

Université de Toulouse

MASTER 2 PROFESSIONNEL GEOMATIQUE

Parcours Professionnel :

« **S**cience de l'**I**nformation **G**éoréférencée pour la **M**îtrise de
l'environnement et l'**A**ménagement des territoires » (**SIGMA**)

RAPPORT DE STAGE

Application de la télédétection à moyenne résolution
spatiale et des Systèmes d'Information Géographique
à la définition d'indicateurs environnementaux relatifs
au risque de grippe aviaire au Vietnam



TOLLIS Sébastien



Septembre 2009

FICHE DE STAGE 2009

Master SIGMA	
NOM ETUDIANT :	TOLLIS
PRENOM ETUDIANT :	Sébastien
Tél. portable	06 12 25 32 07
E. mail	seb.tollis@gmail.com
TITRE DU STAGE :	Application de la télédétection à moyenne résolution spatiale et des Systèmes d'Information Géographique à la définition d'indicateurs environnementaux relatifs au risque de grippe aviaire au Vietnam
DATE DE DEBUT DE STAGE :	15/03/2009
DATE DE FIN DE STAGE :	11/09/2009
NOM TUTEUR-ENSEIGNANT :	Sheeren David
Université	INP ENSAT
SOCIETE :	CIRAD- Maison de la télédétection
Adresse	500 rue Jean-François Breton
Code postal	34093
Ville	Montpellier
Pays	France
Site Web	http://www.teledetection.fr/
NOM DU MAITRE DE STAGE :	TRAN Annelise – DESVAUX Stéphanie
Fonction	Chercheur
Mel	Annelise.tran@cirad.fr stephanie.desvaux@cirad.fr
Téléphone	04 67 54 87 36
Fax	
MOTS-CLES :	Télédétection, SIG, Séries Temporelles, Analyse spatiale, Vietnam, H5N1, Riz, Eau
RAPPORT CONFIDENTIEL	NON
OBJET (description de la mission confiée au stagiaire) :	<p>Intégré au projet international GRIPAVI, ce stage fait partie des nombreuses études symboles de la prise de conscience du risque H5N1 et notamment, son évolution vers une pandémie et une mutation. Détecter des facteurs environnementaux d'émergence de et de propagation du virus semble importants.</p> <p>Par le prisme de la géomatique, cette étude tente de définir ces indicateurs notamment les durées d'inondation et l'intensité des récoltes. La télédétection par série temporelle MODIS à basse résolution permettra de détecter ces indicateurs et l'analyse spatiale à l'aide d'un SIG, d'en déterminer l'importance.</p> <p>Le traitement statistique spatial viendra en appui des résultats pour valider ou inféoder les hypothèses voire les affiner.</p>

A maman...

Remerciements

Je souhaiterai remercier ceux qui ont contribué, de près comme de loin, voire très loin, à l'achèvement de cette année, un peu spéciale...

Ma maman, en premier, qui m'aura toujours beaucoup apporté et dont les conseils me manqueront. Elle n'aura malheureusement pas la chance de me voir rentrer dans cette vie active qui la préoccupait tant. Je t'aime, tu me manques...

Ma femme et mon fils qui ont souffert autant que moi de la séparation et dont le soutien sans faille m'a fait avancer. Je suis passé à côté de pleins de moment mais c'est du passé maintenant.

Ma famille, et notamment mon père, ma grand-mère et mon frère, avec qui nous avons traversé de durs moments : il faut avancer maintenant, une vie reste à faire.

Mes encadrant(e)s également avec qui j'ai noué des relations plus fortes que de simple stagiaire/tuteur. Merci Stéphanie pour ton accueil au Vietnam, ton aide, tes conseils, ta disponibilité et tous les bons moments partagés. Remercie également Olivier et Baptiste en espérant que nous garderons contact dans le futur. Merci Anne Lise pour ta patience algorithmique à toute épreuve, ton enseignement et ta gentillesse de tous les instants. Enfin, merci à David Sheeren d'avoir accepté de m'encadrer, sa disponibilité et son enseignement à l'ENSAT.

Les équipes vietnamiennes du CIRAD et de CARGIS pour leur aide dans mon adaptation au pays et les échanges nombreux. Je pense notamment à M. Pham Van Cu, Hai et Dong.

Certains enseignants du master SIGMA pour leur investissement et la qualité de leurs cours : M Lacombe, S. Le Corre, P. Bergougnoux

La machine à café de la MtD, pour son soutien dans les grandes nuits de rédaction. Accessoirement, les personnes qui sont toujours autour : Jean, Jennifer, Tristan, Robert Melon, Gabi, Joseph

Aux irréductibles stagiaires de la MtD, Maxou, Thibo, John, Elodie pour les franches rigolades, le soutien, les sorties et l'ensemble des bons moments passés ensemble.

Aux amis rencontrés au Vietnam notamment JP, Arthur, Lucas, Antoine et Jean pour les soirées Ruau et les week ends escalades à la baie d'Halong. Que l'esprit de Mickael et de la Wave 110 soit avec vous....

Table des matières

INTRODUCTION	6
I. CONTEXTE, OBJECTIFS ET ORGANISATION DU STAGE	8
A. CONTEXTE SCIENTIFIQUE	8
1. <i>Le CIRAD</i>	8
2. <i>La Maison de la Télédétection (MtD)</i>	8
3. <i>L'UMR TETIS et UPR AGIRs</i>	9
4. <i>Le projet GRIPAVI</i>	10
B. CONTEXTE DE L'ETUDE	11
1. <i>Zone d'étude</i>	11
2. <i>Les liens Grippe Aviaire et environnement</i>	14
C. OBJECTIFS DU STAGE ET MOYENS MIS EN OEUVRE	16
1. <i>Objectifs</i>	16
2. <i>Moyens à disposition</i>	17
3. <i>Méthodologie générale et gestion du projet</i>	20
II. LA TELEDETECTION EN SERIES TEMPORELLES DANS LA DETECTION DES CHANGEMENTS D'OCCUPATION DES SOLS AU NORD VIETNAM	24
A. LES IMAGES SATELLITES	24
1. <i>Le système MODIS/TERRA</i>	24
2. <i>Les synthèses temporelles</i>	25
3. <i>Méthodologie</i>	27
B. ACQUISITION, PREPARATION ET TRAITEMENT DES IMAGES SATELLITES	34
1. <i>Nettoyage des images nuageuses</i>	34
2. <i>Mise en forme des images</i>	37
3. <i>Traitements des données</i>	42
C. VERIFICATION TERRAIN ET RESULTATS FINALS	44
1. <i>Zones de recherche et procédure</i>	44
2. <i>Résultats et classification SPOT</i>	47
3. <i>Correction et classification finale</i>	50
III. L'APPORT DE L'ANALYSE SPATIALE PAR SIG POUR LA DEFINITION D'INDICATEURS ENVIRONNEMENTAUX LIES AU RISQUES DE H5N1 AU VIETNAM	55
A. ACQUISITION ET ORGANISATION DE LA BASE DE DONNEES	55
1. <i>Description des données spatiales disponibles</i>	55
2. <i>Sélection des couches</i>	57
3. <i>Nettoyage des données</i>	58
4. <i>Jointure et architecture du SIG</i>	60
B. ETUDES SPATIALES : PROBLEMES ET SOLUTIONS.....	62
1. <i>Méthodologie d'application</i>	62
2. <i>Recherche d'outils et d'extensions d'Arcgis adaptées</i>	67
C. RESULTATS ET PERSPECTIVES	68
CONCLUSION ET BILAN PERSONNEL	74

Introduction

« *Le risque, objet social, se définit comme la perception du danger, de la catastrophe possible* ». Parce qu'ils ont une expression spatiale, les risques intéressent le géographe, et c'est précisément en l'analysant par le prisme de l'espace et de la représentation cartographique que celui-ci se singularise.

Nombre de maladies animales et zoonotiques ont connu ces dernières années des évolutions dans leur distribution ou leur intensité. En dépit des progrès de la vaccination et des thérapeutiques, la plupart des maladies majeures transfrontalières à transmission directe sont toujours présentes. Plusieurs maladies vectorielles, encore récemment cantonnées aux zones tropicales, sont aujourd'hui aux portes de l'Europe.

Le H5N1 ou Influenza aviaire hautement pathogène, fait partie de ces maladies contextuelles qui effraient, donc, dont les risques sont étudiés. La méconnaissance des possibles mutations du virus notamment vers une transmission interhumaine, sa haute pathogénicité et le risque de pandémie ont fait prendre conscience de l'importance d'étudier le H5N1 afin de l'enrayer ou tout au moins de freiner sa propagation.

Les recherches sur le sujet se développent grâce, en particulier, au perfectionnement des outils statistiques et informatiques et au développement des systèmes d'information notamment géographiques (SIG).

Lorsque l'opportunité d'intégrer, dans le cadre de mon stage de fin d'étude SIGMA, un des programmes de recherche portant sur le H5N1 s'est présentée, j'ai tout de suite accepté. Que se soit par l'acquisition d'une plus grande maîtrise des outils et des méthodes auxquels nous avons été formés, le désir d'intégrer un programme international intégralement en anglais ou l'ouverture vers des techniques et des raisonnements différents émanant des chercheurs vietnamiens, ce stage synthétisait ma conception d'une première expérience enrichissante et valorisable.

L'effet des changements récents qui affecte le virus, conséquence de la perturbation et de l'anthropisation de l'environnement, est encore mal évalué. Les outils de la géomatique, télédétection et SIG, semblent adaptés à la définition d'indicateurs environnementaux relatifs au H5N1.

Ce rapport est la synthèse, en trois parties, de notre travail au sein du programme GRIPAVI du Cirad. Nous l'avons souhaité chronologique pour calquer à la logique que nous avons tenté de mettre en place.

La première partie dressera les contextes, tant scientifique que ceux concernant la zone d'étude, la synthèse des recherches sur les liens environnement/H5N1 ainsi que les objectifs, moyens et méthodes de cette étude.

La seconde partie traitera de l'apport de la télédétection en série temporelle à basse résolution MODIS pour la détection des zones en eaux et des temps d'inondation ainsi que des rizières et l'intensité des récoltes.

La dernière partie comportera la mise en place d'un SIG et les difficultés inhérentes à cette dernière, et exposera la méthode ainsi que les résultats du croisement des couches environnementales avec les données épidémiologiques.

I. Contexte, objectifs et organisation du stage

A. Contexte scientifique

Notre stage s'est déroulé au sein de cadres institutionnel, physique et intellectuel assez diversifiés pour qu'il soit important de les présenter.

1. Le CIRAD

Le Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD) est l'institut français de recherche agronomique pour le développement des pays du Sud et de l'outre-mer français (<http://www.cirad.fr/fr/index.php>). Il a été fondé en 1984 dans une optique de développement durable, privilégiant le partenariat pour ses recherches. Il intervient dans plus de 50 pays du monde.

Sa mission première est de "contribuer au développement rural des régions chaudes, par des recherches et des réalisations expérimentales, principalement dans les secteurs agricoles, forestiers et agroalimentaires". Ainsi il s'est fixé comme enjeux de lutter contre les pauvretés et les inégalités en matière de nutrition, d'alimentation et de santé ; de respecter l'exploitation des ressources végétales et animales et de mobiliser l'action publique pour combattre les inégalités structurelles.

2. La Maison de la Télédétection (MtD)

En 1984, le campus scientifique de Montpellier « Agropolis », reconnu comme plate-forme mondiale de recherche agricole pour le développement, accueillait le Pôle régional de la Télédétection, ainsi que le Laboratoire Commun de Télédétection, associant l'ENGREF et le CEMAGREF. Pour répondre au développement prometteur des sciences géographiques, l'Etat et la Région Languedoc-Roussillon, aidés par l'Union Européenne, ont décidé de réunir sur un site unique, des équipes issues de différents organismes de recherches d'Etat et exploitant les outils de la géomatique, en privilégiant la valorisation de l'information satellitaire. La Maison de la Télédétection en Languedoc-Roussillon est inaugurée en septembre 1994 (<http://www.teledetection.fr/>). Elle regroupe des équipes du CEMAGREF, de l'ENGREF, du CIRAD et de l'IRD, soit actuellement environ 70 permanents et une vingtaine de personnes en séjours de moyenne ou longue durée.

3. L'UMR TETIS et UPR AGIRs

L'unité mixte de recherche TETIS (Territoire, environnement, télédétection et information spatiale) rassemble une centaine de personnes, affiliées au Cemagref, au Cirad et à l'Engref. L'UMR a pour finalité la maîtrise de l'information spatiale au service des démarches de connaissance et de gestion des milieux (agrosystèmes, écosystèmes) et des territoires (espaces ruraux).

L'UMR mène des recherches sur les méthodes d'analyse et représentation spatiale des systèmes agricoles et territoriaux selon 4 axes (<http://tetis.teledetection.fr>) :

- acquisition de l'information spatialisée (acquisition, traitement)
- compréhension de la structure spatiale et de la dynamique des systèmes (biophysique, sociale, technico-économique)
- gestion de l'information spatialisée (Système d'Information (S.I.), diffusion)
- mobilisation de l'information pour la gestion et la gouvernance (appropriation des S.I., apport aux démarches de développement, impact sur la gestion et la gouvernance).

L'Unité de Recherche AGIRs (Animal et Gestion Intégrée des Risques) du CIRAD a pour objectifs d'identifier et d'évaluer au travers de dispositifs de surveillance (observatoires) les risques sanitaires, réels ou perçus, liés à l'animal pouvant pénaliser la santé et par conséquent le développement et la sécurité des sociétés du Sud ; de comprendre et quantifier les facteurs structurels et fonctionnels et les mécanismes déterminants l'apparition et la gravité de ces risques liés aux animaux sauvages et domestiques ; sur la base de modèles et de scénarios, de concevoir et d'expérimenter des outils intégrés de prévention, de gestion et d'intervention pour surveiller, prévenir et limiter les causes et les conséquences de ces risques. Des chercheurs de l'UPR AGIRs travaillant sur l'application de la géomatique à l'épidémiologie sont ainsi accueillis au sein de l'UMR TETIS : c'est le cas de mes encadrantes AnneLise Tran et Stéphanie Desvaux ainsi que de mon stage.

Au Vietnam, nous avons été accueillis l'équipe du Professeur Pham Van Cu en charge du Centre International de recherche avancée pour les changements globaux (International Centre for Advanced Research on Global Changes, ICARGC) à l'Université Nationale du Vietnam, Hanoi, Vietnam et ex-vice directeur du Centre pour la Recherche Appliquée en télédétection et SIG (CARGIS). Concrètement, un chercheur, M. P.N.Hai et un étudiant, M. N.P.Dong, encadrés du le professeur M Pham Van Cu, ont collaboré scientifiquement sur l'étude.

Cette collaboration a permis d'avoir un encadrement technique spécialisé sur place et également de collaborer avec une équipe qui avait déjà une bonne connaissance du terrain.

4. Le projet GRIPAVI

Le projet intitulé **Ecologie et épidémiologie de la grippe aviaire dans les pays du Sud (GRIPAVI)** a été mis en place par le Fonds de solidarité prioritaire du ministère des Affaires étrangères et européennes. Les pays partenaires et bénéficiaires de ce projet sont l'Afrique du Sud, l'Ethiopie, Madagascar, le Mali, la Mauritanie et le Vietnam.

Le projet s'est donné trois objectifs :

- Une meilleure connaissance du mode de transmission des virus de la grippe aviaire
- Une analyse des méthodes de lutte grâce à la modélisation de la diffusion des virus
- La mise en place de méthodes de surveillance adaptées.

Ce projet vise également à renforcer les compétences des services vétérinaires en matière d'épidémiologie et de diagnostic et comprend la formation scientifique et professionnelle des cadres et techniciens des pays impliqués. Les connaissances ainsi acquises permettront de mieux adapter les méthodes de lutte dans les pays en développement afin d'éviter l'apparition ou de contrôler la dissémination du virus grippal à potentiel pandémique. Il concernera cinq écosystèmes susceptibles de jouer des rôles variés dans l'entretien et la dissémination de l'influence aviaire dont le Vietnam.

Le CIRAD travaille au Vietnam en santé et production animales dans le cadre d'un consortium de recherche, Prise, Pole de Recherche sur les risques liés à l'intensification des systèmes d'élevage. Prise est signataire de la convention générale du projet Gripavi. Prise regroupe 4 structures de recherches vietnamiennes qui sont ou seront impliquées dans le projet :

- le NIVR (National Institute of Veterinary Research), dépend du Ministère de l'Agriculture
- l'Université d'Agriculture de Hanoi et notamment le Centre d'études interdisciplinaires du développement rural.
- Le NIAH, l'Institut National de Production Animale.
- l'IPSARD, l'institut de Stratégie et de Politique pour le Développement Agricole et Rural.

Description de l'observatoire

Le Vietnam est composé de 64 provinces. Le projet Gripavi a choisi de ne s'intéresser qu'à l'écosystème du Nord Vietnam et notamment de la région du Delta du Fleuve Rouge. A noter que 2 autres projets internationaux aux objectifs

proches s'intéressent eux au delta du Mékong, dont les paramètres environnementaux, climatiques et humains diffèrent significativement.

Pour des raisons de gestions administratives du projet, il est difficile de travailler sur un nombre important de provinces.

Les critères suivants ont été appliqués pour choisir les provinces :

- provinces à risque où le virus a circulé à plusieurs reprises lors des vagues précédentes (selon données publiées et selon des informations de terrain) ;
- avoir la possibilité d'étudier le risque d'introduction via la Chine,
- avoir des zones de production qui concentrent un nombre important d'élevages à cycle long, pouvant servir de réservoir au virus dans un contexte de vaccination.

Bac Giang est une des provinces sélectionnées et sera un de nos terrains d'étude privilégiés. C'est une province dont la moitié est en zone de delta et l'autre moitié en zone de collines. La province est traversée par une route principale venant de Lang Son, province où les volailles arrivant illégalement de Chine traversent la frontière. Bac Giang est représentative des systèmes de production familiaux traditionnels et semi-commerciaux de la zone du Delta. Bac Giang a déclaré des foyers de grippe aviaire en 2005 et en 2007.

Questions de recherche abordées sur l'observatoire Vietnam

Dans le dispositif Gripavi, le Vietnam rassemble différents éléments uniques. Notamment, la possibilité d'étudier:

- les dernières vagues de foyers de façon rétrospective d'un point de vue épidémiologique mais également socio-économique du fait de l'ampleur de cette crise,
- les risques d'introductions régulières depuis la Chine du Sud par le commerce transfrontalier,
- la mise en œuvre (sans équivalent) d'une campagne de vaccination de masse,
- la présence simultanée de différentes souches de virus aux comportements différents.

Ainsi, plusieurs équipes de chercheurs apporteront des regards croisés sur un même phénomène.

B. Contexte de l'étude

1. Zone d'étude

Intégrée au sein du projet GRIPAVI, notre étude a trouvé logiquement, comme zone d'étude au Vietnam, les écosystèmes du delta du Fleuve Rouge et de la province de Bac Giang (Figure 3). Découpé en provinces puis en districts, eux même intégrant des communes sous lesquelles se situent des villages, le maillage

administratif du pays tiendra une place prépondérante dans le choix de nos échelles d'analyse.

a) Contexte socio-économique et politique

Le Vietnam, avec 85,1 millions d'habitants en 2007, et une densité moyenne de 257 habitants /km² (Bureau Général des Statistiques), est le 3ème pays le plus peuplé d'Asie du Sud Est après l'Indonésie et les Philippines.

La population, en majorité rurale (73%) se concentre principalement dans les deux zones de Deltas, delta du Fleuve Rouge au nord et Delta du Mékong au sud.

Avec un revenu per capita qui ne dépasse pas 690 \$ par an et 30% de sa population en dessous du niveau de pauvreté en 2006 selon les chiffres de la Banque Mondiale, le Vietnam reste un pays pauvre.

Néanmoins, le pays est en plein développement économique : le taux de croissance atteint 8,48 % en 2007 (prix constants 1994). Le PIB a doublé en 10 ans, mais l'écart ville-campagne se creuse.

Depuis les années 1980 le Doi Moi ou « rénovation économique » a consisté en une série de réformes politiques visant à modifier la base de l'économie planifiée socialiste préexistante en faveur des mécanismes de libre marché. Le Vietnam a en outre été admis en janvier 2007 au sein de l'OMC.

Grâce au Doi Moi et à un long processus de réformes, la paysannerie a, petit à petit, accédé à la propriété individuelle du foncier et à ses outils de travail. Les agriculteurs ont pu acquérir une plus grande autonomie en établissant eux même leur rotations, assolements, et systèmes d'élevage. Ils sont donc non seulement parvenus à l'autosuffisance alimentaire, mais ont également diversifié leurs systèmes de culture et d'élevage ainsi que leurs sources de revenus extra agricole. Le Vietnam est devenu le second pays exportateur de riz et d'autres produits tels que le café, le caoutchouc, le poivre etc...

b) Contexte épidémiologique

« L'influenza aviaire (IA), provoquée par le virus de l'influenza de Type «A» est une maladie animale qui peut infecter plusieurs espèces d'oiseaux d'élevage (poulets, dindons, cailles, pintades, etc.) ainsi que les oiseaux d'ornement et les oiseaux sauvages, certaines souches entraînant un taux de mortalité élevé. Ce virus a également été isolé chez des mammifères dont l'homme, le rat et la souris, le vison et le furet, le porc, le chat, le tigre et le chien. » (OIE, Organisation mondiale de la santé animale 2008).

Les oiseaux sauvages aquatiques constitueraient le réservoir naturel des virus de grippe introduiraient le virus dans de nouvelles zones par voies migratoires, en contaminant notamment les volailles par contact direct ou

indirect. Ainsi, le virus se transmettrait soit par contact direct avec un oiseau infecté, soit indirectement par une exposition à des matériaux contaminés par des fèces infectés ou des sécrétions respiratoires.

Le Vietnam a déclaré ses premiers foyers d'influenza aviaire hautement pathogène H5N1 fin décembre 2003. Entre 2003 et 2005 le pays a connu 3 vagues de foyers, nécessitant l'abattage d'environ 50 millions de volailles et entraînant 93 cas humains dont 42 mortels.

Fin 2005, la diffusion de la maladie était telle que le Gouvernement a décidé de lancer une campagne de vaccination massive des élevages de volaille. Cette campagne se poursuit encore aujourd'hui et même si le Vietnam envisage d'opter pour une stratégie vaccinale modifiée à très court terme, notamment en termes de prise en charge financière, l'arrêt de la vaccination n'est lui pas encore à l'ordre du jour. (Desvaux 2008).

L'année 2006 a marqué une accalmie dans le nombre de foyers enregistrés et aucun cas humain n'a été à déplorer. Malheureusement, fin 2006, une quatrième vague de foyers a démarré, notamment sur des canards non vaccinés du Sud du pays. En mai 2007, c'est une cinquième vague qui démarre et qui a d'abord touché essentiellement le Nord avec 8 nouveaux cas humains en 2008 (dont 5 cas mortels). Depuis des cas sporadiques continuent d'être déclarés régulièrement avec une augmentation en février et mars 2008 (Figure 1).

Le total des cas humains pour le Vietnam s'élève désormais à 106, dont 52 cas mortels.

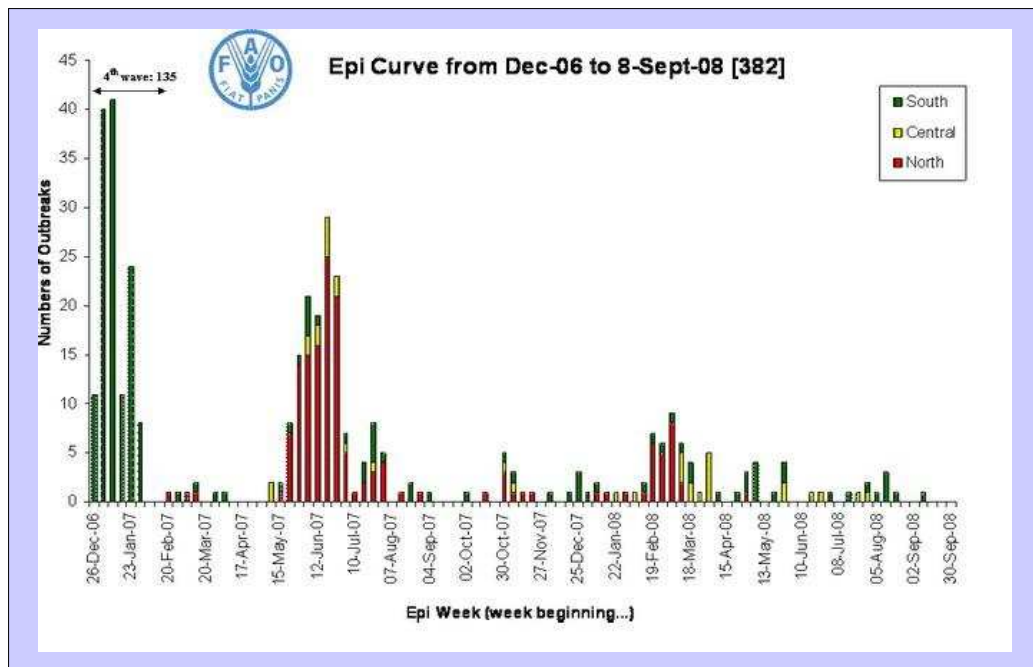


Figure 1 : Courbe épidémiologique des cas de H5N1 déclarés au Vietnam entre décembre 2006 et septembre 2008, FAO

c) Le contexte écologique

Le Vietnam, qui s'étire sur 1700 km est bordé par la Chine au Nord, le Laos et le Cambodge à l'ouest, et la mer de Chine à l'est (figure 1). Le relief est aux trois quarts formé de montagnes et de plateaux.

Le Vietnam, du fait de son extension en latitude présente une forte variété climatique.

L'observatoire Gripavi se situe dans la région du Nord, à cheval sur 2 régions agro-écologiques : le delta du fleuve rouge et la région Nord Est.

Dans la région du Delta du Nord, les oiseaux sauvages subissent une pression de chasse importante et souffrent d'une perturbation et d'une perte de leur habitat.

D'autres oiseaux sauvages présents en Asie du Sud Est, et en particulier les espèces d'eau telles que les canards, sont très rares voir absentes dans les zones d'étude.

Un nombre plus important d'échassiers peut être observé sur la côte du Delta mais ces espèces occupent généralement un habitat qui ne se mélange pas à celui des volailles domestiques (marécage, mangrove etc...).

d) Le système de production avicole

En 2006, la production de volaille était estimée autour de 215,565,000 par le Département de Production Animale du Ministère de l'Agriculture et du Développement Rural. Les poulets représentent 73% de la population aviaire et les oiseaux d'eaux, principalement des canards 27%.

On divise généralement les élevages de la façon suivante :

- L'élevage traditionnel: il s'agit de fermes de moins de 50 oiseaux par cycle ou en continu, sans organisation par bande et partiellement commercialise localement. Cette production représente la vaste majorité des élevages au Vietnam.
- L'élevage semi-commercial: il s'agit de fermes ayant plus de 50 oiseaux par cycle avec un niveau de biosécurité de nul à moyen et montrant un niveau d'intégration au marché et un réseau de commercialisation croissant.

2. Les liens Grippe Aviaire et environnement

Plusieurs études récentes mettent en évidence le lien entre environnement et présence du virus H5N1. Par environnement, nous entendons tant l'humain (riziculture, élevage de volailles, infrastructures,...) que le naturel (eau, température, altitude, migration des oiseaux,...). Il semblerait que le croisement

des pratiques socio-économiques dans des environnements naturels soit à la base de l'émergence et du développement du virus H5N1.

