



HAL
open science

Etude des relations entre la luminosité de l'eau et le paludisme dans l'État d'Amazonas en Amazonie brésilienne

Alexandre Wiefels

► **To cite this version:**

Alexandre Wiefels. Etude des relations entre la luminosité de l'eau et le paludisme dans l'État d'Amazonas en Amazonie brésilienne. Sciences de la Terre. Université Montpellier, 2016. Français. NNT : 2016MONTT185 . tel-01690562

HAL Id: tel-01690562

<https://theses.hal.science/tel-01690562>

Submitted on 23 Jan 2018

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par l'Université de Montpellier

Préparée au sein de l'école doctorale GAIA

Et de l'unité de recherche UMR 228 Espace-Dev

Spécialité: Sciences de la Terre et de l'Eau

Présentée par Alexandre Wiefels

**Etude des relations entre la luminosité de
l'eau et le paludisme dans l'État
d'Amazonas en Amazonie brésilienne.**

Soutenue le 09/12/2016 devant le jury composé de

Mme Frederique Seyler, IRD
M. Laurent Durieux, IRD
M. Frederic Frappart, Legos - OMP
M. Samuel Corgne, Université de Rennes
M. Christovam Barcellos, Fiocruz

Directrice de thèse
Co-Directeur de thèse
Rapporteur
Rapporteur
Examineur





UNIVERSITÉ
DE MONTPELLIER

THÈSE

de

DOCTORAT DE L'UNIVERSITÉ DE
MONTPELLIER

Etude des relations entre la luminosité de l'eau et le paludisme dans l'État d'Amazonas en Amazonie brésilienne.

Thèse préparée au sein de l'école doctorale GAIA,
spécialité Sciences de la Terre et de l'Eau.



Accueillie par l'UMR 228 Espace pour le
Développement (Espace-Dev).



Financée par le Conseil National de Développement
Scientifique et Technologique du Brésil (CNPq) au
travers de l'Institut National de Science et
Technologie pour l'Innovation en Maladies Négligées
(INCT-IDN).



Sous la direction de Mme Frederique Seyler, Dr, IRD Montpellier et
Mr Laurent Durieux, Dr, Ambassade de France à Brasília.

Résumé

Le paludisme trouve dans la région amazonienne des conditions favorables à la transmission de la maladie, par un moustique vecteur qui est essentiellement *Anophèles darlingi*, se reproduisant en milieu aquatique. Ce moustique est connu pour sa grande adaptabilité aux conditions environnementales et en Amazonie, il est réputé être plus spécialement trouvé près des fleuves d'eau blanche (chargés en sédiments). La relation entre la présence du moustique et la couleur des eaux a été peu étudiée à l'échelle régionale. La présente étude a utilisé 11 années d'images MODIS à 250 m de résolution et un pas de temps mensuel, dont il a été extrait un indicateur de luminosité des eaux. D'autre part, ce travail de thèse exploite les données épidémiologiques du système de surveillance épidémiologique brésilien du paludisme. L'objectif principal est d'évaluer les corrélations entre la dynamique saisonnière de la luminosité des eaux et l'incidence parasitaire du paludisme sur différentes zones de l'État d'Amazonas afin notamment de comprendre l'influence de la couleur des eaux sur la présence du vecteur et donc la transmission du paludisme. Les résultats obtenus permettent de montrer que les notifications concernant le paludisme dans l'État d'Amazonas sont en effet corrélées aux eaux blanches, mais que les eaux noires ont aussi une corrélation avec l'incidence du paludisme, d'une façon sensiblement différente, à la fois dans le temps et dans l'espace. Ces résultats pourront être utiles à l'amélioration de notre compréhension des risques épidémiologiques dans cette région ainsi qu'à la mise en place de programmes de surveillance plus efficaces, même si le facteur étudié, i.e. la couleur des eaux, n'est qu'un facteur parmi beaucoup d'autres qui influent sur le risque d'infection paludéen.

Abstract

Malaria in the Amazon region finds favorable conditions for the transmission of the disease by the mosquito vector *Anopheles darlingi*, which breeds in water. This mosquito is known for its great adaptability to environmental conditions. In the Amazon it is deemed to be especially found near rivers of white water (loaded with sediments). The relationship between the presence of the mosquito and water color has been little studied regionally. This study used 11 years of MODIS 250 m resolution and a monthly time base, which enable extracting a reflectance index of water. Secondly, this thesis uses the epidemiological data of the Brazilian system of epidemiological surveillance of malaria. The main objective is to evaluate the correlation between the seasonal dynamics of the reflectance of water and parasite incidence of malaria on different areas of the State of Amazonas in particular to understand the influence of water color on the presence of the vector and therefore the transmission of malaria. The results obtained show that notifications of malaria in the state of Amazonas are indeed correlated with white water, but the black water also have a correlation with the incidence of malaria in a substantially different way, since in the former case, the correlation is related to the flood, while in the case of black water, the presence of *Anopheles* and the flood are disconnected. These results may be useful in improving our understanding of epidemiological risks in the region and the establishment of more effective compliance program, even if the factor of interest, ie the color of the water is one factor among many others that affects the risk of malaria infection.

Remerciements

Je commence par remercier du plus profond de mon âme ma belle épouse Beatriz Nunes Maciel que j'aime tellement d'avoir accepté le pari du voyage et des sacrifices, mais aussi des agréables surprises pour m'accompagner dans l'aventure de la thèse loin de chez soi avec Sofia, notre fille de moins de 1 an à l'époque et par la suite avec Helena, la petite dernière qui nous est arrivée en cours de route.

Je remercie ma directrice de thèse Frédérique Seyler pour ses interventions très efficaces au long de ma thèse et pour toutes les aides qu'elle m'a apportées pour en arriver à ce jour.

Je remercie mon co-directeur de thèse Laurent Durieux pour son invitation à réaliser cette thèse au sein de l'UMR 228 Espace pour le Développement (Espace-Dev) à Montpellier, pour son accueil, ses enseignements et son soutien.

Je remercie Carlos Medicis Morel pour avoir cru en moi et m'avoir accueilli à l'Institut National de Science et Technologie pour l'Innovation en Maladies Négligées (INCT-IDN) à la Fondation Oswaldo Cruz (Fiocruz) à Rio de Janeiro viabilisant mon doctorat. Suite à cela je remercie le gouvernement brésilien pour avoir réalisé l'important programme de financement d'études à l'étranger, le programme « Sciences sans Frontières » (CsF) par le biais du Conseil National de Développement Scientifique et Technologique du Brésil (CNPq) qui a financé ma thèse.

Je remercie particulièrement les chercheurs de l'IRD et de l'UMR Espace-Dev Morgan Mangeas pour son aide et enseignements, Nadine Dessay pour son soutien tout au long de cette thèse, Vincent Herbreteau et Emmanuel Roux pour leur

Remerciements

engagement et pour avoir animé le groupe d'études sur les sujets d'environnement-santé dans l'UMR. Je remercie aussi Christian Depraetere pour son accueil en sol Occitan et ses contributions scientifiques.

Je remercie tous mes collègues de thèse qui étaient là au quotidien pendant des années, plus spécialement Claudio Almeida qui, avec toute sa famille, a adopté ma famille pendant cette période, Bruna Wolfarth, partenaire de recherche, décisive dans ma thèse et avec qui beaucoup d'autres travaux seront réalisés et Missifany Silveira pour le bout de chemin productif parcouru ensemble pendant cette thèse.

Je remercie les chercheurs Helen da Costa Gurgel du département de géographie de l'Université de Brasilia pour son soutien et excellentes idées, Marcos Lacerda de la Fundação de Medicina Tropical de Manaus pour ses interventions scientifiques et spécialement Roberto Reis de l'Instituto Nacional de Infectologia Evandro Chagas pour son support décisif quand j'en avais besoin.

Sommaire

Introduction Générale	1
CHAPITRE 1 Contexte de la recherche	5
1.1 Le paludisme des rivières	6
1.2 Etat d'Amazonas et rôle de l'eau dans l'écologie du moustique Anophèles ..	9
1.3 L'eau dans l'État d'Amazonas	14
1.4 La télédétection de la couleur de l'eau	21
1.5 Conclusion partielle	25
CHAPITRE 2 Analyse critique des données de surveillance épidémiologique du Paludisme 28	
Introduction	28
2.1 Le système de Surveillance épidémiologique de Paludisme Brésilien	29
2.2 La qualité des données du SIVEP-Malaria	37
2.2.1 Analyse temporelle des données manquantes	40
2.2.2 Détection des données aberrantes	42
2.2.3 Détection des données incohérentes	44
2.2.4 Analyse spatiale de la base de données	48
2.3 Résultat du nettoyage des données	50
2.4 Utilisation Scientifique des données SIVEP-Malaria.....	52
2.5 Conclusion partielle	54
CHAPITRE 3 Dynamique de la luminosité des eaux dans l'état d'Amazonas	57

Sommaire

3.1	Introduction.....	57
3.2	Matériel et méthode.....	58
3.2.1	Les images MODIS.....	59
3.3	Résultats et Discussion	84
3.3.1	Résultats de la validation du masque eau/ non eau.....	84
3.3.2	Analyse statistique de la cartographie dynamique et interprétation des résultats	88
3.4	Conclusion Partielle.....	94
CHAPITRE 4 Le paludisme des rivières de l'État d'Amazonas – Brésil		97
4.1	Introduction.....	97
4.2	Matériel et Méthode.....	98
4.2.1	Incidence parasitaire.....	99
4.2.2	Les secteurs censitaires	100
4.2.3	Les zones habitées	103
4.2.4	Les données épidémiologiques du SIVEP-Malaria	105
4.2.5	Croisement et traitement des données	106
4.3	Résultats	109
4.3.1	IPM moyen.....	109
4.3.2	Corrélation entre IPM et indicateur de luminosité	111
4.3.3	Analyse des corrélations entre IPM et classes de luminosité pour tout l'État d'Amazonas	112

Sommaire

4.3.4	Corrélation entre IPM moyen sur la période et moyennes mensuelles des classes de luminosité	116
4.3.5	Discussion.....	128
4.3.6	Conclusion partielle.....	131
	Essai de synthèse et considérations finales	133
Références	141	

<i>Figure 1-1 Niches écologiques de l'Anophèles: A) Varzea inondée ; B) Forêt d'igapós inondée (extrait de: Tadei et al. 2009)</i>	11
<i>Figure 1-2 Carte du bassin Amazonien brésilien et état d'Amazonas</i>	15
<i>Figure 1-3 Carte des principaux fleuves affluents du Rio Amazonas dans l'état d'Amazonas</i>	16
<i>Figure 1-4 Rencontre des fleuves Rio Negro et Rio Solimões. (Source: Wikimedia Commons par Mario Ortiz)</i>	17
<i>Figure 1-5 Routes et hydrographie de l'État d'Amazonas</i>	19
<i>Figure 1-6 Routes sur le territoire brésilien et sur l'État d'Amazonas</i>	22
<i>Figure 1-7 Le transport des communautés riveraines amazoniennes (photo: Martinelli 2014)</i>	26
<i>Figure 2-1 Cas positifs pour l'État d'Amazonas entre 2003 et 2015</i>	31
<i>Figure 2-2 Cartes de l'IPA par commune. Produit du SVS/MS pour le public</i>	32
<i>Figure 2-3 Relation cas-suspect/cas-positif de paludisme en région amazonienne pour 2010</i>	34
<i>Figure 2-4 Flux de patients dans les postes de santé urbains et ruraux à Anapu-PA</i>	36
<i>Figure 2-5 Classifications des signatures des données manquantes par variable et par jour</i>	40
<i>Figure 2-6 Cycles épidémiologiques et données manquantes pour la variable date de naissance</i>	42
<i>Figure 2-7 Age des patients</i>	46
<i>Figure 2-8 Schéma de l'organisation du nettoyage des données SIVEP-Malaria</i>	50
<i>Figure 2-9 Cartographie des concentrations d'erreurs du SIVEP Malaria de 2003 à 2014 suite au nettoyage de données</i>	51
<i>Figure 2-10 Production scientifique à partir de données SIVEP Malaria</i>	53
<i>Figure 3-1 Tuiles MODIS sur l'état de l'Amazone au Brésil</i>	60
<i>Figure 3-2 Nuages pour MOD09A1 et MYD09A1 en janvier 2006 sur la tuile Modis h11v09 (500 m)</i>	65
<i>Figure 3-3 Image composée mensuelle avec valeurs A) Minimum, B) Moyenne et C) Maximum des pixels</i>	66
<i>Figure 3-4 Images des composantes du système de couleurs Teinte – Luminosité – Saturation pour la région de Coari</i>	70
<i>Figure 3-5 Comparaison entre SWI et SWIb pour la classification par seuil des eaux du Lac de Coari</i>	71
<i>Figure 3-6 Multiplication de l'indice SWIb par le NDVI sur le Lac de Coari</i>	74
<i>Figure 3-7 Indices et création du masque des eaux</i>	76
<i>Figure 3-8 Comptage du nombre de superposition du pixel Eau sur 12 images de crue des fleuves</i>	78

<i>Figure 3-9</i> Résultat: masque de l'étendue maximum des eaux dont les pixels se répètent au moins deux fois en 11 ans. Les limites municipales apparaissent en jaune	79
<i>Figure 3-10</i> Composante Luminosité (HLS) et composition RGB - 143 de MODIS 2006/01 sur le Lac Coari	80
<i>Figure 3-11</i> Histogramme de la fraction Luminosité (Lightness) du système de gestion des couleurs HLS	81
<i>Figure 3-12</i> – Flux de traitement des images MODIS pour l'obtention de l'indice de luminosité.....	82
<i>Figure 3-13</i> Classification du masque de l'eau à partir de MODIS (500 m).....	85
<i>Figure 3-14</i> Classification ESA/CCI (150 m)	86
<i>Figure 3-15</i> Résultats de la validation du masque eau/ non eau (150 m) sur la région de Manaus à l'encontre des fleuves Rio Negro et Rio Solimões.	87
<i>Figure 3-16</i> Graphique en boîte à moustaches des moyennes mensuelles des classes d'eau pour l'état d'Amazonas.....	89
<i>Figure 3-17</i> Histogrammes de la répartition des classes d'eau moyenne (de 1 à 10) par mois pour tout l'état de l'Amazonas.....	90
<i>Figure 3-18</i> Moyenne mensuelle de la luminosité de l'eau entre les classes 1 et 5 pour la région de la confluence des fleuves Rio Negro et Rio Solimões.....	91
<i>Figure 3-19</i> Variance de la luminosité des eaux aux alentours de Itacoatiara-AM au long du Rio Amazonas; région caractérisée par de vastes zones d'inondation	92
<i>Figure 3-20</i> Variance de la luminosité des eaux à l'encontre des fleuves Japurá; Juruá et Solimões.....	93
<i>Figure 3-21</i> Variance de la luminosité des eaux à la confluence des fleuves Rio Negro; Rio Solimões et Rio Madeira.....	94
<i>Figure 4-1</i> Schéma méthodologique des étapes du traitement des données.....	99
<i>Figure 4-2</i> Produit du projet Global Human Footprint de la NASA pour la région centrale de l'État d'Amazonas	103
<i>Figure 4-3</i> Buffer de 15 km des marges des plans d'eau pour l'État d'Amazonas	104
<i>Figure 4-4</i> Secteurs censitaires découpés par le buffer des eaux et postes de santé se trouvant dessus	107
<i>Figure 4-5</i> IPM moyen des zones d'étude et postes de santé à l'origine de la donnée épidémiologique.....	108
<i>Figure 4-6</i> Temps entre les symptômes de paludisme et la notification au poste de santé	109
<i>Figure 4-7</i> Rassemblements de secteurs à IPM moyen supérieur à 2.6	111

<i>Figure 4-8 Cercle des corrélations entre classes de luminosité des eaux et IPM pour la zone d'étude avec le coefficient de corrélation le plus fort dans l'analyse de corrélation entre 2003 et 2014</i>	113
<i>Figure 4-9 Carte de chaleur du regroupement hiérarchique des corrélations entre les classes de luminosité des eaux et l'IPM par zone d'étude. (Obs: les lignes constituent les différentes zones d'étude - secteurs)</i>	112
<i>Figure 4-10 Regroupements effectués</i>	112
<i>Figure 4-11 Cartographie de la classification par regroupement hiérarchique des corrélations entre les classes de luminosité des eaux et l'IPM entre 2003 et 2013.</i>	113
<i>Figure 4-12 Regroupement: I-1, 1)Rio Juruá, 2)Rio Solimões, 3)Rio Madeira, 4)Rio Purús</i>	113
<i>Figure 4-13 Regroupement I-2 ; 1) le Rio Uneiuxi, 2) le Rio Javari, 3) le Rio Ituxi, tributaire du Purús, 4) le Rio Purús, 5) Rio Solimões.</i>	114
<i>Figure 4-14 Regroupement I-4 - fleuve Purús (1 et 2) sur des zones inondables en période de crue et affluents du Rio Madeira (3 et 4) en zone de lacs.</i>	114
<i>Figure 4-15 Regroupement I-5 : 1) Manaus, 2) Benjamin Constant, 3) Lábrea (3) et 4) Humaitá.</i>	115
<i>Figure 4-16 Regroupement I-3 : 1) Rio Jutai, 2) Rio Solimões proche à la frontière, 3) Rio Madeira en aval de Humaitá, 4) Archipel des Anavilhanas.</i>	116
<i>Figure 4-17 Carte de chaleur du regroupement hiérarchique des corrélations entre la moyenne mensuelle des classes de luminosité des eaux et celle de l'IPM par zone d'étude. (Obs: les lignes constituent les différentes zones d'étude)</i>	117
<i>Figure 4-18 Regroupement des classes de la classification hierarchique</i>	118
<i>Figure 4-19 Regroupement II-1</i>	119
<i>Figure 4-20 Regroupement II-2</i>	120
<i>Figure 4-21 Regroupement II-3</i>	121
<i>Figure 4-22 Regroupement II-4</i>	122
<i>Figure 4-23 Regroupement II-5</i>	123
<i>Figure 4-24 Regroupement II-6</i>	124
<i>Figure 4-25 Regroupement II-7</i>	125
<i>Figure 4-26 Regroupement II-8</i>	126
<i>Figure 4-27 Cartographie de la classification par regroupement hiérarchique des corrélations entre les moyennes mensuelles de l'IPM et celles des classes de luminosité des eaux.....</i>	Erro! Indicador não definido.

Introduction Générale

Le paludisme est endémique de la région amazonienne située dans la région nord du Brésil qui concentre 99% des cas dans le pays. L'Amazonie brésilienne possède toutes les conditions pour la prolifération du paludisme. Cette maladie se transmet principalement par le moustique *Anophèles darlingi* qui est éminemment adaptable à son environnement et se reproduit dans les différents plans d'eau comme les fleuves et les lacs.

Le cycle épidémiologique du paludisme se calque sur le cycle annuel des eaux des fleuves amazoniens. Les fleuves amazoniens ont été caractérisés depuis Sioli et Klinge (1962) par la couleur de l'eau. Les fleuves noirs provenant du plateau des Guyanes sont chargés en matière organique dissoute et présentent un pH faible, les fleuves blancs provenant de la cordillère des Andes sont chargés d'alluvions et ont un pH alcalin et les fleuves clairs, provenant des plateaux de la région centrale du Brésil, se caractérisent par leur faible charge en sédiments argileux et un pH alcalin.

Quelques études ont montré, pour différentes localités de l'État d'Amazonas, la relation entre l'incidence du paludisme et des variables hydrologiques tel que la hauteur d'eau (Wolfarth et al. 2013; Coutinho 2015) ou plus localement la couleur de l'eau (Tadei et al. 1998). Cependant, les variations saisonnières de la qualité (couleur) de l'eau des fleuves à l'échelle régionale ainsi que leurs relations avec l'incidence du paludisme est encore mal connue.

La question centrale de ce travail est ; Existe-t-il une relation entre la couleur des eaux telle que définie par Sioli et Klinge (1962) et les incidences du paludisme en

Amazonie ? Notre objectif est de caractériser cette relation si elle existe à l'échelle régionale.

Notre méthode est basée sur la mise en relation de deux jeux de données : une base de données épidémiologiques d'une part et un jeu d'images MODIS, permettant de caractériser la couleur de l'eau à un pas de temps mensuel, sur une période de temps couverte par les deux bases de données, soit entre 2003 et 2014. La base de données épidémiologiques, par ses caractéristiques qui seront discutées au chapitre 2 a également conduit à choisir le cadre géographique de l'étude : l'Etat d'Amazonas.

Le travail est ainsi structuré de la manière suivante:

Le Chapitre I présente les principales facettes de la complexité géographique, écologique, géochimique et sociale de l'État d'Amazonas. Ce chapitre introduit ensuite la caractérisation physique/optique de l'eau par des capteurs optiques de télédétection.

Le chapitre II fait une analyse approfondie du Système de Vigilance épidémiologique du Paludisme du Brésil (SIVEP-Malaria) pour mieux comprendre son fonctionnement et ses sources d'incertitudes afin de pouvoir en faire une utilisation scientifique.

Le chapitre III décrit la méthode développée pour suivre les variations de la couleur de l'eau au pas de temps mensuel sur une période de 11 ans dans l'État d'Amazonas à partir d'un indice original de luminosité de l'eau.

Le chapitre IV présente l'étude des corrélations entre les dynamiques épidémiologiques du paludisme et celles de la luminosité des eaux de l'État d'Amazonas afin de comprendre l'influence des changements de couleur de l'eau sur l'incidence du paludisme dans différentes zones choisies comme représentatives de ces dynamiques croisées.

CHAPITRE 1 Contexte de la recherche

Ce premier chapitre expose le contexte de cette thèse en décrivant les principaux facteurs formant le cadre des questions de recherche posées et des méthodes adoptées.

L'objectif général est de caractériser la relation, si elle existe, entre l'incidence du paludisme et la luminosité de l'eau dans l'État d'Amazonas entre 2003 et fin 2014.

Le Système de Surveillance Epidémiologique du paludisme (SIVEP-Malaria), qui existe depuis 2003, rassemble des informations sur des patients suspectés ou avérés d'infection par le paludisme au travers de 43 variables pour les 9 États Brésiliens de la Région Amazonienne. L'État d'Amazonas est l'État du Brésil où le paludisme existe à l'état endémique sur la plus grande surface. C'est aussi l'État pour lequel la base de données SIVEP-Malaria est la plus complète. Enfin c'est là que se trouvent les plus grands fleuves du bassin amazonien caractéristiques des eaux noires et des eaux blanches. Pour cette raison, l'État d'Amazonas a été choisi comme zone d'étude sur la période de 2003 à 2014..

L'objectif général se décline en plusieurs objectifs spécifiques qui forment la structure de ce travail :

1. Quelles sont les limitations et les sources d'erreur présentes dans la base de données SIVEP-Malaria
2. Comment réaliser une extraction totalement automatique cohérente dans le temps et dans l'espace des zones en eau sur un grand ensemble de données qui

représentent une période de plusieurs années sur une très grande région présentant une forte variabilité spatio-temporelle

3. Comment caractériser de façon stable et relativement peu coûteuse en temps calcul la luminosité de l'eau

4. Comment caractériser les relations entre incidence des cas de paludisme et luminosité de l'eau.

1.1 Le paludisme des rivières

Le terme paludisme des forêts est le terme utilisé pour parler d'une population en situation de risque paludique et habitant le milieu forestier (Kar et al. 2014). Tout comme il existe les "peuples des eaux" (Martinelli 2014; Diegues 2001; Diegues 2002) (Figure 1-3) nous parlerons de paludisme des rivières. C'est un paludisme qui accompagne les populations riveraines des fleuves selon la saisonnalité des eaux et selon leur déplacement dans une région où les routes sont les fleuves.

Les flux migratoires qui ont peuplé la Région Amazonienne sont le fruit des explosions économiques successives qui s'y sont déroulées au fil des siècles et des incitations des gouvernements. Ces flux ont emprunté les fleuves comme axe de pénétration, au bord desquels se sont constituées les villes (Becker 2013). C'est pourquoi les fleuves dessinent le paysage mais aussi la régionalisation de ce territoire. Pour cette raison également, en Région Amazonienne, le paludisme est souvent étudié par grande région fluviale (Suárez-Mutis & Coura 2007; Santelli et al. 2012; Pantoja & Andrade 2012; Cabral et al. 2010; Assis & Gurgel 2008; Santos et al. 2007; Kffuri 2014; Wolfarth et al. 2012; Saraiva et al. 2009; Melo & Pinheiro 2012; Assis et al. 2008; Magris et al. 2007; Katsuragawa et al. 2008). L'État d'Amazonas

possède 9.829,6 km de routes pour 1.570.745 km² de surface, soit 0,006 km/km² de routes (Figure 1-1).

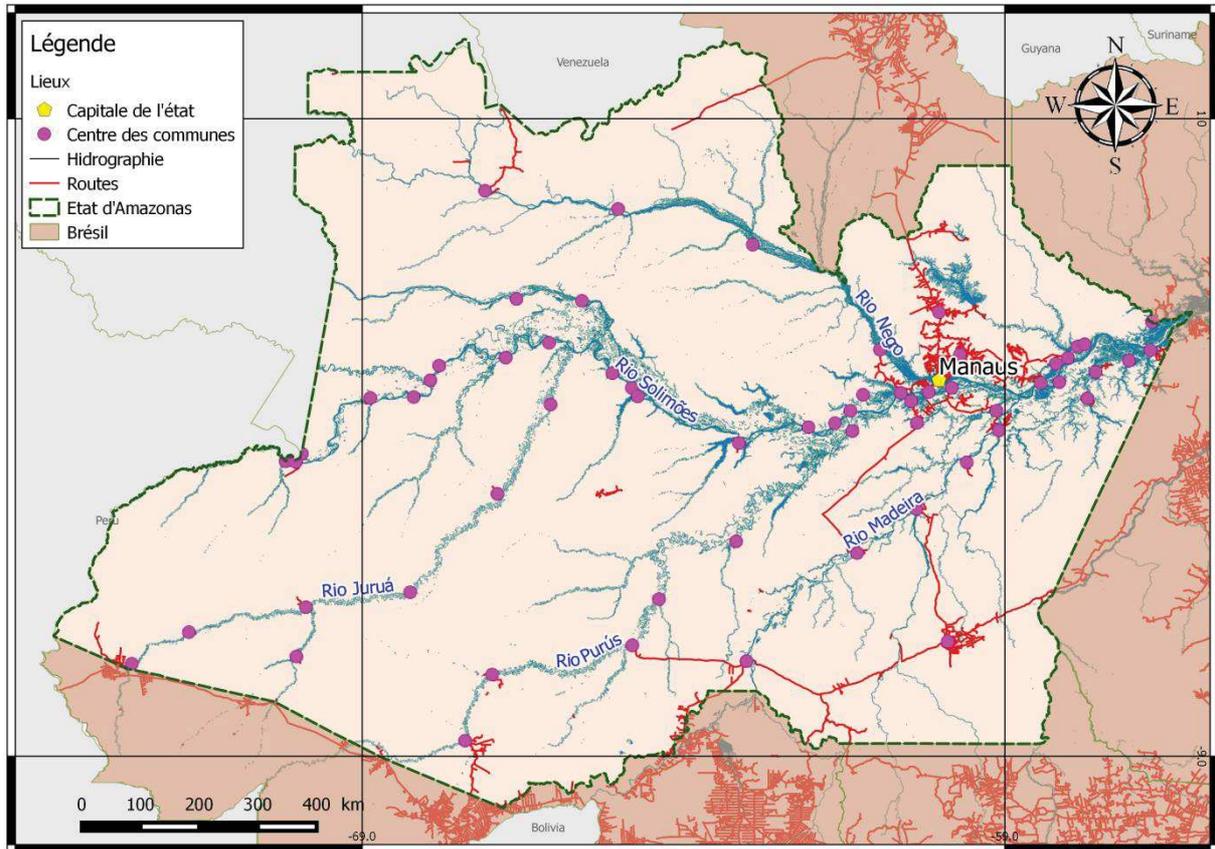


Figure 1-1 Routes et hydrographie de l'État d'Amazonas

En comparaison, l'État voisin et deuxième plus grand État du pays, l'État du Pará possède 0,07 km/km² de route et l'État de São Paulo, le plus industrialisé, en a 0,32 km/km². La Figure 1-6 montre la différence entre les réalités régionales du réseau routier et met en évidence la condition d'État fluvial de l'État d'Amazonas en comparaison du reste du pays.



