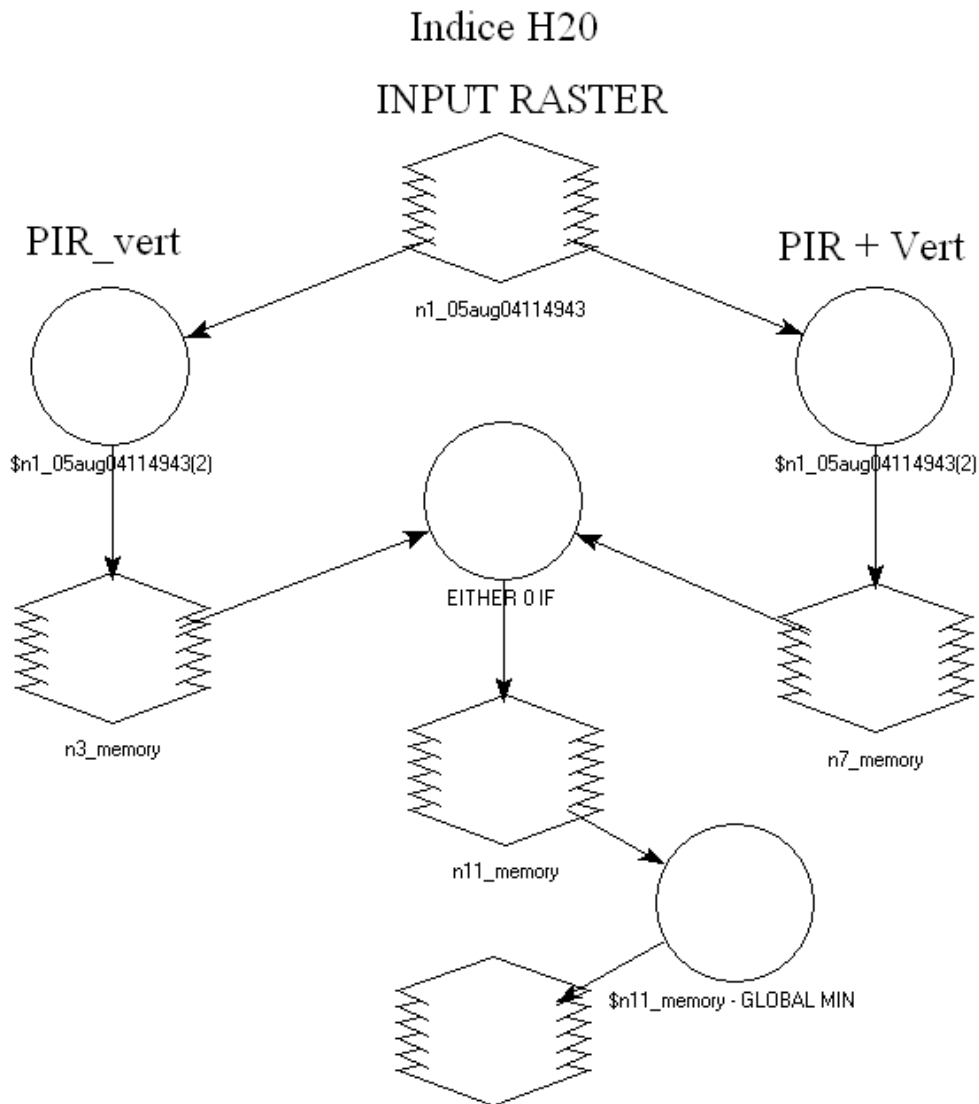
P_{ij} : entrée de la matrice de cooccurrence

N_g : niveaux de gris de l'image

$$\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * i$$

$$\sigma^2 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * (i-\mu)^2$$

Annexe 6 : Modèle Erdas de l'indice de l'eau



Details de \$n11_memory-Global min :

$$\frac{(\$n11_memory - GLOBAL MIN (\$n11_memory))}{(GLOBAL MAX (\$n11_memory) - GLOBAL MIN (\$n11_memory))} * 255$$