Ainsi la présence du virus pourrait être liée non seulement aux migrations d'oiseaux sauvages, comme nous l'évoquions précédemment, mais également à certaines pratiques humaines notamment agricoles : « Plutôt que les poulets, les canards, les hommes et les rizières seraient les principaux facteurs à l'origine des foyers du virus H5N1 hautement pathogène de la grippe aviaire en Thaïlande et au Vietnam,... » (FAO, 2008)

En Asie du sud-est on observe un agrosystème traditionnel associant la riziculture et l'élevage de canards : suivant un système de rotation, les récoltes achevées, les canards sont déplacés de rizières en rizières où ils se nourrissent de résidus de grains de riz, d'insectes et de vers. (Gilbert et al. 2006).

Les mêmes chercheurs en Thaïlande (Gilbert, Xiao, et al. 2007) et (Pfeiffer 2007) au Vietnam ont démontré une forte corrélation entre l'intensité des récoltes de riz et le nombre de canards d'élevage. Et parallèlement, les concentrations de canards dans les rizières augmenteraient la probabilité d'apparition de cas de grippe.

Les rizières sembleraient donc constituer un facteur important dans la persistance du virus dont les canards seraient un des réservoirs potentiels. De plus, des observations montrent que ces animaux pourraient ainsi le transmettre à de nombreux autres oiseaux. (Pfeiffer 2007; Gilbert et al. 2006).

Les chercheurs ont aussi découvert une certaine saisonnalité dans l'émergence de la maladie, induite, en partie, par les pratiques humaines.

Lors de la saison sèche, les volailles et les oiseaux sauvages se regroupent autour de points d'eau augmentant ainsi les risques. (Gilbert et al. 2006)

En Thaïlande, le plus grand nombre de détection de virus coïncide avec les périodes températures les plus basses (octobre à février), époque à laquelle les oiseaux du Nord et du Centre de l'Asie migrent vers la Thaïlande. On constate alors une coexistence entre espèces d'oiseaux sauvages et volailles dans les rizières et autour de points d'eau.

C'est aussi pendant cette saison qu'on trouve de nombreux troupeaux de canetons (septembre à octobre), qui sont en pleine croissance en novembre-décembre, au cours des récoltes.

Les rizières sont non seulement un habitat favorable pour les canards mais également pour les oiseaux sauvages. Il y a donc un fort effet de saisonnalité dans le nombre de contacts entre volailles et oiseaux sauvages. (Gilbert, Xiao, et al. 2008)

Certains chercheurs ont même poussé leurs investigations jusqu'aux pratiques religieuses et festives qui en découlent. Ainsi, la saison sèche coïncide également avec les festivals de fin d'année, et le nouvel an chinois où il y a une

forte demande en volailles. Il y a transport, abattage et vente de volailles correspondant aux pics du nombre de cas déclarés. (Pfeiffer 2007; Minh et al. 2009).

Le plus préoccupant dans cette crise demeure les nombreux problèmes socioéconomiques qu'elle génère. Pour prendre le cas du Vietnam, notre terrain d'étude, une majeure partie de la population est liée directement au système agricole et l'industriel. Ainsi, 33% de la consommation mondiale de riz est produite dans le pays et la volaille constitue l'essentiel de la consommation de viande du pays : ce virus représente donc une menace importante pour la santé humaine mais également pour l'économie du pays.

En résumé, cette maladie représente un problème économique liés aux mortalités et abattage chez les volailles et un problème de santé publique avec un passage à l'homme ponctuel

Un des enjeux fondamentaux de la maîtrise du virus est la compréhension de l'écologie de la maladie, et particulièrement des facteurs qui favorisent la réapparition de nouveaux cas.

C. Objectifs du stage et moyens mis en oeuvre

Etude pionnière au sein de GRIPAVI, notre stage reste le seul, sur les cinq sites que compte le projet, à analyser les données épidémiologiques par le prisme de la géomatique. Dans cette optique, des moyens tant logiciels, données que méthodologique ont été mis en place.

1. Objectifs

Les objectifs initiaux du stage nous ont été présentés ainsi ::

- A l'échelle régionale, extraction de paramètres environnementaux pour la caractérisation des zones en eaux (eaux permanentes, eaux de rizières, eau d'inondation sur les communes du Nord Vietnam à partir d'une série temporelle de données MODIS.
- Etude cas témoin à l'échelle locale (village), calcul d'indicateurs environnementaux relatifs à l'occupation du sol, la distance aux routes ou au réseau hydrographique (données issues de relevés terrain, cartes topographiques, images SPOT). Les villages cas et témoins étant déjà géoréférencés.

Le premier objectif consistera donc à utiliser la télédétection à partir de série temporelle MODIS pour détecter certains paramètres environnementaux et notamment les zones et les temps d'inondation ainsi que les rizières et l'intensité des récoltes pratiquées dessus. L'ensemble du Nord Vietnam est concernée par ces traitements.

En effet, les zones d'eaux sont une source possible de réservoir du virus du fait de l'utilisation de certaines surfaces pour le pâturage des canards et du fait de la présence d'oiseaux sauvages susceptible de pouvoir excréter du virus.

Les zones d'eaux à caractériser, sont a priori :

- les zones d'eaux permanentes (étang de pisciculture, les rivières etc...)
- les différents types de rizières : il s'agira de différencier les rizières en fonction du nombre de jours où l'eau est présente et de l'occurrence du phénomène d'inversion des indices eau/végétation.
- les zones inondables

Finalement chaque pixel devra être caractérisé par une durée de présence d'eau et par le nombre de récolte de riz.

Le second objectif se définira plus comme la construction d'une base de données SIG qui sera utilisées pour les analyses de statistiques spatiales dans une étude cas-témoin au niveau village pour les foyers de 2007 de la province de Bac Giang située dans le Nord Vietnam.

2. Moyens à disposition

a) Séries temporelles MODIS et images SPOT

Afin d'atteindre les objectifs fixés pour la partie télédétection, nous disposons de deux types d'images satellites :

- 44 images MODIS Terra téléchargées gratuitement et reprojctées à l'aide de l'outil MRTWeb couvrant l'ensemble du Nord Vietnam sur l'année 2007. Il s'agit du produit série temporelle MOD09A1.
- Un lot de 6 images SPOT acheté au centre SPOT de Hanoï par le projet, composés de 3 images SPOT2 et de 3 images SPOT4 couvrant la province de Bac Giang.

Cependant, les besoins ne concernent pas uniquement les images satellites mais également le logiciel capable d'effectuer les traitements dont nous avons besoin pour détecter les zones inondables ainsi que les rizières. Notre choix s'est porté sur le logiciel ENVI développé par la société ITT.



➤ **ENVI 4.3**

Envi comprend en un seul package des outils de traitement d'image de télédétection optique et radar particulièrement avancés, mais néanmoins interactifs et faciles d'utilisation : analyse spectrale, correction géométrique,

analyse topographique, analyse radar, fonctionnalités SIG raster et vecteur, support d'une large gamme de formats images, et encore bien d'autre possibilités.

Toutes les méthodes de traitement d'images de correction géométrique, radiométriques, de démixage radiométrique, de classification et de mise en page cartographique sont présentes. D'autres outils relatifs à la visualisation et à la modélisation de données topographiques sont aussi disponibles.

Envi constitue le choix naturel des professionnels de part ses performances, sa facilité d'utilisation et la qualité de ses algorithmes. Aussi, il est aussi important de signaler que le logiciel ENVI est conçu en langage IDL (Interactive Data Language- Research System) et offre donc des moyens de programmation évoluée : c'est cette dernière qualité qui a orienté notre choix afin de développer l'automatisation du traitement de séries temporelles importantes. En effet, contrairement à ces concurrents, les traitements ne sont pas limités à un nombre maximal d'images et ne nécessite pas d'être appelés un par un dans l'algorithme : nos 44 images peuvent ainsi être traitées et appelées en un seule fois en remplaçant la terminologie du nom par une *.

b) SIG du Vietnam et données épidémiologiques

Au Vietnam, deux lots de données ont été préalablement achetés à la maison de la cartographie du Vietnam et mis à notre disposition.

Le premier a été produit en 2002 alors que le second en 2004 a bénéficié d'une mise à jour jusqu'en 2008. Le laboratoire vietnamien avec lequel nous collaborons sur ce projet et qui nous apporte une aide précieuse tant dans la connaissance du terrain, des outils que dans la traduction du vietnamien, nous a également fourni son lot de données.

Nous nous retrouvons finalement avec trois lot de couche de données qu'il nous a fallu sélectionner, nettoyer et assembler pour en faire notre base spatiale à laquelle les données épidémiologiques seront croisées et analysées.

A ce propos, nous disposons également de données épidémiologiques issues des travaux de recherche des épidémiologistes du projet GRIPAVI. Ces données sont hétérogènes tant par leurs sources, leurs précisions que par les échelles auxquelles elles font références.

La première table « case_control_2007 » est le fruit des enquêtes de terrain des épidémiologistes du projet. Ses données sont très précises, à l'échelle du village où elles ont été collectées. Les villages cas sont ceux sur lesquels des cas de H5N1 ont été découvert. Les villages contrôle se situe à proximité et n'ont pourtant pas été touché par le virus. C'est la comparaison des deux types de villages qu'il est intéressant d'analyser.

La seconde table « outbreak » est issue des recherches sur internet, des communes ayant déclarées des cas d'infection au H5N1. Ces données sont d'une

fiabilité plus difficile à contrôler et sont agrégées au niveau communal. Par contre, elle présente l'avantage de répertorier les cas sur les différentes vagues d'épidémie qu'a connu le Vietnam et cela à l'échelle du pays.



- **ARCGIS 9.2**

ARCGIS est un Système d'Information Géographique (SIG) universel. Il est doté d'une application SIG complète, avec une logique et des outils SIG sophistiqués permettant d'accomplir quasiment toutes les tâches du SIG compilation des jeux de données SIG, création de bases de données SIG, mais aussi création des cartes et des modèles analytiques et documentation des travaux et méthodes.

La gamme de produits ARCGIS a été conçue pour s'adapter à ces besoins en constante évolution et proposer une plateforme SIG complète et évolutive.

ARCGIS propose une structure évolutive permettant la mise en œuvre du SIG pour un seul utilisateur ou de nombreux utilisateurs sur des postes bureautiques, des serveurs, sur Internet et sur le terrain. ARCGIS 9 permet d'élaborer un système d'information géographique complet grâce à un ensemble intégré de logiciels SIG. Il est constitué des éléments suivants :

ARCGIS Desktop : une suite intégrée d'applications SIG professionnelles,

ARCGIS Engine : des composants qui peuvent être incorporés par des développeurs afin de personnaliser des applications SIG,

Applications SIG pour serveur : ArcSDE®, ArcIMS® et ARCGIS Server,

Applications SIG nomades : ArcPad®, ainsi qu'ARCGIS Desktop et ARCGIS Engine pour les tablettes PC.

ARCGIS est basé sur une bibliothèque commune de composants logiciels SIG modulaires, appelés ArcObjects™.

Le choix de ce logiciel s'est basé sur la grande bibliothèque d'outils initialement intégrée dans ArcToolbox ainsi que l'importante communauté de développeur ayant créé des extensions utiles pour répondre à nos besoins notamment ceux en épidémiologie.

c) Sortie terrain de validation

La validation terrain a permis de valider la première phase de l'étude avant de pouvoir en intégrer les résultats dans la seconde phase de construction d'un SIG. Elle a été une étape primordiale dans le travail..

Ainsi, l'étude de terrain a semblé la solution de validation la plus précise et appropriée au traitement algorithmique de détection de l'intensité des récoltes de

riz sur la zone du Nord Vietnam. Elle se décomposera en plusieurs processus complémentaires qui auront pour but commun, la collecte d'information capables de déterminer une occupation des sols la plus proche possible de la réalité.

Un questionnaire a ainsi été créé et une zone délimitée. Durant huit jours, des relevés GPS appuyés par des photographies ainsi que questionnaires auprès des paysans et interviews des chambres régionales d'agriculture, ont été entrepris.

Finalement, leurs traduction par une classification sur image SPOT a été comparée au résultat de la classification obtenue par traitement algorithmique sur série temporelle MODIS.

3. Méthodologie générale et gestion du projet

Comme nous venons de l'évoquer, le but de cette étude demeure de permettre la détection, à partir de série temporelle MODIS, des zones inondables et de leurs durées d'inondation ainsi que des rizières et de leur nombre de cycle de culture et d'intégrer ces données environnementales, ainsi que d'autres provenant d'autres sources, dans un SIG pour les croiser aux données épidémiologiques. Une exploration de la relation entre ces variables et les données épidémiologique sera ensuite réalisée à l'aide de la cartographie par le biais des statistiques spatiales. . Les résultats du stage n'incluant par les statistiques spatiales, seule une analyse statistique univariée sera entreprise pour explorer cette relation.

Il semble important donc de profiter des différentes potentialités des capteurs satellites pour mieux caractériser l'occupation du sol. Dans ce contexte, la méthodologie établie est l'utilisation conjointe et complémentaire, des algorithmes de détection des zones et des durées d'inondation de L. Chamailé (Chamailé 2008) ainsi que celles des rizières et de l'intensité des récoltes de Xiao (Xiao et al. 2006).

La première étape consistera en une adaptation de l'algorithme de détection des zones inondées de L. Chamailé auquel nous apporterons une amélioration par la reconstitution des données parasitées à l'aide d'un processus d'interpolation temporelle. Nous obtiendrons ainsi la carte des durées et zones d'inondation du Nord Vietnam.

La seconde étape sera l'application de l'algorithme de détection des rizières et des récoltes de Xiao amélioré par l'utilisation des résultats de l'étape précédente. Une carte de rizières et du nombre de récolte sera ainsi dressée.

La validation, étape ultime de la partie télédétection de l'étude, s'établira par le biais de recherches terrain aboutissant à une classification référence sur

des images SPOT que nous confronterons à celle MODIS à l'aide d'une matrice de confusion.

Notre algorithme validé, et les différentes variables extraites, il s'agira alors de créer un SIG à partir de différentes couches mises à notre disposition et de les ordonner en fonction des différents besoins d'analyse multi scalaires souhaités.

La nature des données épidémiologiques dont nous disposons, nous oblige à adopter une démarche allant du particulier vers le général.

Ainsi la province de Bac Giang, offrant une précision des données allant à l'échelle kilométrique du village, servira de zone test à l'ensemble des traitements mis en place.

Si ces derniers se révèlent concluants, une extrapolation à l'échelle du delta du Fleuve Rouge sera envisagée malgré des données épidémiologiques agrégées à l'échelle communale.

Enfin, dans un souci de pérennité de notre démarche un jeu de données sera préparé à l'échelle du Vietnam pour les transpositions éventuelles des études futures.

C'est par le biais des notions de distance et de densité que seront analysés les relations entre environnement et les foyers déclarés de H5N1.. La constitution de données sous la forme de table est normalement ce qui nous est demandé mais l'intérêt pour l'étude, nous a poussé à tenter des représentations cartographiques des résultats ainsi que des ébauches de traitements statistiques. De même, les résultats attendus pour ce stage se limitent à l'échelle de l'étude cas-témoin de la province de Bac Giang : nous avons cependant tenté les mêmes analyses sur le Great Delta avec l'équipe partenaire vietnamienne....

Le projet s'articulera de la manière suivante :

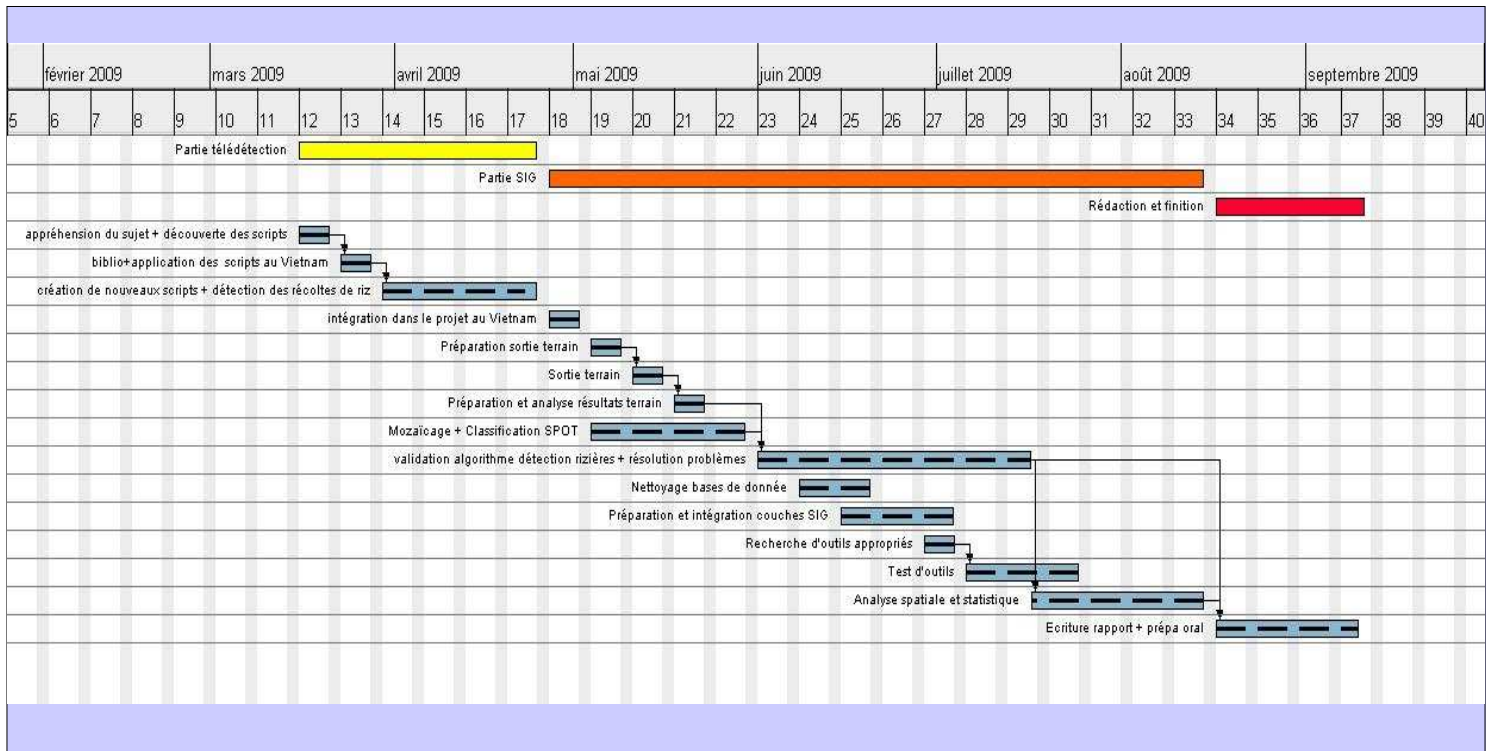
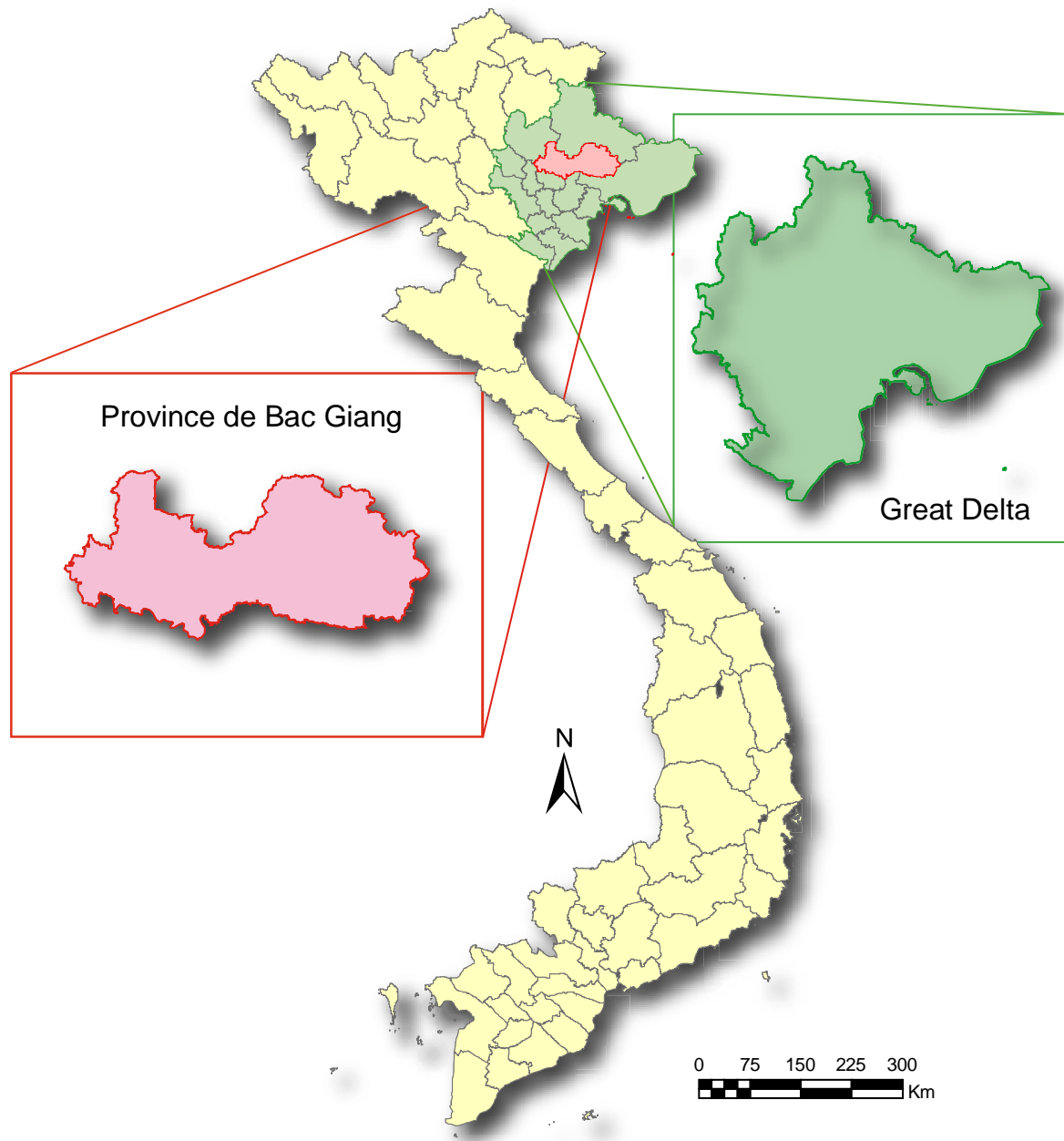


Figure 2 : Gestion du projet sous Gantt

Zones d'étude Projet GRIPAVI Vietnam



Réalisation: S. TOLLIS

Figure 3 : Les zones d'étude du projet

II. La télédétection en séries temporelles dans la détection des changements d'occupation des sols au Nord Vietnam

La télédétection constitue ainsi un outil intéressant dans le cadre d'études sur la grippe aviaire, par l'apport de nombreuses informations quant à l'occupation du sol, au temps d'inondation, aux températures etc. Couplé à des logiciels de traitement d'image, des traitements algorithmiques peuvent automatiser la recherche d'indicateurs environnementaux tels que les rizières et leurs intensités de récolte qui serviront à nos analyses spatiales futures.

A. Les images satellites

Nous avons choisi les images satellites en fonction de leurs potentialités pour l'identification de l'occupation du sol à une échelle nationale et en fonction de certains critères pratiques tels que le coût ou la facilité d'acquisition.

Il nous a d'abord semblé nécessaire d'acquérir des images à basse résolution spatiale et à forte occurrence temporelle afin d'exploiter la détection de changements phénoménologiques. Plusieurs capteurs produisent des images adaptées à l'observation des surfaces terrestres : NOAA-AVHRR, SPOT-VEGETATION, ENVISAT-MERIS, TERRA-MODIS.

Notre choix s'est rapidement porté vers les images issues du capteur Modis/Terra qui dispose de caractéristiques techniques beaucoup plus intéressantes que ces concurrents pour nos besoins. En effet, Modis/Terra produit des images qui peuvent atteindre la résolution spatiale de 250m, contre 1km pour ses concurrents, en gardant une résolution temporelle journalière.

1. Le système MODIS/TERRA

Lancé le 18 décembre 1999 par la NASA, le satellite américain TERRA a commencé à acquérir des images le 24 février 2000. Il se trouve sur une orbite héliosynchrone polaire à 705 km de la surface de la Terre. La mission de TERRA est orientée sur l'étude de l'évolution de la végétation et l'extension des déserts. Pour cela, il propose une offre variée de capteurs ayant des caractéristiques spécifiques afin de permettre l'observation des différents systèmes terrestres (atmosphère, surface



Figure 4 : Le satellite TERRA

terrestre, océans...).

Avec son couloir de visée d'une largeur de 2330 km, le capteur MODIS (Moderate Resolution Imaging Spectroradiometer) embarqué sur TERRA et AQUA permet de voir chaque point de la Terre tous les 1 à 2 jours dans 36 bandes spectrales de 0,4 à 14,4 μm .

Il fournit des informations sur la température de surface des terres et des océans, la productivité primaire, la couverture terrestre, les nuages, les aérosols, la vapeur d'eau, et les feux.

La liste des produits proposée par ce capteur est grande et variée. Selon les produits, la résolution spatiale varie de 250m à 1km au nadir.

La synthèse temporelle permet d'obtenir théoriquement des images d'une plus grande qualité: elle reste le produit qui semble la plus adaptée à nos besoins.

2. Les synthèses temporelles

Produit proposé par MODIS de résolution 500 mètres, les dimensions de l'image sont de 2400 lignes par 2400 colonnes couvrant une aire de 10° par 10° (en latitude/longitude). Les pixels ont une valeur correspondant à l'estimation de la réflectance qui pourrait être mesurée en l'absence de diffusion ou d'absorption par l'atmosphère.

Les synthèses temporelles de 8 jours Modis sont élaborées à partir d'un algorithme particulier qui, pour générer une composition, a besoin de plusieurs observations. Le chevauchement des fauchées du capteur permet d'enregistrer plusieurs enregistrement par jour pour un même pixel, quatre au maximum. En théorie, sur une période de 8 jours, nous pouvons donc obtenir au maximum trente deux enregistrements. En réalité, la présence de nuages et la couverture spatiale réel du capteur, réduit ce nombre d'enregistrement entre trente deux et zéro.

Première étape de l'algorithme : filtrer les données en fonction de leur qualité, des nuages et de la géométrie de visée. Seuls les pixels de bonne qualité, sans nuages, sont alors retenus pour la composition. Le nombre de valeurs acceptables sur les 8 jours, pour chaque pixel, est généralement inférieur à 10 et varie souvent entre 1 et 5. Une sélection s'opère ensuite pour extraire une seule valeur par pixel. La méthode dépend alors du nombre de pixels retenus après le filtre de qualité :

- 0 pixel acceptable : méthode **MVC Maximum Value Composite**. Elle consiste à attribuer au pixel de la synthèse sa plus forte valeur enregistrée sur les 8 jours permettant ainsi de réduire considérablement les effets dus aux perturbations

atmosphériques. Inconvénient : chaque valeur retenue se trouve dans une configuration angulaire a priori différente des autres. Or les réflectances dépendent de la configuration angulaire. Il devient alors difficile de comparer les valeurs pour chaque synthèse temporelle.