Figure 1-2 Routes sur le territoire brésilien et sur l'État d'Amazonas



Figure 1-3 Le transport des communautés riveraines amazoniennes (photo: Martinelli 2014)

1.2 Etat d'Amazonas et rôle de l'eau dans l'écologie du moustique Anophèles

En 2015 le Brésil compte 143.415 cas de paludisme dont 36 suivis de décès, un chiffre qui décroît d'année en année (World Health Organization 2015).

Le paludisme est endémique de la région amazonienne située dans la région Nord du Brésil et qui concentre 99% des cas dans le pays. (Sampaio et al. 2015; Mourao et al. 2014; Oliveira-Ferreira et al. 2010; Braz et al. 2006; Katsuragawa et al. 2008). L'État d'Amazonas concentre 42% des cas de paludisme, suivi de l'État de l'Acre avec 25% (Brasil, Ministério da Saúde 2015).

L'Amazonie brésilienne possède toutes les conditions pour la présence du paludisme (Stefani et al. 2013; Tadei et al. 1988; Santos et al. 1981). Cette maladie est

transmise principalement par le moustique *Anophèles darlingi*. L'*Anophèles* se reproduit en milieu aquatique et est doté d'une importante adaptabilité à différents environnements présents dans cette région (Tadei et al. 2016). Les principales espèces de plasmodium responsables des infections dans l'État d'Amazonas sont *P. falciparum* (16%) et *P. vivax* (84%).

Le paludisme entretient un lien étroit avec les fleuves Amazoniens et son cycle épidémiologique se calque sur le cycle annuel des eaux (Assis et al. 2008; Wolfarth et al. 2013; Tadei et al. 1988). La crue des fleuves du bassin Amazonien entre novembre et juillet fait monter le niveau de 10 à 15 mètres à Manaus, capitale de l'Etat d'Amazonas, située à la rencontre des eaux des deux grands fleuves Negro et Solimões. La plupart des centres urbains subissent des inondations annuelles dues à cette crue. Le Rio Negro, avec des eaux noires chargées de matière organique dissoute, prend sa source sur le plateau des Guyanes à 2250 km de Manaus. Le Rio Solimões, chargé d'alluvions et caractérisé par une teinte laiteuse qui lui vaut d'être le principal représentant des fleuves dits « d'eau blanche », prend sa source dans les Andes à 5400 km de Manaus.

La saison de la crue provoque le débordements de ces fleuves sur les varzeas (lacs d'inondation d'eau blanche), et les igapós (lacs d'inondation d'eau noire)(Figure 1-1) et influence le cycle de vie d'une grande variété d'espèces animales et végétales (Ianniruberto et al. 2015; Gualtieri et al. 2015; Moquet et al. 2016). Les variations biogéochimiques modifient la bio-écologie des écosystèmes (de Barros et al. 2007; Barros & Honório 2007; Magris et al. 2007; Basurko et al. 2011; Confalonieri et al. 2014) et sont en grande relation avec la reproduction des *Anophèles* (Tadei et al. 2009).



Figure 1-4 Niches écologiques de l'Anophèles: A) Varzea inondée ; B) Forêt d'igapós inondée (extrait de: Tadei et al. 2009)

Les gîtes larvaires de l'*Anophèles* sont variés et peuvent se trouver dans des zones à ciel ouvert, dans des ruisseaux d'eau courante, en eaux stagnante et ombragée, sur des végétations aquatiques comme le *Pistia sp.*, *Ludwigia*, *Cabomba*, *Tonina*, *Miconia*, et *Rollini* ou dans des flaques d'eau et fossés (Tadei, B. Thatcher, et al. 1998).

On sait aussi que ce moustique est moins présent en saison sèche. Pendant cette période il cherche des niches plus fraîches proches des eaux plus profondes. On les retrouve ainsi souvent loin des berges dans les varzeas et igapos accrochés aux feuillages et aux troncs flottants (Deane, Causey, and Deane 1948). Les parties toujours en eau de ces lacs d'inondation sont ainsi des gîtes larvaires pour les *Anophèles* tout au long de l'année.

Le cycle de vie depuis l'œuf jusqu'au spécimen adulte varie selon la température de l'environnement. En conditions normales, cela peut varier entre 12 et 16 jours (Forattini 1962; Santos et al. 1981). Cette période peut être de 7 jours à 31°C ou a 20 jours à 20°C (Williams & Pinto 2012). La température est relativement constante

tout au long de l'année en bordure des fleuves (amplitude diurne d'une dizaine de degrés mais moins de 2° de variation sur la moyenne journalière sur l'année), et ne constitue pas une variable discriminante

Plusieurs études relèvent que la pluie est une variable essentielle (Amanajás et al. 2011; Kar et al. 2014; Wolfarth et al. 2013; Gurgel 2006). Non seulement elle conditionne l'extension du milieu aquatique nécessaire à la phase larvaire du cycle de vie du moustique mais on observe aussi une plus grande longévité du moustique dans des conditions plus humides (Olson et al. 2009; Hahn et al. 2014).

Il est intéressant de noter également que pendant le pic de crue les conditions nécessaires à la prolifération du moustique ne sont pas réunies, à cause de l'écoulement rapide de l'eau qui traîne les larves de moustiques loin des lieux de ponte vers les fleuves. L'épidémie de paludisme débute en période de décrue, principalement à partir du mois d'août (Confalonieri 2003; Motta 2015; Katsuragawa et al. 2010; Gil et al. 2003; Wolfarth et al. 2013), mais c'est après les premières grandes pluies que les taux de transmissions augmentent (Braz 2013; Angelo 2015) car les gîtes larvaires sont inondés progressivement faisant éclore les œufs alors que le courant est encore trop faible pour les arracher des lieux de ponte (Assis et al. 2008).

Différentes études définissent qu'un pH de l'eau entre 6,5 et 7,3 constitue une niche écologique idéale pour les moustiques *Anophèles darlingi* (Hiwat et al. 2011), mais également qu'ils se reproduisent plus aisément dans des milieux plus alcalins (Tadei, B. D. Thatcher, et al. 1998). Ils sont néanmoins présents dans toutes les communautés du Rio Negro, un fleuve présentant un pH acide (Tadei et al. 2009; Cabral et al. 2010). Il est avancé que la période des pluies peut produire des

variations de pH importantes créant les conditions pour les gîtes larvaires de l'*Anophèles darlingi* dans des milieux jusque-là trop acides (Hiwat et al. 2011; Cabral et al. 2010; Giglioli 1938).

Les études locales le long des fleuves indiquent qu'il existe différents modèles saisonniers du paludisme en Amazonie dû à la diversité de milieux qu'on y trouve (Becker 2013; Wolfarth et al. 2013; Confalonieri et al. 2014; Basurko et al. 2011). Selon la région et ses caractéristiques hydrologiques, l'épidémie de paludisme a des caractéristiques différentes. Chaque fleuve a sa propre dynamique. La plus grande présence de larves d'*Anophèles* est observée pendant la baisse du niveau des eaux pour l'amont du fleuve Amazone (Vittor et al. 2009). Pour le moyen Orénoque, on observe une augmentation de la présence des moustiques après le pic des pluies, quand le fleuve atteint son maximum de débit (Magris et al. 2007). Pour le Rio Purus, la plus grande concentration de cas de paludisme se trouve dans la région de la plaine où le fleuve perd de la vitesse et où la nature est plus préservée. Quand le fleuve Purus croise les routes entre l'Acre et le Pérou, le paludisme est caractérisé comme un paludisme de frontière, majoritairement importé d'autres états ou pays (Assis et al. 2008). Les routes faciliteraient le transport de la maladie dans cette région (Santos et al. 2007). Santos et al., (2007) montre en effet l'existence de différentes dynamiques de contagion et de transfert de l'agent contagieux du paludisme selon l'organisation territoriale des axes de transport.

On peut observer aussi que la construction des usines hydroélectriques sont un cas à part de modèle de dispersion des épidémies car leurs chantiers génèrent un grand nombre de cas de paludisme comme a Santo Antônio ou Jirau sur le Rio Madeira (Katsuragawa et al. 2008; Katsuragawa et al. 2009), plus spécifiquement chez les migrants provenant de régions non-endémiques attirés par les opportunités créées

par le chantier et qui sont souvent exposés à une situation sanitaire déficiente. Le manque d'immunité de ces populations peut contribuer à la forte parasitémie (grandes quantités de parasites par millilitre de sang) qui mène à des paludismes très graves (Parise 2009; Gomes et al. 2011; Fernandes et al. 2010).

1.3 L'eau dans l'État d'Amazonas

Le bassin amazonien est le plus grand bassin hydrographique au monde avec une superficie de 6.915.000 km² et un débit total estimée à 209 000 m³/s (Goulding et al. 2003). L'État d'Amazonas avec une surface proche de 1.571.000 km² est entièrement situé dans ce bassin (Figure 1-2). Cet État dispose d'un réseau fluvial navigable de 22.000 kilomètres.

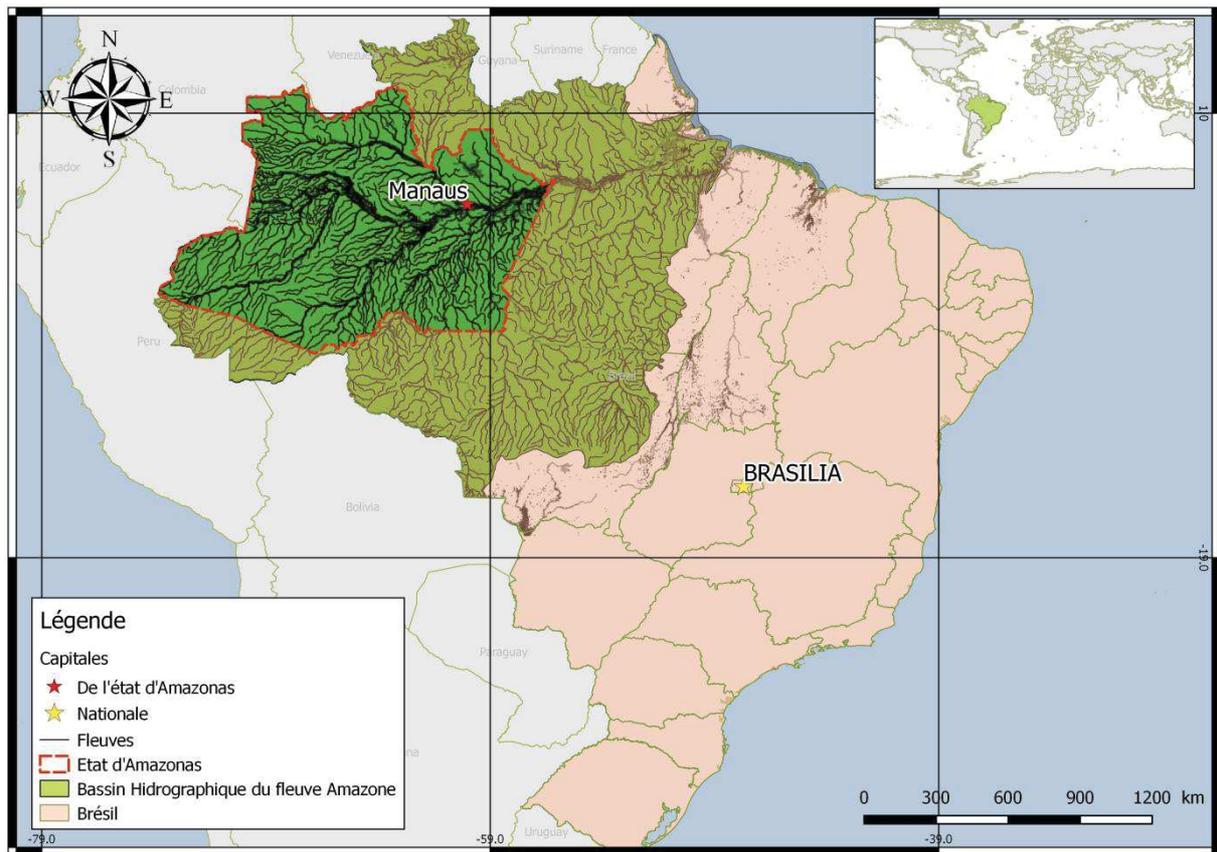


Figure 1-5 Carte du bassin Amazonien brésilien et état d'Amazonas

Ce réseau se substitue au réseau routier quasi-inexistant pour la plus grande partie du territoire. Les principaux affluents du Rio Amazonas dans l'État d'Amazonas sont le Rio Negro, le Rio Solimões, le Rio Madeira, le Rio Japurá, le Rio Juruá et le Rio Purus (Figure 1-3).

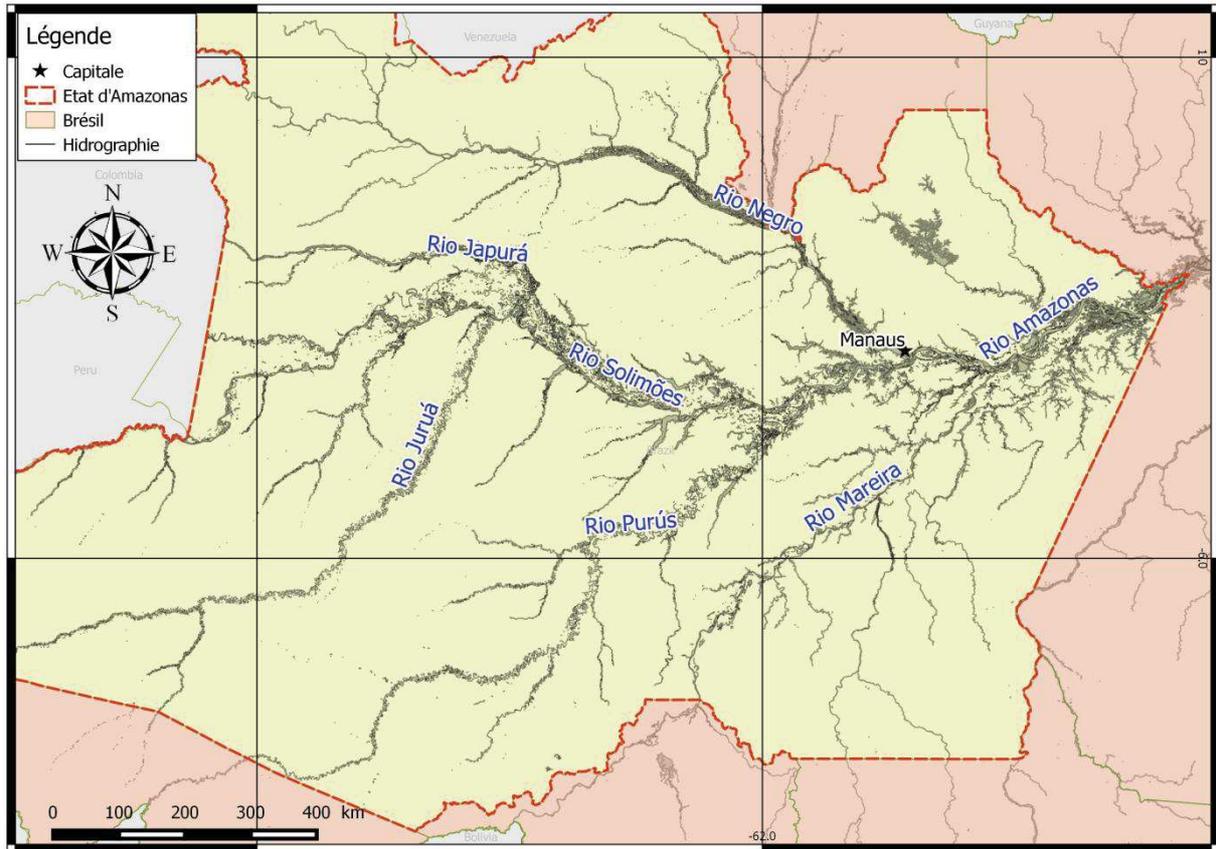


Figure 1-6 Carte des principaux fleuves affluents du Rio Amazonas dans l'état d'Amazonas

La composition des eaux des bassins de drainage sont le produit de l'interaction physique et chimique qui survient au long du processus d'infiltration et de ruissellement (Martinez et al. 2009; Kilham & Roberts 2011; Stallard & Edmond 1983). Sioli et Klinge (1962) et Sioli (1968), pionniers des études des fleuves amazoniens, classifient en trois grandes catégories les fleuves amazoniens selon leurs caractéristiques hydrographiques: fleuves d'eaux blanches, fleuves d'eaux noires et fleuves d'eaux claires. Le Rio Negro est l'exemple le plus connu de fleuve d'eaux noires du bassin amazonien et le Rio Solimões est celui connu pour ses eaux blanches. Ces deux fleuves se rencontrent devant Manaus pour former le fleuve Amazonas (Figure 1-4).



Figure 1-7 Rencontre des fleuves Rio Negro et Rio Solimões. (Source: Wikimedia Commons par Mario Ortiz)

Les eaux claires ne sont pas connues pour être présentes de façon significative dans l'état d'Amazonas (l'exemple typique est celui du fleuve Tapajós dans l'État de Pará). Le faible apport en nutriments et le pH acide des eaux font de la région du Rio Negro une région moins propice à la diversité biologique que le Rio Solimões et est, par conséquent, moins habitée par l'homme.

Le Rio Negro - Le Rio Negro est le principal représentant des eaux noires du bassin amazonien. Il est l'un des trois plus grands fleuves du monde. Son débit est supérieur à tous les fleuves européens rassemblés. Il est aussi l'un des principaux affluents du Rio Amazonas, responsable de 15% du débit de celui-ci dans

l'Atlantique. Le Rio Negro draine une surface qui correspond à 10% des 7 millions de kilomètres carrés du Bassin Amazonien.

Le Rio Negro tient son nom de sa couleur qui s'apparente à un thé obscur. Avec une transparence entre 1,30 et 2,30 mètres, il provient du drainage du plateau des Guyanes. Les acides fulviques et humiques dissous qui donnent cette coloration sont issus de la lente décomposition de la biomasse pendant des milliers d'années sur les grandes étendues inondables qu'il baigne. Il a une grande quantité de Na^+ (Sodium) et K^+ (Potassium), ce qui lui confère un pH réduit (autour de 4.0) (Furch 1984; Walker 1987; Melfi et al. 2011). La matière organique est aussi produite dans l'eau par le phytoplancton dans la zone euphotique de la colonne d'eau. Le phytoplancton consomme les nutriments et les convertit en matière organique par la photosynthèse (production primaire). Le zooplancton consomme le phytoplancton et crée encore de la matière organique. Dans certains cas il peut y avoir une quantité suffisante de matière organique dissoute dans l'eau pour réduire considérablement la pénétration de lumière dans la colonne d'eau. Mais contre un fond de sable, en zone peu profonde, l'eau peut paraître limpide car elle transporte peu de sédiments. La matière organique est principalement issue des zones inondées périodiquement. Ces zones inondées sont appelées par les populations autochtones "igapós" quand elles sont recouvertes de forêts. Les forêts des zones d'igapós sont appelées "caaigapó" (forêts inondées).

Les forêts du bassin du Rio Negro sont parmi les moins peuplées et les mieux préservées au Monde. Les eaux noires du fleuve n'offrent pas les conditions nécessaires à l'agriculture. L'acidité des eaux ne favoriserait pas les insectes comme les moustiques, tenus comme bien plus présents dans les fleuves d'eau blanche. Les

forêts du bassin du rio Negro sont aussi, en comparaison avec les autres, pauvres en animaux terrestres et aquatiques (Oliveira & Dalcy 2001; Junk 1983).

Le Rio Solimões - Le Rio Solimões est le principal représentant des eaux blanches dans le Bassin amazonien. Ce fleuve est le tronçon du Rio Amazonas entre la triple-frontière Brésil-Pérou-Colombie et la ville de Manaus. Il possède des eaux navigables sur 1.620 km avec une profondeur comprise entre 8 et 20 mètres. Sa période de crue s'étale de février à juin.

Les eaux du Rio Solimões, du Rio Purus et du Rio Madeira sont dites d'eaux blanches car elles sont chargées en sédiments. Elles proviennent des Andes, une chaîne de montagne relativement jeune où les processus érosifs sont actifs, notamment grâce au climat régional. Le pH de ces eaux est proche de 7 et elles sont classées comme bicarbonatées car elles sont riches en Ca^{2+} (Calcium) et HCO_3 (Bicarbonate) et pauvres en matières organiques (Sioli 1968; Konhauser et al. 1994; Gaillardet et al. 1997; Queiroz et al. 2009).

Une autre caractéristique des fleuves d'eaux blanches est l'instabilité de leur cours naturels par l'action simultanée de la sédimentation et de l'érosion. Ce sont des fleuves en méandres qui coulent dans leurs plaines d'inondation. Ce mouvement constant érode les marges de terrains argilo-sableux mal consolidés qui provoquent pendant les crues des glissements de terrain qui, à leur tour, sont une des sources de matière argileuse rencontrée en suspension dans les eaux.

Aux bordures des fleuves blancs on peut apercevoir des étendues de sols fertiles de plus en plus exploités pour l'agriculture. Ce sont les varzeas les plus élevées qui subissent l'inondation annuelle pour un temps moindre et permettent l'agriculture

d'espèces de cycle court (niébé, riz, maïs, melon d'eau, etc.), les fibres (malva, jute) et des légumes (choux, concombres, etc.) (Fajardo et al. 2009).

Les eaux claires - Les fleuves d'eaux claires ou limpides se caractérisent par leur faible charge en matières en suspension. Les principaux représentants de ce type d'eau sont le Rio Tapajós et le Rio Xingú qui proviennent des plateaux de la région centrale du Brésil, et se trouvent en dehors de l'État d'Amazonas. La région centrale du Brésil est une région plate et régulière avec une faible érosion. On observe pour ces fleuves une charge sédimentaire pendant la saison des grandes pluies, qui entraîne des sols plutôt rocheux ou sableux avec peu d'argile et génèrent des plages pendant les périodes sèches.

Ces trois catégories primaires de couleur d'eau sont accompagnées de diverses variantes issues des mélanges entre les eaux des fleuves provenant de différentes régions et par d'autres types de variation. Un exemple typique de cette transformation est ce qui se produit dans le cours supérieur du Rio Branco, grand affluent du Rio Negro. Il porte son nom car il descend des montagnes de la frontière avec le Venezuela et transporte pendant les périodes de crue une grande quantité de matériel en suspension, ce qui lui donne une teinte blanche. Pendant les périodes sèches ses eaux deviennent transparentes. Aussi, les petites rivières amazoniennes, les igarapés, subissent des variations importantes de la qualité des eaux. Leur composition est similaire à celles des grands fleuves néanmoins les variations sont beaucoup plus abruptes et peuvent être perçues dans un laps de temps d'une journée et même de quelques heures.

Différentes études font référence lorsqu'on parle du débit solide (matières en suspension) ou débit dissous des fleuves amazoniens; certaines études se

spécialisent plutôt sur le débit dissous (Sánchez et al. 2015; Moquet et al. 2016; Gaillardet et al. 1997) et d'autres sur le débit solide (Martinez et al. 2007, 2009; Filizola 2011). D'autres aussi détaillent les origines des sédiments (Dunne et al. 1998; Meade et al. 1985).

Des études en télédétection produisent des connaissances sur la relation qui existe entre les propriétés optiques et les paramètres de la qualité de l'eau. Il existe encore peu d'études sur la qualité de l'eau à l'échelle régionale. Il existe aussi très peu d'études qui font une relation entre la variation de la couleur, le niveau des eaux en Amazonie et des données épidémiologiques (Basurko et al. 2011; Wolfarth et al. 2013; Confalonieri et al. 2014).

Les données d'observation de couleur de l'eau de cette recherche sont produites par télédétection et la méthode utilisée est le produit de l'application de la connaissance des interactions physiques entre la radiation électromagnétique et les éléments présents dans l'eau.

1.4 La télédétection de la couleur de l'eau

La télédétection de la couleur de l'eau est l'observation des propriétés optiques apparentes de l'eau telles que la réflectance, l'irradiance ascendante, la transparence ou le coefficient d'atténuation (Danson & Plummer 1995). Ces mesures sont des indicateurs indirects des caractéristiques physiques, chimiques et biologiques de l'eau (Ritchie et al. 2003).

Le grand défi de travailler sur la région amazonienne est de s'abstraire de l'abondance de nuages présents une grande partie de l'année (Samanta et al. 2012; Schroeder et al. 2008).

En télédétection, les méthodes pour classer les eaux en fonction de leurs couleurs diffèrent selon les caractéristiques des capteurs, principalement selon leurs résolutions spatiale, spectrale ou radiométrique. Elles diffèrent aussi selon la taille de la zone d'étude et la récurrence des observations prévues. Il faut aussi considérer les caractéristiques physiques, spécifiques aux objets de la zone d'étude.

Les propriétés des différentes eaux superficielles subissent de grandes variations au cours de l'année en fonction du type limnologique, des saisons, de l'activité biologique et de l'impact humain. Ainsi, il n'est pas possible de définir simplement les types d'eau comme plus ou moins transparentes. Les caractéristiques optiques des eaux sont définies par les composants optiquement actifs dans l'eau (Richardson & LeDrew 2006). Dans les eaux continentales les principaux composants sont la matière en suspension, la matière organique dissoute et le phytoplancton (Arst 2003).

L'identification et la quantification des éléments présents dans l'eau, par télédétection est possible car ces éléments ont des relations particulières avec le rayonnement électromagnétique (REM). Mise à part la production chlorophyllienne, la réflectance des caractéristiques optiquement actives de l'eau se produit majoritairement sur les bandes du visible des images de télédétection (Bitencourt-Pereira 1992; Pompêo et al. 2015; Palmer et al. 2015).

Lathrop, Lillesand, et Yandell (1991) ont testé l'utilité des images multi-spectrales pour le suivi de la turbidité des eaux continentales et ont conclu que le quotient

rouge/bleu donne les meilleurs résultats dans la recherche des eaux transparentes. A mesure qu'augmentent les sédiments minéraux en suspension, augmente la réflectance de l'eau dans le visible et le proche infra-rouge (PIR) avec un pic sur le rouge. Au contraire, à mesure qu'augmente la concentration de matière organique dissoute dans l'eau, c'est l'absorption qui augmente (Kutser et al. 2005).

L'eau pure ne contient pas de matière organique ou inorganique. On retrouve dans la littérature les valeurs des coefficients d'absorption, diffusion et d'atténuation des molécules d'eau dans la bande spectrale allant du 250 à 800nm (Jensen 2007).

En résumé, les trois grands groupes de composants optiquement actifs de ces eaux sont :

- Les particules inorganiques ou organiques en suspension dans l'eau (sédiments; argiles...)
- La matière organique dissoute dans l'eau (DOM - acide fulvique et acide humique)
- Les pigments en relation avec la présence d'organismes vivants (phytoplanctons) comme la chlorophylle-a

Les eaux turbides ou avec des matériaux en suspension reflètent la radiation électromagnétique selon les caractéristiques de ces éléments (i.e. calcium et magnésium). Les eaux limpides (avec une certaine profondeur; afin d'éviter l'influence du réfléchissement des fonds) absorbent la plupart des rayons électromagnétiques (Lyzenga 1981; Novo; Hansom; and Curran 1989; Novo; Steffen; and Braga 1991).