- 1 pixel : ce dernier est retenu pour la synthèse.
- Entre 1 et 5 : méthode **CV-MVC (constraint-view angle Maximum Value Composite)**. Les 3 observations ayant la plus forte valeur sont comparées et celle ayant le plus petit angle de visée et étant plus proche du nadir est choisie pour la synthèse.
- + de 5 : méthode **BRDF -C (Bidirectional reflectance distribution function composite)**. Toutes les valeurs retenues sont utilisées pour interpoler la valeur de réflectance au nadir. Cela permet d'obtenir des valeurs ayant les mêmes configurations angulaires (au nadir).

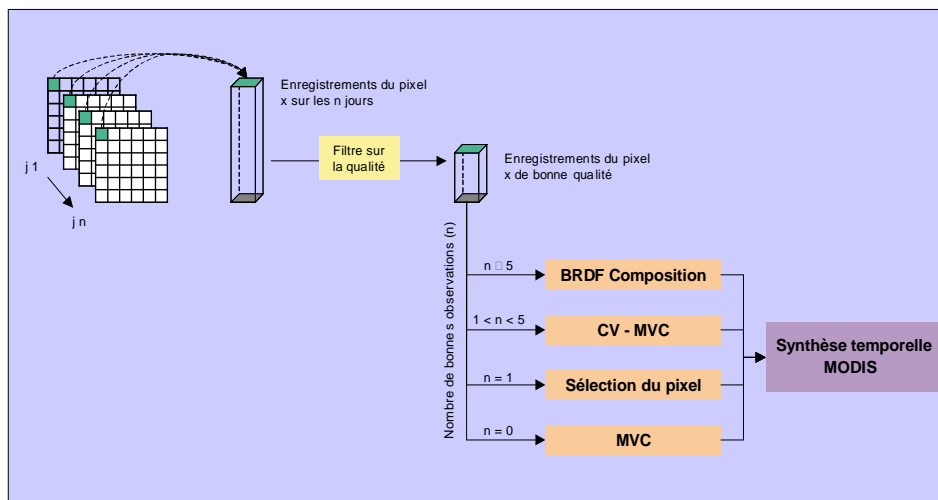


Figure 5 : Méthode de constitution des synthèses temporelles MODIS

Au final, chaque pixel contient la meilleure valeur possible observée durant une période de 8 jours c'est-à-dire un faible angle de visée, une absence de nuages ou d'ombres de nuages, et d'aérosols.

Les bandes 1 à 7 avec lesquelles nous avons travaillé ont donc une résolution de 500mX500m en projection sinusoidale d'origine que nous avons reprojété en UTM Zone 48N, projection correspondante aux autres données géographiques du SIG.

3. Méthodologie

L'apport de la télédétection dans notre étude réside dans sa capacité à extraire des paramètres environnementaux qui seront croisés, par la suite, aux données épidémiologiques et analysés afin de détecter une hypothétique relation expliquant les zones d'émergence et de propagation du virus H5N1.

Notre démarche fait suite aux travaux de recherche de Xiao (Xiao et al. 2006; Xiao 2005) relatifs à la détection des récoltes de riz par le changement de comportement d'indice de végétation et d'eau. Facteur présumé important de l'émergence et de la propagation du H5N1 (Pfeiffer 2007; Gilbert, Xiao, et al. 2007; Gilbert et al. 2006; Minh et al. 2009; Gilbert, Xiao, et al. 2008), l'intensité des récoltes de riz constitue une donnée essentielle à localiser et cartographier afin de la croiser aux données épidémiologiques. En effet, ces auteurs ont mis l'accent sur le fait que l'intensité des récoltes de riz favorise la réoccurrence du virus au même titre que deux autres facteurs principaux que sont la densité humaine et celle des canards.

Le chercheur américain Xiao Xiangming a mis au point un algorithme de détection des rizières et des récoltes à l'aide d'images satellites à base résolution MODIS en observant le comportement de plusieurs indices dans le temps. Partant du principe que la culture du riz passe d'une étape de repiquage à une croissance jusqu'à la maturité pour finir par une récolte, Xiao a analysé ces phases du point de vue de la réflectance.

En effet, lors de la phase de repiquage, le riz est jeune donc peu développé et peu dense laissant ainsi apparaître le sol et par conséquent l'eau qui caractérise ces milieux. Peu à peu, sa croissance fait que le sol est de moins en moins visible et au contraire, qu'il prend le dessus en offrant une couverture végétale importante. Finalement, la phase de récolte se traduit par un arrachage des plants de riz et donc, une réapparition du sol et de l'eau.

D'un point de vue spectral, ces phases du cycle du riz se traduisent par des réflectances plus ou moins importantes en fonction des bandes spectrales par lesquelles elles sont observées. Des indices ont ainsi été créés pour détecter un objet plutôt qu'un autre. Xiao a essentiellement travaillé avec trois d'entre eux : le LSWI (Land Surface Water Index), le NDVI (Normalized Difference Vegetation Index) et l'EVI (Enhanced Vegetation Index). Si le premier est apprécié pour sa capacité à faire ressortir l'humidité des sols, les deux derniers sont plutôt indiqués pour caractériser la végétation, le dernier étant une amélioration du second comprenant une correction atmosphérique par l'utilisation de la bande bleue.

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}$$

$$\text{LWSI} = \frac{\rho_{\text{nir}} - \rho_{\text{swir}}}{\rho_{\text{nir}} + \rho_{\text{swir}}}$$

$$\text{EVI} = 2.5 \times \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + 6 \times \rho_{\text{red}} - 7.5 \times \rho_{\text{blue}} + 1}$$

Figure 6: Les trois indices utilisés par Xiao pour la détection de l'intensité des récoltes de riz

Ainsi, lors d'une récolte, le LWSI prend des valeurs supérieures à celles des deux autres indices, traduisant ainsi l'inondation dont font l'objet les rizières. Mais cette phase n'est que provisoire et surtout moins longue que celle qui suit. En effet, très vite les valeurs des indices de végétation deviennent supérieures à celle de l'indice de présence d'eau et ceci durant plusieurs mois, symbolisant la croissance du riz et par conséquent du couvert végétal. Puis à nouveau, une chute du NDVI ou de l'EVI couplée à une croissance du LWSI sont observées, traduction de la récolte et de la réapparition du sol humide de la rizière.

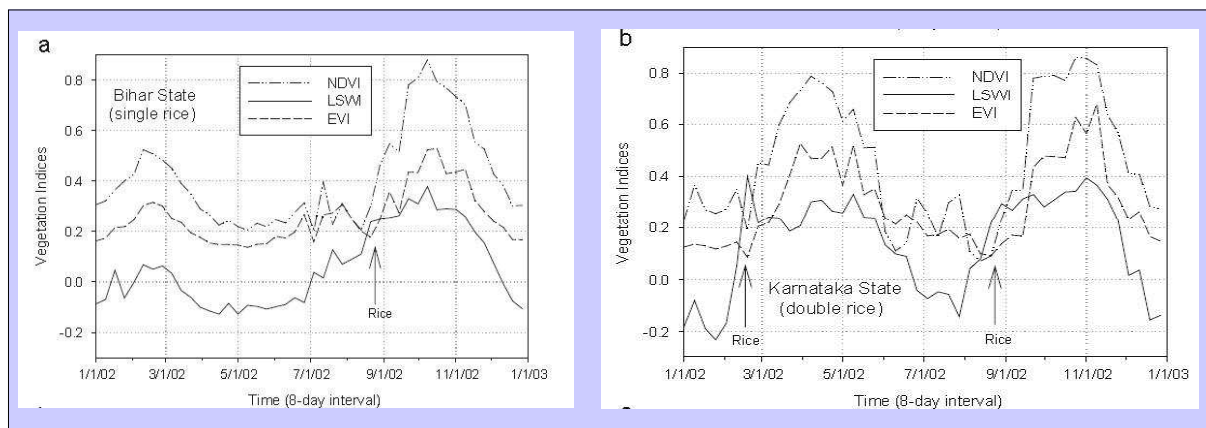


Figure 7: Deux exemples issus des recherches de Xiao
a) zone à 1 seule récolte de riz par an b) zone à 2 récoltes annuelles

La notion de temporalité apparaît prépondérante dans l'observation de ces mutations phénoménologiques justifiant ainsi l'utilisation de séries temporelles. Cependant, la détection des rizières passe inévitablement par celle des zones en eau, élément indispensable à la culture du riz mais doit cependant être distinguée des zones d'eau permanentes telles que les rivières, lacs, mares et autres objets du réseau hydrographique.

A cette fin, notre méthode prend également en compte les résultats de Lise Chamailé, menée dans le cadre de la réalisation du stage de fin d'étude SIGMA, concernant la détection des zones en eau et des durées d'inondation. Etape intermédiaire de notre algorithme d'occupation du sol, la méthode de détection

des temps d'inondation n'a constitué qu'une application et une amélioration du procédé mis en place par Lise Chamaillé.

Partant du postulat que les travaux de recherche bibliographique de cette dernière concernant les indices de l'eau ont été validés, nous n'avons pas cherché à tester plusieurs d'entre eux mais nous sommes contents d'appliquer le MNDWI qui semble être le plus adapté à nos recherches.

Par contre, le recul que nous avons sur son travail, a permis d'améliorer sa méthode par application d'un algorithme d'interpolation temporelle permettant de reconstituer les zones masquées par la couverture nuageuse très présente toute l'année sur cette partie du globe. Contrairement à L. Chamaillé, nous nous sommes également servis du fichier d'indice de qualité, téléchargeable en même temps que les images avec l'outil MRTWeb (<http://mrtweb.cr.usgs.gov/>). Apportant des informations notamment sur la présence ou non de nuage, il a permis d'apporter une condition supplémentaire améliorant la création du masque des nuages. Le seuil de détection des nuages a également été affiné.

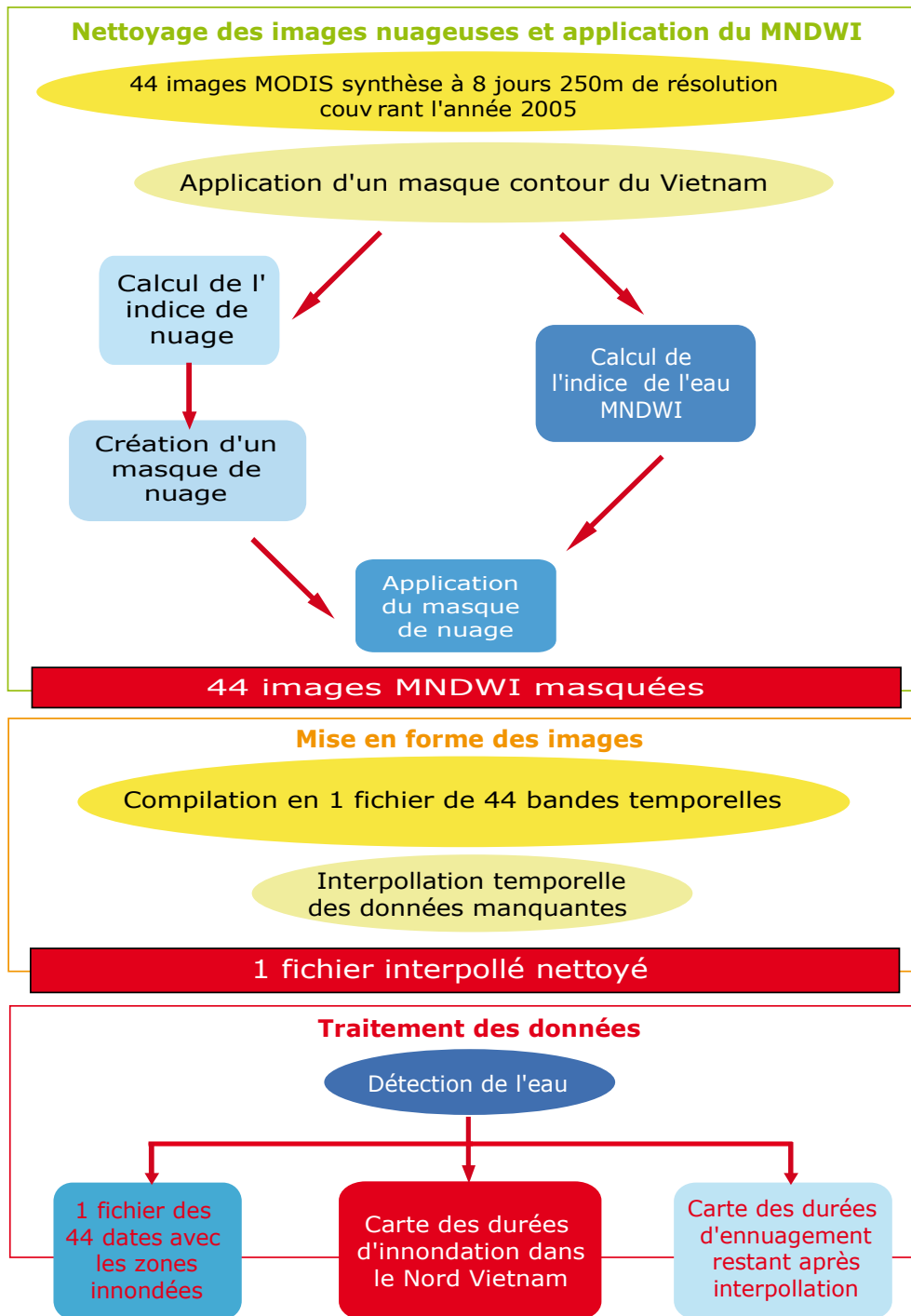


Figure 8 : algorithme amélioré de la méthode de L. Chamailé

Au final, nous établissons une cartographie plus précise et complète des zones touchées par la présence de l'eau et de ses durées d'occupation. Sans cette interpolation, il devient difficile et imprécis de tirer des conclusions globales tant le Vietnam présente des périodes de couverture nuageuse importantes.

Cartes des durées d'inondation dans le Nord Vietnam

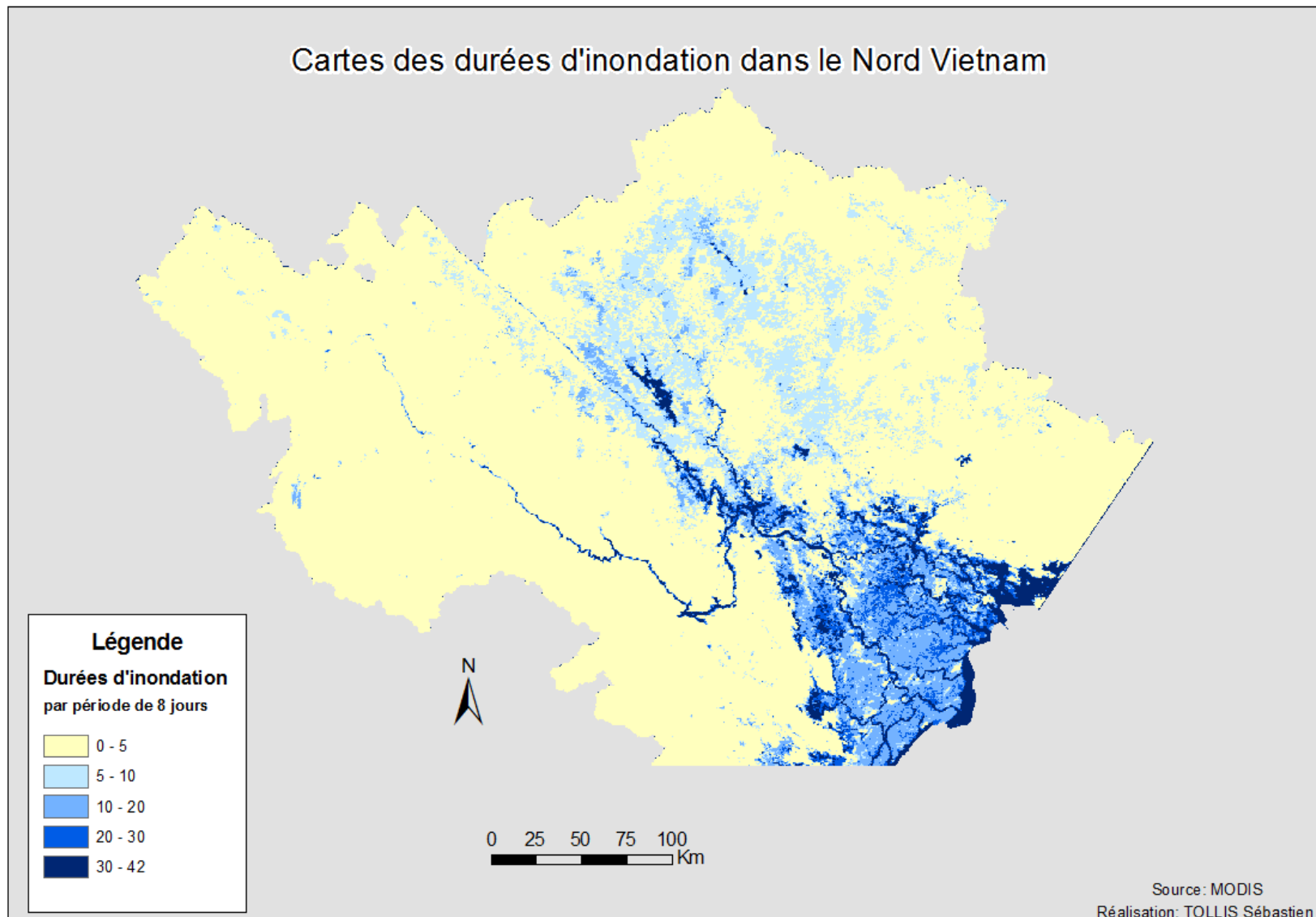


Figure 9 : Carte des durées d'inondation du Nord Vietnam

Notre démarche se situe donc au croisement des méthodes de Xiao et L. Chamailé. Si nous avons apporté des améliorations sur les aspects techniques des différentes détections, la méthodologie a également été modifiée afin de répondre à une logique d'organisation des données pérenne et transposable.

Sur la partie purement technique, les principales modifications concernent l'utilisation de l'algorithme vu plus haut afin de déterminer les zones d'eaux permanentes là où Xiao n'utilise que le simple NDVI.

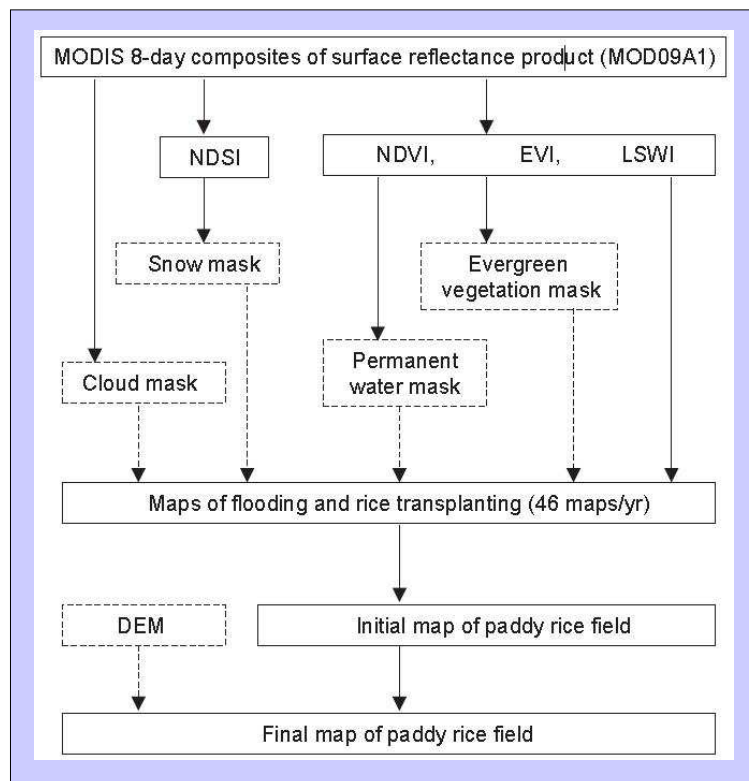


Figure 10 : algorithme de Xiao pour déterminer l'intensité des rizières

Les améliorations apportées à l'algorithme développé par L. Chamailé ayant déjà été décrites, c'est autour de l'organisation des divers traitements algorithmiques que les efforts ont été concentrés.

Si cette dernière a privilégié l'allégement des traitements en calculant très tôt son indice MNDWI réduisant ainsi l'ensemble de la série temporelle à une seule bande spectrale, notre souci a plutôt été de créer une base sur laquelle des traitements futurs s'appliqueraient facilement. En effet, en ne calculant nos indices que tard dans l'algorithme, nous créons une série temporelle certes lourdes (44 images de 7 bandes spectrales puis 7 fichiers, un par canal, de 44 bandes temporelles) mais nettoyée par interpolation et qui se prête donc à tout traitements notamment les calculs d'index. C'est un choix qui se discute mais qui semble cohérent par rapport aux traitements que notre étude nécessite.

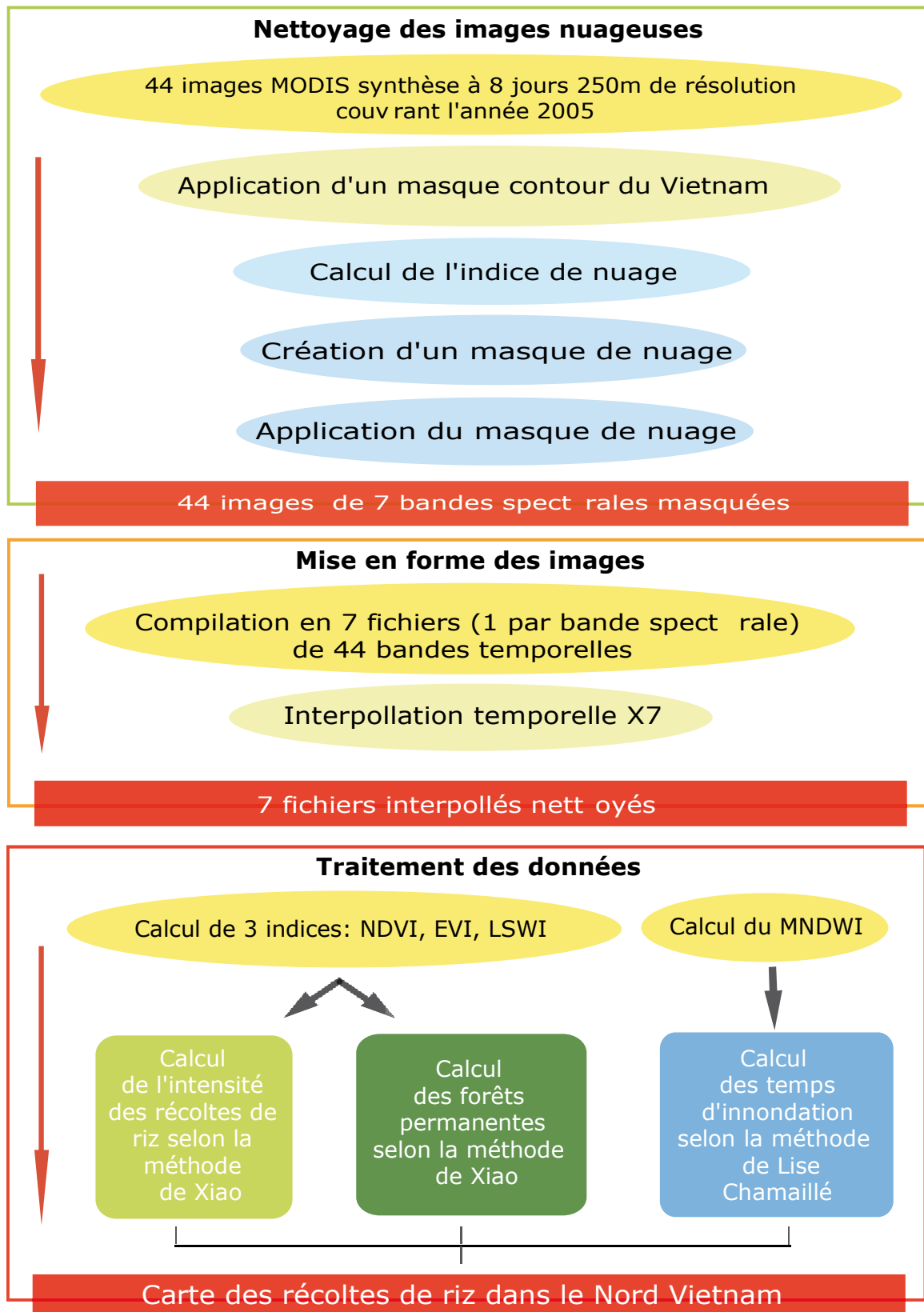


Figure 11 : notre algorithme de détection de l'intensité des récoltes

B. Acquisition, préparation et traitement des images satellites

Comme nous l'évoquions précédemment, la recherche des zones de présence d'eau et des durées d'inondation n'est qu'une étape dans notre processus de détection des rizières et du nombre de récolte mais elle reste néanmoins indispensable. Suivant les recherches de Xiao sur la cartographie des zones de cultures du riz (Xiao et al. 2006; Xiao 2005), nous avons développé sous IDL (Interactive Data Language), le langage de programmation d'ENVI, une série d'algorithmes ayant pour objectif final de bien distinguer les rizières des eaux permanentes, de la forêt ainsi que des autres cultures.

Pour cela nous disposons d'une série temporelle de 44 images à 8 jours d'intervalle composées de 7 bandes spectrales (bleu, Vert, Rouge, Proche Infrarouge, Moyen Infrarouge), de résolution spatiale 500x500mètres.

Numéro de bande	Longueur d'onde Réfléchi (nm)	Résolution (mètres)	Utilisation
1	620-670 (rouge)	250	Transformation de l'occupation des sols, chlorophylle de la végétation
2	841-876 (NIR)	250	Quantité de nuage, transformation de l'occupation du sol
3	459-479 (bleu)	500	Difference sol/vegetation
4	545-565 (vert)	500	Vegetation
5	1230-1250 (MIR)	500	Difference feuille/canopée
6	1628-1652 (MIR)	500	Difference Neige/nuage
7	2105-2155 (MIR)	500	Propriétés des nuages et des terres

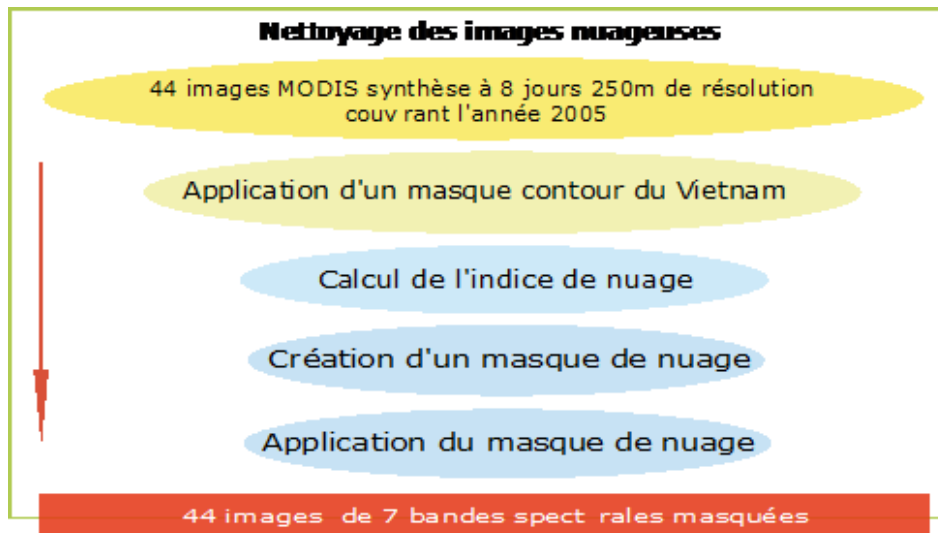
Tableau 1: les 7 bandes spectrales MODIS utilisées

1. Nettoyage des images nuageuses

Etape préliminaire, laborieuse mais néanmoins indispensable à tout travail scientifique, la principale fonction à laquelle a répondu cette phase demeure la résolution du problème de couverture nuageuse.

L. Chamailé a pu s'en rendre compte dans son étude sur le Cambodge (Chamailé 2008), la couverture nuageuse est extrêmement présente en Asie du Sud-Est et encore plus durant la saison des pluies. Un nombre important d'images sont inexploitable ce qui limite donc les traitements possibles ainsi que la validité des résultats. Une solution était donc nécessaire.

Nous avons mis en place une méthodologie qui représente la première partie de notre algorithme final.



Comme pour toute étude, il était primordial de délimiter la zone d'étude. Notre réflexion portant sur le Vietnam et sur certaines de ses entités administratives internes, nous avons commencé par créer et appliquer un masque des contours du pays correspondant aux frontières nationales.

Le terrain d'étude étant borné, le traitement de résolution du problème nuageux pouvait être mis en place. Il s'organise en trois étapes :

- Déterminer les propriétés des nuages pour les détecter
- Créer un masque
- Application du masque

Le choix du produit synthèses MODIS s'était justifié par l'intégration de premières corrections radiométriques, notamment atmosphériques, présageant d'une qualité de pixel optimale. Il est cependant constaté sur ces images certains défauts ou bruits dégradant leurs qualités et notamment, la présence de nuage.

Suivant les travaux de L. Chamailé, nous avons créé un indice de nuage à l'aide des bandes bleues et vertes qui possède la particularité de maximiser la réflectance des nuages au détriment des autres surfaces. Il est également important de différencier l'eau des nuages car si, la bande bleue fournit une réflectance plus importante des nuages, la verte en fait de même pour l'eau. La formule de cet indice résout ainsi l'ensemble de ces problèmes.

$$\text{Indice de nuage} = \frac{\text{Bande bleue} - \text{Bande verte}}{\text{Bande verte} - \text{Bande bleue}}$$

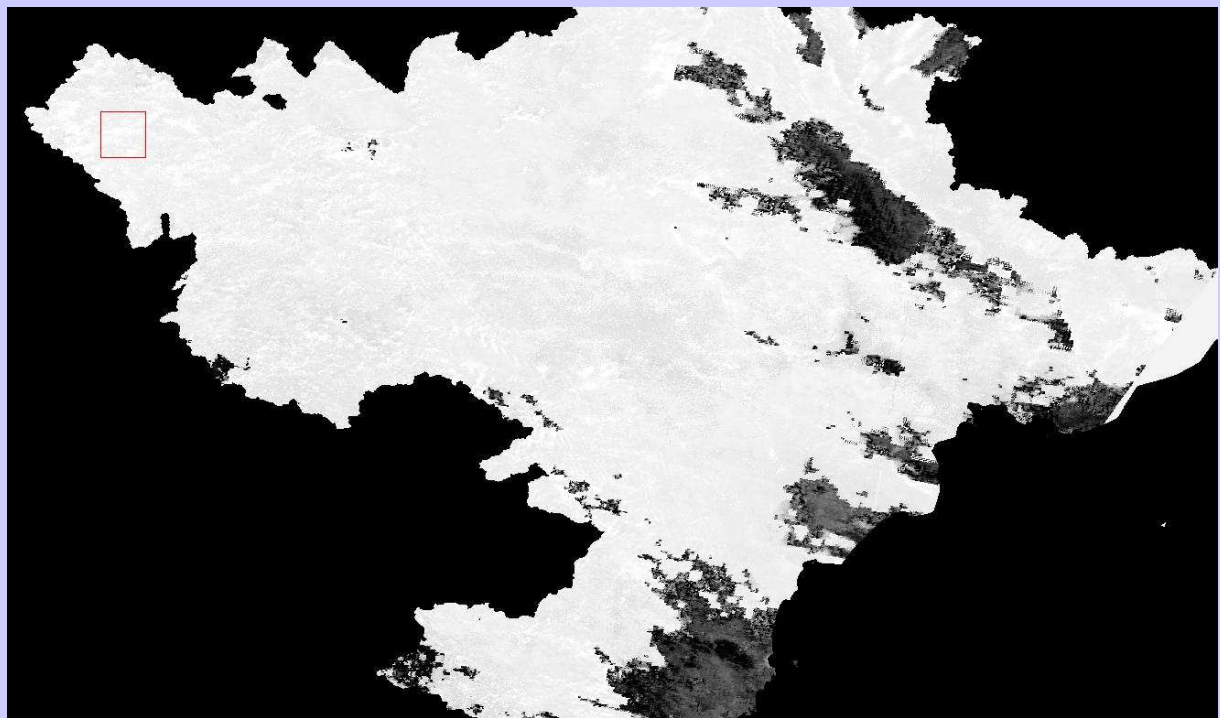


Figure 12 : illustration de l'importance de la couverture nuageuse et de l'application de l'indice de nuage

Suite à ce traitement, 44 images d'une bande spectrale, issue du croisement des bandes bleue et verte, sont créées faisant apparaître les nuages dans une réflectance très soutenue (Figure 11).

Les nuages ainsi surexposés, un masque peut être créé. Contrairement à L. Chamailé, nous n'avons pas fixé qu'une condition pour la détection de ces derniers mais deux. En effet, nous avons intégré le fichier qualité MODIS afin d'évaluer chaque image en plus d'un seuil qui a été affiné à -0,2 au lieu de -0,13.

Dans une première boucle, l'algorithme fait le tri à partir du seuil que nous avons fixé définissant les pixels nuageux et ceux qui ne le sont pas. Puis, à l'intérieur des pixels classés comme nuageux, il prend en considération les données qualité MODIS. Ce sont deux conditions qui doivent être réunies pour qu'un pixel soit classé comme nuage.

En effet, dans les canaux MODIS, chaque pixel a pour valeur un nombre décimal correspondant à une valeur binaire codée sur 16 bits. Chaque bit code une information sur la qualité d'enregistrement du pixel ou sur les traitements effectués sur celui-ci.

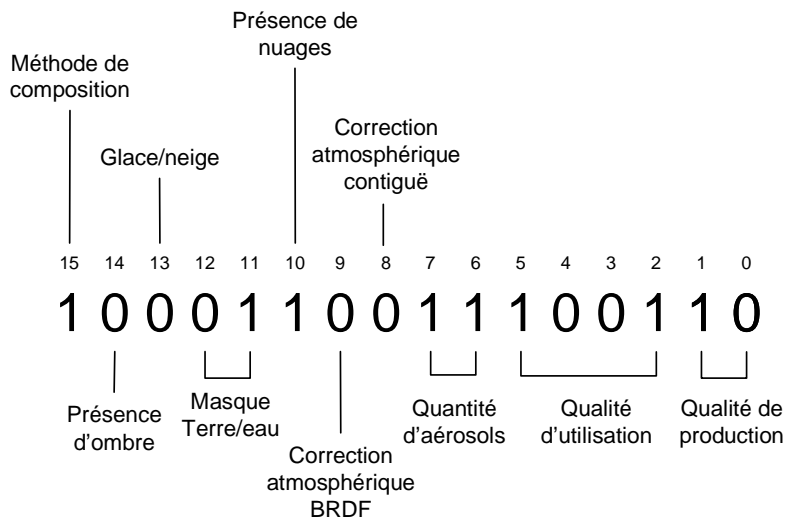


Figure 13 : Codage de la qualité du NDVI des images Modis

Nous obtenons, au final, des fichiers masque de nuage où les pixels nuageux prennent la valeur 0 quand ceux non nuageux, la valeur 1.

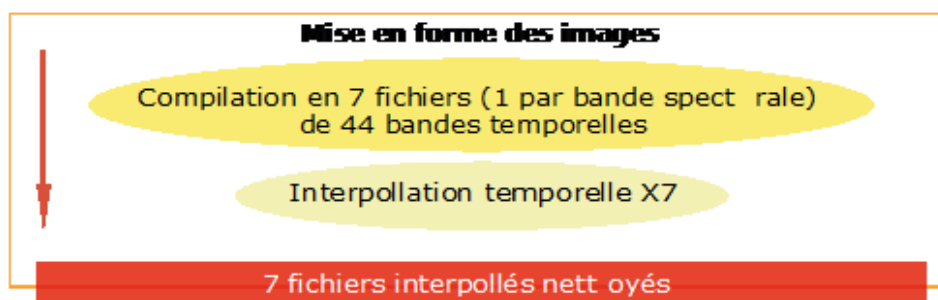
Dernière étape de cette phase de nettoyage des données, l'application du masque sur les images précédemment découpées aux limites du Vietnam. Les pixels nuageux obtiennent la valeur 255, ceux en dehors du Vietnam la valeur 100 et tous les autres entre -1 et 1.

Le résultat final est 44 images de 7 bandes spectrales masquées.

2. Mise en forme des images

Si l'étape précédente a permis de contourner le problème de la couverture nuageuse en masquant les pixels dégradés, elle n'en a pas moins résolu le problème de perte d'information occasionné.

Une méthode de reconstitution des données a été mise en place mais elle nécessite, pour cela, de travailler sur des séries temporelles.



C'est l'outil "Layer Stacking" du logiciel ENVI, modifié en programmation sous IDL, qui a été utilisé pour réaliser les séries temporelles. Il permet normalement de compiler, par exemple, les 44 canaux MNDWI provenant des 44 images MODIS dans un seul fichier en

attribuant une date à chaque bande : c'est ce que nous avons réalisé pour l'algorithme de détection des durées d'inondation. Le problème, c'est que dans notre configuration, nous ne sommes plus en présence de 44 images d'une seule bande mais 44 images masquées de 7 bandes. En développant un outil pour ENVI sous IDL, nous avons créé 7 fichiers, un par bande spectrale, regroupant 44 bandes non plus spectrales mais temporelles : en sommes nous avons inversé l'ordre entre les fichiers et les bandes. En obtenant cette organisation, le calcul d'indice ou d'autres traitements est grandement facilité notamment lorsqu'il s'agit de données présentes en quantité pouvant s'étaler sur des durées importantes.

Les séries temporelles permettent alors la visualisation des profils temporels de chaque pixel à l'aide de l'outil « z profil » qui nous sera d'une grande utilité au moment de comparer les comportements dans le temps de différents indices caractérisant le cycle du riz.

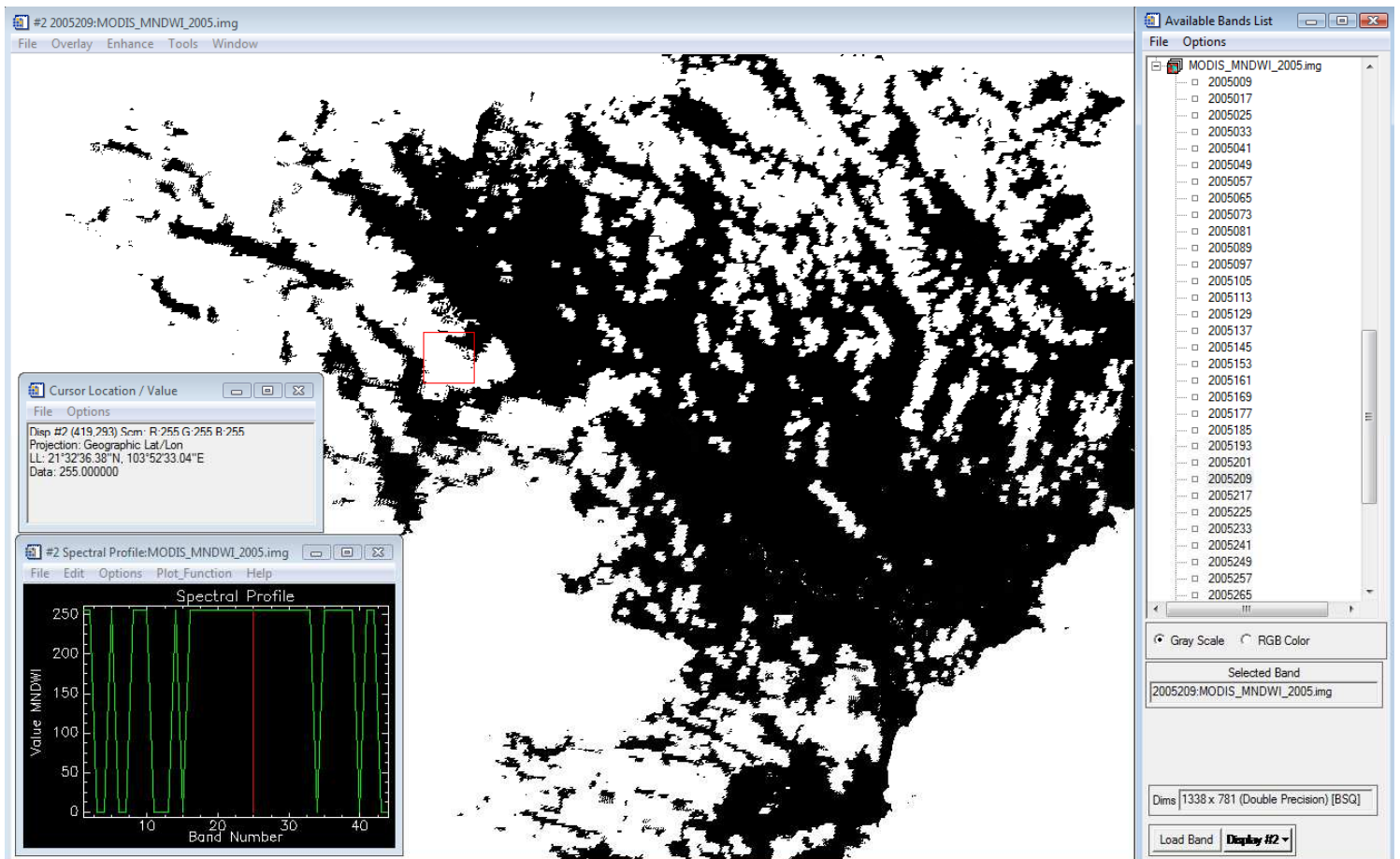


Figure 14 : exemple de profil temporel sous ENVI

La capture d'écran ci-dessus illustre parfaitement nos propos : la courbe du profil temporel ne cesse d'être affectée par des montées brutales de la valeur du MNDWI que nous pouvons apparenter aux périodes de couverture nuageuse. Il

est ainsi aisé de comprendre l'importance de la mise en place d'une méthode de reconstitution des données manquante : l'interpolation temporelle.

Cette démarche se justifie par l'impossibilité de travailler avec des données trop parasitées qui entraînent des erreurs dans la caractérisation des différents stades d'évolution. Les pixels contaminés par les nuages présentent des profils temporels similaires avec une montée anormale des valeurs des pixels. Lors de la classification, cela se traduit par le regroupement au sein d'une même classe de ces pixels "perturbés". Les différentes occupations du sol, dont le riz, ne peuvent en aucun cas être distinguées.

Nous avons donc créé un batch sous IDL suivant les procédés suivants :

- Pour la première date, pour tous les pixels identifiés comme nuageux (valeur 255 délivré lors de l'application du masque de nuage), les 2 premières valeurs considérées comme correctes du même pixel sur les dates supérieures sont utilisées et une interpolation linéaire est réalisée suivant la formule :

$$y = (y_2 * (x_3 - x_2) + (y_3 - y_2) * (x - x_2)) / (x_3 - x_2)$$

où

y est la valeur du pixel nuageux (valeur 255)

y₂ est la valeur du même pixel correct sur la première date supérieure

y₃ est la valeur du même pixel correct sur la deuxième date supérieure

x est la date (ou n° de bande) du pixel nuageux

x₂ est la date (ou n° de bande) du pixel correct sur la première date supérieure

y₃ la date (ou n° de bande) du pixel correct sur la deuxième date supérieure

- De la deuxième date à l'avant dernière, pour tous les pixels nuageux, la valeur du même pixel sur la date juste inférieure et la première valeur correcte sur une date supérieure sont conservés et une interpolation linéaire est également réalisée suivant la même formule que précédemment.
- Pour la dernière date, pour tous les pixels nuageux, les valeurs du même pixel sur les 2 dates juste inférieures sont utilisés et l'interpolation linéaire est une nouvelle fois exécutée.

Marion Houlès, dans le cadre de son rapport de fin de stage SIGMA (Houles 2006), a réalisé un très intéressant schéma retraçant l'ensemble des étapes de l'interpolation temporelle : nous le lui avons emprunté afin d'illustrer nos propos.

Interpolation

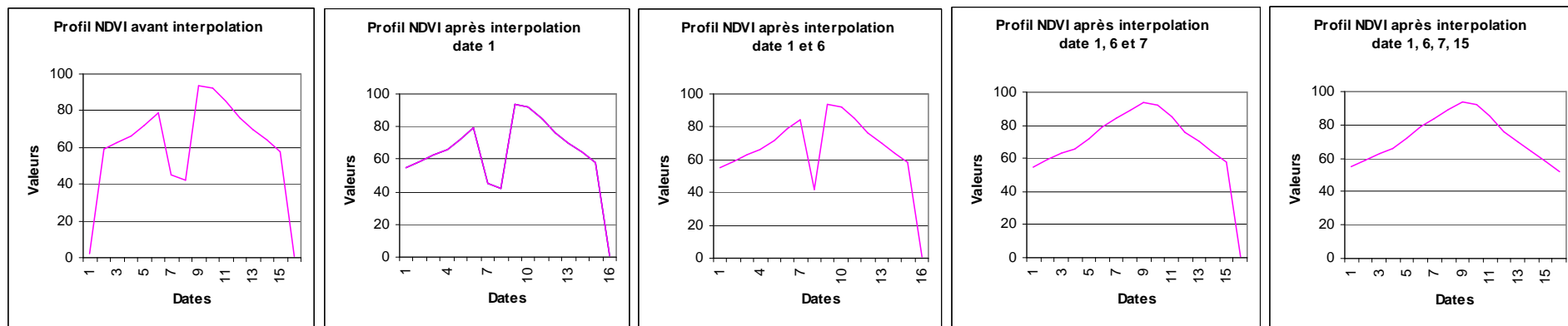
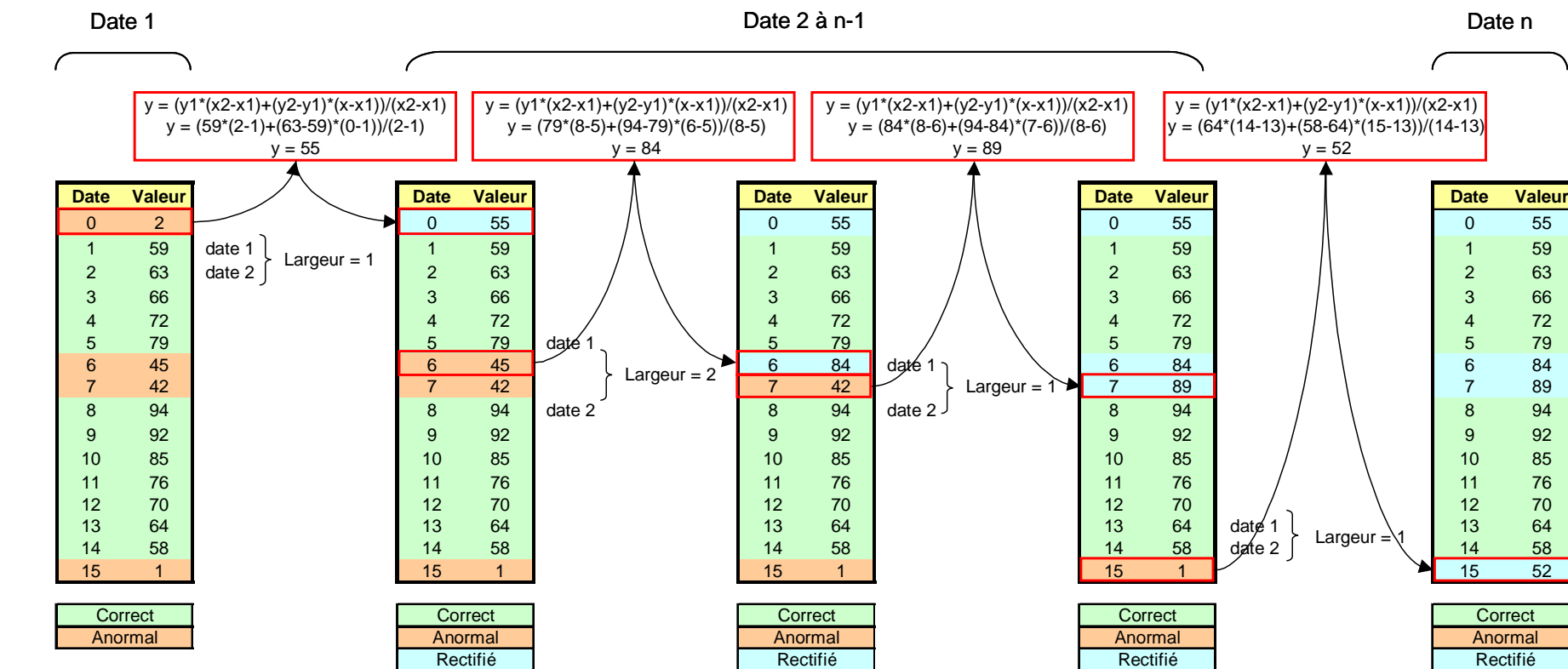
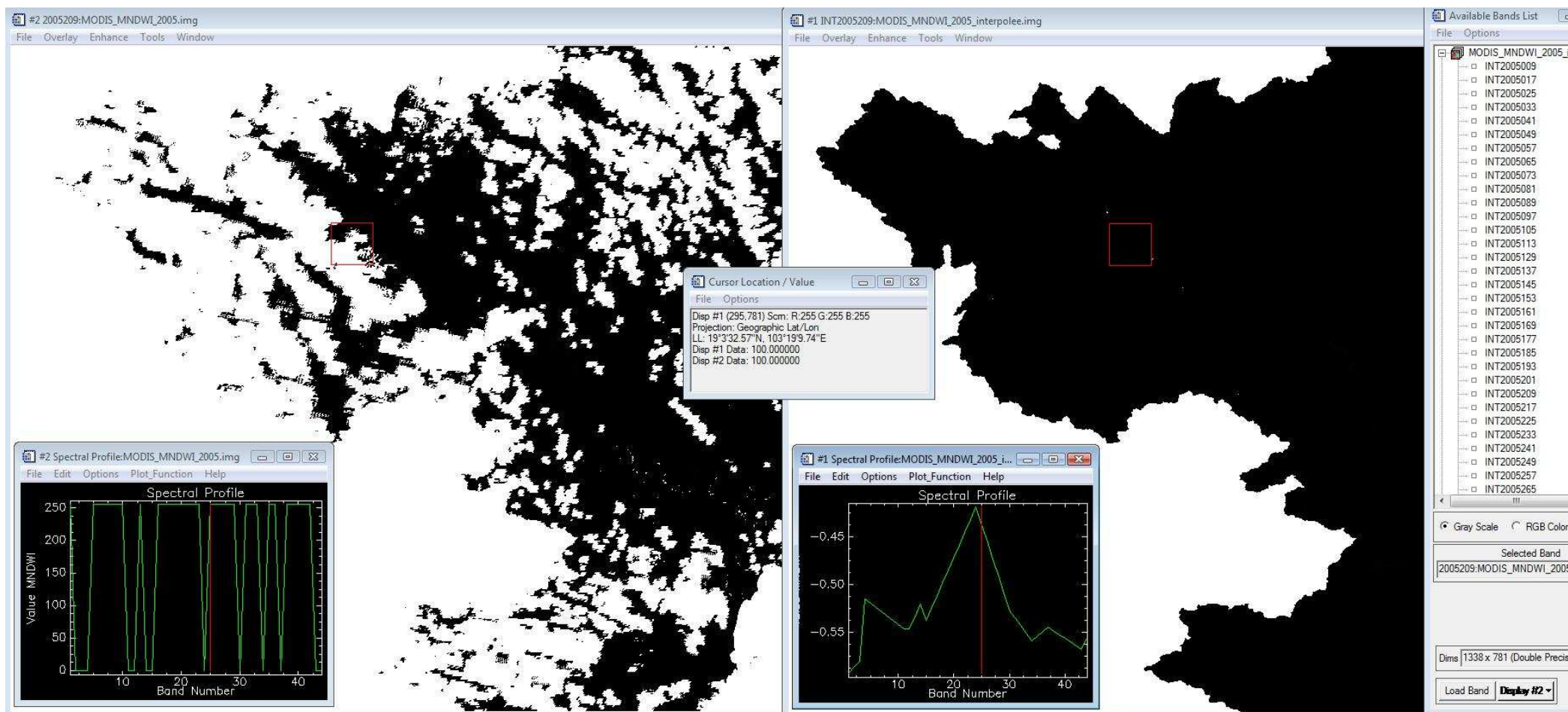


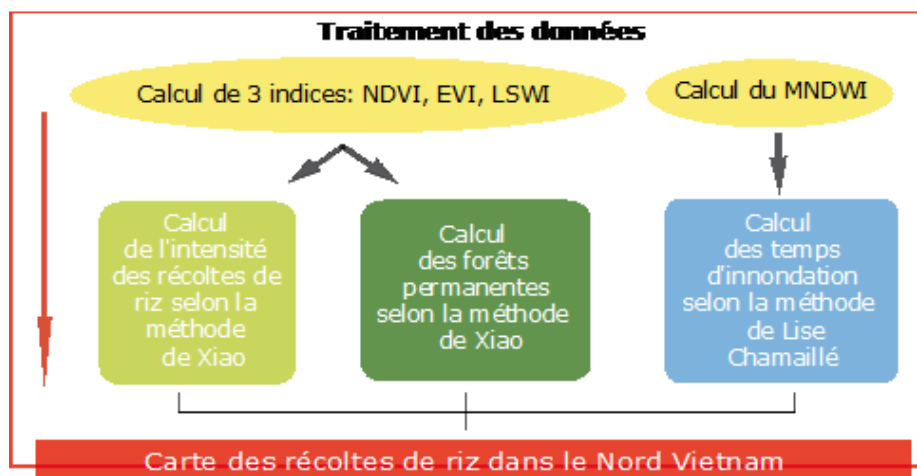
Figure 15 : schéma de l'interpolation temporelle, M Houllès



Cette capture d'écran permet d'appréhender concrètement le résultat de l'interpolation temporelle réalisée sur une série temporelle MODIS. Les deux profils temporels sont issus du même pixel mais celui de gauche se situe avant le traitement et celui de droite, après. Il apparaît très clairement que le profil de gauche est affecté par des perturbations nuageuses qui font monter la valeur du pixel à 255 (valeur que nous avons défini pour les nuages). Il est intéressant de noter l'occurrence et la durée importante de ces périodes de couverture nuageuse démontrant toute l'importance de ce traitement.

3. Traitements des données

Dans l'esprit du cheminement voulu pour notre algorithme, la dernière étape se réalise sur des séries temporelles nettoyées. Comme nous l'évoquions en méthodologie, l'avantage de notre démarche reste que les images MODIS traitées sont prêtes pour toutes futures études et traitements sur le Vietnam:



Sans perdre le fil de notre démarche qui reste de réaliser un algorithme de détection de l'intensité des récoltes de riz, le calcul des indices peut intervenir. Issu de l'héritage des travaux de Xiao et Chamailé, cette phase associera les deux méthodes qui semblent complémentaires.