La recherche pour l'amélioration des algorithmes de classification de l'eau à partir de l'association et du seuillage de différents indices génère des résultats toujours plus performants (Zhang et al. 2007) mais pas encore universels.

Deux problématiques émergent :

1. la séparation entre zone en eau et zone « hors d'eau » dénommé ci-après masque de l'eau.

2. La caractérisation de la dynamique des valeurs de luminosité

Au cours des dernières décennies, un certain nombre de techniques de télédétection ont été utilisées pour la détection et l'extraction des littoraux/rivages et les eaux intérieures servant à surveiller et détecter les changements dans l'espace et dans le temps pour différentes régions du globe. Mais chaque méthode est très spécifique du contexte géographique.

On retient que la méthode la plus simple pour réaliser un masque d'eau est de définir un seuil sur une bande infra-rouge pour distinguer le contraste entre les pixels qui réfléchissent le rayonnement électromagnétique (REM), et ceux qui ne la réfléchissent pas pour cette longueur d'onde, dont, principalement des pixels d'eau claire et d'eau noire. Les pixels d'eau chargée de sédiments réfléchissent de manière plus significative (Overton 2005; Powell et al. 2008).

Quelques travaux ont été réalisés pour la cartographie automatique des eaux à l'échelle globale qui observent les disparités locales, mais ces travaux ne sont pas conçus pour le multi-temporel (McFeeters 1996; Feng et al. 2015; Verpoorter et al. 2014; Sharma et al. 2015). Aussi, les travaux de Berry et al. (2005) et de Bartsch et al. (2012) sont particulièrement importants pour ce travail puisqu'ils sont à l'origine du

produit Land Cover (LC) *Water Bodies* (Land Cover CCI 2016) utilisé pour la validation de ce travail.

L'extraction journalière des plans d'eau par les images MODIS (haute résolution temporelle) est pratiquée couramment dans le but de surveiller les crues des fleuves et cartographier les inondations au niveau local (Martinis et al. 2013; Ticehurst et al. 2014; Hu et al. 2015). Les principales méthodes d'extraction des plans d'eau sont décrites dans l'article de Nath and Deb (2010).

1.5 Conclusion partielle

Ce travail de thèse se focalise sur le paludisme des rivières dans l'état d'Amazonas, les rivières et fleuves étant la principale voie de communication et espace habité à l'intérieur de l'état. Les méthodologies de télédétection à mettre en œuvre concernent 1) d'une part la mise au point d'une méthode de classification permettant d'isoler une classe « eau » du reste de l'image, méthode suffisamment robuste pour identifier l'eau dans toute sa diversité de qualité et de coloration, à l'échelle régionale de l'État d'Amazonas et sur une profondeur temporelle de plus de dix ans, donc tenant compte des variations intra et interannuelles ; 2) une classification de l'eau en fonction de sa luminosité, comme un indicateur des deux grands types d'eau présents dans les fleuves Amazoniens, d'eau noire ou d'eau blanche.

Les justifications pour étudier le paludisme des rivières dans l'État d'Amazonas sont multiples :

- Le paludisme est endémique de la région amazonienne (99% des cas dans le pays) dont 42% dans le seul état d'Amazonas.

- Le paludisme entretient un lien étroit avec les fleuves Amazoniens, mais les gîtes larvaires de l'*Anophèles* sont très variés : eau courante, eau stagnante, végétation aquatique ou flaques d'eau et fossés

- Le moustique *Anophèles* est moins présent en saison sèche mais peut être présent sur les lacs tout au long de l'année. La température est propice partout dans l'État. Le pic de crue n'est pas favorable, mais les taux de transmission augmentent à la montée de crue et lors de la décrue.

- Un pH de l'eau légèrement alcalin est favorable à la reproduction du moustique *Anophèles darlingi* mais il est néanmoins présent dans toutes les communautés du Rio Negro, un fleuve plutôt de pH acide, mais qui peut voir des variations de pH importantes en saison des pluies.

Beaucoup d'autres facteurs que la qualité de l'eau interviennent dans la transmission et l'occurrence du paludisme.

La question de la relation entre l'exposition au risque de transmission du paludisme et la qualité de l'eau dans l'État d'Amazonas est une question ouverte.

CHAPITRE 2 Analyse critique des données de surveillance épidémiologique du Paludisme

2.1 Introduction

Les systèmes de surveillance épidémiologiques sont conçus pour le suivi et le contrôle de maladies par le gouvernement (Bayat et al. 2001) mais pas pour des applications scientifiques. L'OMS définit un *système de surveillance épidémiologique* comme :

« La collecte, l'analyse et l'interprétation régulières et systématiques de données sanitaires pour la description et l'observation continue d'un événement de santé en vue de faciliter la planification, la mise en œuvre et l'évaluation des interventions et programmes de santé publique » (World Health Organization 2002).

Le recueil de données pour le Système de Surveillance Épidémiologique du Paludisme (SIVEP-Malaria) a été rendu obligatoire par le gouvernement Brésilien (Brasil 2006) et est décentralisé entre les 805 communes de la région Nord du pays. Celles-ci sont responsables de la saisie d'un grand nombre d'informations sur les patients. La décentralisation se traduit par des différences de pratiques dans les notifications, ce qui est une source potentielle d'erreurs.

Les données du SIVEP-Malaria sont disponibles en ligne et sont de plus en plus utilisées par la recherche scientifique avec une production de plus de 50 publications

entre thèses et articles pour l'année de 2014 contre 2 pour l'année de lancement en 2003.

Il n'existe pas de manuel pour l'emploi de ces données ni d'analyse critique de son contenu. La question qui se pose lors de l'utilisation scientifique de ces données est: Quelles sont les limitations et les sources d'erreurs présentes dans la base de données SIVEP-Malaria?

L'objectif de ce chapitre est de déterminer s'il existe des limitations à l'utilisation du SIVEP-Malaria pour des buts de recherche et quelle serait la méthode appropriée permettant son utilisation.

Ce chapitre décrit donc l'architecture du SIVEP-Malaria, les étapes de nettoyage de données qui ont été mises en œuvre et l'analyse des données erronées.

Ce qui nous conduit à une brève critique de l'indice d'Incidence Parasitaire Annuelle (IPA) du paludisme qui est employé dans la suite de cette thèse. L'IPA est l'indice le plus utilisé pour comparer et estimer les incidences du paludisme entre différentes localités mais, comme pour beaucoup d'indicateurs, son utilisation ne peut se faire sans quelques considérations critiques.

2.2 Le système de Surveillance épidémiologique de Paludisme Brésilien

En 1999 le Ministère de la Santé a lancé le Plan pour l'Intensification des actions de contrôle du paludisme [*Plano de Intensificação das Ações de Controle de Malária (PIACM)*] comme une réponse au projet de l'Organisation Mondiale de la Santé, le *Roll Back Malaria Program* (Braz 2013; Sampaio et al. 2015).

En 2003, en partenariat avec les états et les municipalités de la région amazonienne, le Système de Surveillance de la Santé (SVS) du Ministère de la Santé a créé le Programme National pour la Prévention et le Contrôle du Paludisme [*Programa Nacional de Prevenção e Controle da Malária (PNCM)*] comme une politique permanente pour la prévention et le contrôle de la maladie donnant suite aux progrès réalisés avec le PIACM (Deane 1988; Ferreira & Silva-Nunes 2010).

Le Système de Surveillance/Vigilance Epidémiologique du Paludisme (SIVEP-Malaria) est le principal outil utilisé par PNCM pour améliorer la production, le suivi et la qualité de l'information sur le paludisme. Les données produites dans la région amazonienne sont enregistrées dans le SIVEP-Malaria, alors que ailleurs au Brésil les données sont envoyées au Système d'Information de Maladies et Notification (SINAN) (CONASS 2011).

De 2003 à 2004, l'intégration des communes au programme SIVEP-Malaria est passé de 60% à 100% (Brasil, Tribunal de Contas da Uniao 2003; Braz et al. 2006).

Dans la région Nord du Brésil, tous les cas suspects de paludisme doivent être obligatoirement notifiés par les services de santé dans un délai de 7 jours au SIVEP-Malaria en fournissant les 43 informations du formulaire de notification (Brasil, Ministério da Saúde 2014). La prise en charge gratuite des patients n'est faite que si elle est accompagnée de la confirmation du diagnostic (Costa et al. 2010). La Figure 2-1 illustre le nombre total de cas positifs enregistrés pour l'État d'Amazonas entre les années 2003 et 2015. On y observe des cycles saisonniers mais surtout une tendance à la baisse très nette. Le Ministère de la santé a communiqué sur les effets positifs de la lutte anti-vectorielle et le projet d'élimination du paludisme à l'occasion du plus bas taux d'infection par paludisme des dernières 35 années avec une

réduction en 2015 de 89% par rapport à l'an 2000 (Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. 2016).

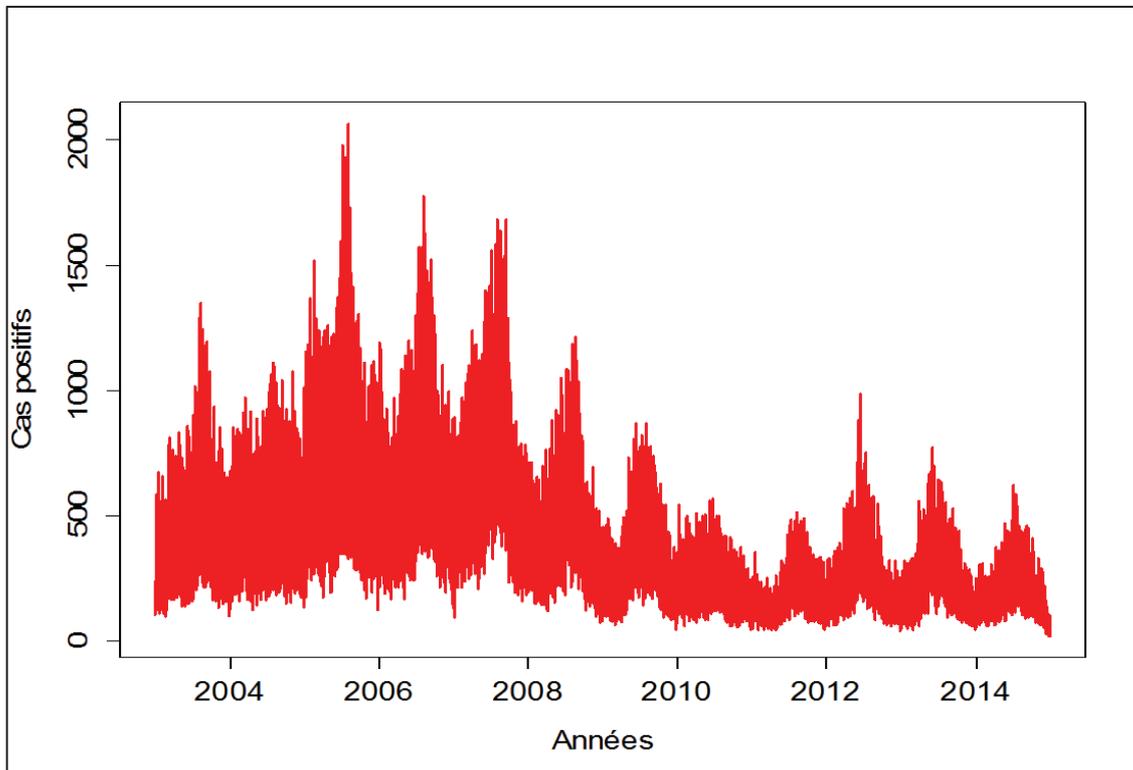


Figure 2-1 Cas positifs pour l'État d'Amazonas entre 2003 et 2015

Critique de l'Incidence Parasitaire Annuelle (IPA)

Les rapports gouvernementaux et la plupart des études scientifiques utilisent l'Incidence Parasitaire Annuel (IPA) comme indicateur de risque paludéen. Il s'agit du nombre de nouveaux cas parmi les populations exposées pendant un an (Equation 2-1).

$$IPA = \frac{\text{Nombre de cas}}{\text{Population}} \times 1000 \quad (2-1)$$

La Figure 2-2 montre le produit cartographique de l'IPA disponible sur le site du Système de Vigilance de la Santé (SVS) du Ministère de la Santé pour les années 2011 à 2014.

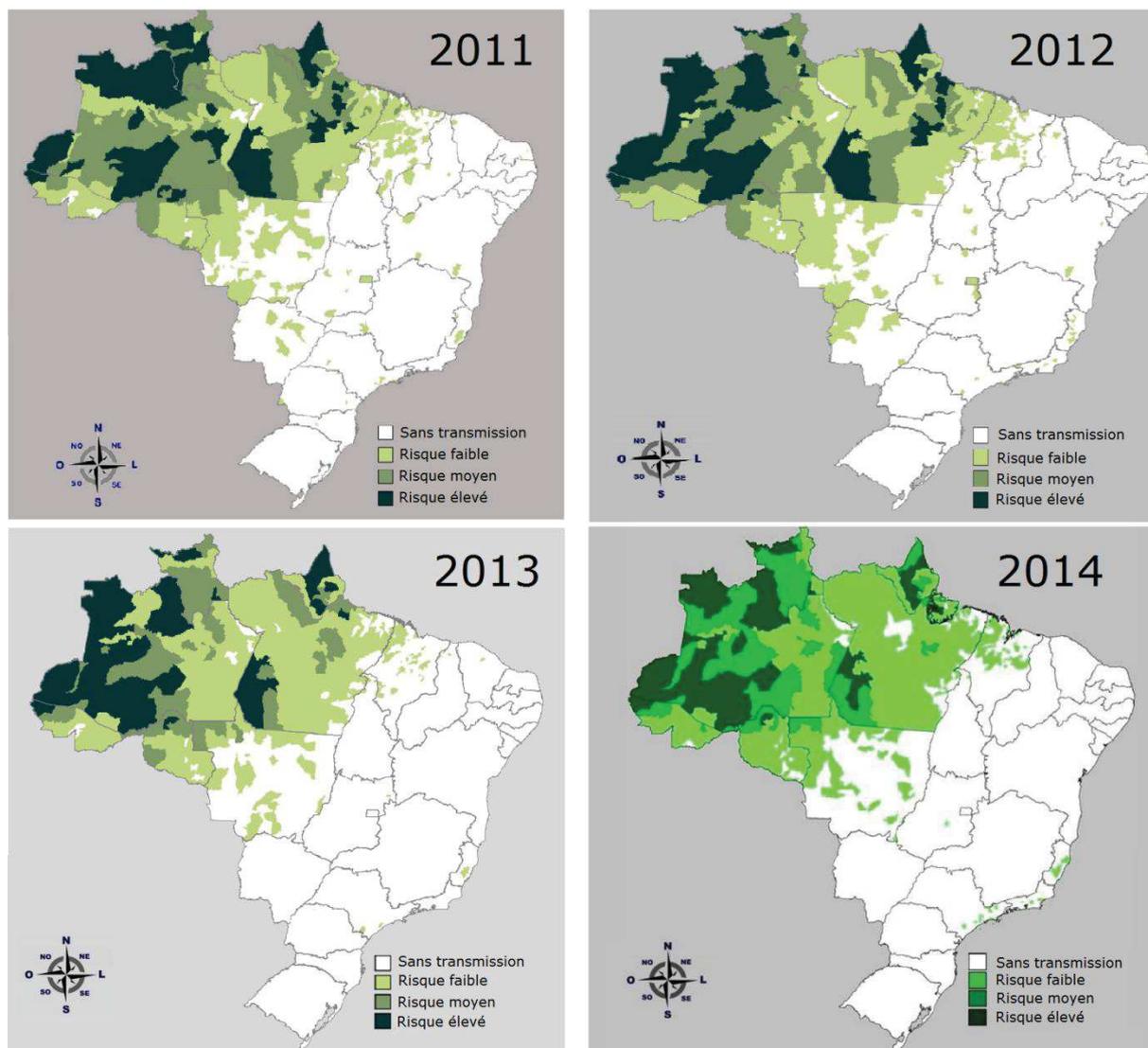


Figure 2-2 Cartes de l'IPA par commune. Produit du SVS/MS pour le public

L'IPA couramment utilisé est calculé sans discernement du type de *Plasmodium* à l'origine de l'infection chez le patient. D'autres variables peuvent avoir une influence sur l'indice, comme l'accès à la prévention par exemple, mais il reste cependant

l'indicateur le plus utilisé en épidémiologie grâce à sa pondération par l'effectif de la population résidente dans le lieu et pendant la période de la mesure et donc potentiellement vulnérable. Le chapitre 4 de cette thèse utilise l'IPA modifié dans sa version mensuelle (IPM) pour étudier le lien entre paludisme et luminosité des eaux.

L'IPA peut être influencé par les spécificités territoriales locales, spécialement les différences d'accès aux services de santé selon les communes. Pour illustrer cette influence, la Figure 2-3 montre la cartographie d'un indicateur de la qualité du système de santé. Cet indicateur est la relation entre le total de patients notifiés dans la commune et le nombre d'infectés (cas-suspects/cas-positifs). La carte réalisée porte sur la Région Nord du Brésil pour l'année 2010. L'hypothèse qui a conduit à la définition de cet indicateur repose sur l'idée qu'une prévention efficace va conduire à une augmentation des cas suspects par rapport aux cas positifs (indicateur tendant vers 1), se traduisant par une qualité élevée du système de santé, alors qu'un indicateur tendant vers 0 (tous les cas suspects sont des cas positifs) indiquerait une prévention défailante et donc une mauvaise qualité du système de santé. Un indicateur tendant vers 0 peut également faire suspecter qu'il ne soit enregistré qu'une partie des cas réels de paludisme. La spatialisation de cet indicateur permet de visualiser les disparités dans l'accès aux soins entre différentes communes et conduit également à une cartographie des biais potentiels dans l'utilisation de l'IPA à l'échelle régionale.

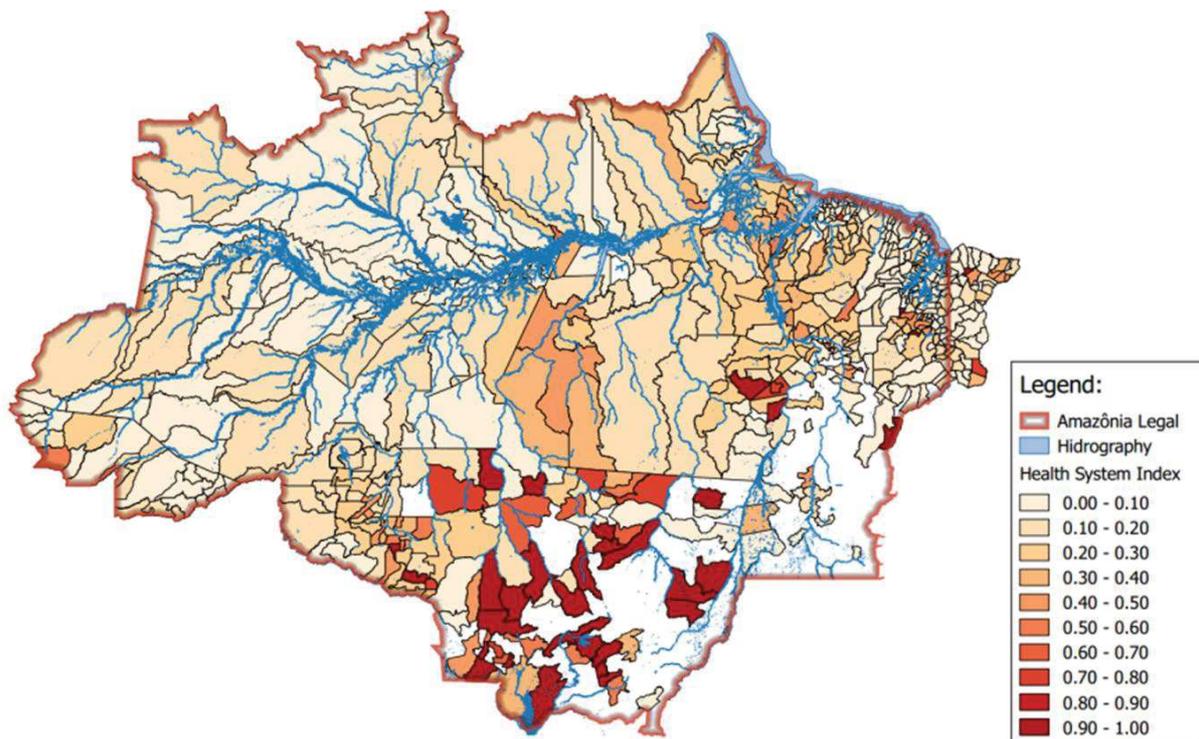


Figure 2-3 Relation cas-suspect/cas-positif de paludisme en région amazonienne pour 2010

Il faut noter, dans cette carte, que pour les communes de l'état du Mato Grosso, au Sud de l'Amazonie Brésilienne, les valeurs de l'indicateur tendent vers 1 ou sont égales à 0. L'indicateur est égal à 0 pour les communes qui se trouvent hors de la zone endémique du paludisme et les postes de santé ont pris l'habitude de n'enregistrer que les cas positifs de paludisme et non les cas suspects, le nombre de cas de paludisme étant généralement très bas.

Il peut être donc être difficile de comparer les valeurs de l'IPA du fait des différences d'accès aux soins. Ce phénomène est observable aussi entre la périphérie et le centre urbain d'une même commune. Une étude sur les impacts de la construction du barrage de Belo-Monte dans l'état du Pará sur le paludisme (Wiefels and Silveira, 2015) donne un exemple de cette complexité. Cette étude s'appuie sur les données épidémiologiques journalières du SIVEP-Malaria, pour analyser les tendances

épidémiologiques pendant la période de réalisation d'un barrage et décrit les différences de flux de patients dans les postes de santé urbains et ruraux pour la ville d'Anapu, voisine d'Altamira dans l'État du Para (Figure 2-4).

En milieu rural, il a été observé que le nombre de cas de paludisme était très proche du nombre total de patients reçus dans l'établissement de santé. En milieu urbain il y avait beaucoup plus de patients suspects que de patients infectés. Il est effet beaucoup plus facile de se rendre au poste de santé et d'avoir accès aux services de soins en milieu urbain. On peut inférer de cette situation qu'une partie des patients suspects n'a pas été notifiée par les postes de santé ruraux.

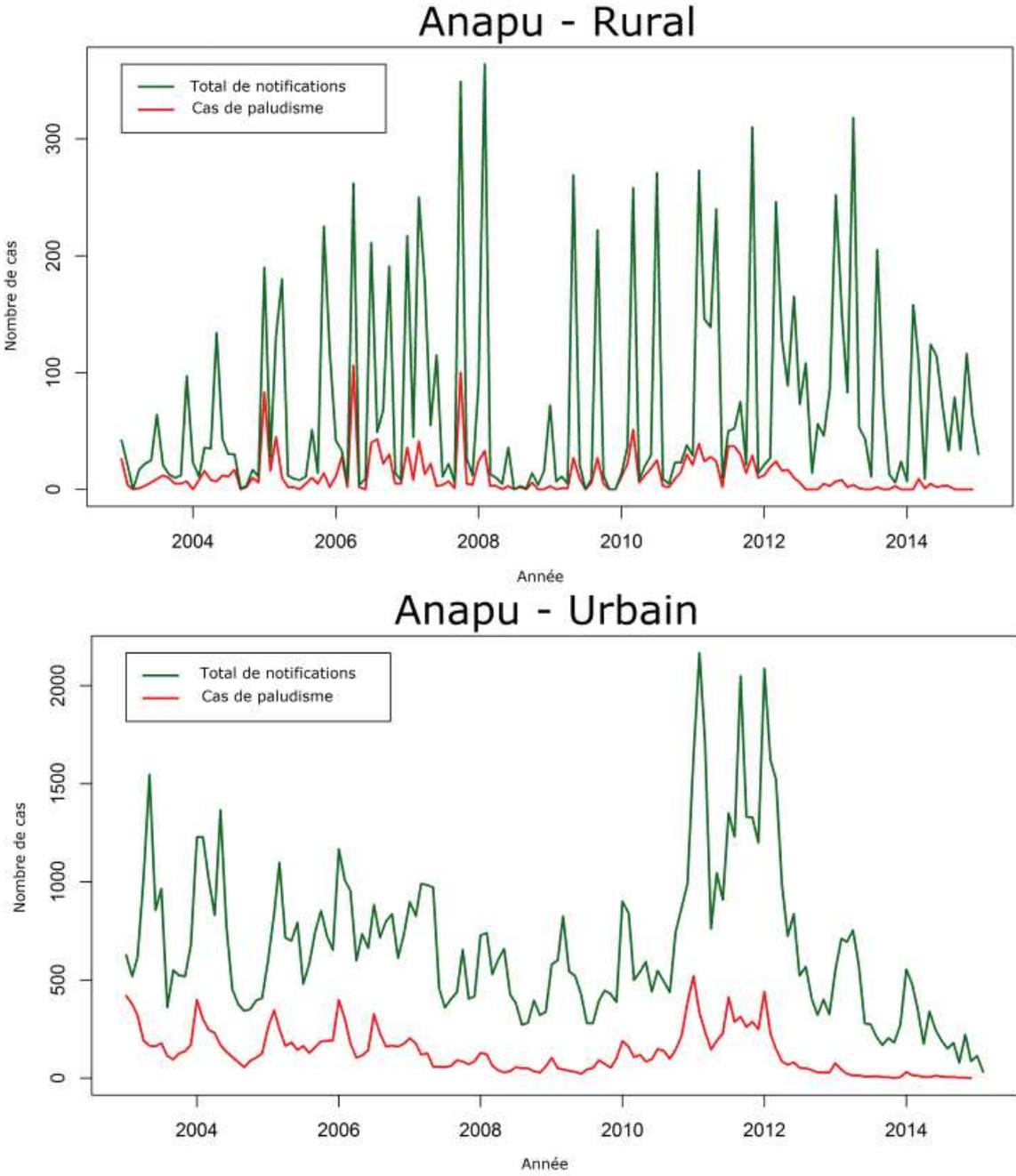


Figure 2-4 Flux de patients dans les postes de santé urbains et ruraux à Anapu-PA

2.3 La qualité des données du SIVEP-Malaria

On a pu constater un grand nombre d'erreurs lors de l'analyse de la base de données SIVEP-Malaria. Les causes de ces erreurs vont de la faute de frappe au manque de précision des déclarations des patients. De plus, en 2013, le Ministère de la Santé du Brésil a décentralisé la gestion de la production des données épidémiologiques en la confiant aux 805 communes de la région Nord, ce qui a généré des pratiques de terrain différentes.

La méthodologie développée pour l'analyse de la qualité de la base de données a été de mettre en œuvre un processus de nettoyage de données inspiré de la littérature (de Jonge & van der Loo 2013; Van den Broeck et al. 2005), de faire l'analyse des données rejetées en listant les principales sources d'erreurs, puis de définir une méthode adaptée pour l'exploitation des données.

La méthodologie de nettoyage des données est réalisée en quatre étapes: 1. Détecter les données manquantes ; 2. Détecter les données aberrantes (outliers) ; 3. Détecter les incohérences ; 4. Apporter les corrections (Van den Broeck et al. 2005; de Jonge & van der Loo 2013; Rahm & Do 2000).

Les données du SIVEP-Malaria sont disponibles en fichiers *Data Base* et maintenues par le Département Informatique du Système de Santé Brésilien (DATASUS) (Brasil, Ministério da Saúde 2015). Elles contiennent des données administratives, personnelles et de laboratoire comme le montre le Tableau 2-1. La base de données ne contient pas de noms ou d'autres informations qui puissent permettre l'identification des individus.