Plutôt que d'utiliser le NDVI pour détecter l'eau, comme le préconise Xiao, nous avons utilisé l'image résultante de l'algorithme de détection des zones en eaux et de leurs durées d'inondation faisant appel au MNDWI. Sur cette même thématique, là où Xiao place un seuil de 10 périodes de 8 jours de présence d'eau par an pour définir les eaux permanentes, nous avons préféré la périodicité de 30 intervalles minimum.

Pour la détection des autres occupations du sol, nous avons suivi les résultats des études de Xiao sans modification.

Les forêts permanentes sont ainsi définies par deux seuils complémentaires articulés par une opération booléenne OU, le ET étant trop restrictif. Ainsi, tout pixel sur lequel le NDVI est supérieur ou égal à 0.7 durant plus de 20 périodes de 8 jours sur l'année OU où le LSWI est strictement supérieur à 0.10 toute l'année est classé en forêt.

La rizière, objectif avoué de notre traitement algorithmique, qu'il fallait distinguer des deux autres surfaces que nous venons de voir, est déterminée par des changements de comportements des profils temporels de trois indices. Ainsi, est défini comme rizière, tout pixel sur lequel le LSWI +0.05 est supérieur ou égal au NDVI ou à l'EVI ET où deux mois après, les valeurs du NDVI et de l'EVI sont

au maximum. En comptant le nombre d'inversions qu'il se produit par an, nous obtenons facilement le nombre de récolte.

Le croisement de l'ensemble de ces résultats et la distinction entre ces trois types de surface opérée, une classification de l'occupation du sol au Nord Vietnam a été établie.

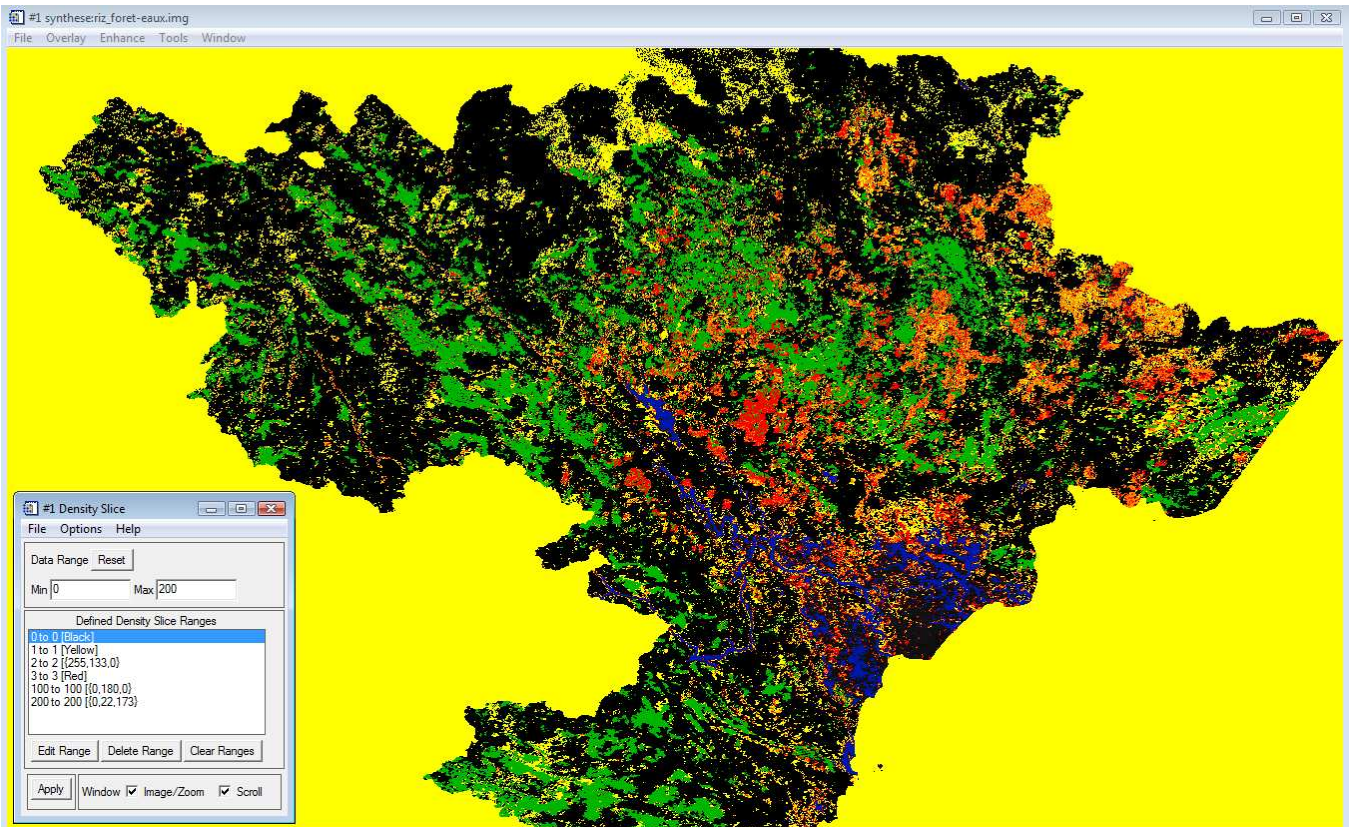


Figure 16 : Illustration du résultat intermédiaire de l'algorithme

Les pixels de valeur 1 sont classés en rizières 1 récolte.

Ceux de valeur 2, en rizière 2 récoltes

Ceux de valeur 3, en rizière 3 récoltes

Ceux de valeur 100, en forêt permanente

Les derniers, de valeur 200 représentent les eaux permanentes

N'étant que provisoire et nécessitant une validation par visualisation de l'équipe vietnamienne et par prospection de terrain, cette classification n'a pas fait l'objet d'une mise en page. Elle demeure, néanmoins, prête à être confrontée aux données terrain et à la classification qui en fera suite.

C. Vérification terrain et résultats finals

La validation par le terrain reste la méthode la plus sûre de confrontation à la réalité, quand elle est possible... En effet, elle demeure lourde à mettre en place et peut être onéreuse notamment lorsque le terrain d'étude est un pays éloigné. Nous avons eu la chance d'avoir cette opportunité.

1. Zones de recherche et procédure

Le projet GRIPAVI ayant ciblé, comme zone d'étude, les écosystèmes du delta du Fleuve Rouge et de la province de Bac Giang, ce choix de terrain de validation s'est tout naturellement imposé à nous. La connaissance du terrain dont sont garants les chercheurs vietnamiens qui collaborent au projet fut également d'une grande utilité.

Un planning est ainsi établi pour 8 jours de terrain durant lesquels de multiples tâches seront à effectuer.

Des relevés GPS avec photos d'appuies et une fiche descriptive par point seront établis sur l'ensemble de la zone. Cette dernière indiquera le type de surface relevé, la distance de la photo à l'objet, l'orientation de l'objet par rapport au point relevé,... Le géoréférencement des photos s'est effectué très simplement, en synchronisant l'heure de l'appareil photo avec celle du GPS puis en créant un fichier de trace où les photos sont positionnées à l'aide du logiciel **GPicSync** suivant la procédure explicitée sur ce lien : <http://code.google.com/p/gpicsync/>.

Un questionnaire est également mis en place et servira à la collecte des informations complémentaires en ce qui concerne les rizières, notamment le nombre de récolte et le type de culture de riz (irrigué, arrosé par les pluies, de zone inondable).

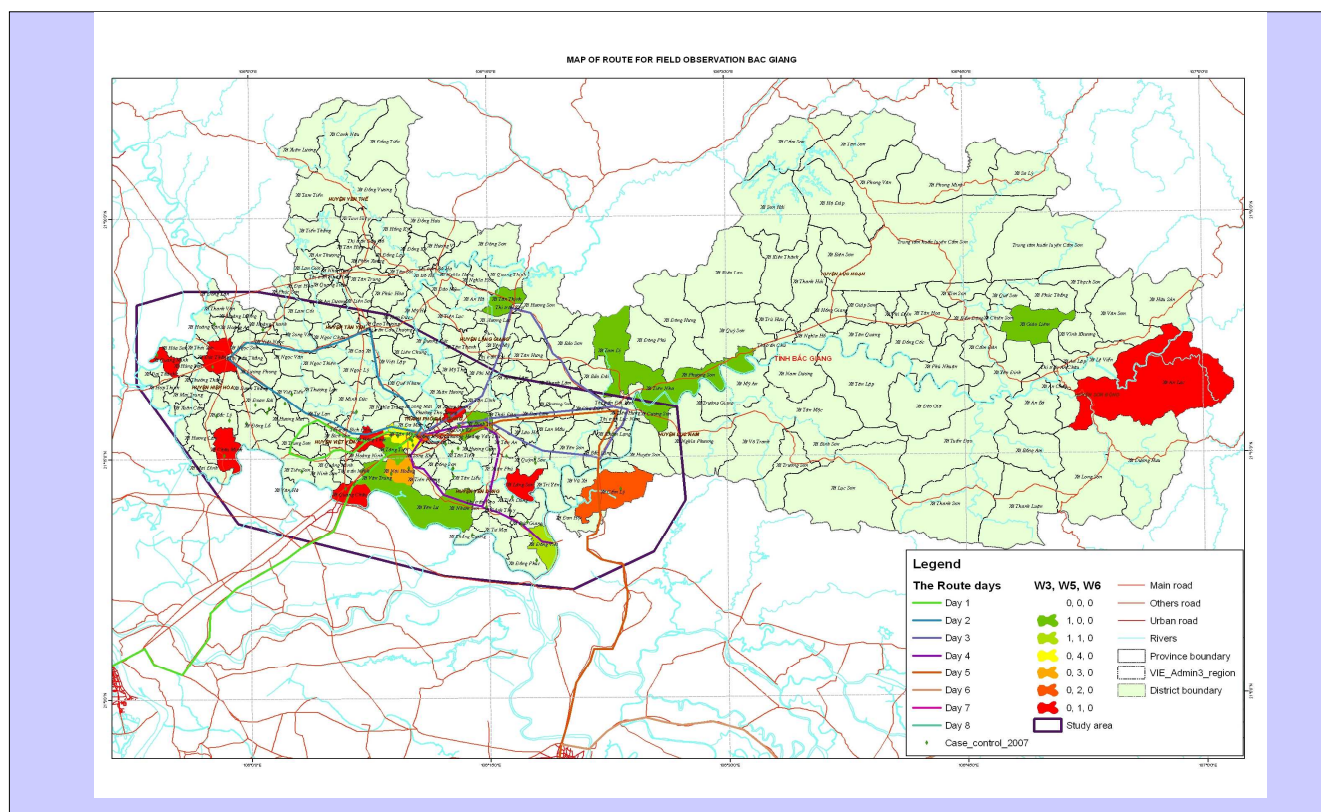
Des interviews des responsables des Offices de l'Agriculture et du Développement Rural sont également prévus afin d'obtenir des informations sur le calendrier des récoltes, le type de riz, le nombre de récolte par an.

L'ensemble de ces procédures sont décrites dans l'annexe « validation terrain ».

Le travail de terrain s'est déroulé du 20 au 28 mai 2009 comme suit :

<i>Jour</i>	<i>Programme</i>	<i>De - vers</i>	<i>Distance (km)</i>	<i>Remarques</i>
1	- 7h: depart de 86 Truong Chinh. -9h: route vers Bac Giang city. -De Bac Giang city vers Viet Yen, observations terrain et retour à Bac Giang city à 17h. Une nuit à Bac Giang	Ha Noi – Bac Giang city – Viet Yen –Bac Giang city	100	Jour de départ (2009-05-20)
2	Départ vers Viet Yen, Hiep Hoa, Tan Yen et retour à Bac Giang city.	Bac Giang city – Viet Yen – Hiep Hoa – Tan Yen - Bac Giang city	80	
3	-7h30 départ vers Lang Giang et Luc Nam district et retour Bac Giang.	Bac Giang city – Lang Giang – Luc Nam - Bac Giang city	70	
4	- Le matin : rdv à l'Office d'Agriculture et du développement rural - Après départ vers Yen Dung et retour vers Bac Giang city.	Bac Giang city – Yen Dung - Bac Giang city	50	Collecte des calendriers de récolte à l'Office d'Agriculture et du développement rural. Bac Giang
5	- Le matin : rdv à l'Office d'Agriculture et du développement rural - Après départ vers Chi Linh et départ pour Hai Duong (1 nuit)	Bac Giang city – Chi Linh – Nam Sach – Hai Duong city - Tu Ky – Gia Loc - Hai Duong city	100	Collecte des calendriers de récolte à l'Office d'Agriculture et du développement rural. Hai Duong
6	Départ vers Hai Phong city – Thuy Nguyen district et départ pour Hai Phong (1 nuit)	Hai Duong city – Hai Phong city – Thuy Nguyen district – Hai Phong city	100	
7	Départ vers An Lao – Vinh Bao et départ vers Thai Binh (1nuit)	Hai Phong city – An Lao – Vinh Bao – Thai Binh city	120	
8	Collecte des dernières informations	Thai Binh city – Nam Dinh - Ninh Binh – Ha Noi city	200	Jour de retour (2009 -05 – 28)
Total			820	

Les cartes qui nous ont servis de “road book”:



MAP OF ROUTE FOR FIELD OBSERVATION HANOI - BAC GIANG - HAI DUONG - HAI PHONG - THAI BINH - NINH BINH

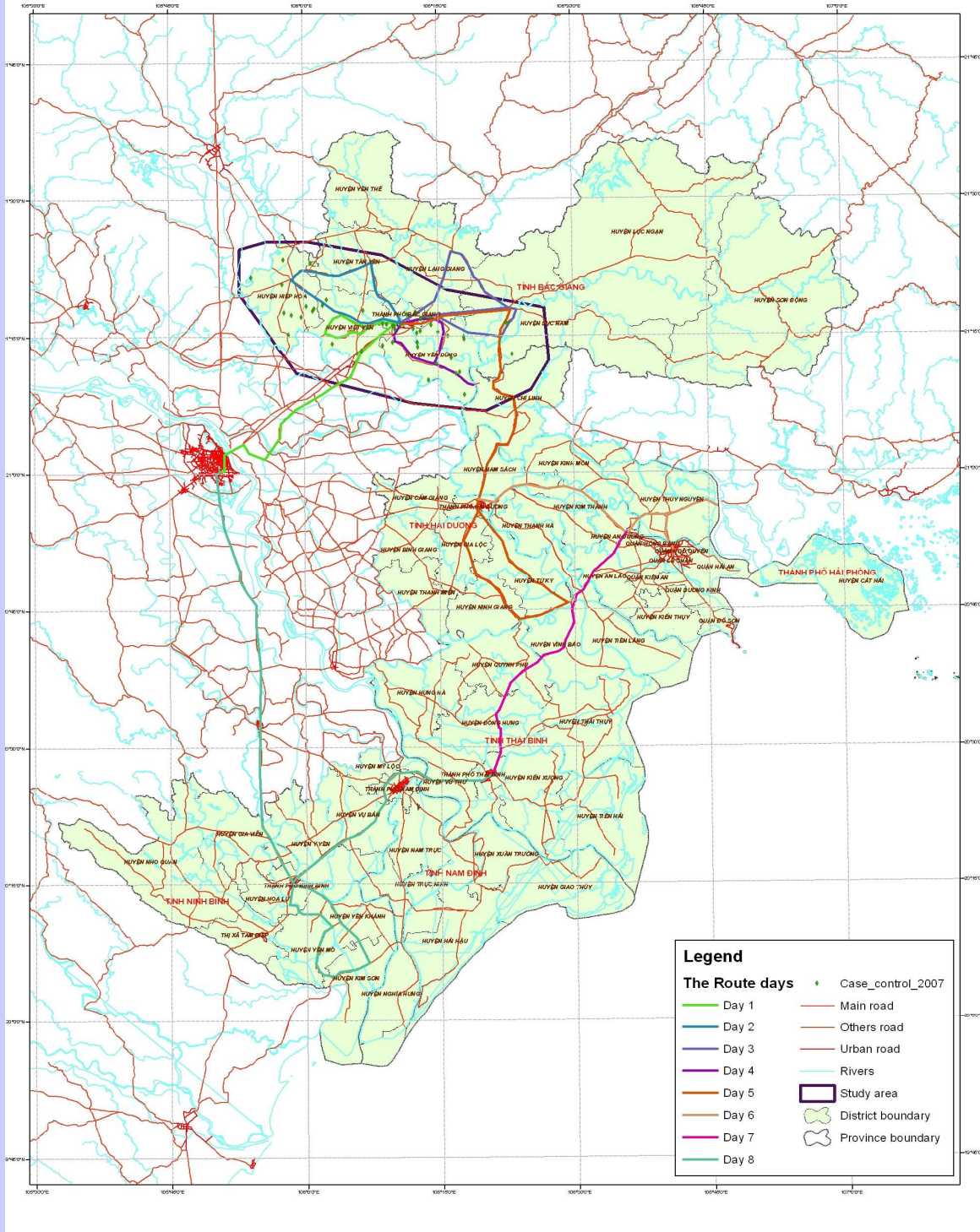


Figure 17 les cartes itinéraires du travail d'observation terrain

L'ensemble des informations apportera un échantillonnage des occupations du sol présents sur la zone qui nous servira ensuite pour la création de ROI (Region Of Interest), étape obligée d'une classification supervisée.

2. Résultats et classification SPOT

Les 8 jours de prospection terrain ont permis la collecte de 73 points d'observation géoréférencés et renseignés. (Cf l'annexe « validation terrain ». Une table a été créée avec l'ensemble des renseignements dont nous disposons : localisation approximative, nom du district, de la province, type d'occupation, nombre de récolte, période de croissance, autre culture, période de croissance de l'autre culture, coordonnées X et Y.

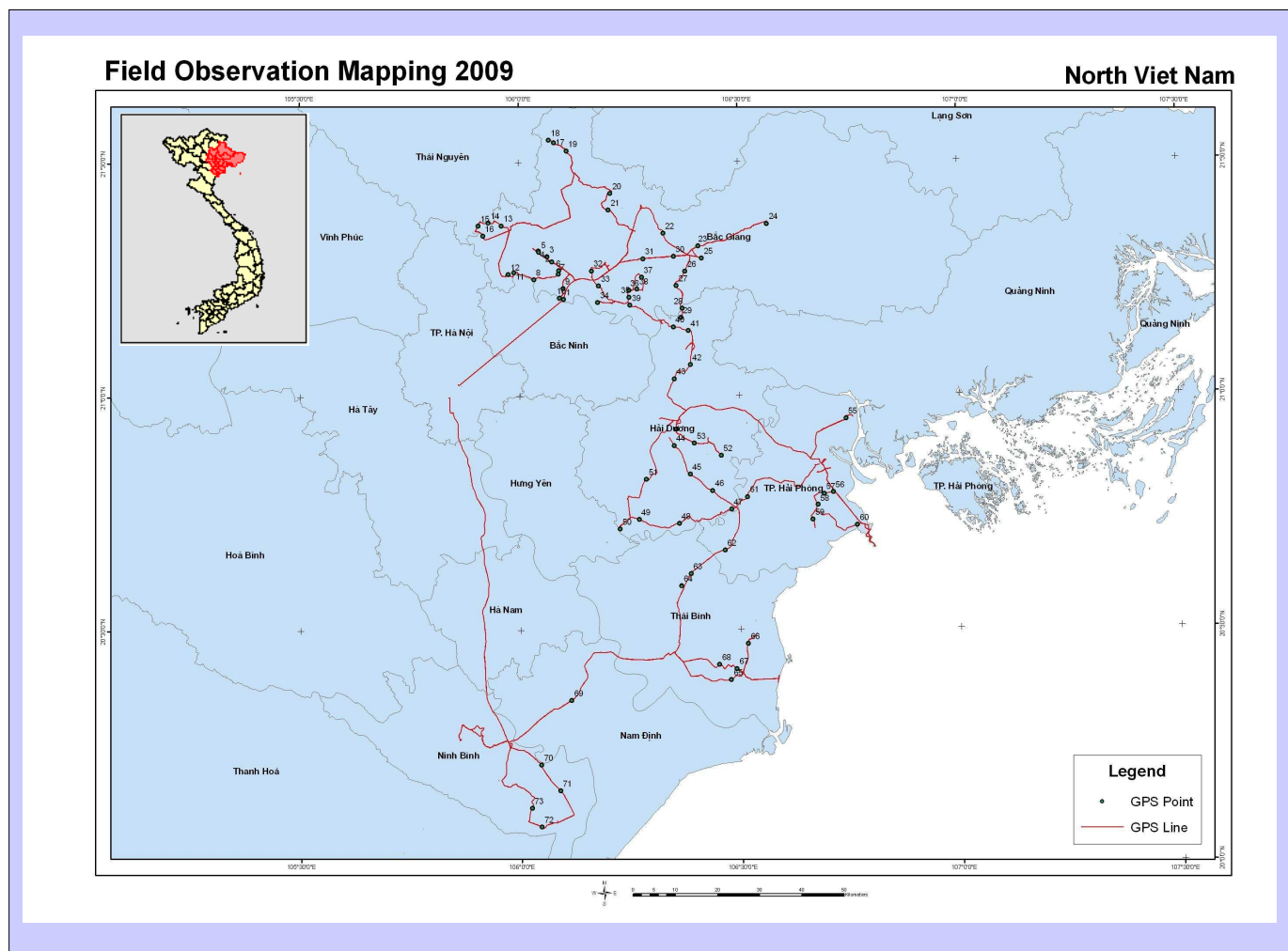
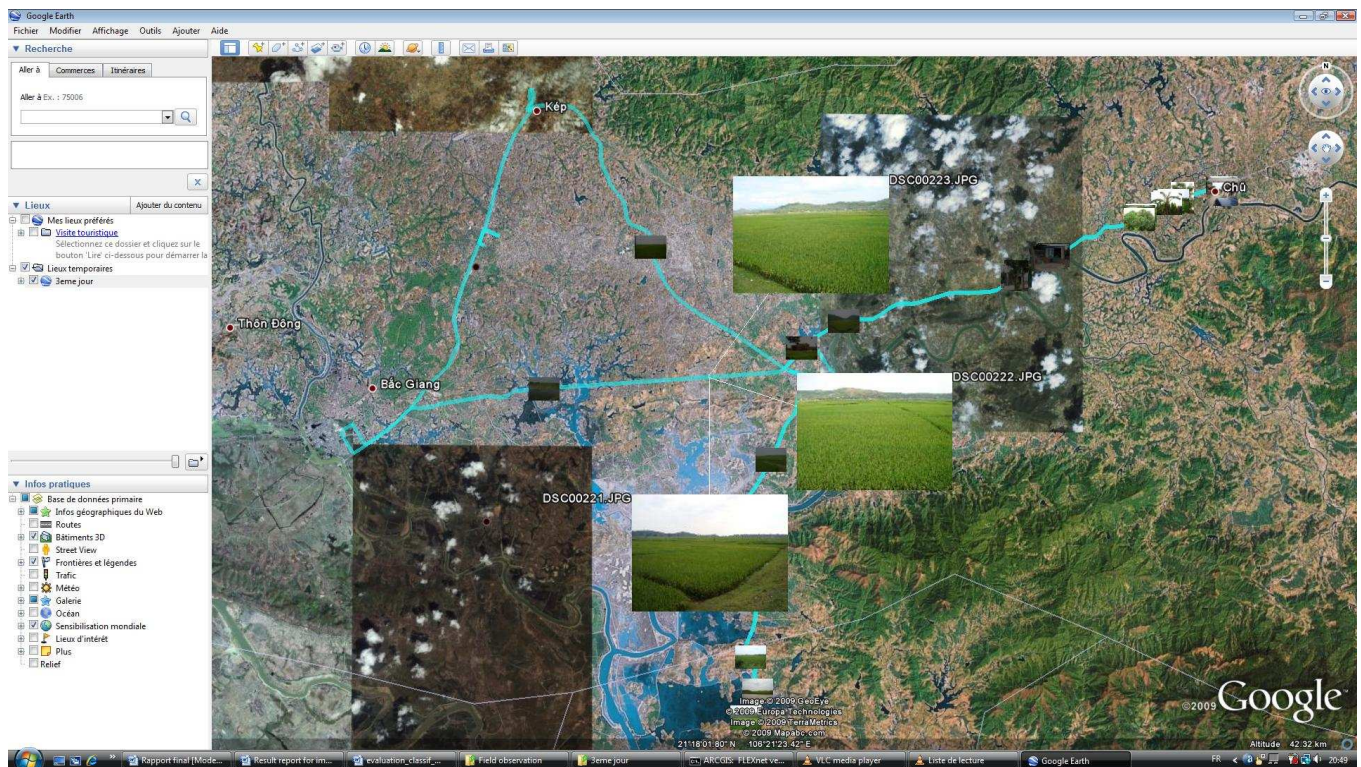


Figure 18 : Relevés GPS du travail d'observation terrain

Intégrés au sein du logiciel ENVI, ces points ont participé à la création de ROI par la connaissance du terrain acquise lors de ces 8 jours mais également par interprétation des pixels environnants celui du point GPS.

L'utilisation de la méthode de géoréférencement des photos a également apporté une amélioration à la dimension d'interprétation de l'environnement proche des points GPS relevés. A titre indicatif, le rendu peut être visualisé sous Google Earth© apportant une résolution plus élevée.



L'idée de départ était d'utiliser ces relevés sur l'image MODIS afin d'effectuer un comparatif à échelle et résolution égale. Le problème réside dans le fait que sur un pixel de 500m par 500m, plusieurs de nos points d'observation sont localisés rendant le choix typologique impossible.

Nous avons donc opté pour une classification sur les images SPOT dont nous disposons. De résolution beaucoup plus fine, les images SPOT se prêtent tout à fait à une classification à l'aide des ROI issus de nos relevés terrains. Cependant plusieurs problèmes demeurent.

Pour commencer, l'ensemble de nos relevés s'étendent sur 5 provinces quant la couverture SPOT dont nous disposons, ne couvre que celle de Bac Giang. Mais le plus problématique reste que nos images SPOT sont prises à différents temps (jour et mois) par différents capteurs. Par exemple, nous disposons de deux images prise le 02/12/2008 par SPOT4, une le 16/10/2008 par le même capteur, deux prises le 12/11/2008 et une le 11/10/2008 par SPOT2. Dernier point négatif, une des images SPOT2 est très nuageuse rendant son utilisation difficile.

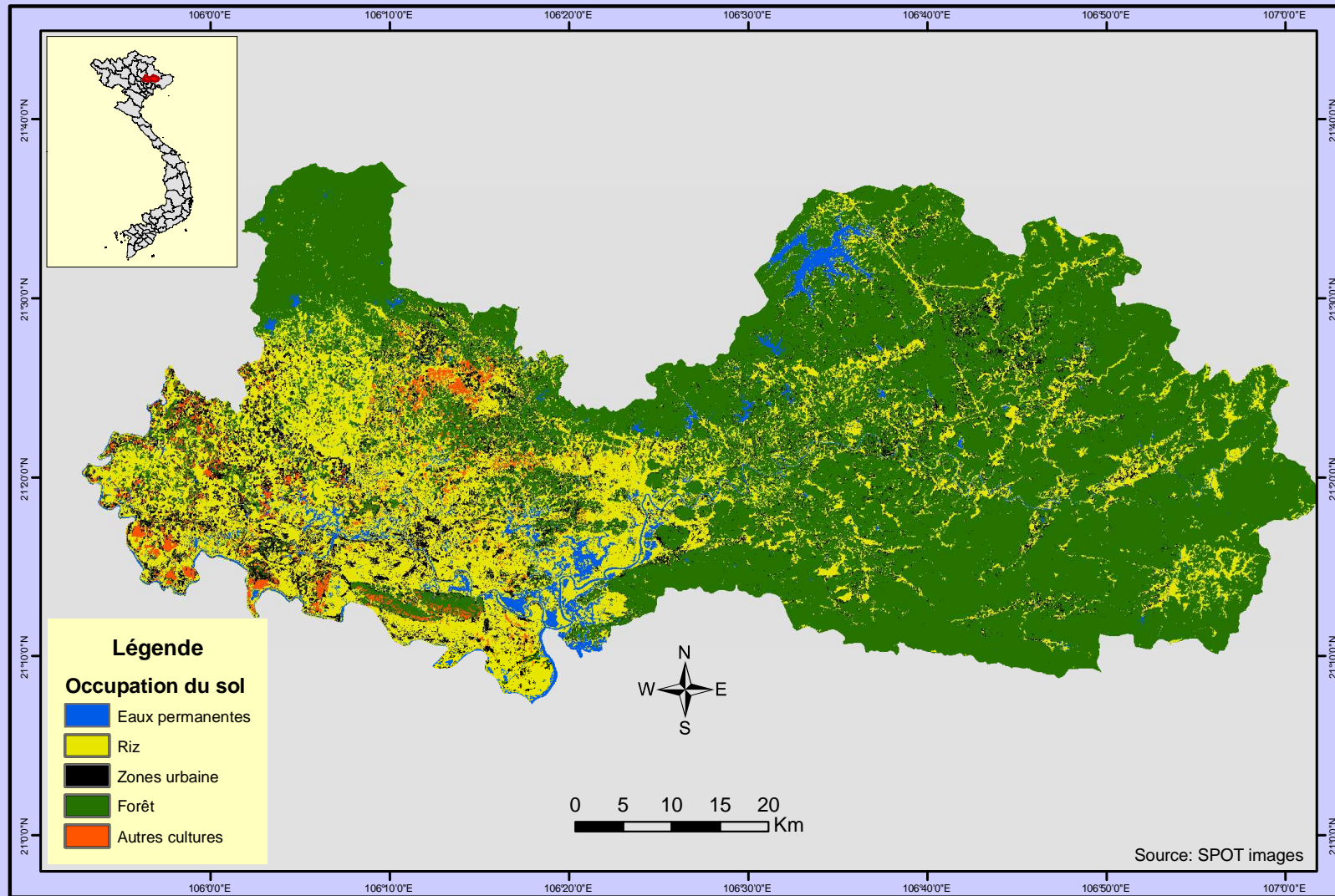
Une mosaïque a donc été nécessaire : les résultats sont meilleurs quand cette opération est réalisée en dernière, après avoir effectué les classifications, plutôt qu'avant.

La classification supervisée Maximum Likelihood a été retenue pour transformer nos ROI en une classification référence pour le test de matrice de confusion qui validera ou infédera notre algorithme de détection de zones de riziculture.

Le résultat sur la province de Bac Giang semble maintenant acceptable : il ne reste plus qu'à le comparer à celui obtenu à partir des images MODIS par l'algorithme sous IDL.

Occupation du sol

Province de Bac Giang



Réalisation: S. Tollis, Nguyen Phan Dong, Pham Ngoc Hai

Figure 19 : Occupation du sol de la province de Bac Giang obtenue par classification des images SPOT

3. Correction et classification finale

Comme nous l'évoquons dans le chapitre précédent, la validation terrain permet un rapprochement de la réalité. Le contact que nous avons pu avoir avec ce terrain que nous avons cartographié à 10 000 kilomètres de là, a révélé les erreurs et les limites de notre méthode sans que nous ayons à comparer les classifications.

Ainsi un décalage est observé sur toutes nos images. Si les limites du Nord Vietnam sont respectées, l'ensemble des pixels à l'intérieur subissent un décalage tout en respectant les alignements. La solution a été trouvée en deux temps : en tâtonnant tout d'abord, en visionnant les images après chaque étape de l'algorithme, par reprojexion ensuite, en créant un batch sous IDL car l'outil MRTWeb du site de téléchargement MODIS semble défailant. La première étape a révélé que c'est l'application, dès le début de l'algorithme, du masque des contours du Vietnam qui créait le décalage : son application en dernière position, suite à l'ensemble des traitements sur l'image entière, semble résoudre ce problème. La seconde solution résout également le problème car c'est bien l'outil de reprojexion qui est la cause du problème. 1 Problème, 2 solutions.

Dès la présentation du résultat de notre classification aux chercheurs vietnamien de CARGIS, ces derniers ont émis des doutes quand à la répartition des différents type d'occupation du sol notamment la forêt qui semble sous représentée. Après vérification des scripts IDL et relecture des travaux de Xiao, le problème est décelé: une erreur d'opérateur booléen est en cause. Les deux conditions sont bien présentes mais articulés avec un ET et non un OU limitant le nombre de pixel concernés. La connaissance du terrain des vietnamiens a été la clé de cette résolution.

Dernière erreur mais la plus grave par rapport aux objectifs fixés par notre étude : le nombre maximal de récolte de riz par an dans le Nord Vietnam est de deux et non de trois comme se peut être le cas dans le sud ou dans d'autres pays d'Asie (Xiao et al. 2006). Le travail de terrain et les résultats des enquêtes ont permis de connaître le calendrier des récoltes et d'ajuster notre algorithme.

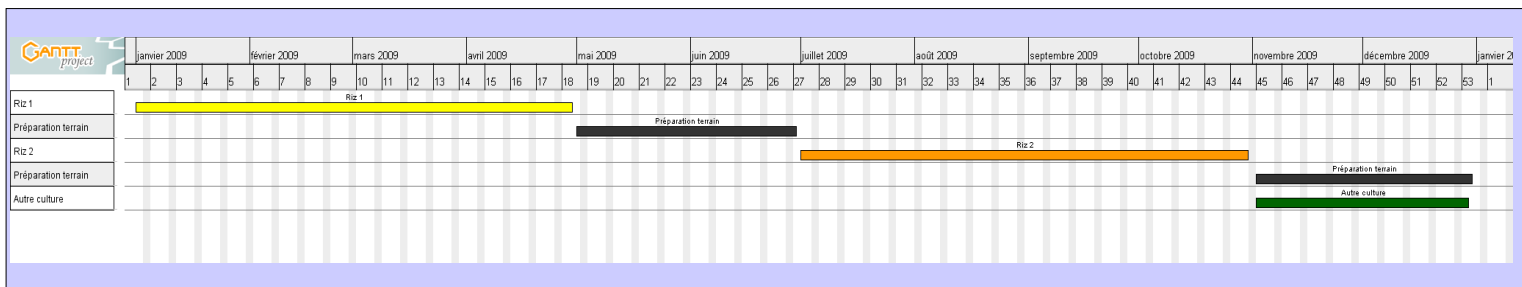


Tableau 2: le calendrier théorique du riz dans le Nord Vietnam

Ainsi, il est possible d'avoir trois récoltes sur une même rizière mais ce sera toujours deux maximum de riz et une, d'autres plantes tel que le liseron d'eau. En effet, contrairement au sud du pays, le cycle du riz s'étale sur 4 mois auxquels il faut ajouter une ou deux périodes de préparation du terrain, la seconde pouvant être remplacée par une autre culture.

Concernant la détection du riz, cette mise au point n'était pas suffisante. En vérifiant sur certains pixels, la signature temporelle des indices, alors qu'il n'y avait qu'une seule inversion $LSWI + 0.05 > NDVI$ ou EVI, notre algorithme en détectait trois car le LSWI peut rester supérieur au NDVI plusieurs semaines

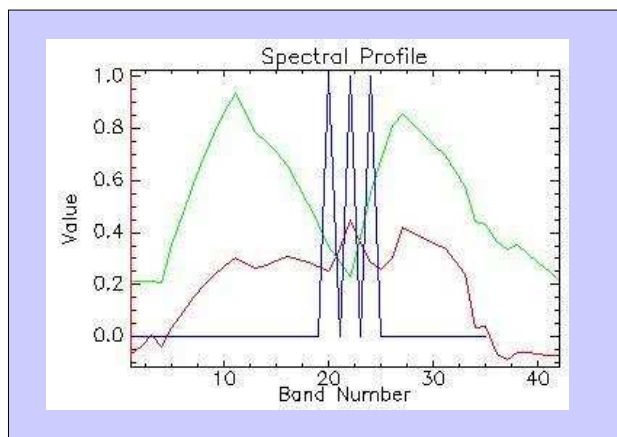


Illustration du problème :
Signature temporelle des indices
NDVI (vert), LSWI(rouge) et de la
détection de plantation de riz (bleu),
sur un pixel classé 3 récoltes de riz /
an.

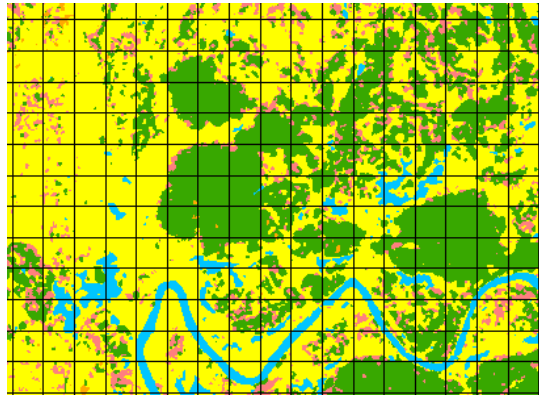
Une solution a été trouvée : l'ajout d'un tableau "nb_weeks_last_flood" qui compte le nombre de semaines écoulées après une première plantation de riz, pour chaque pixel. Ce compteur est mis à zéro la première fois qu'une mise en eau / plantation est détectée ($LSWI + 0.05 > NDVI$) et incrémenté chaque semaine suivante. Quand il devient supérieur à 20 semaines (4 mois riz + 1 mois préparation), une nouvelle détection de plantation de riz est possible.

Ces petits problèmes corrigés, la classification MODIS devient beaucoup plus « vraisemblable ». Tout l'intérêt de la connaissance du terrain ou de la validation par le terrain prend toute sa signification.

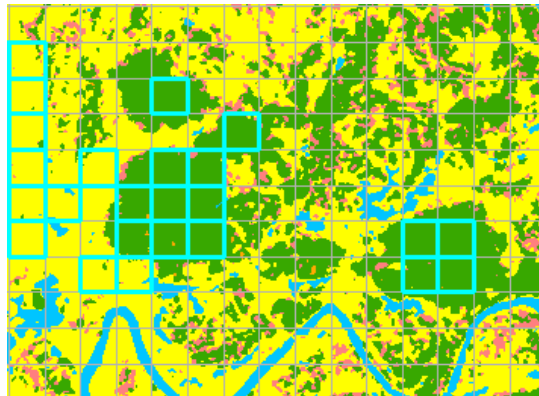
Il est temps de passer à la dernière phase de notre démarche : l'évaluation de la qualité notre classification à partir de la classification de l'image SPOT.

La première étape consiste à sélectionner les pixels homogènes.

A cette fin, sous ArcGIS, la création d'une grille de 500 x 500 m correspondant à la taille des pixels MODIS (Outils Jensen enterprise / Repeating shapes) paraît nécessaire



Nous croisons ensuite les superficies (spatial analyst -> tabulate area) avec la classification SPOT afin de se retrouver à la même résolution et nous sélectionnons uniquement les pixels MODIS homogènes (minimum 90% du même type d'occupation du sol).



A partir des pixels sélectionnés (3017 pixels homogènes), le calcul d'une matrice de confusion paraît cohérent.

		Classification SPOT			
		Forêt	Riz	Eau	Total
Classification MODIS	Forêt	2613	9	3	2625
	Riz	84	206	17	307
	Eau	21	50	14	85
	Total	2718	265	34	3017

Overall accuracy = 0.94%

Kappa coefficient = 0.705

Les résultats sont très corrects et la validation de notre méthode de détection ainsi que la classification qui en a découlée peut être approuvée. Certains détails subsistent comme quelques confusions eau/riz mais cela paraît

normal tant les deux milieux sont proches. L'existence de nombreux pixels non classés apparaît également pouvant nécessiter un ajustement du seuil de détection des forêts ou du riz. Partant du principe qu'il vaut mieux un pixel non classé que mal classé, nous laissons le soin d'améliorer notre algorithme aux études futures.

La prise en compte des contraintes d'altitude (<2000 m) et de pente ($<2^\circ$), comme le suggère Xiao dans sa méthode, peut être une piste intéressante. Dans notre cas de figure, nos zones d'étude étant situées en delta ou dans des zones peu accidentées, l'intérêt d'appliquer ce facteur restait minime.

L'extraction des paramètres environnementaux pour la caractérisation de l'occupation du sol à l'échelle régionale étant achevée, l'analyse spatiale du comportement du H5N1vis à vis d'indicateurs environnementaux à des échelles plus locales peut commencer.

Classification Modis:
carte de l'intensité des récoltes de riz

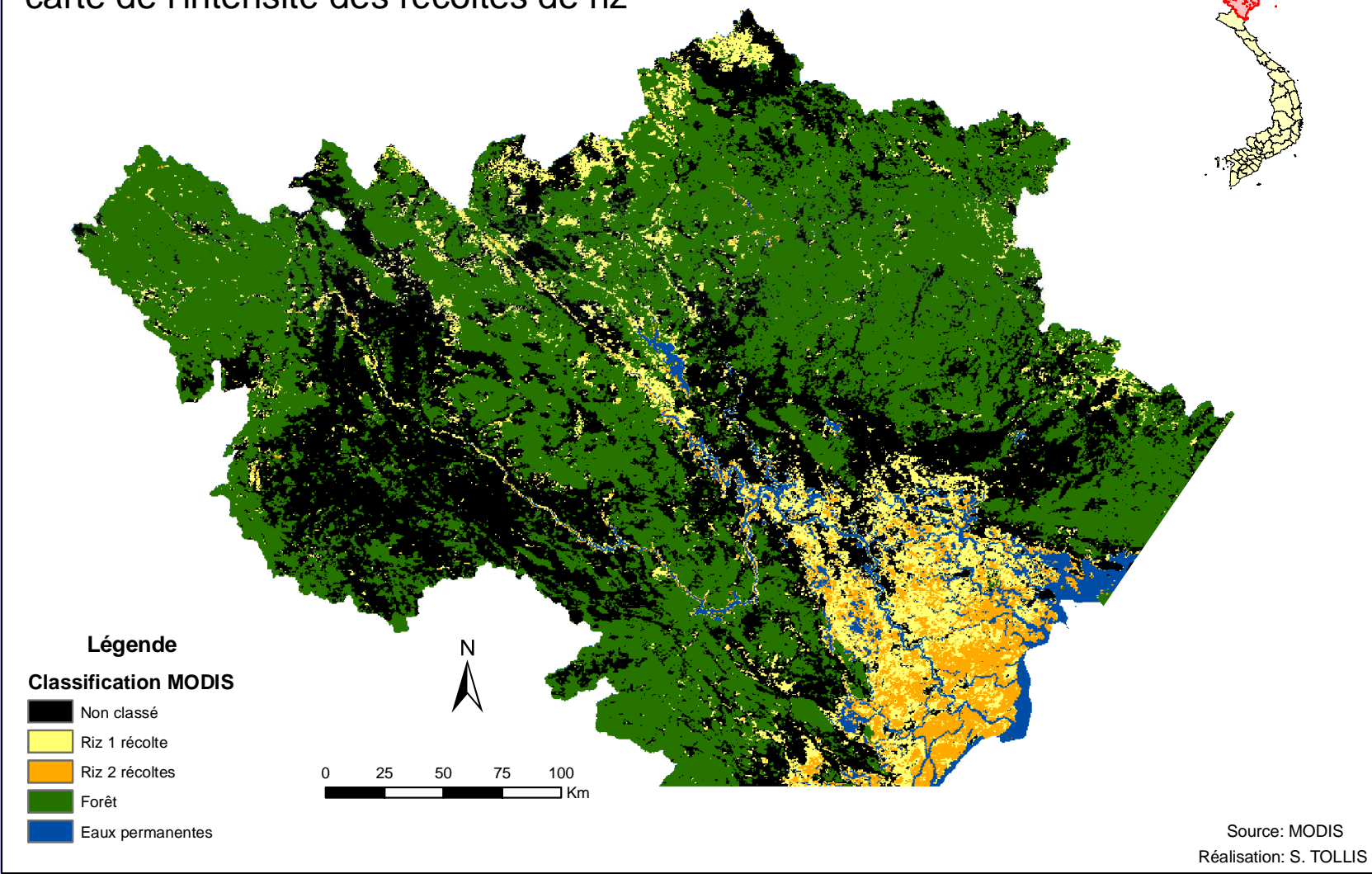


Figure 20 : Résultat cartographique de notre algorithme de détection de l'intensité des rizières

III.L'apport de l'analyse spatiale par SIG pour la définition d'indicateurs environnementaux liés au risques de H5N1 au Vietnam

La phase de détection des rizières et des durées d'inondation étant achevée, nous disposons de toutes les données environnementales nécessaires à la constitution d'un SIG répondant aux besoins d'analyse de notre étude. Partant des hypothèses de travail des épidémiologistes du projet selon lesquelles certains facteurs environnementaux ont une influence sur les lieux d'émergence et de développement du virus H5N1, notre objectif est d'apporter une expertise par le prisme de l'analyse spatiale et ainsi valider ou inféoder les pistes de recherche. Volontairement, nous sommes limités aux indicateurs auxquels les chercheurs du CIRAD semblent accorder le plus de crédibilité à savoir, les réseaux routiers, ferroviaire et hydrographique, l'occupation du sol, la présence ou non de rizières, l'intensité des récoltes sur ces dernières ou les temps d'inondation. Croisées aux données épidémiologiques et notamment les villages et communes avec foyers de H5N1, des analyses statistiques de corrélations et de comparaison ont pu être exploitées. Deux types de calculs nous ont servi à évaluer la pertinence de ces différentes variables: la distance et la densité.

A. Acquisition et organisation de la base de données

Comme nous l'évoquions dans la première partie de ce rapport, à notre arrivée sur notre terrain d'étude, le Vietnam, l'épidémiologiste avec laquelle nous travaillons avait déjà effectué l'achat de deux jeux de données auxquels il convient d'ajouter celui de notre laboratoire partenaire.

1. Description des données spatiales disponibles

a) Les lots du CIRAD

Première distinction : une hétérogénéité des couches. Notre objectif est justement de faire une sélection des couches en fonction de la qualité des données ainsi que de la thématique abordée. C'est essentiellement sur ce point que nous avons travaillé puisque, si les couches de données administratives étaient les mêmes (province, district, commune), celles thématiques demeureraient complémentaires voire différentes.

Seconde particularité de ces deux lots de données : leurs dates de création. Le plus ancien date de 2002 mais sa source n'est pas connue, et, le plus récent de

2004 mais a bénéficié d'une mise à jour jusqu'en 2008 de la part de la Maison de la Cartographie du Vietnam.

S'il est aisé de choisir les couches thématiques selon les besoins scientifiques de notre étude lorsque celles-ci ne sont présentes que dans un des deux sets de données, il l'est beaucoup moins lorsqu'elles le sont dans les deux : sur quels critères choisir plutôt l'une que l'autre ? Comment justifier notre choix ? Posséder un troisième jeu qui ferait office de référent semble représenter la meilleure et plus simple solution.

Données 2002	Données 2008	Type de données	Légende
lep50	/	Polygone	Contours du Vietnam
country			
admin1	VnAdmin1_region	Polygone	Carte des provinces
admin2	VnAdmin2_region	Polygone	Carte des districts
admin3	VIE_Admin3_region	Polygone	Carte des communes
popdens	/	Polygone	Densités de population
rivers	/	Polyligne	Réseau hydrographique principal
riversCIAS	/	Polyligne	Réseau hydrographique secondaire
Lakes	VnRivers2_region	Polygone	Lacs et grands fleuves
LakesCIAS	vnivers1_polyline	Polygone	Lacs et tous les fleuves
/	vntransport1_polyline	Polyligne	Ensemble du réseau routier et ferroviaire
roads	/	Polyligne	Réseau routier principal
roadsCIAS	/	Polyligne	Réseau routier secondaire
/	Vntransport2_point	Point	Gare et ouvrages d'art
railroads	/	Polyligne	Réseau ferroviaire
/	PopulationN_region	Polygone	Zones urbanisées du Nord Vietnam
/	PopulationN_font_point	Point	?
/	PopulationNameN_text	Point	Noms de tous les villages
urb	/	Polygone	Zones urbaines ?
Forest	/	Polygone	Zones forestière
wetlands	/	Polygone	Zone de marais
elev	/	Polygone	Altitude
airp	/	Point	Hauts lieux touristiques
18	10		Nombre total de couche

Tableau 3: Tableau des différentes couches présentes dans les jeux de données du CIRAD

b) Le SIG référent

Le laboratoire CARGIS possède également un SIG du Vietnam qu'il utilise et peut être considéré comme fiable. Son rôle, dans notre cas de figure, sera de

servir de référent pour faciliter le choix des couches finales constituant notre base spatiale. Petit bémol dans son utilisation : son système de projection est le Vn-2000 alors que notre projet est en UTM WGS84 zone 48N. Un décalage de 179m s'opère entre les couches des deux systèmes de projection alors que le datum géodésique soit le même (WGS84). Si les couches thématiques ne nécessitent pas réellement ce test puisqu'elles sont sélectionnées en fonction des besoins de l'étude, de leurs précisions et surtout de leurs plus ou moins grandes mises à jour, celles qui contiennent les limites administratives seront comparées à ce dernier afin de s'assurer de l'exactitude des frontières de nos zones d'étude.

2. Sélection des couches

Suivant la phase d'observation des couches et de comparaison au SIG vietnamien de référence, une sélection de couche à intégrer dans la base de données de notre SIG a été opérée.

Le tableau suivant représente les couches conservées et celles écartées :

Données 2002	Données 2008	Type de données	Légende
lep50	/	Polygone	Contours du Vietnam
eountry	/	Polygone	Contours du Vietnam
admin1	VnAdmin1_region	Polygone	Carte des provinces
admin2	VnAdmin2_region	Polygone	Carte des districts
admin3	VIE_Admin3_region	Polygone	Carte des communes
popdens	/	Polygone	Densités de population
rivers	/	Polyligne	Réseau hydrographique principal
riversCIAS	/	Polyligne	Réseau hydrographique secondaire
Lakes	VnRivers2_region	Polygone	Lacs et grands fleuves
LakesCIAS	vn_rivers1_polyline	Polygone	Lacs et tous les cours d'eau
/	vntransport1_polyline	Polyligne	Ensemble du réseau routier et ferroviaire
roads	/	Polyligne	Réseau routier principal
roadsCIAS	/	Polyligne	Réseau routier secondaire
/	Vntransport2_point	Point	Gare et ouvrages d'art
railroads	/	Polyligne	Réseau ferroviaire
/	PopulationN_region	Polygone	Zones urbanisées du Nord Vietnam
/	PopulationN_font_point	Point	?
/	PopulationNameN_text	Point	Noms de tous les villages
urb	/	Polygone	Zones urbaines ?
Forest	/	Polygone	Zones forestière
wetlands	/	Polygone	Zone de marais
elev	/	Polygone	Altitude
airp	/	Point	Hauts lieux touristiques
5	9		Nombre total de couche

Tableau 4: Sélection des couches pour la constitution du SIG

Dans la globalité, ce sont les couches issues du jeu de données de 2008 qui ont été conservées. Elles présentent l'avantage d'être de conception plus récente

mais surtout d'avoir été mises à jour régulièrement jusqu'en 2008. Elles se superposent parfaitement avec le SIG référant (petit décalage dû au système de projection déjà évoqué plus haut) là où le jeu de 2002 présente des écarts plus importants, mais surtout, les limites administratives ont des formes similaires alors que celles de 2002 étalent des dessins pour le moins imprécis. Elles sont souvent plus détaillées également.

Seule la couche du réseau hydrographique secondaire est conservée pour la densité et la précision de ces objets mais nécessite néanmoins une modification pour intersecter correctement le réseau principal. Sous Global Mapper, nous avons modifié le système de projection pour du "Indian Thailand" et ainsi considérablement réduit le décalage avec les fleuves et autres cours d'eau principaux.

L'ancien jeu de données de 2002 possède également certaines couches non présentes dans celui de 2008 et qui pourront être utiles pour les recherches futures : dans un souci de pérennité du SIG, nous les avons conservé. Il s'agit essentiellement des couches d'altitude et de densité de population.

Le SIG final nommé GRIPAVI comportera donc 14 couches thématiques .shp qu'il a fallu organiser, pour ensuite, pouvoir les croiser avec les données issues des recherches épidémiologiques des vétérinaires du CIRAD.

3. Nettoyage des données

a) Les données épidémiologiques

L'étape du choix des données spatiales étant achevée, un travail de nettoyage des bases de données épidémiologiques était nécessaire afin de garantir les résultats des traitements spatiaux futurs exécutés sous ArcGis. Dans l'ensemble, la majeure partie de ce travail a été effectué a posteriori, les chercheurs du projet travaillant sur ces dernières mais quelques erreurs demeurent.

En utilisant l'outil de non correspondance d'Access©, nous nous sommes rendus compte que deux villages de la table case_control2007 possédaient les mêmes coordonnées. Le système administratif vietnamien très compliqué et imprécis notamment à l'échelon villageois ne nous a pas permis de trouver facilement la localisation des deux villages afin de connaître lequel était bien placé. Nous nous sommes ainsi servis du SIG du laboratoire Cargis ainsi que de Google Earth© et d'un site de cartographie vietnamien pour trouver et modifier les coordonnées.

De même, nous nous sommes aperçus que certaines coordonnées des villages cas_controls paraissaient ne pas correspondre et nous avons eu confirmation de ces erreurs en créant les points à l'aide de la fonction "ajouter des données XY" sous ArcMap©. Ces villages étaient localisés en dehors de la

province de Bac Giang voire même du Vietnam ! Comme pour le village précédent, à l'aide de plusieurs sources, nous avons réussi à retrouver les bonnes coordonnées et localiser correctement ces villages.