Tableau 2-1 Variables de la base de données SIVEP-Malaria

Variable	Définition	Variable	Définition	Variable	Définition
COD_NOTI	Numéro de notification	GESTANTE	Grossesse	UF_INFEC	État fédéral de la notification
DT_NOTIF	Date de notification	NIV_ESCO	Scolarité	MUN_INFE	Municipalité d'infection
TIPO_LAM	Type de notification	RACA	Groupe ethnique	LOC_INFE	Lieu de la notification
UF_NOTIF	État fédéral de la notif.	COD_OCUP	Occupation	DT_EXAME	Date de l'examen
MUN_NOTI	Municipalité de la notif.	PAIS_RES	Pays de résidence	EXAME	Méthode de l'examen
COD_UNIN	Poste de santé	UF_RESID	État fédéral de résid.	RES_EXAM	Résultat de l'examen
COD_AGEN	Code agent de santé	MUN_RESI	Municipalité de résidence	QTD_CRUZ	Parasitaemie
SEM_NOTI	Semaine de notification	LOC_RESI	Lieu de résidence	QTD_PARA	Parasites par mm ³ de sang
DT_DIGIT	Date de digitalisation	DT_SINTO	Date premiers symptômes	HEMOPARASI	Hemoparasites
DT_ENVLO	Date d'entrée dans la BD nationale	DT_TRATA	Date du traitement	EXAMINADOR	Code de l'examineur
DT_NASCI	Date de naissance	VIVAX	Traitement Vivax	ESQUEMA	Plan de traitement
ID_PACIE	Age	FALCIPARUM	Traitement Falciparum	SINTOMAS	Symptômes
ID_DIMEA	Format de l'age	ID_LVC	Consultation de suivi		
SEXO	Sexe	PAIS_INF	Pays de l'infection		

■ Données administratives
 ■ Données personnelles
 ■ Données de laboratoire

Le nettoyage des données a été réalisé et systématisé avec le langage de programmation statistique R (R Development Core Team 2011). Une première sélection a porté sur les notifications de cas d'infection de paludisme ou de cas notifiés à l'intérieur de l'état d'Amazonas (variables UF_NOTIF & UF_INFEC). Cette sélection a permis de prendre en compte des cas enregistrés dans les postes de santé de l'état d'Amazonas mais contractés ailleurs. Un deuxième filtre a permis de retenir seulement les notifications de cas positifs de paludisme (variable

RES_EXAM) excluant les cas suspects. Les cas positifs d'infection, en plus d'être la donnée la plus recherchée, possèdent les données du cadre clinique du patient, ce qui permet d'évaluer l'ensemble des variables pour chaque notification. Ces deux étapes ont conduit à l'extraction de 12 286 851 cas suspects pour 1 801 229 de cas positifs entre 2003 et 2014.

Cette première analyse montre que la collecte de données pour chaque patient n'est pas réalisée partout avec la même rigueur (Tableau 2-2). Certaines variables du formulaire de notification sont systématiquement remplies lorsque d'autres sont systématiquement absentes.

L'analyse des données manquantes permet de déterminer quelles sont les variables utilisables en fonction de la zone géographique ou de la période.

Tableau 2-2 Disponibilité de données par variable entre 2003 et 2004 pour l'état d'Amazonas

Variable	Disponibilité (%)	Variable	Disponibilité (%)	Variable	Disponibilité (%)
COD_NOTI	100	PAIS_INF	99,90	GESTANTE1	79,50
DT_NOTIF	100	QTD_CRUZ	99,80	NIV_ESCO_1	78,60
DT_ENVLO	100	MUN_RESI	99,80	ESQUEMA_1	75,30
TIPO_LAM	100	UF_RESID	99,30	DT_DIGIT	71,40
UF_NOTIF	100	UF_INFEC	99,30	DT_NASCI	70,90
MUN_NOTI	100	MUN_INFE	99,30	RACA	14,10
COD_UNIN	100	LOC_INFE	97,60	FALCIPARUM	14,10
SINTOMAS	100	LOC_RESI	96,60	VIVAX	14,10
RES_EXAM	100	GESTANTE	93,50	EXAME	14,10
ID_PACIE	99,90	DT_SINTO	93,30	EXAMINADOR	14,10
PAIS_RES	99,90	DT_TRATA	90,50	HEMOPARASI	14,10
ID_LVC	99,90	ESQUEMA	89,40	QTD_PARA	11,90
ID_DIMEA	99,90	COD_AGEN	88,80	SEM_NOTI	0
SEXO	99,90	NIV_ESCO	88,60		
DT_EXAME	99,90	COD_OCUP	88,00		

2.3.1 Analyse temporelle des données manquantes

La disponibilité de données n'est pas constante au fil du temps, entre 2003 et 2014 le SIVEP-Malaria a vu une nette amélioration dans la production de données. Les données collectées ont été comptées pour chaque variable par jour (date de notification) et par poste de santé, puis comparées au nombre total de notifications à la même date. Les graphiques de la Figure 2-5, montrent la proportion de données disponibles pour chaque variable au cours du temps, après classement en cinq grands groupes.

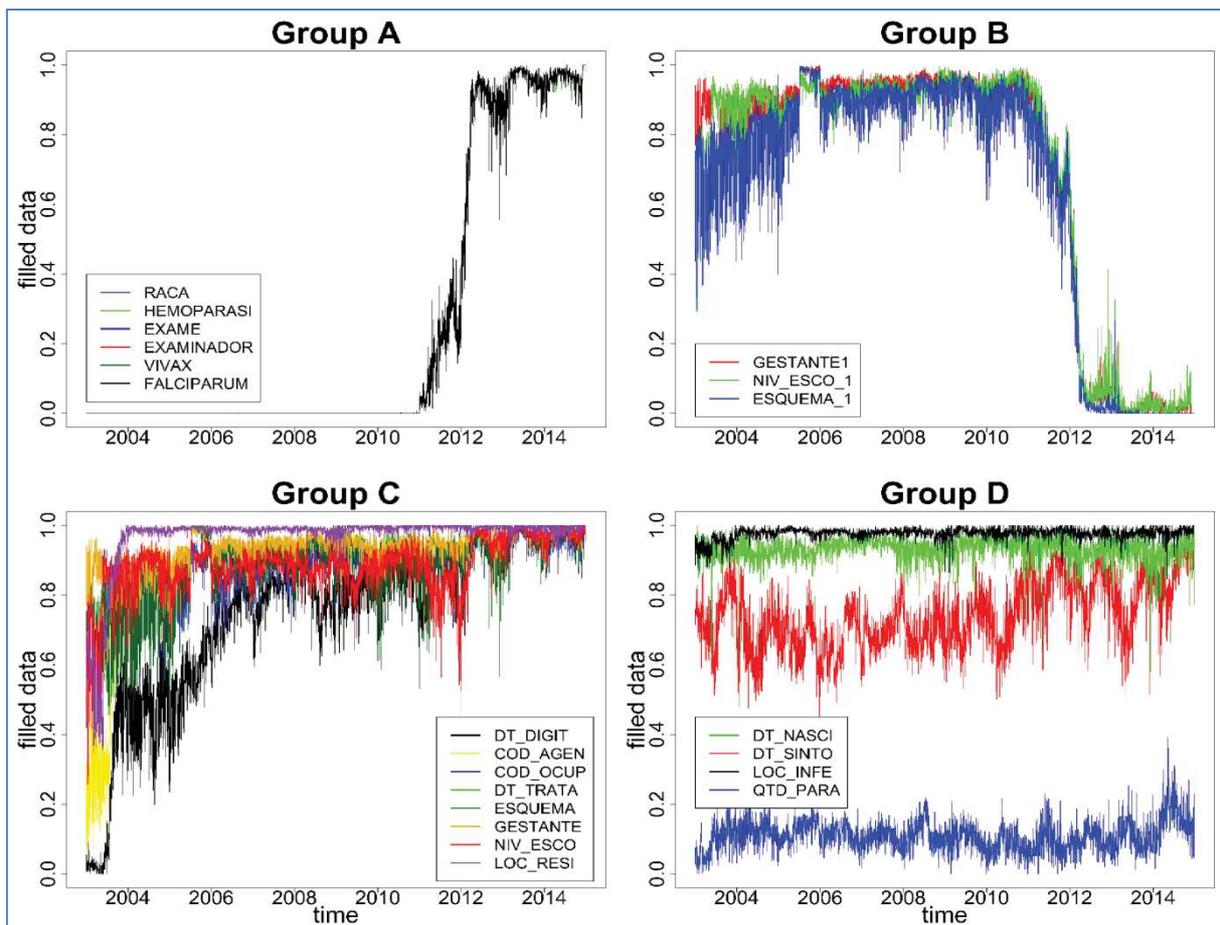


Figure 2-5 Classifications des signatures des données manquantes par variable et par jour

Les résultats sont classés en cinq *Groupes* de *A* à *E*. Le *Groupe E* est composé de six variables dont les valeurs sont pratiquement toujours présentes et pour lesquelles un graphique n'était pas nécessaire. Les variables du *Groupe E* sont la municipalité d'infection, l'État de résidence, l'État de l'infection, la parasitémie, la municipalité de résidence et le pays de l'infection.

Les variables du *Groupe A* ont commencé à être relevées en 2011 et n'ont été systématiquement enregistrés qu'en 2013. Il s'agit des variables groupe ethnique, hemoparasites, méthode d'examen, code de l'examineur, traitement Vivax et traitement Falciparum. En revanche, les trois variables du *Groupe B*, grossesse, scolarité et plan de traitement ont développé une systématisme à partir de 2003, mais ont cessé d'être collectées lorsque le formulaire d'inscription de SIVEP-Malaria a changé en 2010.

Celles-ci ont été transposées au *Groupe C* avec la nouvelle méthode de classification. Le *Groupe C* rassemble les variables avec le comportement le plus attendu. Les variables date de digitalisation, code de l'agent de santé, occupation, date du traitement, plan de traitement, grossesse, scolarité et lieu de résidence gagnent en systématisme de collecte au fil du temps. Le *Groupe D* contient des variables qui ont présenté une tendance saisonnière, avec une valeur moyenne fixe au fil du temps. Ce sont la date de naissance, la date des premiers symptômes, le lieu de l'infection et les parasites par mm^3 de sang.

Cette approche a permis de constater qu'il existe une saisonnalité annuelle dans la production des variables du SIVEP-Malaria, très visible dans le *Groupe D*. Cette saisonnalité est en rapport avec la saisonnalité des cycles épidémiologiques. Par

exemple, la *date de naissance* et les *cas de paludisme positifs* ont présenté une corrélation de Spearman de -0.65. Leurs courbes sont affichées dans la Figure 2-6.

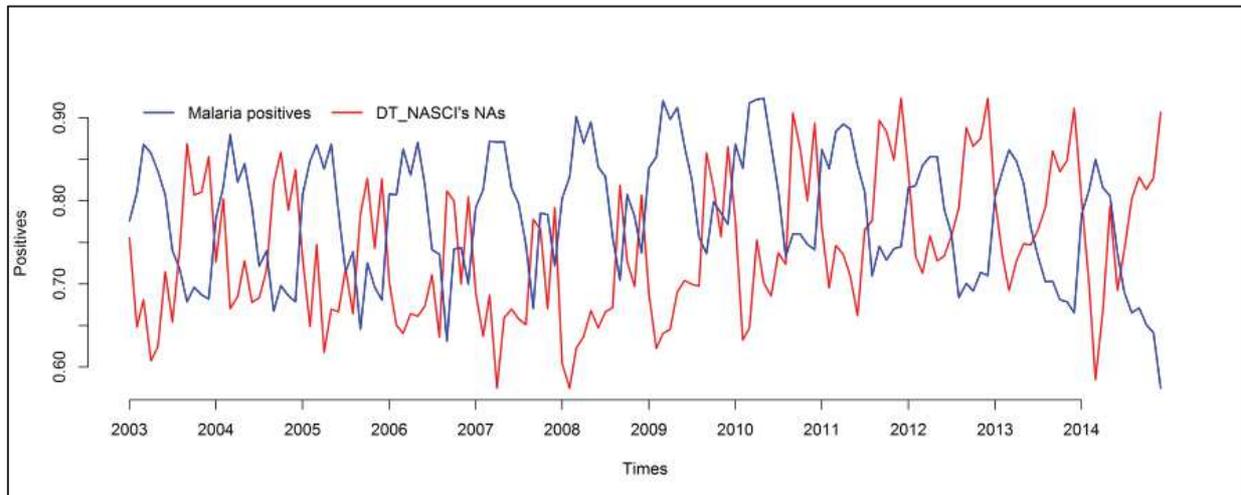


Figure 2-6 Cycles épidémiologiques et données manquantes pour la variable date de naissance

Ceci permet d'inférer que lors des pics d'épidémie, les agents de santé font une collecte de données plus fiable, ce qui diminue la quantité de *dates de naissance* manquantes sur les notifications.

Cette périodicité a été observée pour toutes les autres variables dans des proportions différentes, mais elle est la plus visible pour la variable *date de naissance*. Cette observation suggère que le cycle épidémiologique affecte la notification des autres variables.

2.3.2 Détection des données aberrantes

La deuxième étape du nettoyage de données a été l'identification des valeurs aberrantes (outliers) ou irréalistes. La plupart des données SIVEP-Malaria sont

qualitatives. Pour cette étude, chaque variable a été analysée séparément, en comparant les données fournies à l'intervalle de valeurs attendues.

Sept variables de la base de données sont en format date, parmi lesquelles ont été trouvées les valeurs aberrantes suivantes:

1. la *date de naissance* devrait être dans l'intervalle compris entre l'année "1900" et "2015" mais a présenté 19 occurrences entre l'année "1006 " et "1900" et 22 occurrences entre l'année "2030" et "2091";

2. la *date des symptômes* devrait se situer entre l'année "2000" et "2015" mais a présenté 148 occurrences entre l'année "1002" et "1999" avec un pic autour de "1200";

3. la *date de traitement* qui devrait se situer entre la date "2003-01-01" et "2014-12-31" a enregistré 796 occurrences entre les années "2015" et "9620".

Les *résultats des examens* (avec des valeurs de classes allant de 1 à 11) et la *parasitémie* (avec des valeurs de classes de 1 à 6) ont affiché 17 fois la valeur "0", (ce qui n'était pas une option) pour l'année 2003 et pour les municipalités de Coari et Manaus.

Le code d'enregistrement de la notification, avec des valeurs positives, a révélé 18 occurrences aberrantes (16 x "0", "-1264" et "-682") de 2003 à 2007.

La variable *semaine épidémiologique* est toujours absente, sauf pour un cas en 2003.

La variable *municipalité de résidence* ne dispose que d'une valeur aberrante "539.924", qui n'est pas un code de municipalité valide.

Il y a 20 242 agents de santé enregistrés dans la base de données SIVEP-Malaria pour l'État d'Amazonas. Chacun d'eux est enregistré sous un matricule. Parmi ceux-ci, seuls 15 692 agents sont responsables des notifications entre 2003 et 2014. Quelques 302 236 agents apparaissant comme responsables de notifications ne sont pas enregistrés dans le système.

2.3.3 Détection des données incohérentes

Certaines valeurs sont logiquement ou biologiquement impossibles (i.e. *sexe=masculin* ET *gestation=oui*). Les variables suivantes, les plus fréquemment utilisées ont été analysés en paires afin de vérifier la cohérence des valeurs.

1. La consultation de suivi de l'état du patient (LVC) a été comparée au type de *notification (active/passive)*. La *consultation de suivi* du patient survient des semaines après la confirmation de l'infection par le paludisme (World Health Organization 2012). Il s'agit d'une deuxième consultation pour certifier la guérison et elle est enregistrée comme une nouvelle notification (ce sera un point de vigilance par la suite de distinguer de première consultation afin d'éviter les doublons). Le Tableau 2-3 montre les deux variables; le *type de détection*, appelé TIPO_LAM (1-passif, 2-actif et 3-consultation de suivi) et la *consultation de suivi*, appelée ID_LVC (1-oui, 2-no) qui informent sur la *consultation de suivi*. Il y avait 201 notifications contradictoires, reconnues comme étant des notifications de *consultation suivi* (TIPO_LAM = 3), mais aussi son contraire (ID_LVC = 2).

Tableau 2-3 Consultation de suivi

		ID_LVC	
		1	2
TIPO_LAM	1	16790	1065815
	2	9119	464484
	3	244819	201

TIPO_LAM (1 = 1-passif, 2-actif et 3-consultation de suivi)

ID_LVC (1= oui, c'est une consultation de suivi et 2 = non, ce n'est pas une consultation de suivi)

Ces incohérences sont regroupées dans l'espace et dans le temps autour principalement de Manaus et Coari comme le montre le Tableau 2-4.

Tableau 2-4 Concentration des incohérences pour la variable *consultation de suivi*

Commune	Année de Notification			
	2011	2012	2013	2014
Atalaia do Norte	12	0	0	0
Coari	1	0	0	0
Itacoatiara	1	0	0	0
Lábra	0	1	0	0
Manaus	134	1	43	2
São Gabriel da Cachoeira	5	0	0	0
Tefé	1	0	0	0

2. La variable *format d'écriture de l'âge du patient* est complémentaire à la variable *âge du patient*. Le format d'écriture de l'âge indique si l'âge est enregistré en jours, mois ou années. L'analyse de ces variables (Figure 1-7) montre tout d'abord que les jeunes sont plus présents dans les notifications, mais aussi qu'il y avait 299 valeurs pour des patients âgés entre 101 et 1004 ans.

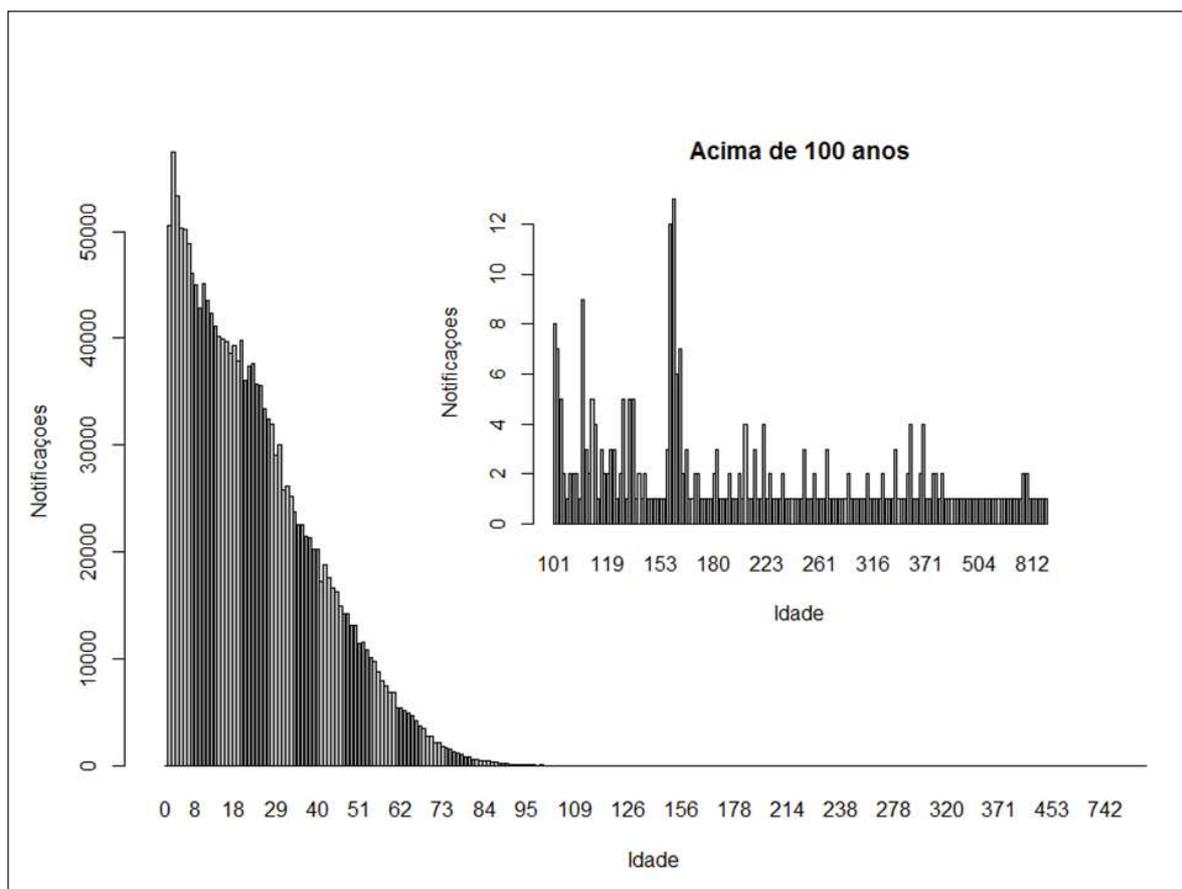


Figure 2-7 Age des patients

Les patients âgés de plus de 101 ans se concentrent surtout dans les communes de Manaus et Barcellos comme le montre le Tableau 2-5.

Tableau 2-5 Concentration des incohérences liées à l'âge des patients.

	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total
Manaus	23	11	19	18	4	10	5	3	20	8	3	5	129
Barcelos	1	0	3	17	7	0	0	0	2	0	0	0	30
Atalaia do Norte	0	0	2	1	3	2	5	0	2	9	0	0	24
Tefé	1	3	5	1	4	2	3	0	0	0	0	0	19
Careiro	0	4	4	5	3	0	0	1	0	0	0	0	17
Porto Velho (RO)	5	0	1	1	2	4	2	0	0	1	0	0	16
Presidente Figueiredo	1	0	2	5	2	2	1	0	1	0	1	0	15

3. L'âge des patients doit être cohérent avec les données de grossesse. Il y avait sept patients de moins de un an considérés enceinte entre le 1er et le 3ème trimestre de la grossesse, 118 notifications de patients âgés de 10 ans avec l'âge gestationnel ignoré, et 38 enceintes, mais avec plus de 55 ans. Les principales communes impliquées dans ces enregistrements sont Coari et Manaus (Tableau 2-6).

Tableau 2-6 Concentration des incohérences de la relation âge-grossesse.

	2003	2004	2005	2006	2007	2008	2009	2010	Total
Coari	9	0	1	0	0	0	0	0	10
Manaus	34	6	10	5	2	1	1	1	60

4. La date du symptôme, la *date de la notification* et la *date de l'examen* sont des variables corrélées car elles doivent être ou successives ou datées du même jour. Entre l'infection du paludisme et la guérison, l'ordre des événements n'est pas toujours respecté : le symptôme, la notification, puis l'examen. Lorsque l'on compare la date de ces événements, certains comportements sont inattendus/incohérents.

Pour les notifications qui ne sont pas des consultations de suivi (1 530 499 notifications), 348 292 patients ont une date de notification identique à la date d'apparition des symptômes, et 1 151 271 patients ayant retardé la notification jusqu'à une semaine après les symptômes ; 13 597 personnes ont notifié avant d'avoir des symptômes, et 6 290 personnes ont notifié jusqu'à 60 jours après les symptômes.

On peut noter également que 1 395 019 (91,15%) des *examens cliniques* ont été réalisés le même jour que la *notification* et 9 276 (0,6%) l'ont été jusqu'à une semaine après. Cependant, 60 *examens cliniques* inattendus ont été effectués avant la *notification en poste de santé*.

Tous ces cas ont été concentrés principalement dans les communes de Manaus et Coari autour de l'année 2003 comme le montre le Tableau 2-7.

Tableau 2-7 Concentration des incohérences dans la séquence des dates

	2003	2009	2010	2011	2012	2013	Total
Manaus	11501	0	4	0	3	2	11.510
Coari	2084	1	0	1	1	0	2.087

2.3.4 Analyse spatiale de la base de données

Le module de localisation géographique de la base de données SIVEP est dans un format de données propre aux systèmes d'information géographiques (un tableur appelé LOCALIDA.DBF). La création et la mise à jour de la liste des lieux est de la responsabilité des communes. Ce fichier contient les coordonnées géographiques, le nom de la ville, le nombre d'habitants, le type d'emplacement, le code de localisation, mais aussi la date de la dernière mise à jour de ces données.

Une première observation en ce qui concerne l'utilisation des coordonnées géographiques de cette base de données est la présence d'une erreur pour les valeurs de latitude. Celle-ci doit, pour être utilisée, être divisée par 10 dû à une erreur présente dans toutes les valeurs de cette variable.

Il existe un total de 525 795 lieux (points SIVEP) dans la base de données mais seulement 5 580 sont géoréférencés, soit seulement 1,06%. L'état d'Amazonas compte 12 552 lieux enregistrés dont 3 593 sont géoréférencés, soit 28,6% du total.

L'État d'Amazonas est divisé en 62 communes. Toutes les communes ne disposent pas de la même organisation pour alimenter le SIVEP en données. Il est possible, par exemple, de savoir quelles communes ont le plus de points SIVEP enregistrés. Pour pondérer cette information afin de pouvoir comparer les communes, on a utilisé

la relation le nombre de points géo-référencés/ nombre d'habitants de la commune (source IBGE), divisé par 1000. Le Tableau 2.8 récapitule les cinq municipalités de l'État d'Amazonas qui possèdent le plus de points SIVEP.

Tableau 2-8 Communes les mieux couvertes en points/habitant

Commune	Nombre de Points	Population	Indice
São Gabriel da Cachoeira	90	42342	23,38104
Japurá	23	5599	21,96821
Tapauá	63	18266	19,87299
Barcelos	10	27273	15,03318
Canutama	14	14944	14,32013

Une minorité de communes enregistrent tous leur points-SIVEP avec des coordonnées géographiques. Néanmoins, cela ne veut pas forcément dire que ces communes possèdent un nombre raisonnable de points pour servir à la surveillance épidémiologique. Les cinq communes avec le plus de points géoréférencés par habitant sont listées dans le Tableau 2-9.

Tableau 2-9 Communes avec le plus de points géoréférencés

Commune	Points Georef.	Population	Indice
Japurá	92	5599	16,43151
Jutaí	136	16977	8,010838
Lábrea	277	42439	6,527015
Guajará	82	15561	5,269584
Carauari	134	27645	4,847169

Sept communes n'ont aucun point géoréférencés. Ce sont les communes de Itapiranga, Urucurituba, Nhamundá, Nova Olinda do Norte, Barreirinha, Boa Vista do Ramos et Silves.

2.4 Résultat du nettoyage des données

Tenant compte des types d'erreurs et de leur source, un nettoyage des données a été réalisé couvrant l'état d'Amazonas, de 2003 à 2014. L'objectif était 1. de connaître les proportions de données erronées par étape de nettoyage et 2. de disposer d'une base de données fiable. Le produit final a été un ensemble de notifications avec les variables du formulaire entièrement remplies et avec les points des unités de santé, des résidences et des lieux d'infection enregistrés et géoréférencés. Le schéma de la Figure 2-8 montre les étapes et les critères utilisés dans le processus de nettoyage des données.

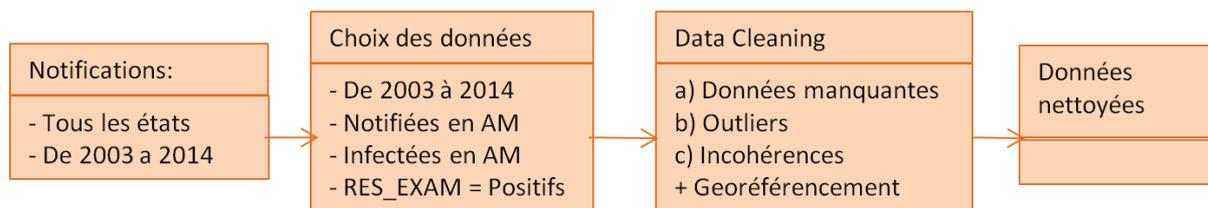


Figure 2-8 Schéma de l'organisation du nettoyage des données SIVEP-Malaria

Le nettoyage des données a commencé en excluant les 10 variables ayant la plus grande quantité de valeurs manquantes et excluant aussi les trois variables qui ont été remplacées (*Groupe B*). Les données d'entrée comptent 1 801 229 notifications positives de paludisme. Dans un premier temps, 49,8% des notifications ont été exclues pour avoir des valeurs manquantes. Moins de 0,01% ont été exclues comme contenant des données aberrantes. Ensuite, 0,02% ont été retirées en raison de leur incohérence. Enfin, après analyse de la localisation géographique, 2,03% des notifications n'ont pas de poste de santé géoréférencé, 37,59% des notifications n'ont pas de géolocalisation pour la résidence des patients et 44,82% des notifications n'ont pas de géolocalisation pour le lieu d'infection.

Le produit du nettoyage des données est résumé Figure 2-9 sous la forme d'une carte de concentration des erreurs par commune. Les données manquantes, les valeurs aberrantes et les incohérences ont été comptées pour chaque notification et le résultat a été apparié avec le géocodage de municipalités de notification respectives pour quantifier la présence de données erronées. Le total des données non valides a été pondéré par le total des notifications pour toutes les années et multiplié par 100 afin de produire cet indicateur. Les valeurs de l'indicateur varient de 4,47 (peu d'erreurs) pour la commune de São Gabriel da Cachoeira à 153,93 (beaucoup d'erreurs) pour Novo Airão.

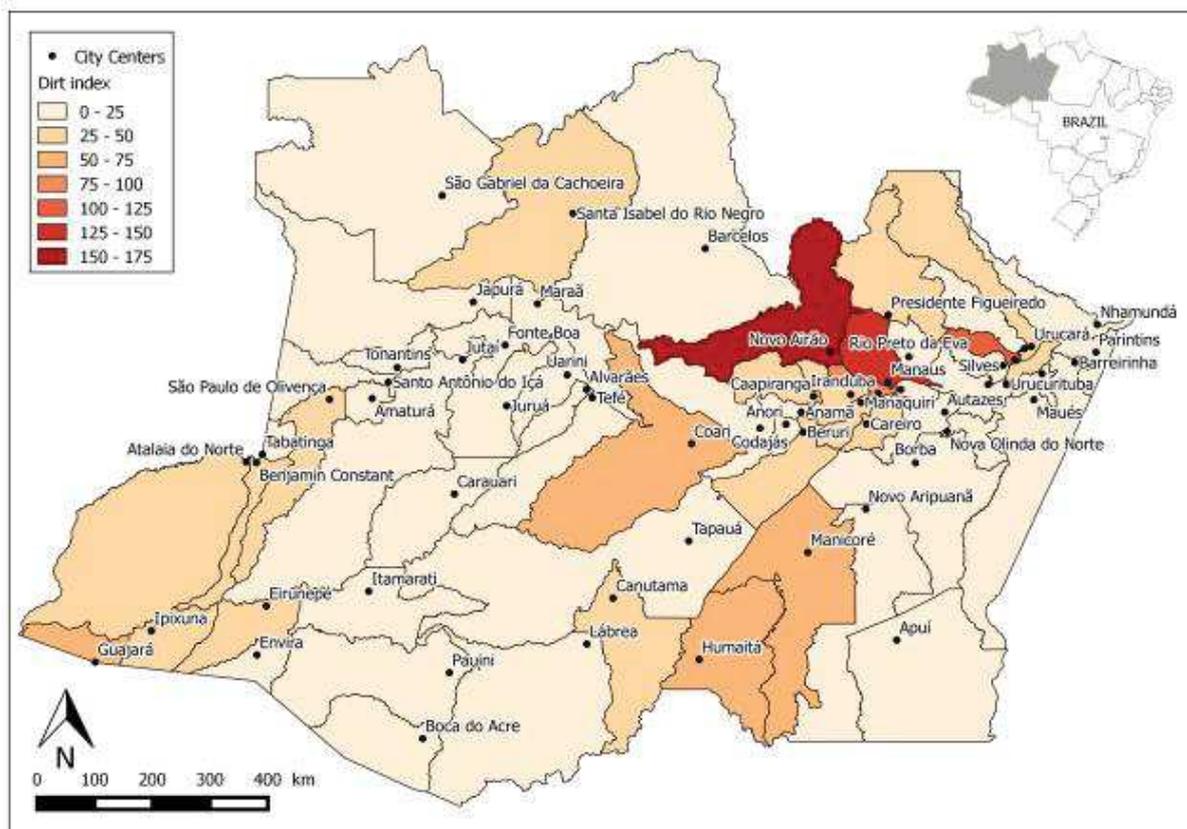


Figure 2-9 Cartographie des concentrations d'erreurs du SIVEP Malaria de 2003 à 2014 suite au nettoyage de données

Suite au nettoyage des données et à la sélection des notifications géo-localisées, 49,17% des notifications s'est révélée exploitable pour une étude basée sur les postes de santé, 31,33% pour une étude basée sur la résidence du patient et 27,70% pour une étude basée sur le lieu d'infection.

2.5 Utilisation Scientifique des données SIVEP-Malaria

Cette analyse critique est inédite pour la base données SIVEP-Malaria et ces observations pourraient remettre en question certaines études ayant utilisé ces données sans les critiquer. Depuis sa création en 2003, cette base de données a été la base de nombreuses recherches épidémiologiques pour la région amazonienne avec un nombre d'études qui augmente constamment au fil du temps. Pour avoir une idée de l'impact de ces connaissances sur la production scientifique, une analyse de la bibliographie scientifique qui utilise les données SIVEP-Malaria a été faite prenant pour modèle la méthode de Herbreteau et al. (2007) pour analyser les articles qui traitent d'un sujet spécifique. A partir du moteur de recherche "Google Scholar" et en utilisant les termes "SIVEP" et "malaria" comme mots-clés, ont été recensés les articles scientifiques, thèses, rapports, livres et autres.

La première analyse montre l'utilisation croissante de ces données dans l'environnement scientifique qui peut être observée Figure 2-10 entre le lancement du programme en 2003 et aujourd'hui. Les articles scientifiques sont nettement en tête des types de productions suivis par les thèses de doctorat.

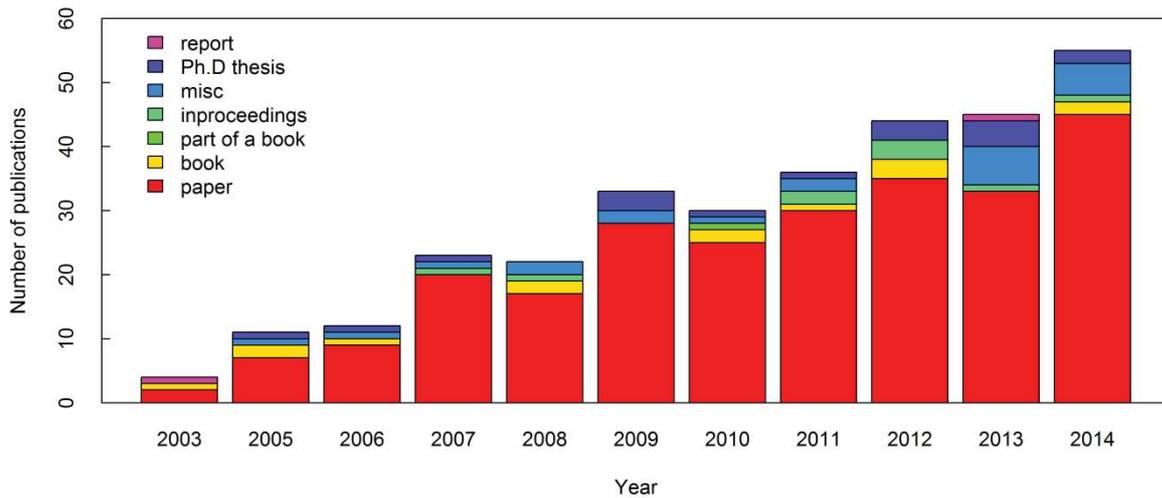


Figure 2-10 Production scientifique à partir de données SIVEP Malaria

Il a été possible de recenser les principales variables SIVEP-Malaria utilisées. Cela a permis la production du Tableau 2-10 qui compare l'usage des différentes variables. Les variables les plus utilisées sont celles qui permettent la production de l'IPA : ce sont essentiellement quatre variables SIVEP: Résultat des examens (RES_EXAM), Municipalité de résidence (MUN_RESI), Date de notification (DT_NOTI), et Consultation de suivi (ID_LVC) en portugais "Lâmina de Verificação de Cura (LVC)".

Tableau 2-10 Variables SIVEP-Malaria les plus utilisées scientifiquement entre 2003 et 2014

Variable	Identifiant Sivep	Utilisation
Résultat de l'examen	RES_EXAM	92
Date de Notification	DT_NOTIF	89
Test de guérison	ID_LVC	59
Lieu de résidence	LOC_RESI	52
Municipalité de résidence	MUN_RESI	26

Sexe	SEXO	23
Scolarité	NIV_ESCO	23
Age	IDADE	20
Lieu d'infection	LOC_INFE	11
Date de l'examen	DT_EXAM	11
Grossesse	GESTANTE	8
Tipe de détection	TIPO_LAM	5
Schema de traitement	ESQUEMA	3
Municipalité d'infection	MUN_INFE	1

Peu de travaux utilisent le SIVEP-Malaria pour une analyse à l'échelle régionale. La plupart utilisent l'échelle municipale. Il est important aussi de noter qu'une grande partie des travaux ne mentionne pas l'utilisation de la variable ID_LVC pour distinguer si le patient est enregistré comme première consultation ou bien comme consultation de suivi du traitement. Cette variable est pourtant clé pour filtrer la base de données afin de ne pas compter deux fois le même patient.

2.6 Conclusion partielle

La base de données du SIVEP-Malaria offre aux autorités une somme importante d'informations. Les 43 variables dont elle dispose sont bien plus nombreuses que les 13 variables essentielles proposées par le Global Health Group (UCSF 2014). C'est un atout mais aussi un défi que de les produire systématiquement avec un minimum d'erreurs et sans surcharger les services de santé locaux. Ainsi, selon l'analyse temporelle de la production des variables de cette base de données, on peut conclure que les notifications sont de plus en plus fiables car le nombre de champs

remplis du formulaire augmentent au fil du temps, parallèlement à la baisse des erreurs. On peut aussi inférer de l'analyse de la base que celle-ci est mieux renseignée en période d'épidémie.

L'échelle régionale reste un défi pour la plus grande partie de la région dû au manque d'informations géoréférencés. L'analyse de cohérence de la base de données permet d'utiliser 49,17% des notifications filtrées d'après la localisation du poste de santé, notifications réparties de façon uniforme sur tout l'état

Une première analyse des disparités d'accès aux soins à partir d'un indice qui varie de 0 à 1 (nombre de cas suspects/nombre de cas positifs) montre que cet indice est faible (variant de 0 à 0,2) indiquant un faible accès au soin, mais assez homogène pour tout l'état.

Le nombre de cas positifs rapportés à la densité de population (IPA) est un indicateur de la prévalence de la maladie, et donc du risque d'exposition. Dans la suite de ce travail cet indicateur du risque d'exposition sera analysé en relation avec la luminosité de l'eau comme indicateur de sa qualité.

Le travail préliminaire sur la base de données exposé dans ce premier chapitre, constituant un mode d'emploi pour des usages scientifiques, très nombreux mais n'utilisant pas forcément ces critères de qualité, a été publié (Wiefels et al. 2016).

CHAPITRE 3 Dynamique de la luminosité des eaux dans l'état d'Amazonas

3.1 Introduction

L'État d'Amazonas est le carrefour des eaux amazoniennes. L'étude de la dynamique de la couleur des eaux de cette région permet d'observer les variations que subissent les principaux fleuves d'eau blanche et d'eau noire (Sioli & Klinge 1962) mais aussi les varzeas (zones d'inondation sur les fleuves d'eau blanche) et igapos (zones d'inondation sur les fleuves d'eau noire). Ces fleuves ont des caractéristiques limnologiques très différentes qui se traduisent par la capacité des eaux à réfléchir l'énergie incidente du soleil. La télédétection du contraste des différentes intensités de luminosité permet de séparer ces deux types d'eau. La création d'un indice de luminosité à partir des bandes du visible est proposée ici comme méthode capable de cartographier ces différences. L'objectif de ce chapitre a été d'établir une cartographie mensuelle de la variation de l'indice de luminosité des eaux pour l'état d'Amazonas en suivant les variations des étendues inondées.

La réflectance est le résultat des effets combinés et simultanés de l'absorption et de la diffusion de l'énergie électromagnétique par tous les composants présents dans la colonne d'eau sur le champ de l'irradiation incidente. De manière générale on peut dire qu'une plus grande absorption spectrale de l'eau cause une diminution de la

réflectance alors qu'une plus grande diffusion génère une augmentation de la réflectance.

Les données de base pour ce travail sont les images issues de l'instrument MODIS, à partir des produits MOD09A1 et MOD09Q1 accumulés sur onze ans, traitées pour obtenir des images composées mensuelles pour l'état d'Amazonas. La répétitivité quotidienne de l'acquisition d'images MODIS a permis de s'affranchir de la plupart des nuages et ombres de nuages des scènes en réalisant ces images composées des meilleurs pixels du mois. De ces images composées mensuelles, à partir de données initiales à 250 mètres et 500 mètres de résolution, ont été extraits le masque de l'eau puis les classifications de la luminosité des eaux, enfin les moyennes mensuelles et la variance inter mensuelle/annuelle de la luminosité.

La luminosité est utilisée ici comme un indicateur de la qualité « blanche » ou « noire » des eaux des fleuves. Nous faisons l'hypothèse que le contraste élevé entre ces qualités d'eau permet leur séparation sous la forme d'une classification basé sur un indice de luminosité.

Les produits et connaissances acquis dans ce chapitre seront utilisés dans le 4ème chapitre pour l'étude de la relation entre la luminosité des eaux et les indices d'incidence du paludisme dans l'État d'Amazonas.

3.2 Matériel et méthode

La base de données SIVEP-Malaria est disponible sur une période de 11 ans (2003 – 2014). Les données du satellite MODIS ont été choisies pour leurs caractéristiques décrites ci-après et parce qu'elles sont également disponibles sur la même période.

Les méthodes de télédétection qui vont être développées dans ce chapitre se focalisent, d'une part sur la détection des plans d'eau, et sur leur luminosité qui est un indicateur de leur qualité.

L'un des objectifs de la chaîne de traitements réalisés est la robustesse, en particulier la reproductibilité des classifications pour tous les mois indépendamment de la saison ou de la qualité de la donnée.

3.2.1 Les images MODIS

Le capteur Moderate Resolution Imaging Spectroradiometer (MODIS) (Justice et al. 1998) répond aux critères choisis :

- Résolution temporelle suffisante pour obtenir des synthèses sans nuages (ou avec peu de nuages). Pour rappel l'atmosphère de la région amazonienne est très souvent chargée de nuages.
- Résolution spatiale qui est un compromis entre la nature du travail à l'échelle régionale (l'État d'Amazonas a près de 2000 Km en longitude et 1000 km en latitude) et la taille des objets d'étude.
- Caractéristiques radiométriques adaptées à la détection de la luminosité de l'eau

L'instrument MODIS fonctionne sur deux orbites journalières (satellite Terra à 10:30 am et satellite Aqua à 1:30 pm). Il fournit une estimation de la réflectance spectrale de la surface au niveau du sol, corrigées pour les gaz et les aérosols atmosphériques (réflectance au niveau du sol) et corrigées géométriquement (Vermote & Vermeulen 1999).

MODIS réalise un balayage de la surface terrestre au long d'une orbite descendante à 705 km du sol avec une largeur de fauchée de 2330 Km. Les images MODIS sont

disponibles pour l'ensemble du globe dans une projection sinusoïdale de 10X10 degrés qui découpe le globe en 648 tuiles. Cette grille est référencée selon l'emplacement des tuiles en horizontal (H) et vertical (V). L'état d'Amazonas au Brésil est à cheval sur cinq tuiles de cette grille, soit les tuiles: H10V09, H11V08, H11V09, H12V08 et H12V09 (Figure 3-1).

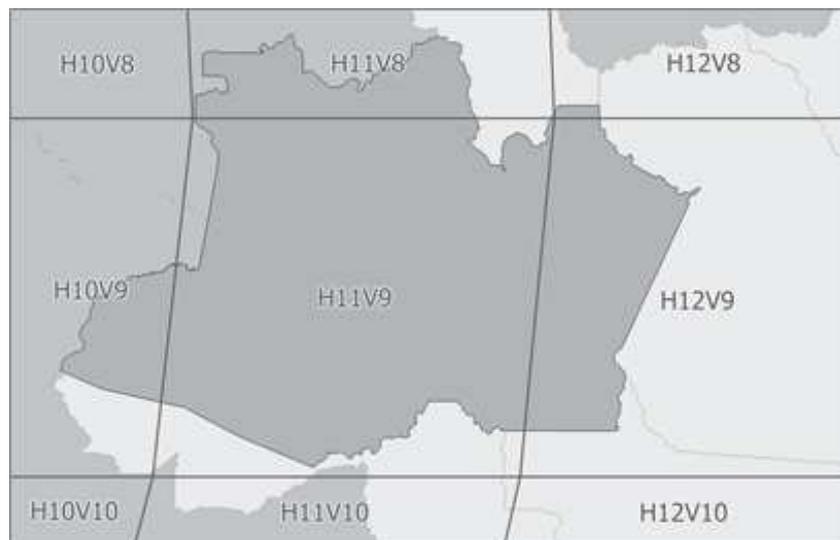


Figure 3-1 Tuiles MODIS sur l'état de l'Amazonie au Brésil

Sa résolution spatiale varie selon les canaux, de 250 mètres pour les canaux 1 et 2 (bleu et vert), de 500 mètres pour les canaux de 3 à 7 et de 1000 mètres pour les canaux de 8 à 36 (Tableau 3.1).

MODIS fonctionne sur 36 canaux de résolution spectrale allant de 405 nm (bleu) à 14.385 µm (infrarouge) dont trois bandes spectrales dans le domaine du visible : la bande 1 dans le rouge (de 620nm à 670nm), la bande 3 dans le bleu (de 459nm à 479nm) et la bande 4 dans le vert (de 545nm à 565nm) dans le vert. Les autres canaux qui ont été utilisés dans cette étude pour réaliser le masque « eau » sont les

canaux 4, 6 et 7 dans l'infrarouge (Tableau 3.1), puisque le rayonnement solaire est entièrement absorbé au-delà de la longueur d'onde de 800nm (infrarouge).

Tableau 3-1 Bandes de réflectance MODIS utilisées

Bande	Produit Modis	Résolution temporelle (jours)	Intervalle (nm)	Résolution spatiale (m)	Nom
1	MOD09Q1	8	620 - 670	250	Rouge
2	MOD09Q1	8	841 - 876	250	Proche infra-rouge
3	MOD09A1	8	459 - 479	500	Bleu
4	MOD09A1	8	545 - 565	500	Vert
5	MOD09A1	8	1230 - 1250	500	Proche infra-rouge
6	MOD09A1	8	1628 - 1652	500	Moyen infra-rouge
7	MOD09A1	8	2105 - 2155	500	SWIR

MODIS propose une quarantaine de produits différents à partir des deux satellites AQUA (MYD) et TERRA (MOD). Les produits proposant l'ensemble des canaux choisis (tableau 3.1) sont labellisés A1, mais ils sont fournis à une résolution spatiale de 500m. Les produits labellisés Q1 ne comportent que deux canaux (1 et 2) mais sont à une résolution de 250m. Les 4 produits MODIS utilisés dans ce travail sont MOD09A1 et MYD09A1 (500 mètres de résolution spatiale, 7 canaux) et MOD09Q1 et MYD09Q1 (250 mètres de résolution spatiale, 2 canaux). Ces quatre produits sont

des compositions temporelles à 8 jours. Chaque pixel de ces produits est un composé du produit journalier MOD09GHK et correspond à la meilleure observation possible sur une période de 8 jours (angle d'observation le plus faible, absence de nuage ou d'ombre de nuage et un minimum de charge d'aérosol).

Les produits à 250m et ceux à 500m sont concordants pour chacune des missions (même date et heure d'acquisition) ce qui permet de combiner les produits à 250m et 500m. Les canaux utilisés dans ce travail sont donc les canaux 1 et 2 des produits Q1 (250m) et les canaux 3, 4, 5 et 7 des produits A1 (500m), qui ont été ré-échantillonnées à 250m de façon à disposer de l'ensemble des canaux à la résolution de 250m.

Le format des produits est celui connu comme *HDF-EOS (Hierarchical Data Format-Missions EOS)*. Chaque pixel a un entête de 16 bits qui renseigne sur sa qualité (Surface Reflectance State - le degré de couverture nuageuse et d'ombre du pixel). Seul l'entête des produits à 500 m a été utilisé pour sélectionner les pixels les meilleurs (combinaison temporelle), assurant le tri le plus sélectif.

3.2.1.1 Méthode de création d'images composites mensuelles

Afin de réduire le nombre d'images totales à traiter, il a été choisi de travailler sur des compositions mensuelles. La production des images composées mensuelles est réalisée à partir des produits MODIS (images composées à 8 jours). La valeur de l'entête indique si le pixel a été marqué en tant que terre, océan profond, océan peu profond, nuage contenant beaucoup ou peu d'aérosol, neige ou feu (Tableau 3-2) (Vermote et al. 2011; DAAC 2013) La lecture du vecteur entête de 16 bits se fait de

droite à gauche La méthode pour obtenir les images mensuelles est composée de deux étapes.

1. identification et exclusion des pixels de nuages ou d'ombre identifiés par l'entête du produit
2. identification des meilleurs pixels entre les huit dates/horaires du mois pour chaque produit suivant les critères de choix des produits MODIS décrit plus bas.

La technique de composition est fondée sur le critère du bleu minimum qui sélectionne les conditions d'atmosphère la plus claire sur la période (Liu & Liu 2013; Luo et al. 2008; Ticehurst et al. 2014).

La Figure 3-2 donne pour exemple les 8 images composites à 8 jours du mois de janvier 2006 pour la tuile h11v09 (4 images MOD09A1 (matin) et 4 images MYD09A1 (après-midi)). Les pixels nuages/ombre/aérosols sont blancs et les pixels non nuages en noir. Les meilleurs pixels de chaque composition à 8 jours (noirs) sont triés pour composer le composite mensuel.

Tableau 32 Interprétation de la qualité du pixel MODIS (Surface Reflectance State)- pour MOD09A1

N° du Bit	Nom du Paramètre	Comb. Bit	Surface Reflectance State (500m)
0 - 1	cloud state	00	clear
		01	cloudy
		10	mixed
		11	not set, assumed clear
2	cloud shadow	1	yes
		0	not set, assumed clear
3 - 5	land/water flag	000	shallow ocean
		001	land
		010	ocean coastlines and lake shorelines
		011	shallow inland water
		100	ephemeral water
		101	deep inland water
		110	continental/moderate ocean
		111	deep ocean
6 - 7	aerosol quantity	00	climatology
		01	low
		10	average
		11	high
8 - 9	cirrus detected	00	none
		01	small
		10	average
		11	high
10	internal cloud algorithm flag	1	cloud
		0	no cloud
11	internal fire algorithm flag	1	fire
		0	no fire
12	MOD35 snow/ice flag	1	yes
		0	no
13	Pixel is adjacent to cloud	1	yes
		0	no
14	BRDF correction performed	1	yes
		0	no
15	Internal snow mask	1	snow
		0	no snow

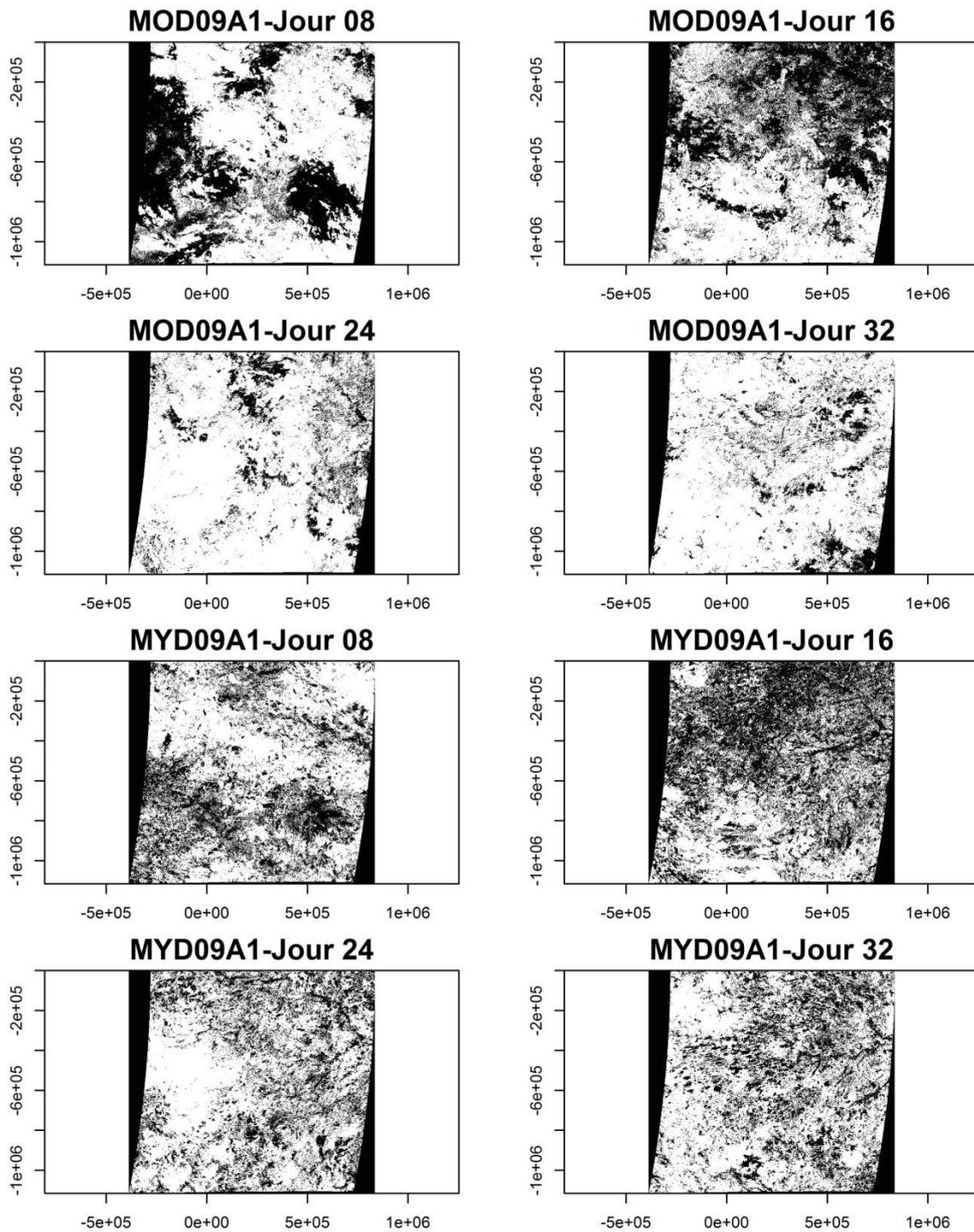


Figure 3-2 Nuages pour MOD09A1 et MYD09A1 en janvier 2006 sur la tuile Modis h11v09 (500 m)

Si plus que 1 pixel sans nuage est identifié pour les 8 produits qui vont servir à faire le composite, le pixel de valeur de réflectance minimum est choisi. En effet, les pixels présentant une valeur de réflectance élevée ont une forte probabilité d'être affectés

par des aérosols ou des nuages en voile qui n'ont pas été détectés par l'algorithme interne de MODIS, ou encore de présenter un effet directionnel de réflectance dû à la variation de l'angle d'incidence (.pic de réflexion dans la direction de rétrodiffusion). La Figure 3-3 illustre les résultats des images composées avec les valeurs minimum, moyenne et maximum des pixels, avant élimination des pixels nuageux. (A) Valeur minimum, B) Valeur moyenne et C) Valeur maximum). L'image A présente le plus fort contraste sur le fleuve car les voiles nuageux et les éventuels effets directionnels ont été éliminés (B et C)

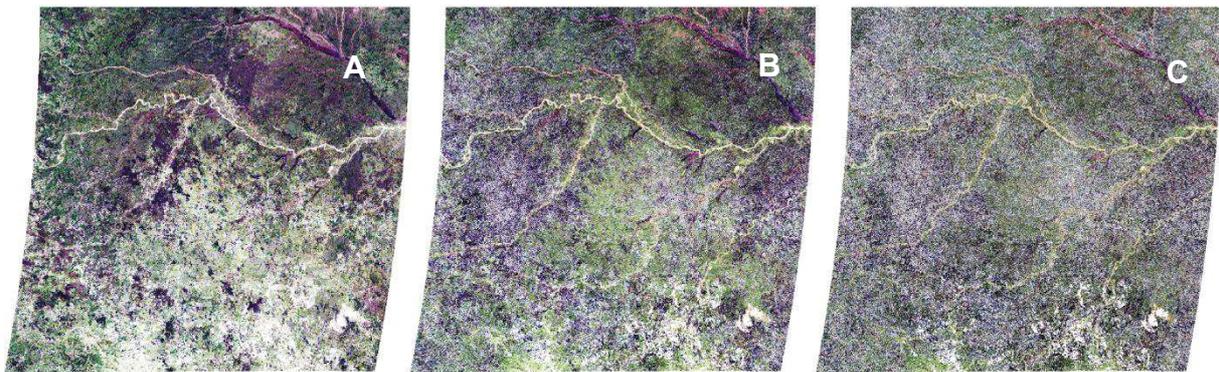


Figure 3-3 Image composée mensuelle avec valeurs A) Minimum, B) Moyenne et C) Maximum des pixels

3.2.1.2 Méthode d'extraction du masque eau/ non eau

Le produit de référence MODIS Water Mask (MOD44W) réalisé à l'aide des données Shuttle Radar Thematic Mapper (SRTM ne propose pas une cartographie dynamique mais un masque terre/eau à 250 mètres de résolution spatiale à l'échelle mondiale sur une seule date (Carroll et al. 2009).