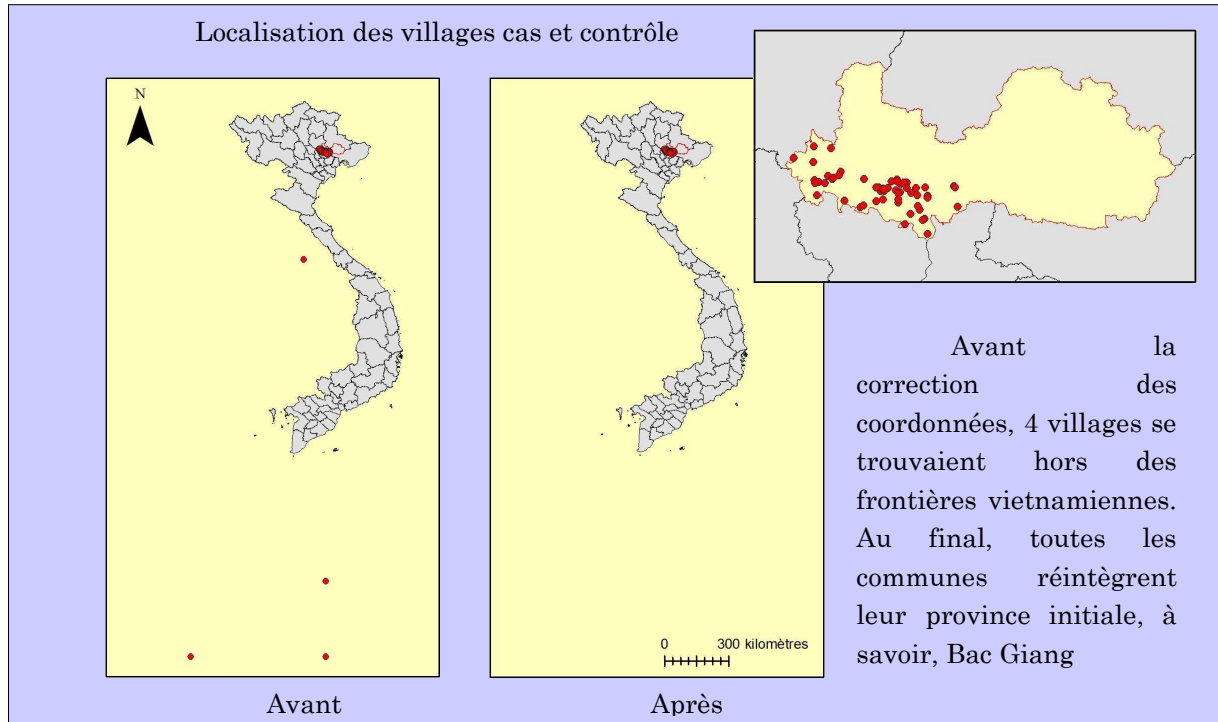


Figure 21: illustration du travail de correction des coordonnées

b) Les données géométriques

Malgré notre première sélection, les données choisies présentaient encore des erreurs qu'il nous a fallu corriger. Ce sont essentiellement des problèmes de trou entre les polygones qui empêchent les découpages de nos couches thématiques aux échelles auxquelles nous désirons travailler.

En cherchant sur le support ESRI ainsi que sur certains forums spécialisés, nous avons trouvés, et essayés, plusieurs méthodes pour résoudre ces problèmes.

- Au sein même d'ArcToolbox©, dans les outils de gestion des données, il suffit d'utiliser la fonction de vérification des géométries puis, si un problème est détecté, de lancer celle de réparation de géométrie. Dans la grande majorité, les problèmes se résolvent mais nous avons tout de même rencontré, des problèmes persistants.
- Dans cette optique, nous avons également testé l'extension ETGeoWizards 9.9© développée par ETSpatialTechniques. Au sein du menu Polygon existe la fonction "Clean Gaps" qui permet de "combler" ces trous. Dans la nouvelle couche créé, les polygones de remplissage sont identifiés par l'attribut "Gap" dans le nouveau

champ créé "ET_Gaps". Il suffit ensuite d'identification des voisins de chaque polygone avec "Polygon / Get Adjacent polygons", de faire une union puis une fusion. La couche est ainsi nettoyée et les traitements peuvent être effectués.

- Dernière solution, utilisée par les chercheurs vietnamiens, se servir d'ArcView3.2© et de sa fonction "Detect Gap" qui créait un fichier de gap. Il ne reste plus qu'ensuite, de faire une union et une fusion pour obtenir des polygones utilisables. Si cette solution semble plus archaïque, elle a le mérite de fonctionner parfaitement, et notamment, dans plus de cas que celle d'ArcGis©, "repair geometry".

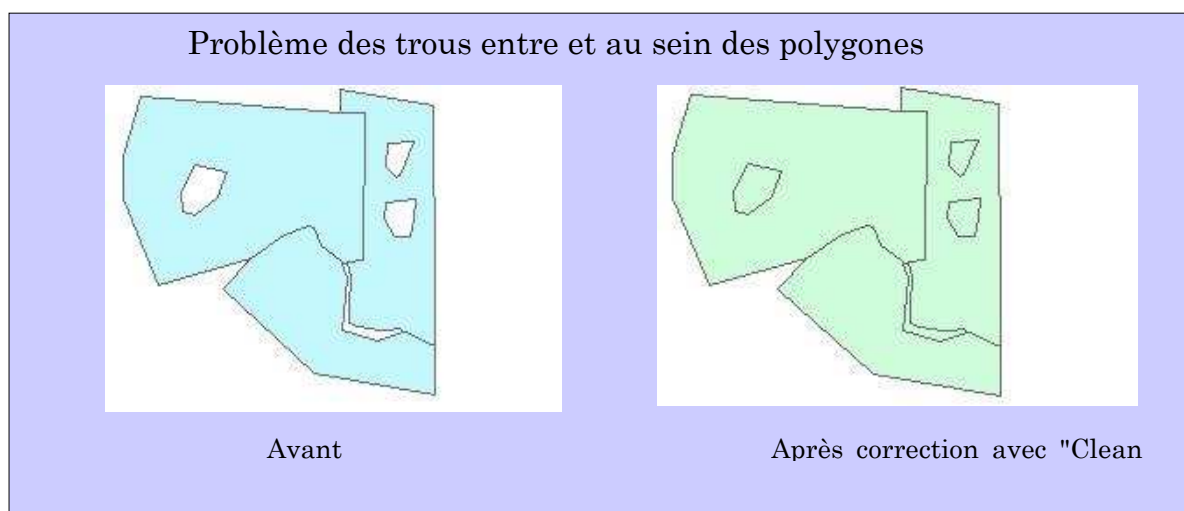


Figure 22: Réparation du problème de gap avec ETGeoWizard©

L'ensemble de ces corrections étaient nécessaire pour partir sur des bases saines et ainsi éviter de multiples problèmes plus difficilement gérables une fois détecté dans les parties traitement et analyses de notre étude notamment.

4. Jointure et architecture du SIG

La phase préparatoire de la création du SIG étant achevée, celle d'organisation pouvait commencer. Calquant à la demande et aux besoins du projet de recherche GRIPAVI, nous avons opté pour une hiérarchie multi-scalaire. Si, dans la plupart des cas, les études s'organisent du global vers le particulier, c'est l'ordre inverse qui a été préférée dans notre analyse.

En effet, nos premiers traitements spatiaux concerneront la province de Bac Giang, située au nord du Vietnam. Les épidémiologistes du projet ayant intensifié leurs enquêtes sur cette zone et y ayant effectué une bonne partie de leurs prélèvements, la plupart des données épidémiologique mises à notre disposition, proviennent de cette entité administrative. Elle représentera donc, notre terrain d'étude test, à partir duquel nous essayerons de mettre en place une démarche transposable sur des territoires plus vaste.

Le cas du "Great Delta" est un peu différent. Si Bac Giang représente une adaptation aux données dont nous disposons, le Great Delta correspond plus aux exigences d'analyse auxquelles, les chercheurs du CIRAD, nous ont demandé de répondre.

Enfin, l'échelle nationale, dernier maillon de notre hiérarchie, a été créée dans un souci de pérennisation du projet et, par conséquent, de notre SIG. Si elle ne correspondra à aucune des analyses auxquelles nous nous prêterons lors de ce stage, elle pourra néanmoins, être une échelle privilégiée des besoins des études futures.

Des données épidémiologiques fournies, nous ne conserverons finalement que deux tables: la "Case_control_2007" et la "Commune_outbreak_may09". La première a été utilisée dans les analyses sur Bac Giang, puisque ses données ne concernent que de cette province, alors que la seconde, sera jointe aux communes du Great Delta.

Nous avons donc créé une geodatabase personnelle nommée GRIPAVI sous ArcCatalog©, au sein de laquelle trois jeux de classe d'entités répondant aux noms de Bac Giang, Great Delta et Vietnam ont été placées. Un découpage de l'ensemble des couches thématiques aux limites des entités géographiques que nous venons d'évoquer a ensuite été effectué à l'aide de l'outil "Clip" d'ArcToolbox©. Au final, nous avons abouti à la configuration suivante (Figure 23):

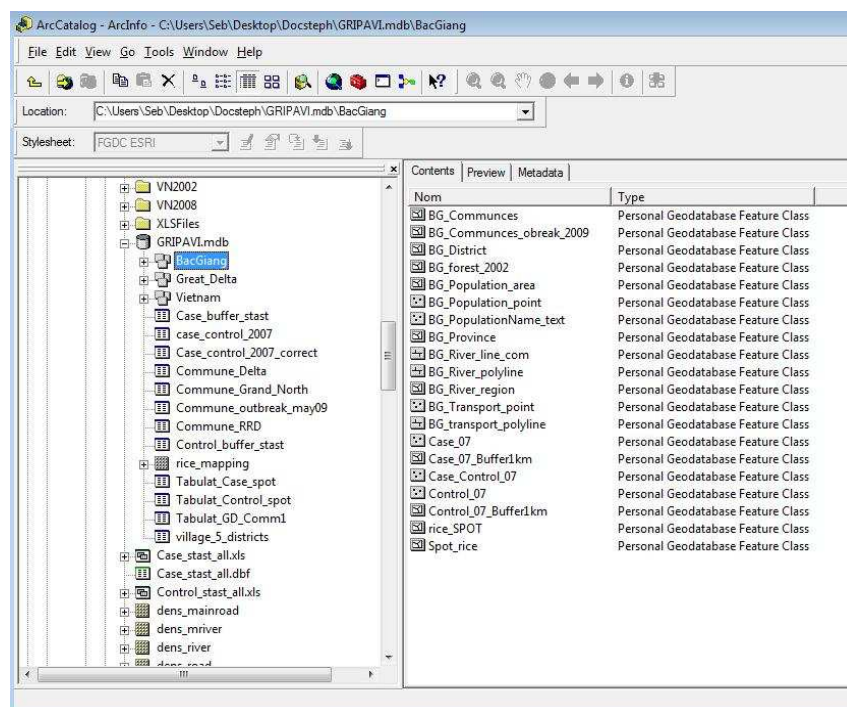


Figure 23: Architecture du SIG sous ArcCatalog

B. Etudes spatiales : problèmes et solutions

L'ensemble du travail préparatoire étant achevé et les données prêtes à être utilisées, l'étape ultime de notre travail pouvait commencer: l'analyse spatiale proprement dite. A cette fin, il était cependant primordial de définir une méthodologie de travail adaptée aux besoins du projet et répondant aux problèmes rencontrés.

1. Méthodologie d'application

L'objectif de ce stage reste de fournir une expertise, par le prisme de l'analyse spatiale, à une étude épidémiologique afin de tenter de découvrir des facteurs environnementaux pouvant expliquer les mécanismes d'émergence, de développement et de propagation du virus H5N1, le tout, avec les outils de la géomatique, à savoir, la télédétection et les SIG. Les besoins hétérogènes des différentes zones d'étude ont conduit à une adaptation de l'architecture de notre SIG.

a) La province de Bac Giang

Premier critère à satisfaire et non des moindres: les besoins de l'étude se situent à l'échelle villageoise. Comme nous l'évoquions précédemment, le système administratif vietnamien est assez compliqué à cette échelle et n'est donc pas cartographié: le village représente une entité à l'intérieur de l'échelon communal que l'on pourrait apparenter aux villages en France et reste donc soumis, assez souvent, à des fusions ou des créations. Le problème réside dans le fait que les données épidémiologiques collectées par les vétérinaires le sont à ce niveau.

Ainsi, la table Case_control_2007 sur laquelle nous travaillons pour Bac Giang, intègre l'ensemble des localisations de 18 élevages ayant subi une infection au H5N1 (cas) et également, les coordonnées de 39 élevages sains, pourtant situés à proximité (contrôle). Protocole très utilisée en épidémiologie, les études de cas/contrôle consistent à choisir aléatoirement le double de contrôles qu'il n'y a de cas, et ensuite, de les comparer par rapport à des variables, dans notre cas, la distance et la densité à certaines variables environnementales. A ce propos, nous précisons que par environnement, nous entendons tant l'humain que le naturel.

Nous avons donc cartographié ces données en créant des points à l'aide des coordonnées présentes dans la table et de la fonction "ajouter des données XY" sous ArcMap.

Etant donné qu'il n'existe pas de couche des limites administratives à l'échelon villageois, il a fallu réfléchir à une solution afin de contourner le problème. Il a été décidé de créer des buffers de la surface approximative d'un village selon les connaissances du terrain de l'équipe vietnamienne, soit d'un

rayon de 500m autour des cas et contrôles. Nous passons ainsi, d'un objet ponctuel à des polygones sous lesquels il devient possible de calculer des densités ou des pourcentages d'occupation des sols.

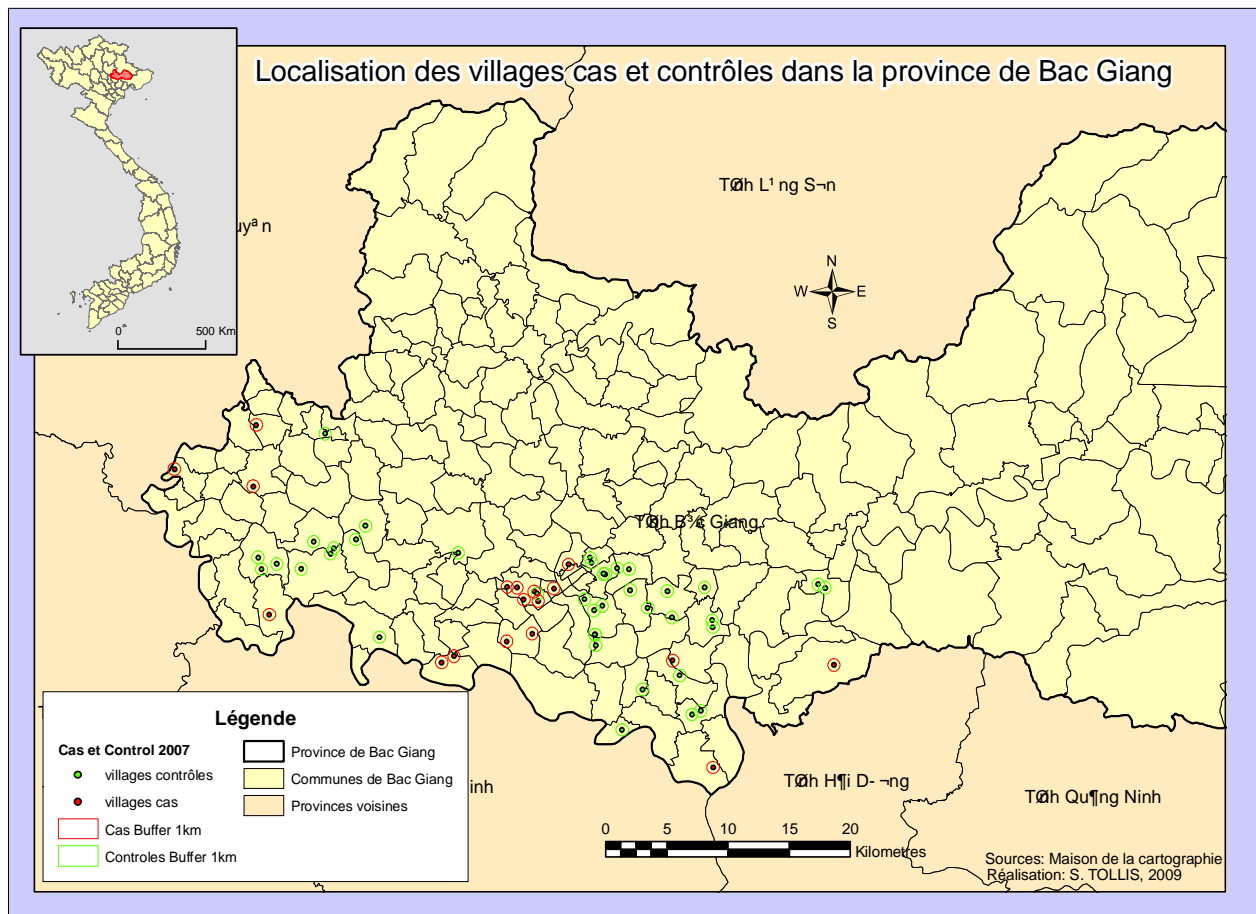


Figure 24: Cartes de localisation des villages cas et contrôles, province de Bac Giang

b) Le Great Delta

Distinction importante par rapport à la zone d'étude précédente, l'échelle d'analyse souhaitée est la commune. Si nous possédons une couche cartographiant tout le Vietnam à cette échelle administrative, les données épidémiologiques sont beaucoup moins fiables mais surtout, ne recensent que les communes cas et pas les contrôles (table "Commune_outbreak_may09").

En effet, ces dernières ne sont que la synthèse du recensement des cas apparus sur internet et dont les sources n'ont pu toutes être vérifiées. Il semble néanmoins que ces données soient acceptables pour notre étude. Privilégiant toujours la technique des cas/contrôles, l'essentiel de notre travail consiste donc à penser une méthode valide afin de déterminer les communes contrôles.

Souvent sélectionné à l'intérieur d'une même unité administrative, nous avons voulu nous abolir des découpages administratifs afin d'effectuer notre choix de communes contrôles en collant le plus possible aux réalités socio économiques

notamment en ce qui concerne la vente de volailles ou leurs transports. Pour contenter les épidémiologistes tout en tenant compte de ces réalités, nous avons essayé de respecter la taille d'une unité administrative, le district, en créant des buffers, autour des communes outbreaks, d'une superficie égale à la superficie moyenne de tous les districts du Great Delta. Nous avons donc calculé cette superficie moyenne et en avons extrait le rayon indispensable à la création des zones tampons (2kms de rayon).

Partant du postulat que les communes contrôles potentielles sont toutes celles qui n'ont pas déclaré de foyers, nous avons élaboré une procédure afin de choisir aléatoirement autant de communes contrôles qu'il n'y a de communes cas (suffisant au vue de la qualité et la précision des données contrairement à la province de Bac Giang).

Nous avons commencé par sélectionner l'ensemble des communes dont le total de cas d'épidémie était égal à zéro (2948 communes). En utilisant la fonction "Generate Random Points" de l'extension Hawth's Tools©, nous avons généré le même nombre de point qu'il y avait de communes potentielles et nous avons finalement retenus, par sélection spatiale, uniquement ceux présents au sein des buffers (1033 points). Après nettoyage des points localisés au sein des cours d'eau notamment, nous avons obtenu 986 points acceptables: nous en avons fait une couche jointe à celle des communes du Great Delta Utilisant une autre fonction d'Hawth Tool©, "Create Random Selection", nous avons sélectionné finalement 237 points et les communes auxquelles ils étaient rattachés. Seul bémol de cette méthode, il est impossible de connaître à l'avance le nombre final de communes puisque, sur une même commune, plusieurs points peuvent être présents: dans notre cas, en ne sélectionnant que 237 points, nous obtenons seulement 200 communes. Il a fallu simplement augmenter le nombre de sélection dans l'outil pour obtenir le nombre d'entité voulu.

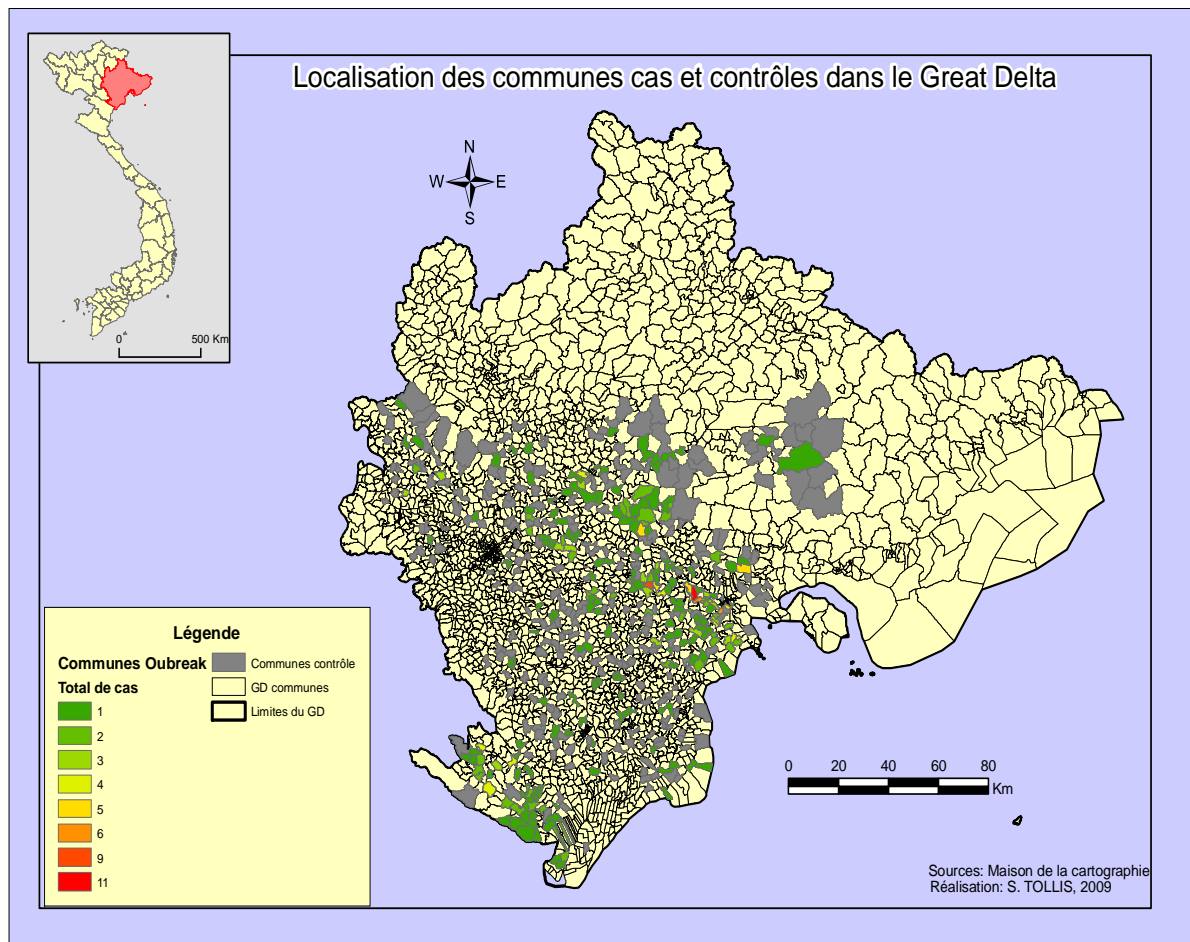


Figure 25: Carte des localisations des communes cas et contrôle dans le Great Delta

c) L'échelle nationale: le Vietnam

Ainsi que nous l'évoquons lors de la hiérarchisation des données, cette dimension scalaire n'a été créée que dans une optique de pérennisation de notre SIG en apportant une réponse plus globale aux futures études sur le sujet. Nous avons intégré l'ensemble des couches mises à notre disposition, ce qui n'a pas la prétention d'être exhaustif, et donc pourra en être agrémenté de supplémentaires afin de répondre avec plus de précision aux besoins ultérieurs: les méthodes seront, encore une fois, adaptés aux exigences du moment.

d) Des démarches différentes mais des objectifs communs

L'originalité de notre démarche réside autant dans les méthodes mises en place, les échelles sur lesquelles elles opèrent que dans la diversité des données épidémiologiques que nous avons à observer. Une constante demeure dans l'ensemble de nos démarches : les calculs par lesquelles nous tentons d'analyser les comportements des croisements issus des tables épidémiologiques avec les données spatiales.

Que se soit pour la province de Bac Giang comme pour le grand delta du fleuve Rouge, nous avons ciblé nos recherches autour de plusieurs thèmes communs:

- Densité des routes nationales
- Densité des routes
- Distance aux routes nationales
- Distance aux routes
- Densité des rivières principales
- Densité des rivières
- Distance aux rivières principales
- Distance aux rivières

Par contre, l'utilisation de nos différents résultats en télédétection est adaptée aux échelles auxquelles ils se prêtent et apportent de la pertinence.

Ainsi, les résultats de la classification obtenue avec l'image SPOT à la couverture limitée, ne concernera que Bac Giang. La résolution plus fine (20m) semble tout à fait adaptée à l'échelle d'analyse du village qui nous est fixée. Nous chercherons à obtenir plus particulièrement:

- Pourcentage de pixels occupés par de l'eau
- Pourcentage de pixels occupés par de la forêt
- Pourcentage de pixels occupés par du riz
- Pourcentage de pixels occupés par des habitations
- Pourcentage des pixels occupés par de l'eau ou du riz

Les classifications MODIS couvrant de plus vastes territoires sont, quant à elles, privilégiées pour le Great Delta où l'unité statistique d'analyse est la commune. La taille des pixels relativement importante (500m X 500m) correspond aux besoins des différentes analyses sur les communes. Les notions de temps et d'intensité de culture sont intégrées et les résultats suivants recherchés:

- Pourcentage des pixels sans interprétation possible
- Pourcentage des pixels occupés par une culture de riz annuelle
- Pourcentage des pixels occupés par deux cultures de riz annuelles
- Pourcentage des pixels occupés par la forêt permanente
- Pourcentage des pixels occupés par de l'eau permanente
- Comptage des durées d'inondation

Afin de parvenir aux objectifs fixés par le projet GRIPAVI, la nécessité de recherche et de test d'outils propres à ArcGis© ou d'extensions développées par des sociétés privées et des communautés d'utilisateurs, nous a semblé évidente.

2. Recherche d'outils et d'extensions d'Arcgis adaptées

Que se soit pour la hiérarchisation des données par découpage, réparation, union, fusion ou pour la conception et réalisation des méthodes d'analyses, les outils proposés par Arctoolbox ou autres extensions ont été d'un apport inestimable. L'obtention des résultats leurs sont également imputables, reste à repérer les plus adaptés.

Pour schématiser, nous sommes en présence d'unités surfaciques, buffer d'un kilomètre de diamètre pour les cas/contrôles de Bac Giang et communes pour le Great Delta, sur lesquelles il nous faut connaître les pourcentages d'occupation du sol, les densités ou les distances par rapport à des objets. Nous avons donc cherché des outils capables de calculer des statistiques au sein même d'un polygone.

Pour les distances, deux techniques s'opposent: soit nous restons en mode vecteur et calculons, à l'aide d'outil, la plus proche distance par rapport à l'objet sélectionné, soit nous commutons en mode raster par le biais de l'outil "Straight Line" du menu distance de Spatial Analyst et obtenons une image raster des distances. Au vue des résultats des deux méthodes, nous avons privilégié la seconde qui a le mérite d'être rapide, d'être cartographiable immédiatement (la méthode vecteur donne des résultats sous forme de table), de permettre le choix de taille de pixel du rendu final (nous avons choisi 20m pour Bac Giang afin d'être à la même échelle que la classification SPOT et 500m pour Great Delta pour se calquer à la classification MODIS) et le rayon d'utilisation autour des objets. Cette méthode semble plus acceptable en statistiques spatiales puisqu'elle permet ensuite de comparer des variables issues d'unités mêmes tailles.

En mode vecteur, nous avons testé les extensions suivantes:

- Jennessent entreprises: "Distance/Azimuth between Matched Features, v. 2.1" et " Nearest Features, v. 3.8b". A ce propos, nous aurions pu travailler selon la méthode développée par Lise Chamailé en utilisant l'outil " Repeating Shapes" et travailler sur une grille hexagonale plutôt que sur des buffers mais cette technique est surtout valable lorsqu'il y a une forte superposition de buffer. De même le test de Morvan définissant l'existence d'une corrélation spatiale se serait révélé positif puisque le H5N1 est une maladie fortement contagieuse donc le temps et la proximité ont une influence dans la zonalité des cas.
- ESRI center: "Minimum Distance 2 Layers" et "Multiple Minimum Distance". Nous avons essayé essentiellement le second qui présente l'avantage d'automatiser la recherche de la distance minimum sur un plus grand nombre d'objets en même temps.