Il a donc été réalisé une méthode permettant de discriminer l'eau du reste de la scène, qui répond aux besoins propres à cette étude : étude régionale, spectre de

réflectance des eaux très large en fonction de leur charge en sédiments et/ou matière organique, et grande variabilité de la surface d'eau libre au cours de l'année.

La méthode utilisée pour la mise au point de l'algorithme de classification eau-non/eau utilisée pour la réalisation du masque est celle connue comme « essayer-erreur ». Plusieurs petites zones ont été choisies comme représentatives de la diversité de couleur de l'eau, correspondant à des variations de surface minimum, et bien cartographiées, ce qui a permis la confrontation de chaque résultat obtenu avec des cartes détaillées des zones choisies : Zone de Coari (mélange d'eaux noires et blanches, plan d'eau circonscrit dans des berges hautes), zone du port Manaus (eau noire, surface du plan d'eau non variable), zone du port de Manacapuru (eau blanche, surface du plan d'eau non variable).

L'algorithme finalement retenu pour la classification eau/non eau est la combinaison de différents indices: l'adaptation de l'indice SWI (Sharma et al. 2015) en un indice SWIb, l'utilisation des indices NDVI (J. W. Rouse et al. 1973) et AWEI (Feyisa et al. 2014) et l'inspiration de la méthode booléenne et de seuillage de l'indice EWV (Ayana et al. 2015).

NDVI - Le Normalized Difference Vegetation Index (J. Rouse et al. 1973; Kriegler et al. 1969) est un indice très utilisé en télédétection pour déterminer la végétation photosynthétiquement active. Il est calculé à partir de deux bandes spectrales, le rouge (R) et le proche infra-rouge (PIR) comme le montre l'Equation 3.1 qui suit.

$$NDVI = \frac{PIR - R}{PIR + R} \quad (3.1)$$

NDWI - Le Normalized Difference Water Index (McFeeters 1996) est un indice utilisé en télédétection pour la surveillance de la teneur en eau des pixels. Il est calculé à partir des bandes spectrales du vert (V) et du proche infra-rouge (PIR) comme le montre l'Equation 3.2 qui suit.

$$NDWI = \frac{V - PIR}{V + PIR} \quad (3.2)$$

EWI - En 2015, Ayana et al. ont mis au point le "Enhanced Water Index" (EWI) ou « Indice d'Eau Amélioré ». Il s'agit d'une combinaison du NDVI et du NDWI qui regroupe les valeurs de NDVI inférieures à zéro et les valeurs de NDWI supérieures à zéro pour retrouver l'eau (Equation 3.3).

$$EWI = \text{if} (NDVI < 0 \text{ AND } NDWI > 0, \text{"Eau"} \text{ else } \text{"Terre"}) \quad (3.3)$$

SWI - L'indice de 2015 appelé "Superfine Water Index" (SWI) (Sharma et al. 2015) mis au point pour classifier les eaux de tout le globe terrestre à partir d'images MODIS. Il a pour but de pallier aux difficultés liées à l'utilisation des indices précurseurs pour l'identification des eaux dans des milieux très différents comme les calottes polaires, permafrosts ou zones marécageuses tropicales. Cette nouvelle méthode consiste à remplacer le «Vert» du NDVI par la composante «Saturation» du système de couleur « teinte-saturation-intensité (Hue – Saturation – Value (HSV) en anglais) qui est une transformée de la composition RGB (rouge-vert-bleu ou red-green blue) qui est utilisée pour les bandes de MODIS (Equation 3.4).

$$SWI = \frac{Saturation_{(RGB)} - 7 \times PIR}{Saturation_{(RGB)} + 7 \times PIR} \quad (3.4)$$

Le SWI offre un contraste très élevé entre les différents types d'eau et les zones non-eau, y compris la neige et la végétation. Il offre un contraste élevé entre les valeurs de saturation et de réflectance dans le proche infrarouge (PIR) des zones non-eau, alors que, dans le même temps, les hautes valeurs du PIR ne parviennent pas à saturation. Le coefficient 7.0 utilisé sert à augmenter les contrastes sans interférer sur la saturation.

SWIb - L'indice SWIb (Equation 3.5) est une modification importante de l'indice SWI. A la place du système de gestion des couleurs HSV d'où l'on extrait la composante « Saturation », on utilise ici, le système Teinte - Saturation – Luminosité (Hue – Lightness - Saturation - HLS) duquel on extrait la composante « Luminosité » (Figure 3-4) qui est la perception de l'intensité de la réflectance pour les trois bandes RGB du visible (Ryerson & Rencz 1999).

$$SWIb = \frac{Lightness_{(RGB)} - 7 \times PIR}{Lightness_{(RGB)} + 7 \times PIR} \quad (3.5)$$

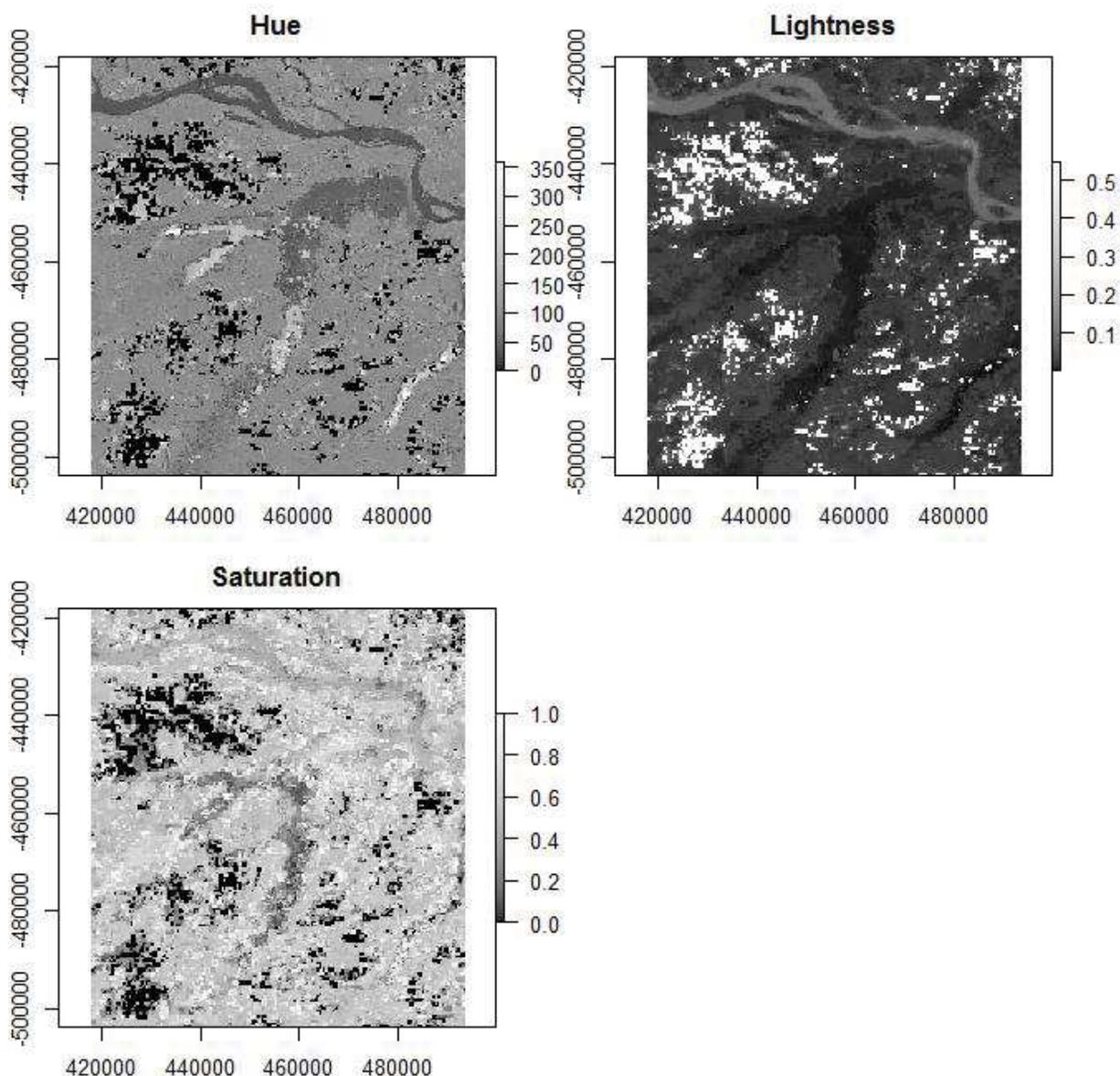


Figure 3-4 Images des composantes du système de couleurs Teinte – Luminosité – Saturation pour la région de Coari

Ce changement est réalisé car la composante « luminosité » est une composante qui offre plus de contraste pour la réflectance entre eau et non eau dans les canaux du visible (*Bandes 1, 4 et 3 de MODIS*) que la composante « Saturation ». En effet la SWI extrait bien les eaux obscures mais ne parvient pas à classifier correctement la diversité des eaux amazoniennes. Pour classifier l'eau "blanche" à partir du SWI il fallait augmenter le seuil d'appartenance de cet indice pour la classe eau, ce qui

générerait énormément de bruit sur l'image car cela intégrait d'autres pixels de classes différentes (Figure 3-5). Ce problème peut être dû aux caractéristiques optiques des fleuves dits blancs qui réfléchissent les matières en suspension produisant une confusion avec les sols nus.

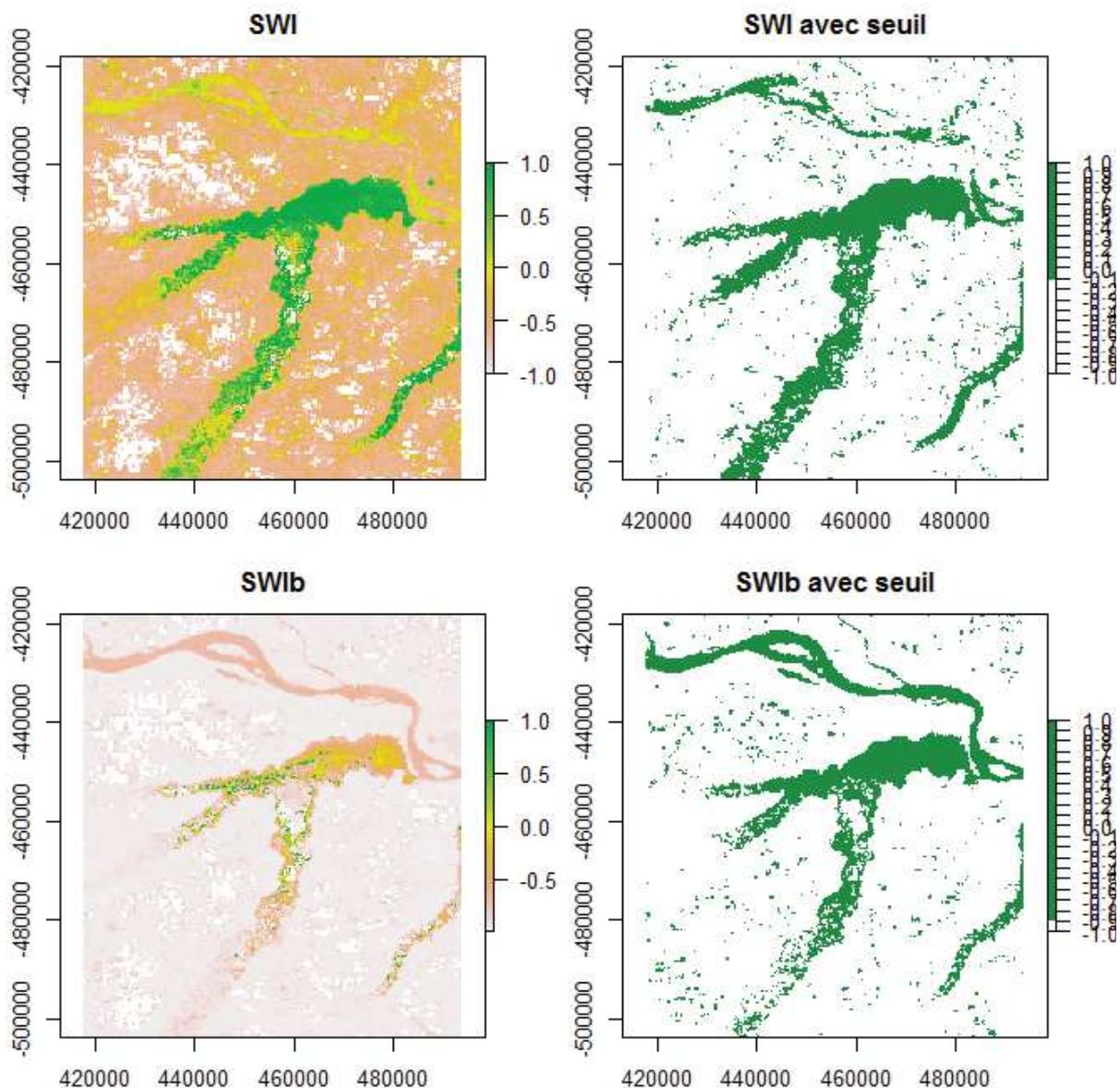


Figure 3-5 Comparaison entre SWI et SWIb pour la classification par seuil des eaux du Lac de Coari

Le SWI et le SWIb ont des valeurs entre 1 et -1. Ce dernier produit des résultats bien plus intéressants, comme le montre la Figure 3-5 sur laquelle on peut observer que le SWI seuillé n'identifie pas bien les eaux blanches du Rio Solimões à l'embouchure du lac de Coari. On trouve sur cette figure, superposés, les deux indices et leur classification par seuillage de l'appartenance des eaux. La classification du SWI est fixé en "Eau > -0,20" et la classification du SWIb a un seuil d'appartenance fixé en "Eau > -0,80". On observe le Lac de Coari et le Rio Solimões. Le lac possède des eaux noires et le fleuve des eaux blanches. La difficulté de classier les eaux blanches (minérales) avec le SWI est surmontée avec le SWIb. Le bruit présent dans le lac a été comblé par la suite avec l'élaboration de la méthode.

AWEI - Feyisa et al. (2014) ont proposé, pour déterminer automatiquement les surfaces inondées, le "Automated Water Extraction Index" (AWEI) en améliorant le contraste spectral des données Landsat TM et en définissant une valeur de seuil optimale pour la classification des eaux de surface (Equation 3.6).

$$AWEI = 4 \times (Vert - MIR) - (0.25 \times PIR + 2.75 \times SWIR) \quad (3.6)$$

La classification du masque des eaux exécutée dans ce travail est élaborée à partir des différents indices définis plus haut. Pris individuellement ces indices ne donnent pas des résultats suffisamment robustes, mais qui, utilisés de manière complémentaire ils se sont révélés efficaces (méthode essai-erreur décrite plus haut).

La conjugaison de ces indices et l'établissement de seuillages a permis de produire le masque de l'eau (Equation 3.7).

$$Eau = (SWIb \times NDVI > -0.556 \text{ OR } AWEI > -0.04) \text{ AND } AWEI > -0.04 \quad (3.7)$$

L'équation proposée ici utilise la complémentarité spectrale des *bandes 1* (620-670 nm) et *2* (841-876 nm) à 250 mètres et des *bandes 4* (545-565 nm) et *5* (1230-1250 nm) à 500 mètres qui composent ces indices. L'équation développée se structure en trois étapes. La première est la multiplication du SWIb par le NDVI. La seconde est la sélection par l'opérateur logique exclusif « OR » de celle-ci avec l'indice AWEI et finalement la sélection par l'opérateur logique inclusif « AND » entre la sélection précédente et l'indice AWEI.

La multiplication entre les indices a été employée comme un moyen d'accentuer le contraste pour les pixels qui en manquaient et renforcer ce que les deux indices ont en commun. La multiplication du SWIb par le NDVI (Figure 3-6) permet la complémentarité des informations contenues dans la composante "Luminosité", dans la bande rouge et la bande du PIR, cette dernière étant commune aux deux indices.

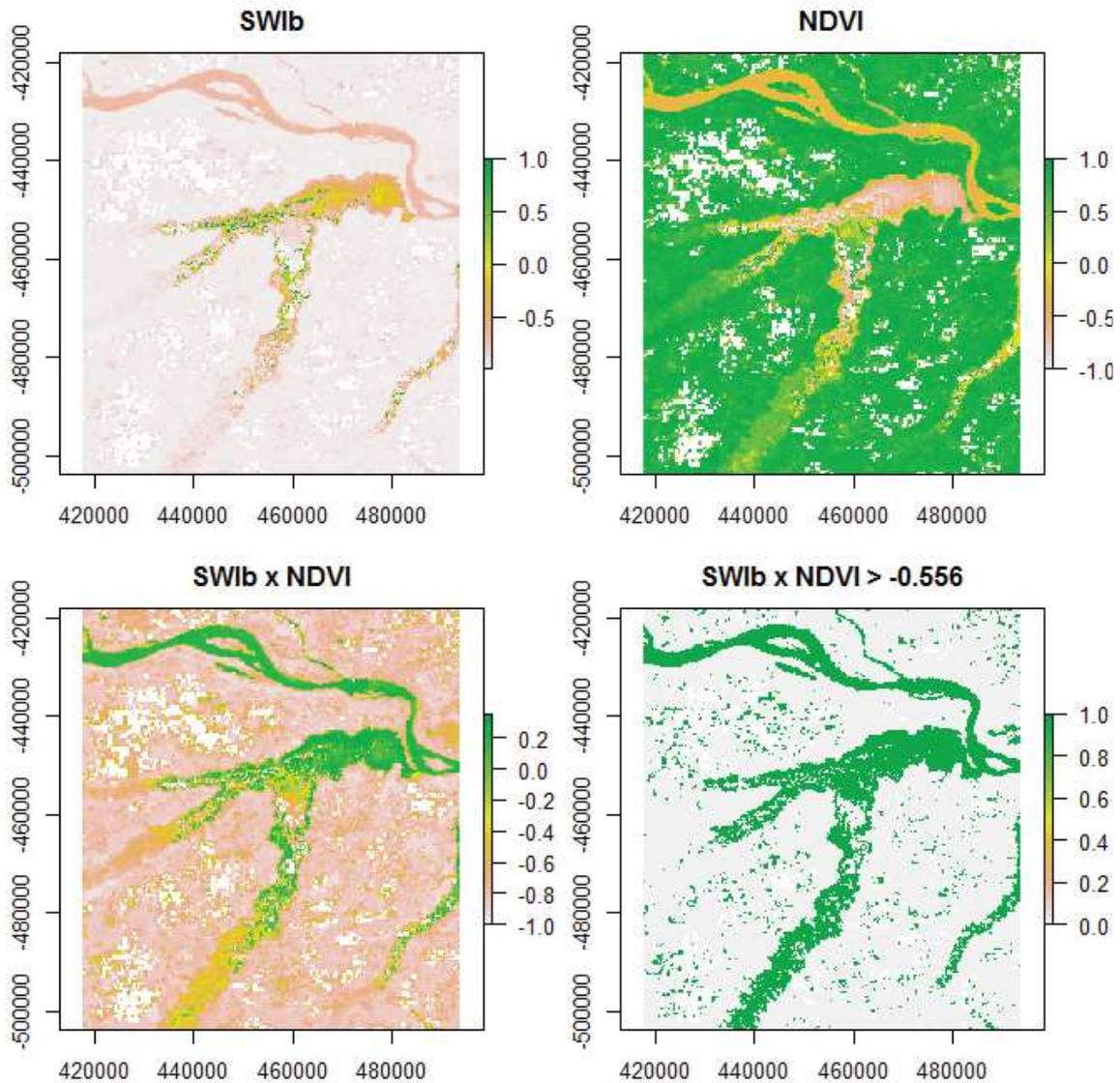


Figure 3-6 Multiplication de l'indice SWIb par le NDVI sur le Lac de Coari

Le deuxième terme de l'équation du masque de l'eau (Equation 3.7) porte sur l'indice AWEI qui est composé des bandes Vert, MIR, PIR et Infra-rouge à ondes courtes (SWIR). Cet indice d'extraction des eaux ne s'est pas montré très efficace à lui seul pour la classification des fleuves amazoniens. Selon le seuil d'appartenance il n'identifiait pas suffisamment de surfaces inondées (Figure 3-7 (3)) ou bien, au contraire, il incluait trop d'éléments externes dans la classification produisant de nombreux artefacts (Figure 3-7 (5)). Néanmoins, son utilisation a permis, dans un

premier temps, d'identifier des pixels d'eau non classés. Le AWEI comble alors les lacunes de classification en se superposant avec la fonction "OR" (+) au résultat antérieur muni d'un seuil d'appartenance de " $>-0,04$ ", issu des essais réalisés.

Pour la dernière partie de l'algorithme du masque d'eau, le AWEI est associé par la conjonction logique « AND » avec le résultat précédent, muni d'un seuil d'appartenance de « $AWEI > -0,4$ », ce qui permet de garder les pixels classifiés comme « eau » et de diminuer le nombre d'artefacts en éliminant les pixels qui ne sont pas communs aux deux images (Figure 3-7 (6)).

Les seuils définis dans cette méthode ont été définis par expérimentation en minimisant les pixels de commission aussi bien que les pixels d'omission.

Une fois cette méthode de classification définie sur les zones test, elle est ensuite appliquée automatiquement à toutes les dates de l'étude.

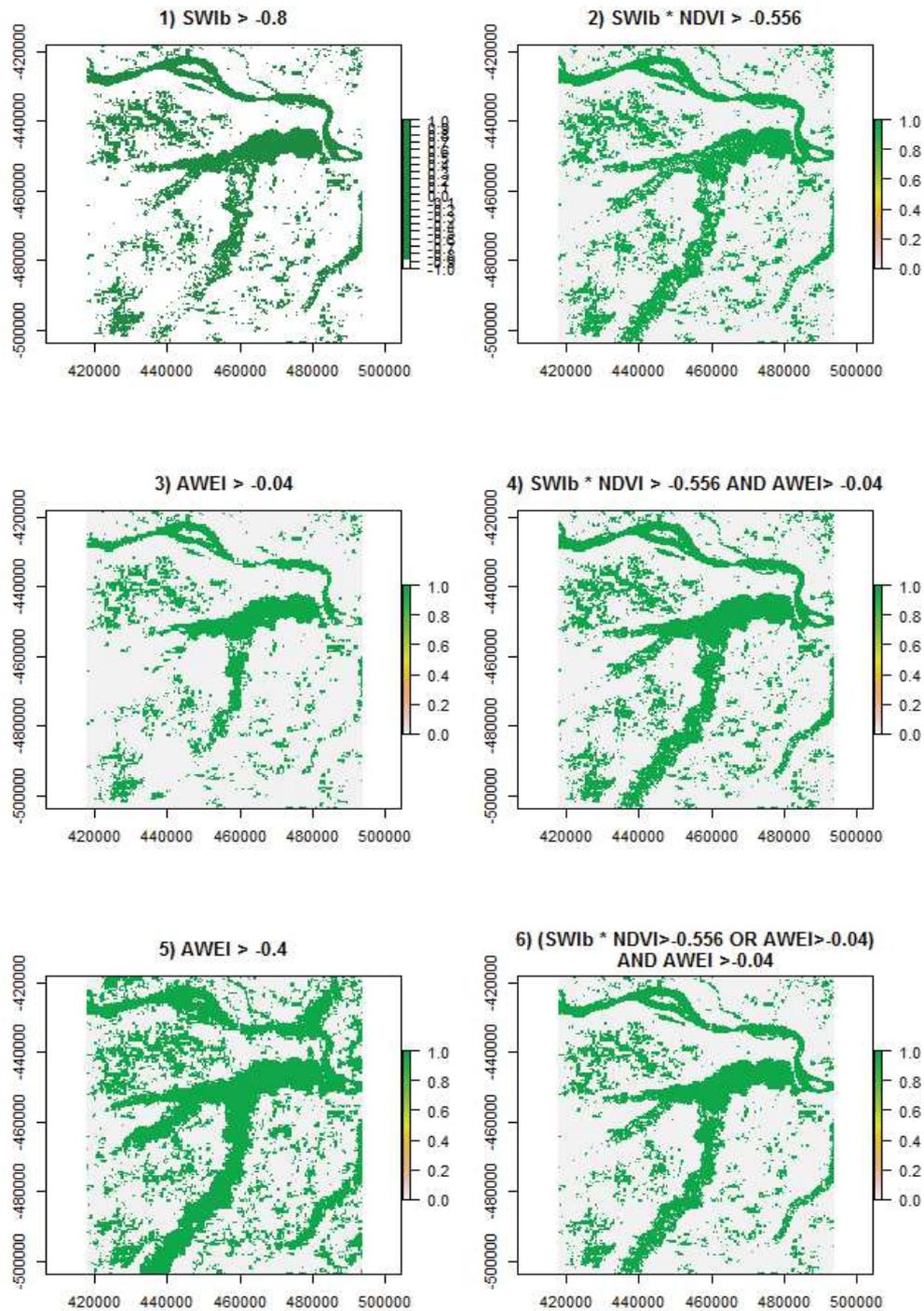


Figure 3-7 Indices et création du masque des eaux

3.2.1.3 Méthode de validation du masque eau/non eau

La validation du masque eau/ non eau a été réalisée par comparaison avec un produit du projet ESA/CCI : Land Cover (LC) *Water Bodies* – v4.0 du Climate Change Initiative (CCI) réalisé par l'Agence Spatiale Européenne (ESA) (Land Cover CCI 2016). Celui-ci est une carte statique des plan d'eau du monde entier à 150 mètres de résolution spatiale, fruit de la compilation et de l'édition de deux différentes classifications Terre/Eau: La classification Envisat ASAR Water Bodies (Bartsch et al. 2012) et la classification Global Inland Water (Berry et al. 2005). La légende de la classification ESA/CCI a deux valeurs: 1-Non-Eau et 2-Eau.