Pour la densité nous avons simplement utilisé l'outil "Density" de Spatial Analyst qui connaît le même mode de fonctionnement que l'outil "Distance" du même module. Nous obtenons également une image raster dont nous prédéfinissons la résolution de sortie ainsi que le rayon d'action.

Dans les deux cas, distance et densité, nous avons défini un rayon d'un kilomètre autour des objets afin de ne pas sous estimer l'attractivité de ces infrastructures sur leurs environnements directs.

C. Résultats et perspectives

Les outils « straight line » et « density » fournissent un résultat sous la forme d'un raster. S'il est aisée, par le biais de l'analyse thématique, d'obtenir un rendu cartographique, visuellement il est difficile d'en tirer des résultats probants.

Il faut avouer que représenter cartographiquement des distances et des densités n'est pas des plus courants.

Nous ne présenterons ici uniquement les cartes qui permettent une interprétation visuelle. Les autres, ne serviront que de supports pour les différents calculs qui serviront aux analyses statistiques.

La carte la plus représentative au niveau des résultats reste celle de la densité routière. Il apparaît clairement que les villages cas sont presque toujours autour des zones à fortes densités d'infrastructures routières. Un axe se dégage d'ailleurs plus que les autres : la route nationale 1A en provenance de la Chine. Sachant que cet axe est suspecté d'être le passage de volailles alimentant le marché noir depuis la Chine, nos résultats vont étayer cette hypothèse.

Un deuxième résultat est significatif graphiquement parlant, même si ce n'est pas au même niveau que le premier : il concerne les densités du réseau hydrographique. Dans l'ensemble, il apparaît que les zones où ce réseau est plus dense, les villages cas ce concentre. Ce phénomène semble moins flagrant chez les villages contrôles.

Nous présentons également un de nos résultats cartographiques à titre démonstratif car s'il ne donne rien visuellement, nous verrons plus tard qu'à l'analyse statistique, c'est un des deux plus importants résultats : il concerne la distance au réseau hydrographique principal.

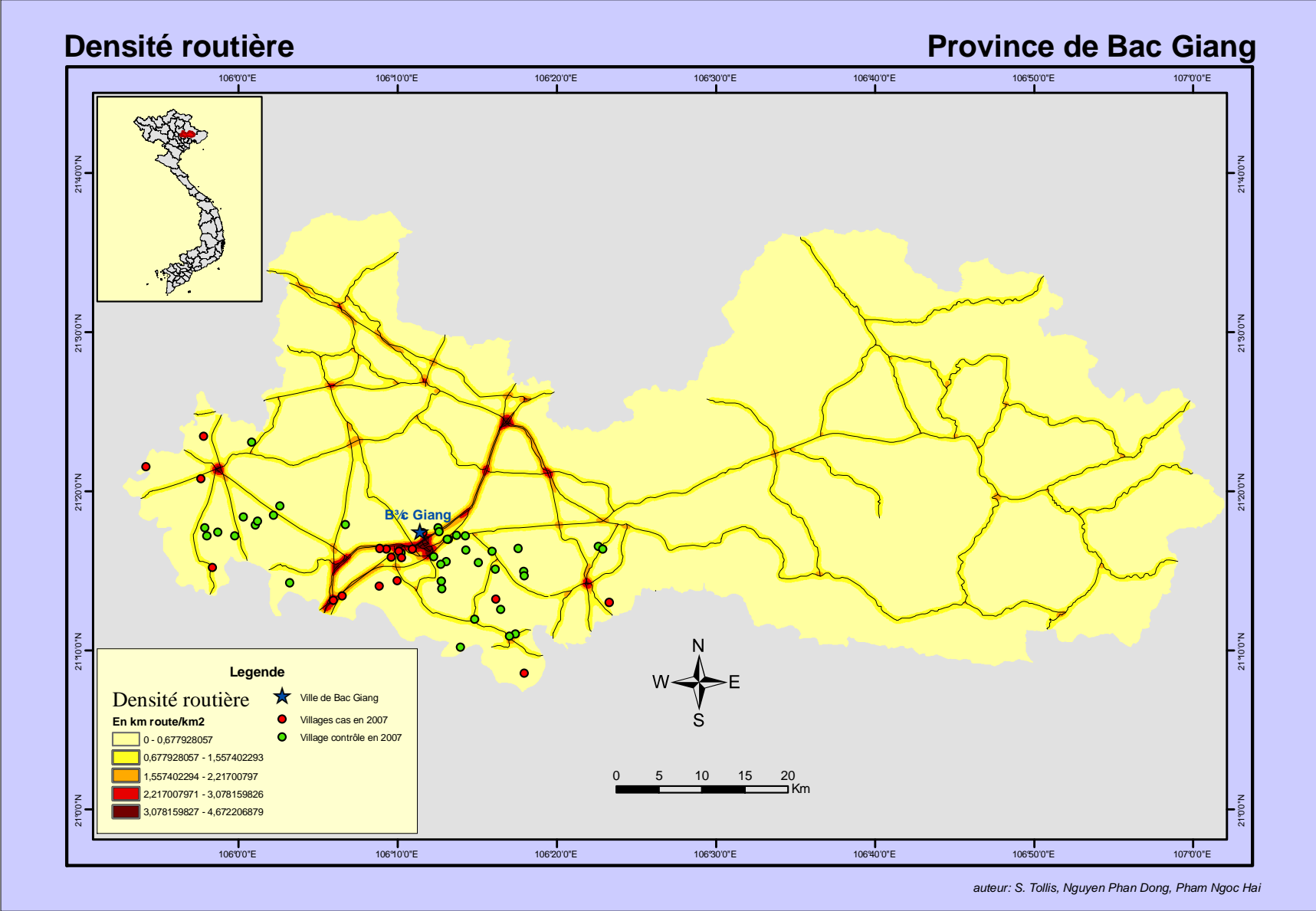
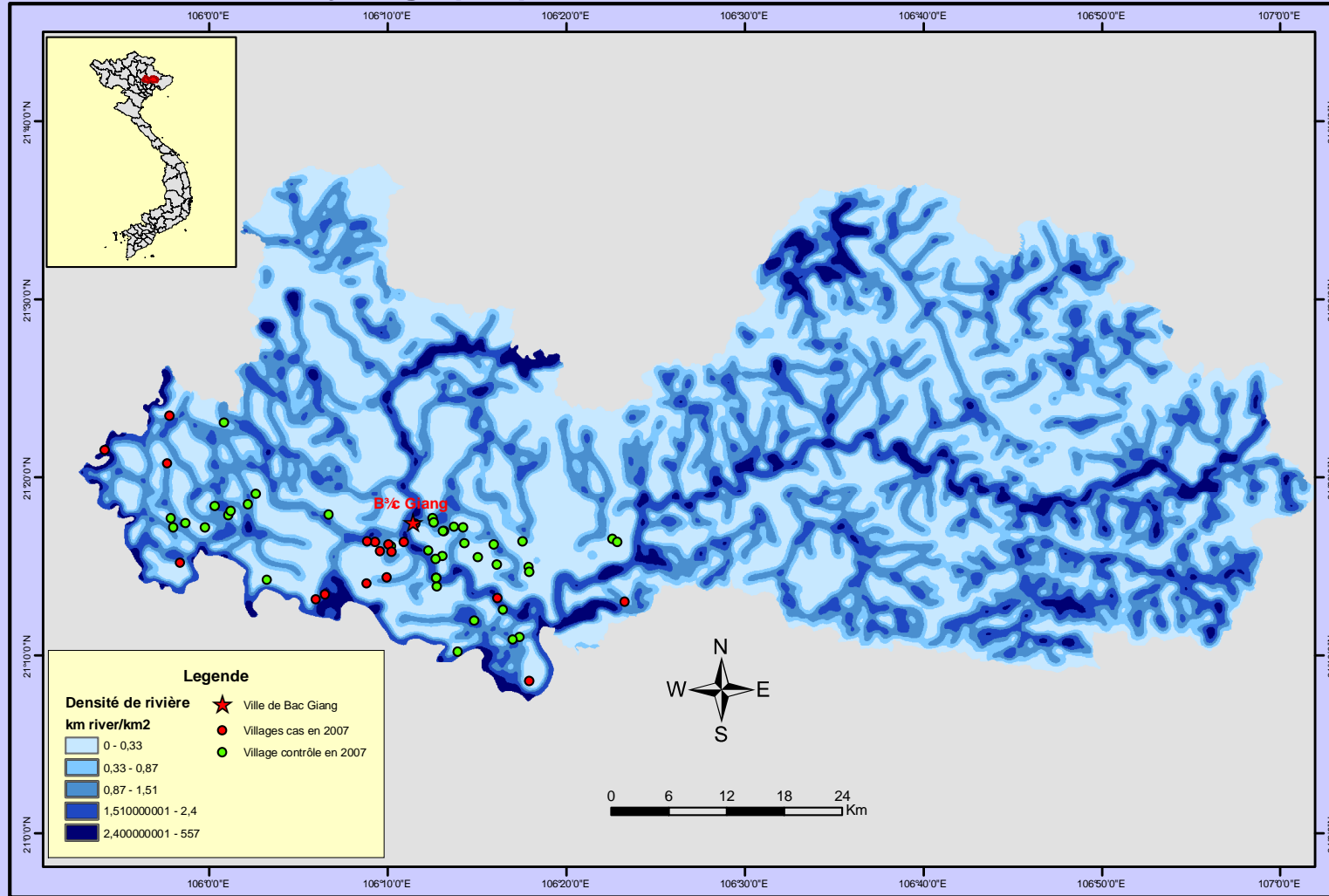


Figure 26 : Carte des densités routière, province de Bac Giang

Densité du réseau hydrographique

Province de Bac Giang



auteur: S. Tollis, Nguyen Phan Dong, Pham Ngoc Hai

Figure 27 : Carte des densités de réseau hydrographique, province de Bac Giang

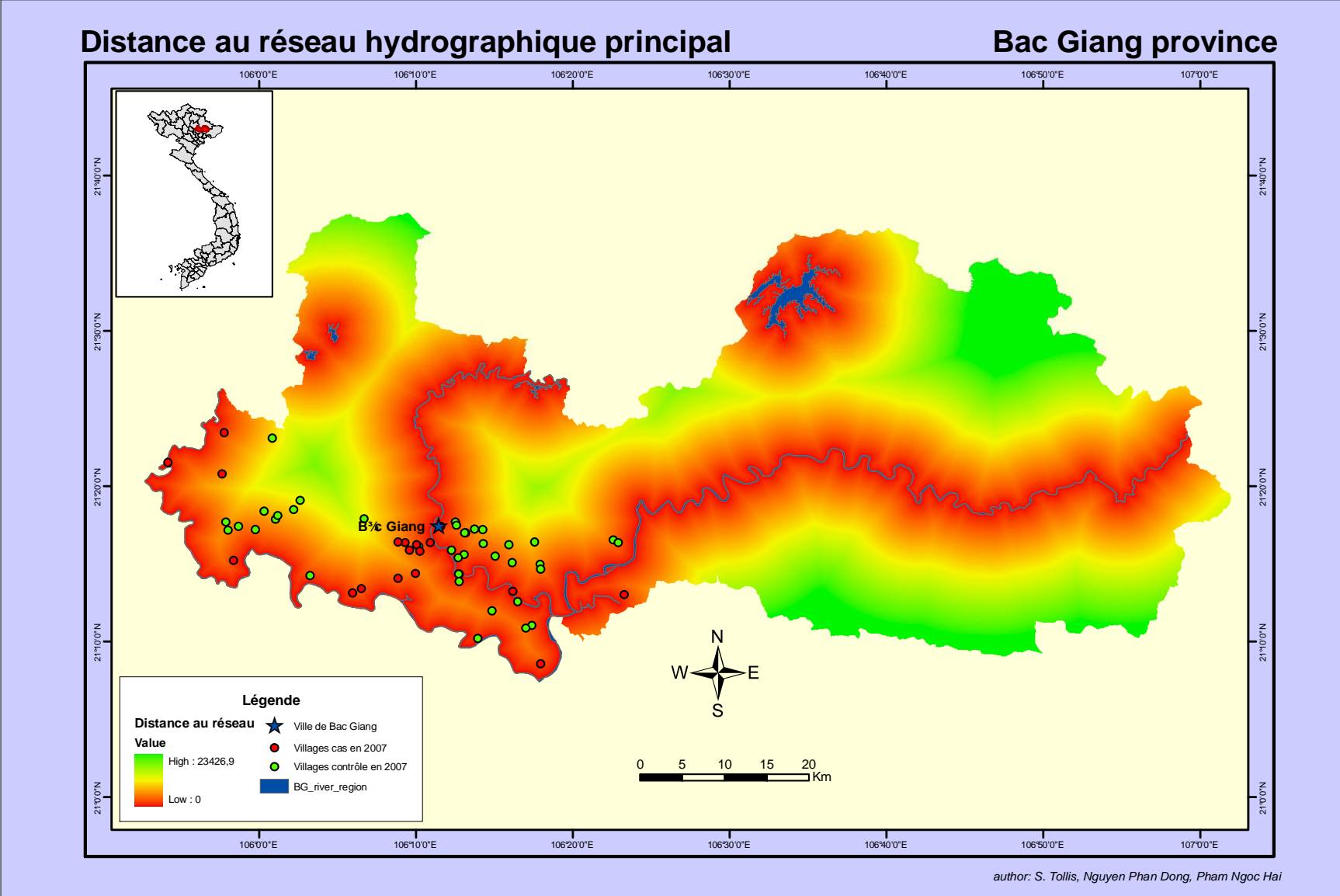


Figure 28 : Carte des distances au réseau hydrographique principal, Province de Bac Giang

Au final, nous obtenons 18 rasters issus des différents traitements.

Afin d'extraire les données de ces derniers, nous utilisons l'outil Zonal Statistics + de l'extension Hawth's Tools qui présente l'avantage de calculer l'ensemble opération en une seule fois (min, max, moyenne, médiane, écart type, somme et compte).

Obtenu sous la forme de table, ces données serviront aux futurs traitements de statistiques spatiales auxquelles nous nous sommes quand même essayés.

Nous avons commencé par l'analyse exploratoire univariée dans laquelle nous avons effectué une comparaison des moyennes pour chaque variable quantitative (t-test) entre les cas et les contrôles.

La province de Bac Giang a servi de zone de test :

	Mean_cas	STD_cas	Mean_control	STD_control	T_test
Dens_mainroad_mean	0.7796283	0.7975761	0.4636649	0.5678936	Cas>cont p=0.0463
Dens_mainriver_mean	0.3052456	0.669956	0.2387852	0.5474132	p>0.05
Dens_river_mean	1.077793	0.7907751	0.7862719	0.5813399	p>0.05
Dens_road_mean	0.9393842	0.9062744	0.5551344	0.6257569	Cas>cont p=0.0340
Dist_mainroad_mean	787.0995	809.1213	1189.854	1019.553	p>0.05
Dist_mainriver_mean	1757.86	1085.717	2731.08	1886.755	cas# cont p=0.0173 cas< cont p=0.0087
Dist_road_mean	752.3276	824.3861	898.7086	750.6376	p>0.05
Per_water	6.49918	14.09123	4.233685	7.644763	p>0.05
Per_forest	11.3092	22.38784	5.632289	8.98565	p>0.05
Per_rice	53.60904	22.09187	53.75944	21.32889	p>0.05
Per_farm	3.82254	7.7749	4.288532	7.912966	p>0.05
Per_residential_area	21.6592	12.61728	22.52132	14.36104	p>0.05
Per_rice_water	60.1082	23.70103	57.99312	23.07353	p>0.05

Nous sommes arrivés finalement à ces conclusions principales de l'analyse exploratoire univariée :

- La densité des routes est plus importante chez les cas que chez les contrôles.
- La .distance moyenne aux rivières principales est plus courte chez les cas que chez les contrôles

Dans un second temps, un modèle de régression logistique sera développé pour modéliser la probabilité d'apparition d'un foyer de H5N1 en prenant en

compte plusieurs variables en même temps et en vérifiant la présence d'interaction entre ces variables. L'autocorrélation spatiale devra être prise en compte à ce stade.

Conclusion et bilan personnel

Bilan du stage

L'utilisation de la télédétection à basse résolution pour caractériser l'occupation du sol est une technique relativement récente tout comme l'est son utilisation dans le domaine de la grippe aviaire. Si le lien entre le virus et son environnement n'est plus à prouver, les indicateurs environnementaux relatifs à ce dernier restent à trouver. Tout l'objectif de cette étude résidait dans cet état de fait.

Par le prisme de la géomatique et donc de l'espace, nous avons tenté de déterminer certains d'entre eux après les avoir détecté.

Les zones en eaux ainsi que les rizières ayant été déterminées comme très probables réservoirs du virus, des algorithmes sur des séries temporelles MODIS à 500m de résolution ont été développés afin de détecter ces espaces.

Les premiers résultats obtenus sont satisfaisants. Les durées d'inondations ont ainsi pu être cartographiées de manière précise et l'intensité des récoltes sur les rizières semble correspondre à la réalité du terrain que nous avons pu observer lors de notre travail de validation dans les provinces du delta du Fleuve Rouge.

Malheureusement le résultat n'est pas parfait et des zones de pixels non classés demeurent malgré le traitement d'interpolation temporelle dont nous avons fait usage et il persiste une légère confusion entre l'eau et les rizières: un affinement des seuils ou l'ajout de conditions supplémentaires sont à prévoir. Le tableau n'est cependant pas complètement noir si nous nous fions aux 94% d'exactitude de la matrice de confusion et au coefficient de Kappa à 0.7.

L'extraction de ces paramètres environnementaux ne constituait pas une fin en soi et s'inscrivait plus, dans une logique d'étoffement de la base de données environnementales du SIG que nous avons à mettre en place et à croiser aux données épidémiologiques du CIRAD.

Les outils que nous avons testés et utilisés semblent fournir des résultats corrects mais la représentation cartographique de ces derniers demeure décevante. Il faut avouer que cartographier des résultats de distance et de densité n'est pas des plus visuel. C'est dans la statistique spatiale que nos résultats donneront le plus de satisfaction et se révéleront probants.

Dans l'ensemble, le bilan de cette première étude épidémiologique couplée aux outils de la géomatique semble bon. Des résultats intéressants ont été obtenus même si des améliorations peuvent être apportées.

Perspectives et évolutions

Cette étude étant tellement pionnière dans le domaine de la géomatique appliquée au H5N1 que tout semble encore à faire. Un autre stage peut aisément être envisagé pour affiner les résultats même si cette dernière est déjà intégrée dans la thèse de Stéphanie Desvaux.

Concernant la partie télédétection, nous l'avons déjà dit mais, des pixels non classés persistent. Un ajustement des seuils de détection du riz et de la forêt semble la meilleure solution. De même, il serait peut être intéressant d'ajouter de contraintes d'altitude et de pente comme le suggère Xiao dans ses études. Nous ne nous en sommes pas servis tant nos zones d'étude semblaient nivelées.

Il serait également peut être judicieux de tenter l'algorithme de détection du riz et de l'intensité des récoltes avec des indices d'eau différents du notre : le NDVI, le NDWI, NDII, NDMI ou simplement le LSWI donneraient peut être de meilleurs résultats. Dans le temps qui nous était impartis, nous nous sommes contentés de suivre les résultats de L. Chamailé sans critiquer les sources.

La partie SIG et notamment le traitement des résultats sous statistiques spatiales demeurent la plus ouverte en terme de perspective. Nous n'avons réellement appliqué de traitement que sur la province de Bac Giang, tenté de mettre ne place une méthodologie pour le Great Delta. et tout reste à faire à l'échelon national.

De même en terme de statistiques spatiales, nous n'avons travaillé qu'en analyse exploratoire univariée. Un modèle de régression logistique pourra être développé pour modéliser la probabilité d'apparition d'un foyer de H5N1 en prenant en compte plusieurs variables en même temps et en vérifiant la présence d'interaction entre ces variables.

Une analyse en composante principale (ACP) pourra être réalisée sur ces données afin de réaliser une typologie des communes basée sur leur occupation du sol. Les clusters obtenus seront ensuite croisés avec les données épidémiologiques.

Des outils tel que GeoDa (<http://geodacenter.asu.edu/>) seront intéressants à utiliser.

Bilan personnel

Je tire un bilan très positif de ces 6 mois passées au sein du projet GRIPAVI tant sur le plan humain, personnel que professionnel.

Humainement, j'ai eu la chance d'évoluer au sein d'une structure, la MtD, jeune, ouverte et dynamique où le relationnel prime autant que le travail. Les personnes sont ainsi détendues ce qui ne les empêche pas d'être compétentes bien

au contraire. Les chercheurs ont demeuré très accessibles, rendant la MtD, un lieu d'échange très intéressant.

Personnellement, j'ai rencontré des personnes, chercheurs, stagiaires, thésards, très attachantes qui ont partagé leurs connaissances, leurs expériences, leurs amitiés. Au-delà du stage, des contacts seront conservés.

Professionnellement, ce stage constitue un choix murement réfléchi. Il m'a apporté l'ensemble des compétences que je souhaitais acquérir notamment celle de la 1^{ère} expérience valorisable qui fait souvent défaut aux étudiants. J'ai ainsi pu approfondir la totalité des domaines que nous avons ébauchés durant les six premiers mois de formation SIGMA : télédétection, SIG, SGBD, gestion de projet. J'ai particulièrement apprécié l'autonomie dont j'ai pu bénéficier notamment dans la gestion des deux chercheurs vietnamiens avec lesquels nous avons collaboré.

Je souhaitais un stage à l'étranger pour parfaire mon anglais et force est de constater que je « speak » bien mieux la langue de Shakespear qu'en arrivant : il est important de ne pas se fermer la possibilité de postuler à des projets anglophones comme c'est souvent le cas pour les projets internationaux.

De même, avoir travaillé sur un projet de cette envergure, pour une institution de cette renommée et sur une thématique aussi conjoncturelle ne peut être que bénéfique pour ma future recherche d'emploi.

Finalement, que du positif se dégage de ce stage et si je n'ai qu'un point négatif à donner, c'est le manque de temps et la frustration de laisser une étude inachevée.

Tables des illustrations

Liste des figures

Figure 1 : Courbe épidémiologique des cas de H5N1 déclarés au Vietnam	13
Figure 2 : Gestion du projet sous Gantt	22
Figure 3 : Les zones d'étude du projet	23
Figure 4 : Le satellite TERRA	1
Figure 5 : Méthode de constitution des synthèses temporelles MODIS	26
Figure 6: Les trois indices utilisés par Xiao pour la détection de l'intensité des récoltes de riz	1
Figure 7: Deux exemples issus des recherches de Xiao	28
Figure 8 : algorithme amélioré de la méthode de L. Chamaillé	30
Figure 9 : Carte des durées d'inondation du Nord Vietnam	31
Figure 10 : algorithme de Xiao pour déterminer l'intensité des rizières	1
Figure 11 : notre algorithme de détection de l'intensité des récoltes	33
Figure 12 : illustration de l'importance de la couverture nuageuse et	36
Figure 13 : Codage de la qualité du NDVI des images Modis	37
Figure 14 : exemple de profil temporel sous ENVI	38
Figure 15 : schéma de l'interpolation temporelle, M Houllès	40
Figure 16 : Illustration du résultat intermédiaire de l'algorithme	43
Figure 17 les cartes itinéraires du travail d'observation terrain	46
Figure 18 : Relevés GPS du travail d'observation terrain	47
Figure 19 : Occupation du sol de la province de Bac Giang obtenue par classification des images SPOT	49
Figure 20 : Résultat cartographique de notre algorithme	54
Figure 21: illustration du travail de correction des coordonnées	59
Figure 22: Réparation du problème de gap avec ETGeoWizard©	60
Figure 23: Architecture du SIG sous ArcCatalog	61
Figure 24: Cartes de localisation des villages cas et contrôles, province de Bac Giang	63
Figure 25: Carte des localisations des communes cas et contrôle dans le Great Delta	65
Figure 26 : Carte des densités routière, province de Bac Giang	69
Figure 27 : Carte des densités de réseau hydrographique, province de Bac Giang	70
Figure 28 : Carte des distances au réseau hydrographique principal, Province de Bac Giang	71

Liste des tableaux

Tableau 1: les 7 bandes spectrales MODIS utilisées	34
Tableau 2: le calendrier théorique du riz dans le Nord Vietnam	50
Tableau 3: Tableau des différentes couches présentes	56
Tableau 4: Sélection des couches pour la constitution du SIG	57

Bibliographie

Les rapports

- Brun, Julien. 2004. *Etude du potentiel des indices de végétation de l'imagerie MODIS pour l'observation de l'évolution intra- et inter-annuelle de la couverture du sol: application à la région genevoise*. Genève: Université de Genève.
- Chamaillé, Lise. 2008. *Apport de la télédétection pour la détection de l'eau et la caractérisation des propriétés de l'eau. Application de la télédétection à l'étude des facteurs de risques liés à l'eau pour la survie du virus H5N1 dans l'environnement*. Montpellier: Maison de la Télédétection, Septembre.
- Coste, Silvan. 2007. *Mise en place d'un SIG et réalisation d'un atlas du littoral du Languedoc Roussillon*. Nîme: BRL ingénierie, Septembre.
- Houles, Marion. 2006. *Indicateurs spatiaux et temporels pour l'identification du domaine cultivé en milieu soudano-sahélien*. Montpellier: Maison de la Télédétection, Septembre.
- IFEN. 2005. *Application de la télédétection à l'étude des zones humides: identification des prairies, des roselières, des peupleraies et des gravières*.
- Kouane, Adonis, et al. 2008. La télédétection et les sig comme moyen de hiérarchisation des localités dans une région endémique de Bilharziose: cas de la région de l'Agneby (sud-est de la Côte d'Ivoire 22, no. 1. *European Journal of Scientific Research*: 66-83.
- Tep, Vottana. 2008. *Mise en place d'un SIG d'appui aux recherches agronomiques et à la gestion agricole sur un domaine d'expérimentation scientifique*. Roquefort sur Souzou: INRA, Septembre.
- Vintrou, Elodie. 2008. *Analyses de séries temporelles d'images satellite à basse résolution spatiale pour la détection des changements d'occupation du sol au Mali*. Montpellier: Maison de la Télédétection, Septembre.

Les articles de revue

- Desvaux, Stéphanie. 2008. *Field surveillance model for HPAI in Vietnam*. Bangkok.
- Gilbert, Marius, et al. 2006. Free-grazing ducks and highly pathogenic avian influenza, Thailand 12, no. 2. *Emerging Infectious Diseases* (Février): 227-234.
- Gilbert, Marius, Xiangming Xiao, et al. 2007. Avian influenza, domestic ducks and rice agriculture in Thailand, no. 119. *ScienceDirect*: 409-415.
- Gilbert, Marius, Xiangming Xiao, et al. 2008. Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia 105, no. 12. *PNAS* (Mars 25): 4769-4774.
- Minh, Phan, et al. 2009. Spatio-temporal epidemiology of highly pathogenic avian influenza outbreaks in two deltas of Vietnam 2003-2007. *Preventive Veterinary Medicine*.
- Pfeiffer, Dirk. 2007. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Science Direct*.

- Xiao, Xiangming. 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images, no. 95. Science Direct: 480-492.
- Xiao, Xiangming, et al. 2006. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images, no. 100. Science Direct: 95-113.

Les sites internet

- <http://avian-influenza.cirad.fr/projects/research/gripavi>
- <http://mrtweb.cr.usgs.gov/>
-

—

Annexes