La validation a été effectuée en comparant le masque ESA/CCI avec un masque issu de notre chaîne de traitement, à la même date et ré-échantillonné à 150m. Une matrice de confusion a été produite.

3.2.1.4 Création d'un masque eau-non eau d'étendue maximale

Certains pixels résiduels, principalement issus des ombres des nuages proviennent de la création de l'image composée mensuelle qui utilise les pixels ayant une valeur de réflectance minimum, créant une confusion entre pixels d'ombre de nuages et pixels d'eau.

Les pixels mal classés ont la particularité de ne jamais se reproduire au même endroit entre les dates, mais surtout entre les années, différemment des pixels d'eau. Ainsi, pour extraire ces pixels mal classés, un nouveau masque est produit représentant la surface inondée maximale, à l'intérieur duquel les pixels de la classe « Eau » auront une probabilité plus importante de réellement être de l'eau. Ce masque rassemble l'étendue maximum des eaux sur les 12 mois ou le niveau des eaux est le plus haut pour toutes les années disponibles. La méthode consiste à

compter le nombre de fois où le pixel « Eau de crue » est superposé. Les images de crue sélectionnées sont celles des dates: 2003/07, 2004/07, 2005/07, 2006/06, 2007/06, 2008/06, 2009/07, 2010/07, 2011/07, 2012/07, 2013/06 et 2014/06.

A partir des images superposées, les pixels ayant le score de répétition le plus haut sont sélectionnés. La Figure 3-8 illustre la démarche : les pixels à plus fort score de répétition sur les 12 mois de crues sélectionnés apparaissent en vert. Un grand nombre de pixels isolés qui ne possèdent pas d'équivalents pour les autres mois sont des artefacts. Les pixels classés « Eau » au moins deux fois ont été retenus. Le produit de ce masque est illustré dans la Figure 3-9.

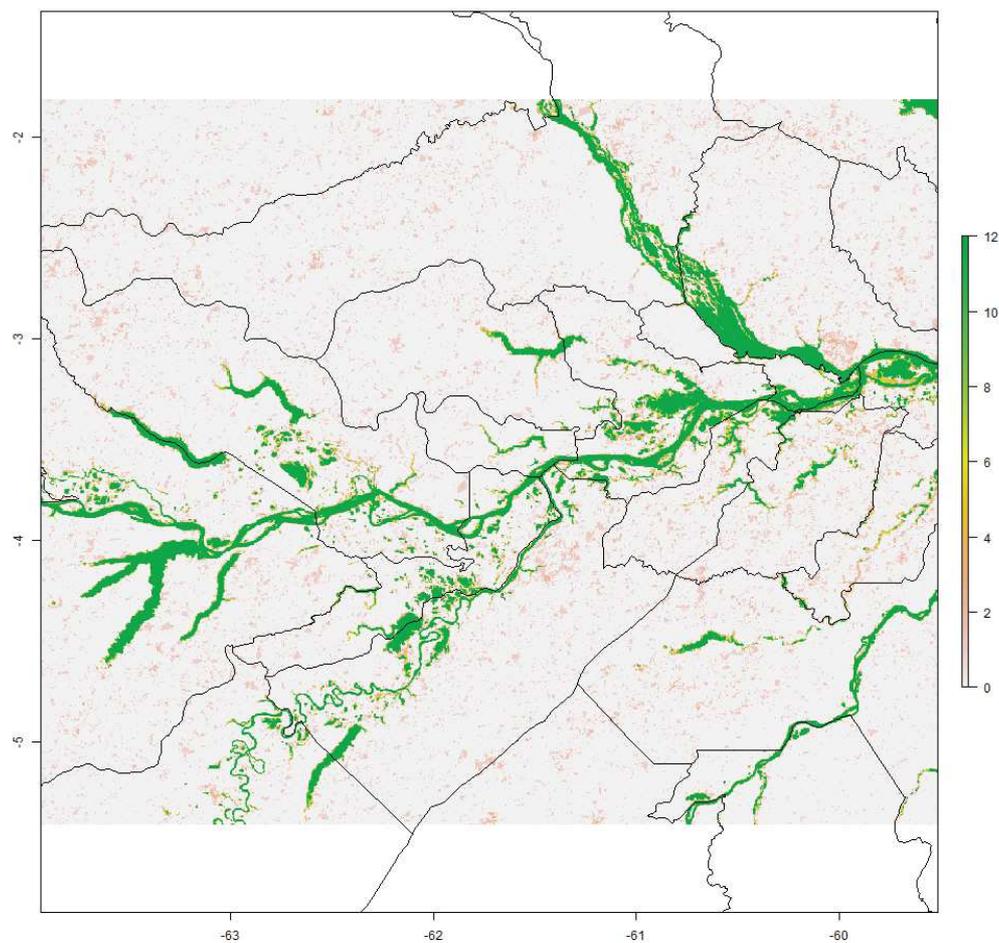


Figure 3-8 Comptage du nombre de superposition du pixel Eau sur 12 images de crue des fleuves

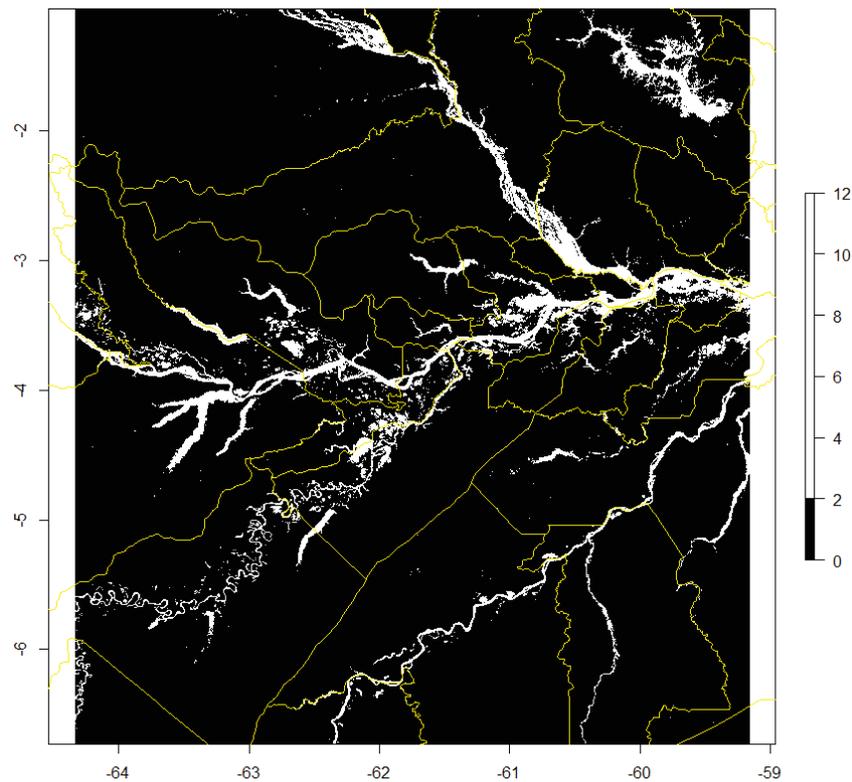


Figure 3-9 Résultat: masque de l'étendue maximum des eaux dont les pixels se répètent au moins deux fois en 11 ans. Les limites municipales apparaissent en jaune.

3.2.1.5 Méthode de création d'un indice de luminosité des eaux

L'indice de luminosité des eaux est calculé mensuellement pour les pixels du masque eau/non eau mensuel.

Comme exposé dans le chapitre 1, les caractéristiques responsables des teintes noire ou laiteuse des fleuves dits noirs et blancs amazoniens peuvent être observées dans les bandes du spectre visible et donc par les différentes bandes de MODIS: Bande 1 (Rouge – 250m), Bande 4 (Bleu - 500m) et Bande 3 (Vert - 500m).

Etant donné le contraste espéré pour les différents systèmes limnologiques amazoniens, la méthode employée ici pour les identifier vise la production d'une image d'intensité de la luminosité provenant de l'ensemble des trois bandes du visible.

L'intensité de la luminosité de la lumière visible a été extraite en changeant de référentiel de représentation des couleurs : passage du système RVB (Rouge – Vert – Bleu au système Teinte – Luminosité – Saturation (Hue-Lightness-Saturation), pour lequel la composante « Luminosité » exprime l'intensité recherchée (Figure 3-10).

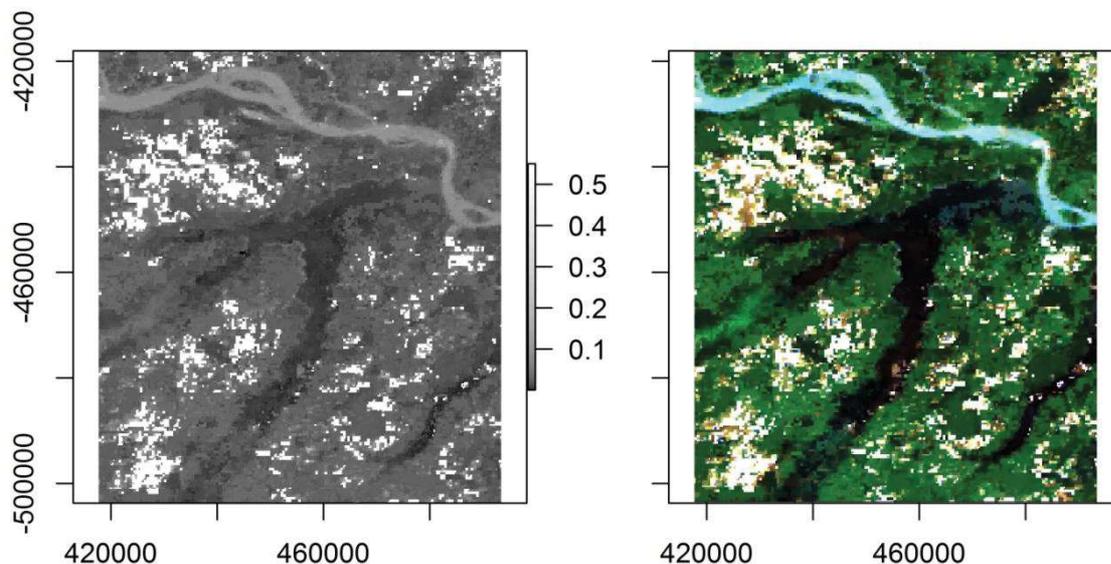


Figure 3-10 Composante Luminosité (HLS) et composition RGB - 143 de MODIS 2006/01 sur le Lac Coari

La composante « Luminosité » a des valeurs comprises entre 0 et 1, surtout concentrées entre 0 et 0.3 (histogramme des valeurs présenté Figure 3-11). Les valeurs comprises entre 0 et 0.3 ont été réparties en dix intervalles égaux: [0 - 0.03; 0.03- 0.06; 0.06 - 0.09; 0.09 - 0.12; 0.12 - 0.15; 0.15 - 0.18; 0.18 - 0.21; 0.21 - 0.24; 0.24 - 0.27; 0.27 - 0.30].

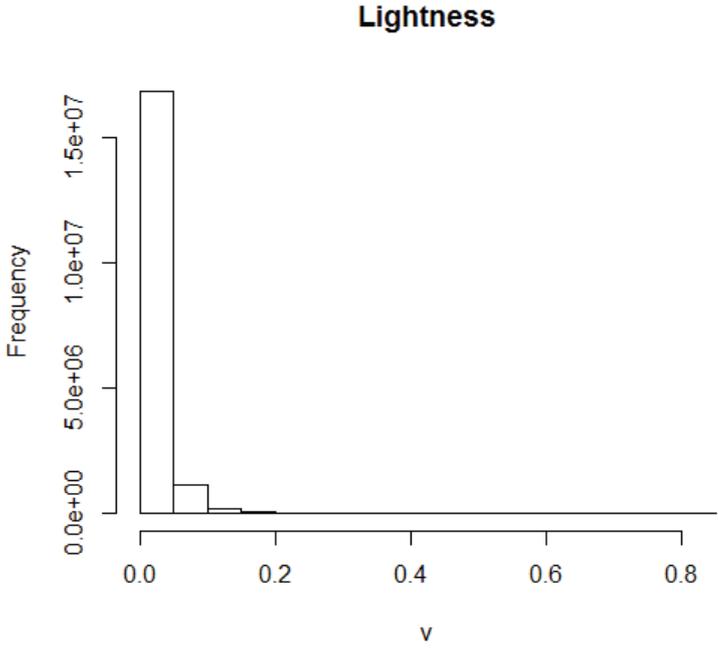


Figure 3-11 Histogramme de la fraction Luminosité (Lightness) du système de gestion des couleurs HLS

3.2.1.6 Le flux de traitements

Le flux de traitements s'est constitué selon les étapes de la Figure 3-12 qui suit.

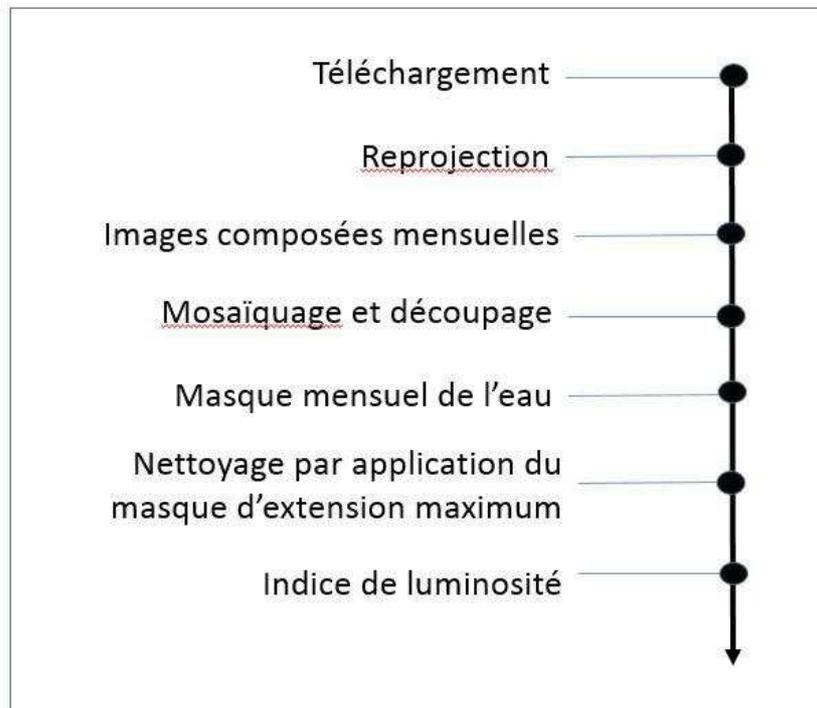


Figure 3-12 – Flux de traitement des images MODIS pour l'obtention de l'indice de luminosité

Etape 1 : Téléchargement des images

La chaîne de traitement commence par le téléchargement des données MODIS. 11 années de données ont été téléchargées. L'ensemble de la base de données est constituée de 35.227 images.

Le site *Reverb | ECHO* de la Nasa (<http://reverb.echo.nasa.gov/>) permet de produire une requête d'images. Le résultat de la requête peut être sauvegardé sous la forme d'une liste de liens *http* qui donnent ensuite accès aux images.

Le téléchargement de piles d'images a été programmé en R (R Development Core Team 2011) permettant un gain de temps considérable et l'organisation automatique des images par produit et par tuile dans des dossiers informatiques.

Etape 2 : Reprojection

Les images MODIS disponibles en projection sinusoidale ont été reprojctées en projection UTM pour coïncider avec le référencement géographique de la base de données SIVEP-Malaria et avec les données géographiques utilisées dans ce travail. L'algorithme R « reprojectHDF() »; issu de la librairie de fonctions " ModisDownload" développée par Babak Naimi (2011) basé sur l'outils manuel « MODIS Reprojection Tool » (Land Processes DAAC 2011) a été utilisé.

L'ensemble de la pile d'images est ré-échantilloné à 250m.

Etape 3 : Images composées mensuelles

La méthodologie de création d'images composites mensuelles décrites plus haut est ensuite appliquée. Ce traitement aboutit à un total de 132 images composées mensuelles sur 11 ans avec une résolution spatiale composée de pixels de 250 mètres et de pixels de 500 mètres et avec les bandes spectrales de 1 à 7.

Etape 4 : Mosaïquage et découpage de la zone d'étude

Pour chacun des 132 mois de l'étude; est réalisée une mosaïque des cinq tuiles MODIS qui couvrent la région par date et le découpage pour l'État d'Amazonas.

Etape 5 : Création du masque sur l'eau

Le processus de classification du masque eau/non eau peut alors être développé d'après l'Equation 3.7 pour produire le masque mensuel qui enregistre les différentes étendues d'eau chaque mois.

Etape 6 : Nettoyage du masque eau / non eau

Le nettoyage des artefacts du masque eau/non eau est effectué par application du masque d'étendue maximale de l'eau.

Etape 7 : Création des images « indices de luminosité »

La classification par intervalles de valeurs de la composante « Luminosité » est réalisée pour chacune des images composées mensuelles dans la limite des masques des eaux des mois respectifs.

3.3 Résultats et Discussion

3.3.1 Résultats de la validation du masque eau/ non eau

Le résultat de la matrice de confusion entre le masque eau/ non eau des eaux (ré-échantillonné et à la même date) et le produit ESA/CCI a donné 72% de pixels identiques entre les jeux de données et les erreurs de commission et d'omission qui sont récapitulées dans le Tableau 3-3.

Tableau 3-3 Matrice de confusion entre le masque eau/ non eau des eaux (ré-échantillonné et à la même date) et le produit ESA/CCI en nombre de pixels.

		ESA/CCI		
		Non Eau	Eau	Total
Masque d'eau MODIS	Non Eau	13048264	121593	13169857
	Eau	143627	314316	457943
	Total	13191891	435909	
		98.91%	72.11%	

Le masque eau/non eau a une erreur d'omission de 27.89% et une erreur de commission de 32.95% par rapport à la classification ESA/CCI. La Figure 3-12 et la Figure 3-13 illustrent la classification du masque eau/non eau en comparaison de la classification Eau/Terre de ESA/CCI.

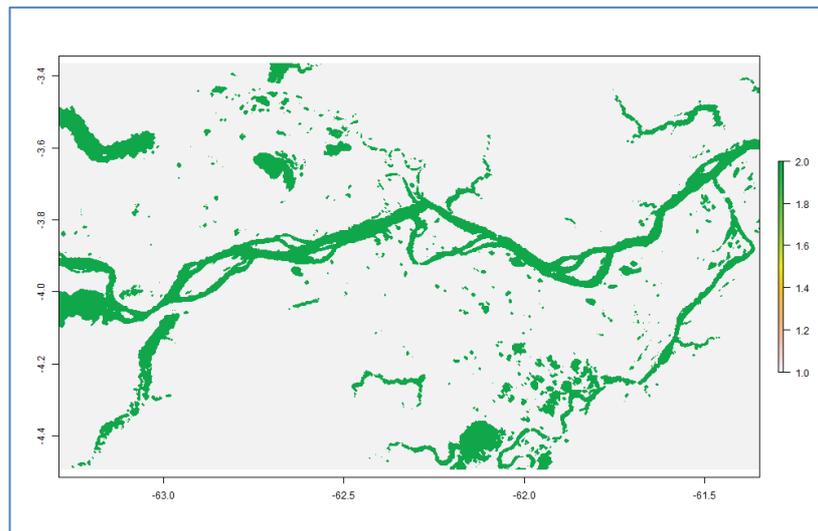


Figure 3-13 Classification du masque de l'eau à partir de MODIS (500 m)

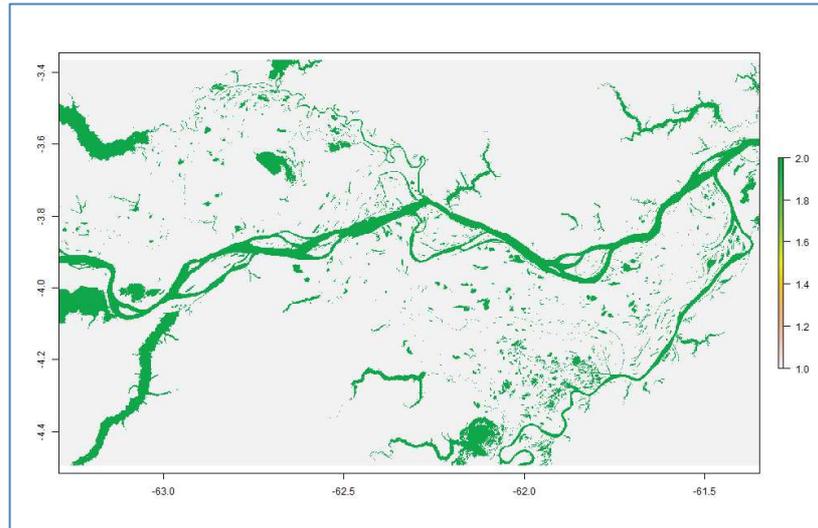


Figure 3-14 Classification ESA/CCI (150 m)

Les confusions sont représentées dans la Figure 3-14 : (les pixels classés correctement (« *Eau* » et « *Non-Eau* ») ; les pixels d'omission (classés « *Non-Eau* » alors qu'ils sont classés *Eau* dans l'image de référence) et les pixels de commission (classés « *Eau* » alors qu'ils sont classés « *Non-Eau* » dans l'image de référence).

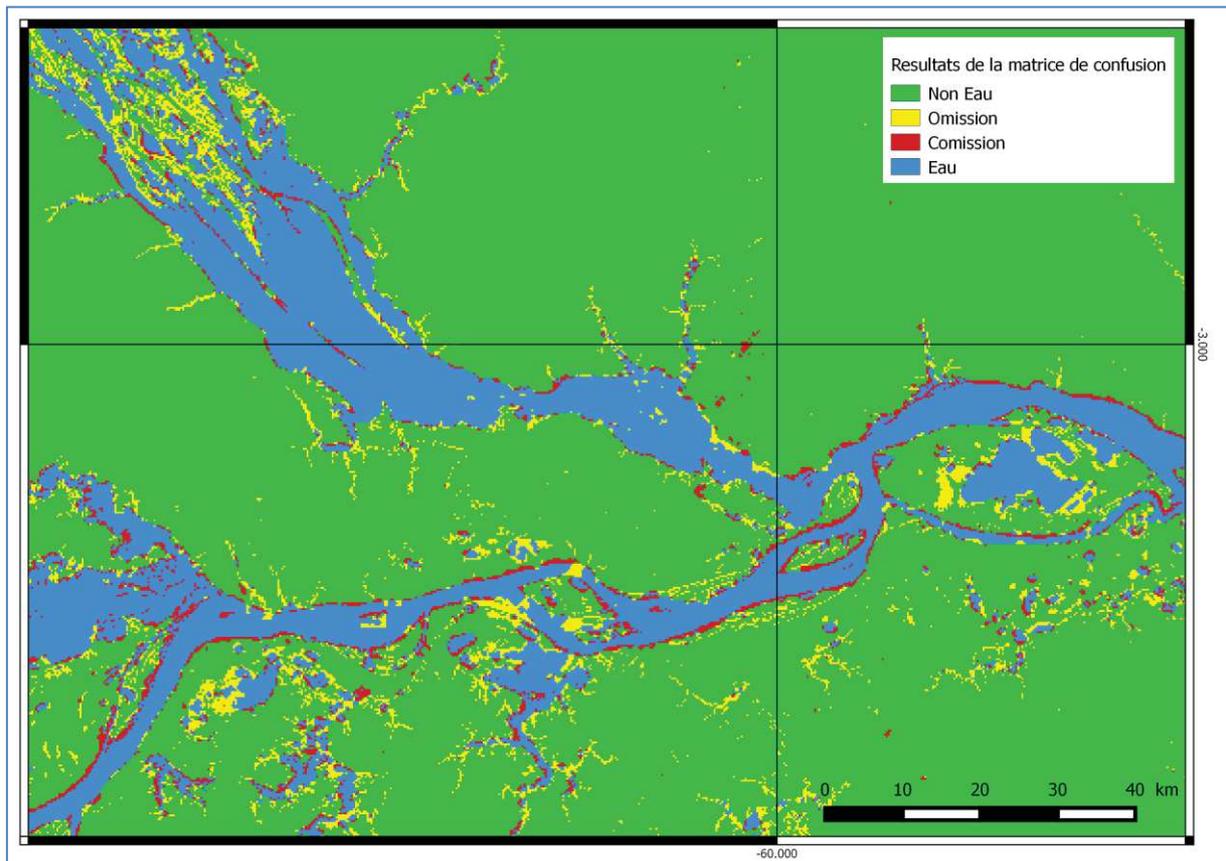


Figure 3-15 Résultats de la validation du masque eau/ non eau (150 m) sur la région de Manaus à l'encontre des fleuves Rio Negro et Rio Solimões.

Le résultat de la validation permet d'estimer la qualité du masque eau/non eau comme correcte (72% de pixels classifiés correctement); Les pixels de commission (en rouge) sont adjacents aux pixels correctement classés eau (en bleu). Les pixels d'omission (en jaune) sont majoritairement dus à la différence de résolution des pixels. Il s'agit surtout de bras de fleuves (igarapés) invisibles à la résolution de MODIS. On peut inférer que la même raison a joué pour les pixels de commission.

3.3.2 Analyse statistique de la cartographie dynamique et interprétation des résultats

Le résultat de la chaîne de traitements décrite plus haut est une cartographie dynamique qui permet l'observation des changements mensuels et saisonniers de la couleur des eaux pour l'état d'Amazonas. La première observation est que cette variabilité saisonnière de la couleur de l'eau est beaucoup plus grande que ne le laisse penser de prime abord la répartition « statique » en fleuves d'eau blanche et d'eau noire. Elle décrit un paysage des eaux Amazoniennes extrêmement dynamique et changeant.

Pour rendre compte de la dynamique de la couleur des eaux sur onze ans, un graphique en boîte à moustaches (box-plot) de la moyenne mensuelle de chaque classe d'intensité pour toute la zone d'étude a été réalisé (Figure 3-15) graphique permet d'appréhender l'intensité moyenne mensuelle de la luminosité des eaux caractérisant la totalité de l'état d'Amazonas. En ordonnée se trouvent les classes d'intensité de luminosité de 1 à 10. La classe 1 est la moins réfléchissante (noire) et la 10 la plus réfléchissante (blanche). Les mois d'étiage affichent les intensités les plus fortes alors que les mois de crue affichent des intensités de luminosité basses. Le mois de novembre est le mois avec la luminosité moyenne la plus élevée alors que le mois de juin celui avec les valeurs les plus basses. Les mois de mai, juin, juillet, août qui sont les mois où les niveaux du fleuve sont les plus élevés (Wolfarth et al. 2013) sont classés en intensité 1.

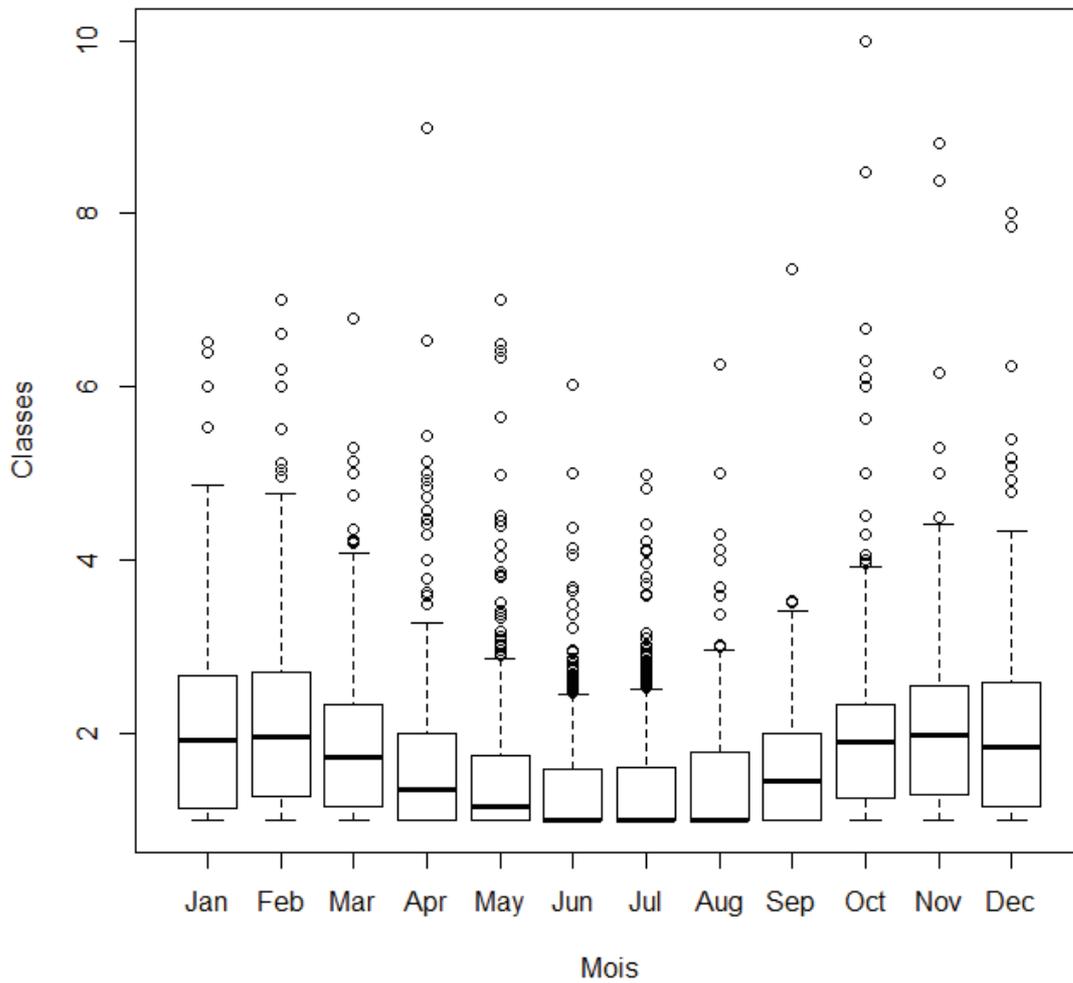


Figure 3-16 Graphique en boîte à moustaches des moyennes mensuelles des classes d'eau pour l'état d'Amazonas

Les mois de transition (avril et septembre) sont également bien identifiés. La variance autour de ces moyennes représentée dans le graphique « boîte à moustache » est mieux représentée dans la Figure 3-16,

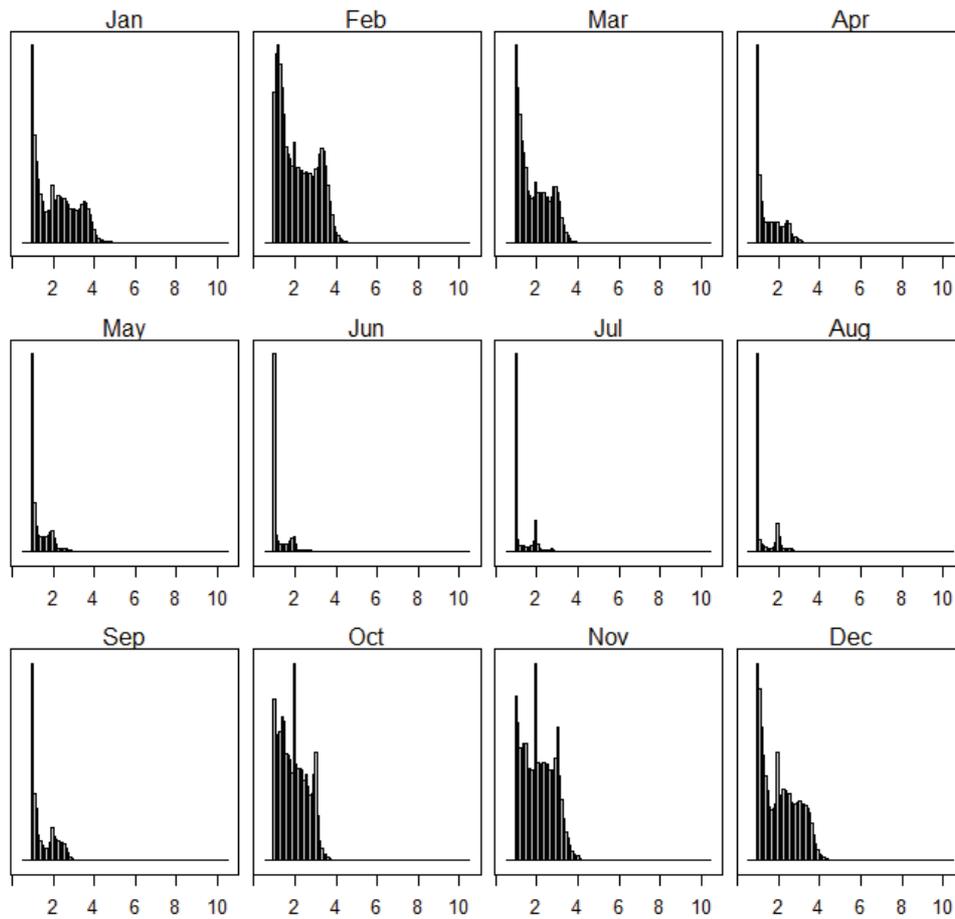


Figure 3-17 Histogrammes de la répartition des classes d'eau moyenne (de 1 à 10) par mois pour tout l'état de l'Amazonas

La représentation cartographique de la moyenne mensuelle entre les classes 1 et 5 d'intensité de la luminosité de l'eau des principaux paysages limnologiques ainsi que leur dynamique inter-mensuelle est représentée Figure 3-17 pour la région de confluence des fleuves Negro et Solimões.

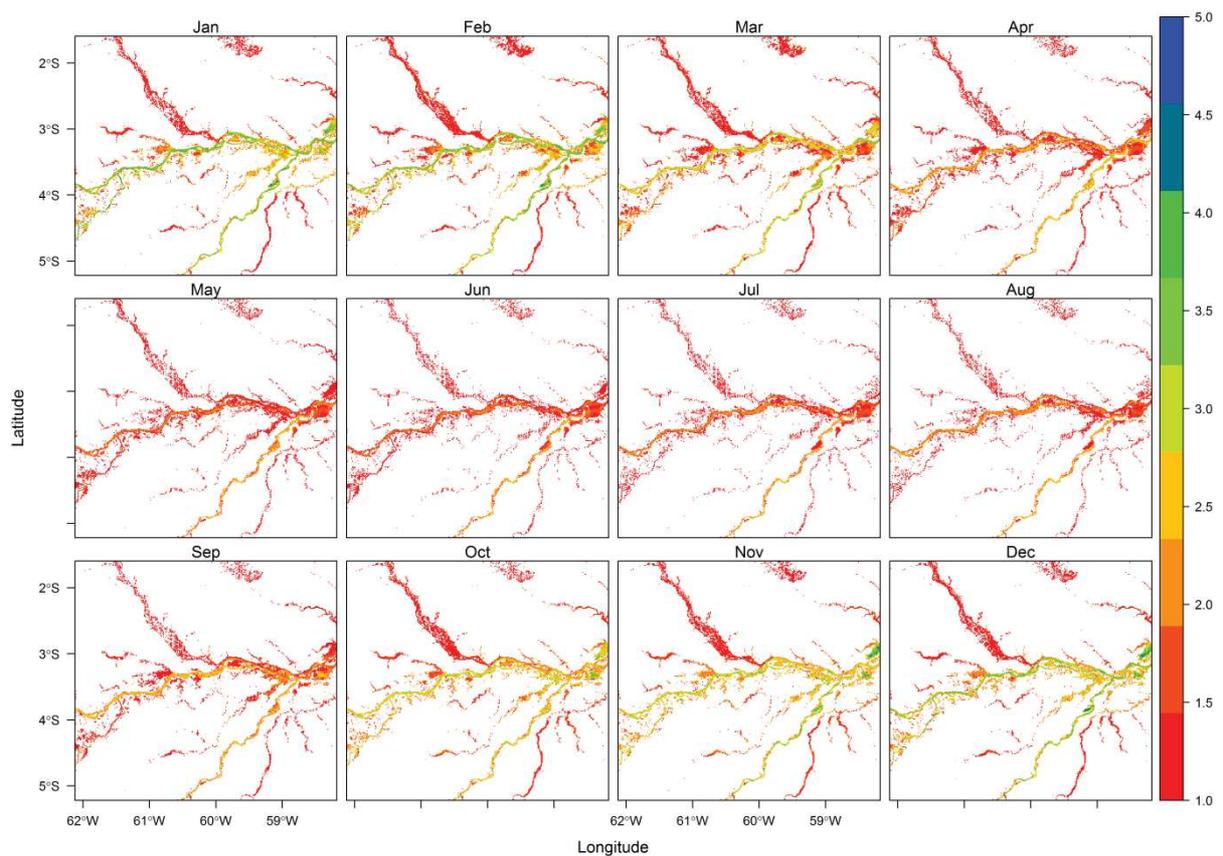


Figure 3-18 Moyenne mensuelle de la luminosité de l'eau entre les classes 1 et 5 pour la région de la confluence des fleuves Rio Negro et Rio Solimões

Les moyennes mensuelles spatialisées permettent d'identifier les grands ensembles de paysage et les prédominances saisonnières. Une autre approche permet de mettre en évidence les zones qui sont les plus variables au cours de l'année et celles qui sont les plus stables. Il s'agit d'une image représentant la variance calculée à partir des valeurs mensuelles sur les 11 années pour chaque pixel. Les trois zooms qui suivent de la Figure 3-18 à la Figure 3-20 illustrent cette variance, qui est un indicateur de la dynamique de la qualité des systèmes aquatiques.

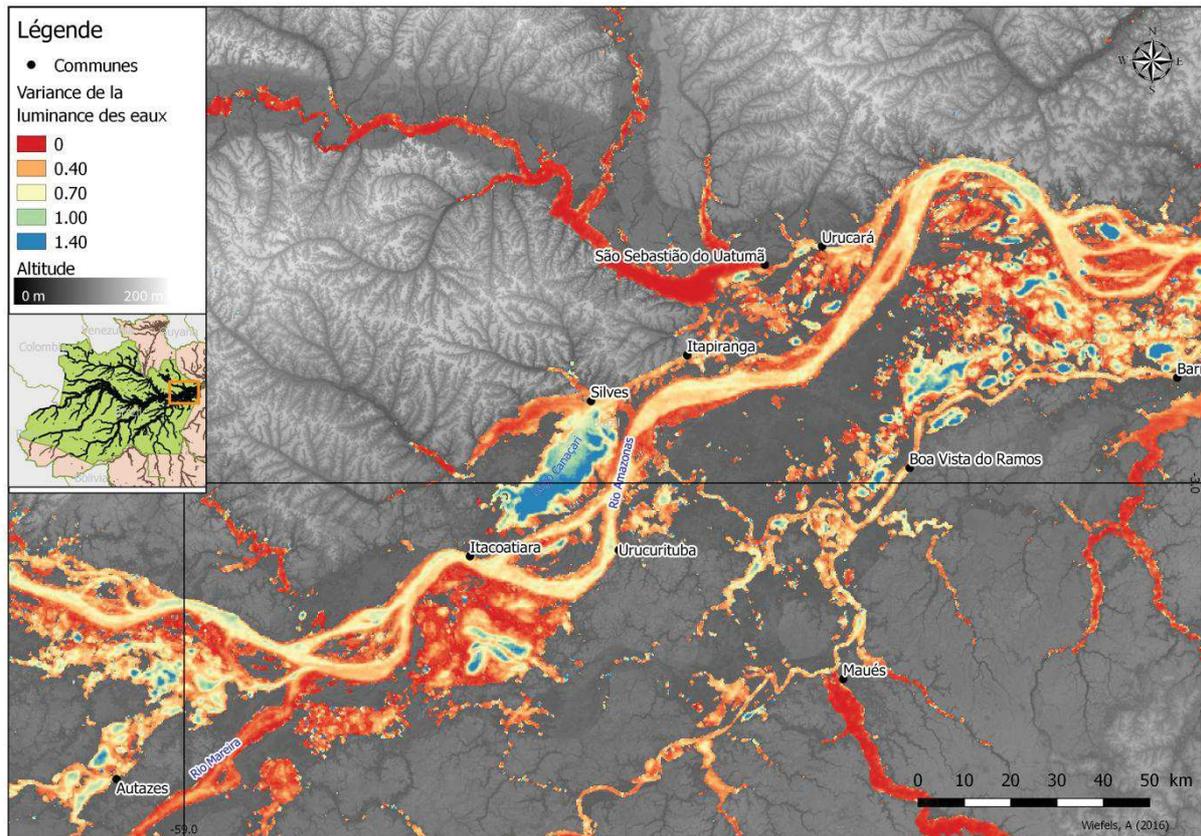


Figure 3-19 Variance de la luminosité des eaux aux alentours de Itacoatiara-AM au long du Rio Amazonas; région caractérisée par de vastes zones d'inondation

La Figure 3-18 propose un zoom autour des grandes zones d'inondation en aval de Manaus. Ces « varzeas » sont isolées du fleuve pendant la période sèche et la variation de couleur y est maximale (bleu).

La Figure 3-17 illustre la variation de l'intensité de la luminosité pour la zone de confluence des fleuves Japurá et Juruá sur le Rio Solimões. On y observe la variabilité des eaux (pixels en bleu) des fleuves Solimões et Juruá qui se mélangent avec les eaux moins variables du Japurá et qui entourent l'île qui se forme a cet emplacement.

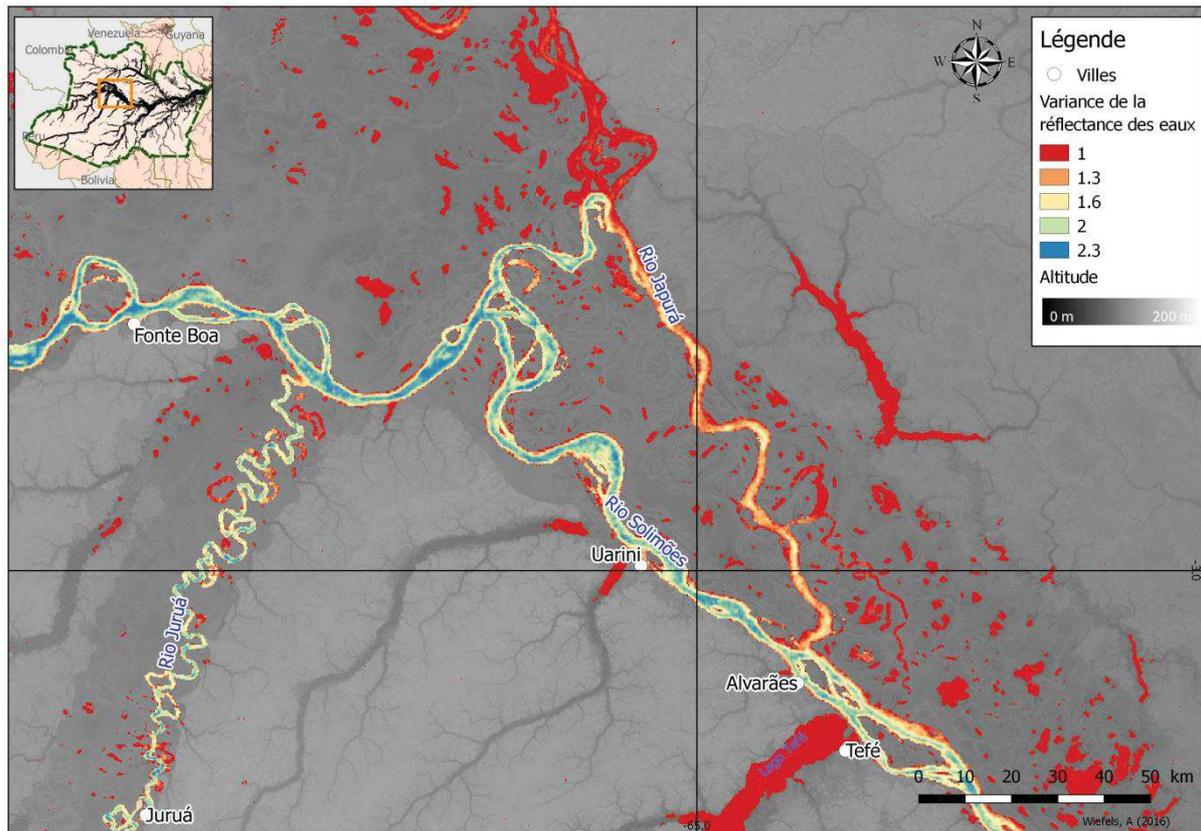


Figure 3-20 Variance de la luminosité des eaux à l'encontre des fleuves Japurá; Juruá et Solimões

La Figure 3-19 est un zoom sur la rencontre des eaux en face de Manaus. A cet endroit, les eaux du Rio Solimões et celles du Rio Negro se rencontrent mais ne se mélangent pas sur une certaine distance en raison des différences de couleur et de densité des eaux des deux fleuves, formant deux flux parallèles. Ce qui est observé sur la figure est cette fois la différence de variabilité de la couleur des eaux qui suit le même schéma que la couleur jusqu'à la ville de Itacoatiara : stabilité du Rio Negro et des affluents du Rio Solimões, variabilité de ce dernier et des principales zones d'inondation. On peut noter que le Rio Madeira affiche des valeurs de variance très élevées; ce qui signifie que sa luminosité subit des changements importants au fil des saisons.

A côté de ces observations il y a les lacs et igarapés qui enregistrent des faibles variations de luminosité dans le temps; même s'ils enregistrent d'importantes variations en terme de surfaces inondées.

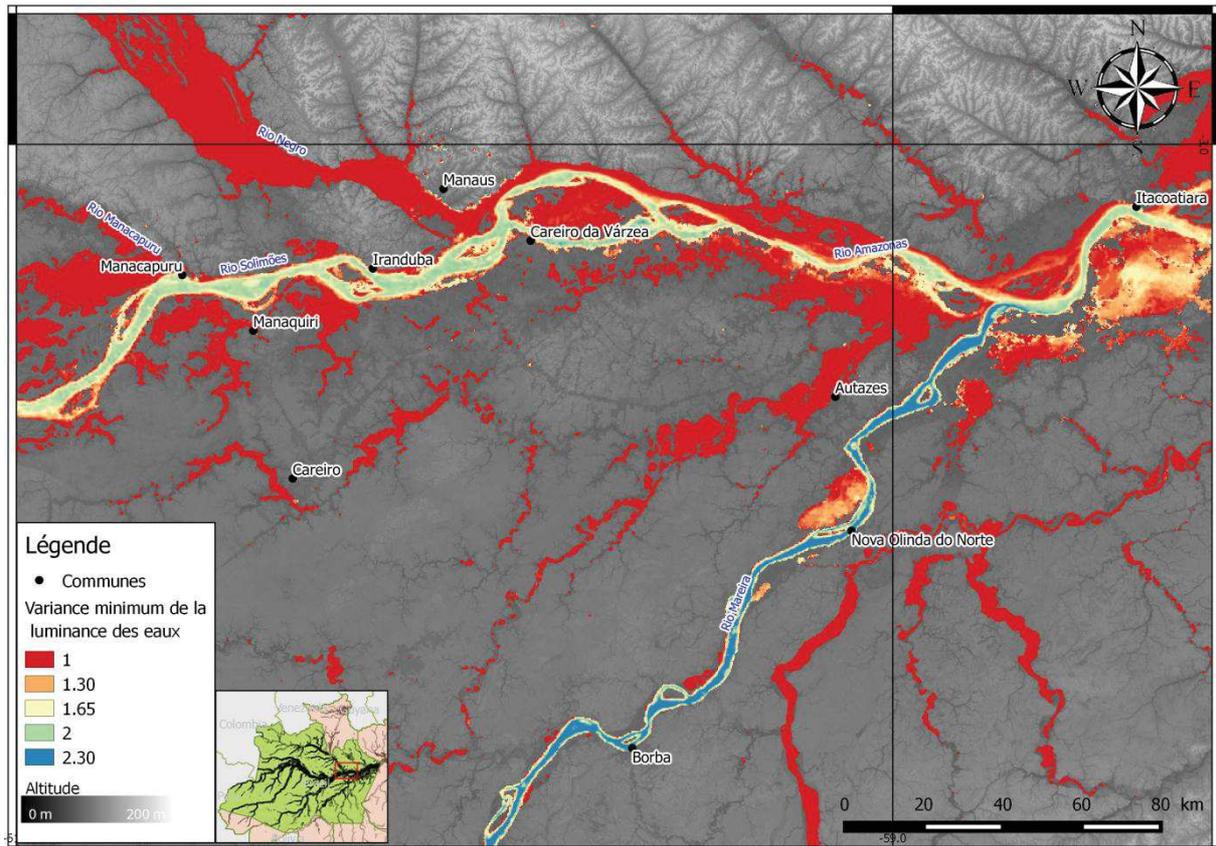


Figure 3-21 Variance de la luminosité des eaux à la confluence des fleuves Rio Negro; Rio Solimões et Rio Madeira

3.4 Conclusion Partielle

Cette étude a exigé plusieurs étapes de production de données et la définition de méthodologies appropriées aux contraintes écologiques et physiques locales ainsi qu'à l'échelle de traitement et à la profondeur temporelle de l'étude.

Les méthodologies développées peuvent être appliquées pour différents objectifs. Par exemple, une approche semblable a été réalisée pour l'Australie avec l'objectif de mesurer la fraction de l'eau pour tout le territoire à partir d'images MODIS pendant une période de 10 ans (Ticehurst et al. 2013; Ticehurst et al. 2014; Guerschman et al. 2011) ; l'enjeu étant l'identification de crues rapides à une échelle intra-mensuelle dans une région où il y a peu de nuages.

Le résultat de l'étude présentée ici est une cartographie mensuelle de l'intensité de la luminosité de l'eau en dix classes d'intensité échantillonnant la diversité des eaux entre les extrêmes « eau blanche » et « eau noire ».

Sur une analyse multi-temporelle basée sur la télédétection passive, la couverture nuageuse dans ces régions est très souvent une contrainte; L'utilisation des images MODIS permet de pallier cette contrainte. Cependant le flux d'images à traiter est considérable, et la chaîne de traitement est de ce fait rendue complexe par les contraintes, en particulier de couverture nuageuse.

L'approche proposée est basée sur des indicateurs classiques en télédétection pour la cartographie des eaux continentales, mais qui ont dû être revisités et traités de manière complémentaire pour renforcer le caractère reproductible de l'approche (Zhang et al. 2007; Sharma et al. 2015). Cette revisite a valu la création de l'indice SWIb proposé dans ce travail. L'indice SWI s'est en effet montré inefficace pour la reconnaissance des eaux turbides chargées en sédiments. Le SWIb s'est avéré plus adapté pour les réalités de l'état d'Amazonas.

L'intensité de la luminosité des eaux est traité ici comme un indicateur des modifications globales dans le système fluvial avec l'objectif de retrouver les grandes classes limnologiques et de les analyser dans le temps et dans l'espace.

Dans les résultats de la classification des niveaux d'intensité de l'indice de luminosité on retrouve le cycle saisonnier des hautes eaux et basses eaux caractéristique du bassin Amazonien et l'effet diluant des crues qui ont leur pic aux mois de juin-juillet ; en effet la crue, par une augmentation du débit liquide fait diminuer le débit solide. Au contraire les étiages voient une concentration des matières en suspension et ce qui fait croître l'intensité de la luminosité. Les fleuves dits « d'eau blanche » enregistrent pendant la période de crue des luminosités beaucoup plus faibles.

Cette constatation laisse supposer que l'indice de luminosité ne sera pas efficace pour séparer les eaux noires des eaux de crue « diluées » ou claires. Ce défaut de discrimination entre ces deux types d'eau n'a pas été anticipé, car il n'existe pas de fleuves dits « d'eau claire » dans l'état d'Amazonas, le représentant le plus connu de ce type d'eau étant le Rio Tapajos qui se trouve entièrement en dehors de l'état.

L'automatisation du traitement a permis de traiter une quantité considérable d'images sur une longue période. Cette automatisation peut permettre aussi la reproduction de l'ensemble du traitement avec des adaptations de configurations. Il serait souhaitable de produire d'autres indicateurs descriptifs des caractéristiques optiquement actives des eaux pour avoir une étude plus approfondie de la dynamique saisonnière et mensuelle des fleuves amazoniens. Cela permettrait de compléter les études sur l'implication de la qualité des eaux sur l'environnement au fil du temps et des événements climatiques pour finalement appliquer ces connaissances dans une étude d'environnement santé.

CHAPITRE 4 Le paludisme des rivières de l'État d'Amazonas – Brésil

4.1 Introduction

La variable environnementale étudiée dans cette thèse est l'intensité de la luminosité de l'eau utilisée comme un proxy de la qualité de l'eau, représentatif des grandes catégories d'eau présentes en Amazonie tel que décrit par Sioli (1968).

Pour résumer les études existantes, rappelées au chapitre 1, qui décrivent la relation entre couleur de l'eau et présence des moustiques, on peut rappeler:

- Que la majorité de ces études sont des études entomologiques.
- Que le principal vecteur du paludisme, *Anophèles darlingi*, est éminemment adaptable à son environnement (Tadei, B. D. Thatcher, et al. 1998; Tadei et al. 2016) ; mais que les niches écologiques semblent être principalement situées sur les plans d'eau semi-stagnants, partiellement ensoleillés et riches en matière organique (Consoli & Lourenço De Oliveira 1994; Katsuragawa et al. 2008; Carnevale et Robert Carnevale et Robert ..
- Qu'il a été noté une augmentation de l'incidence du paludisme (Cabral et al. 2010) dans le bassin du Rio Negro, malgré le fait que ces eaux ne soient pas réputées favorables au développement des larves du moustique, à la différence des fleuves d'eaux blanches.

Les développements méthodologiques exposés au chapitre précédent ont permis une cartographie dynamique, mensuelle, de l'intensité de la luminosité des eaux de surface de l'État d'Amazonas, indicateur de la transparence des eaux, du plus transparent ou noir (eau diluée ou chargée en matière organique) au plus réfléchissant (chargées en sédiments).

L'objectif de ce chapitre est d'analyser la relation entre les cas de paludisme et la luminosité des plans d'eaux de l'État d'Amazonas. Pour cela nous nous proposons de comparer au pas de temps mensuel sur onze années (2003 – 2014) l'incidence parasitaire de différents secteurs censitaires de l'État d'Amazonas et leur relation avec la luminosité des plans d'eau qui s'y trouvent.

L'État d'Amazonas possède un réseau de transport essentiellement fluvial. La grande majorité des lieux habités sont au bord des fleuves et des lacs. C'est aussi l'État où les données épidémiologiques du SIVEP-Malaria sont les plus complètes et les mieux géoréférencés comme le montre le *Chapitre 2* de cette thèse. On y trouve aussi les fleuves Rio Negro et Rio Solimões, principaux représentants des eaux noires et des eaux blanches du bassin Amazonien.

4.2 Matériel et Méthode

La méthode employée suit le schéma de la Figure 4-1 : dans un premier temps on obtient la spatialisation de l'incidence parasitaire mensuelle (IPM) par secteur censitaire en croisant les données épidémiologiques du SIVEP-Malaria et les données issues des recensements de population, mais aussi en découpant les zones peuplées de la région d'étude. Dans un deuxième temps on produit l'indice de luminosité des eaux pour, dans un troisième temps, produire l'étude de corrélation entre l'IPM et la luminosité des eaux par secteur censitaire.

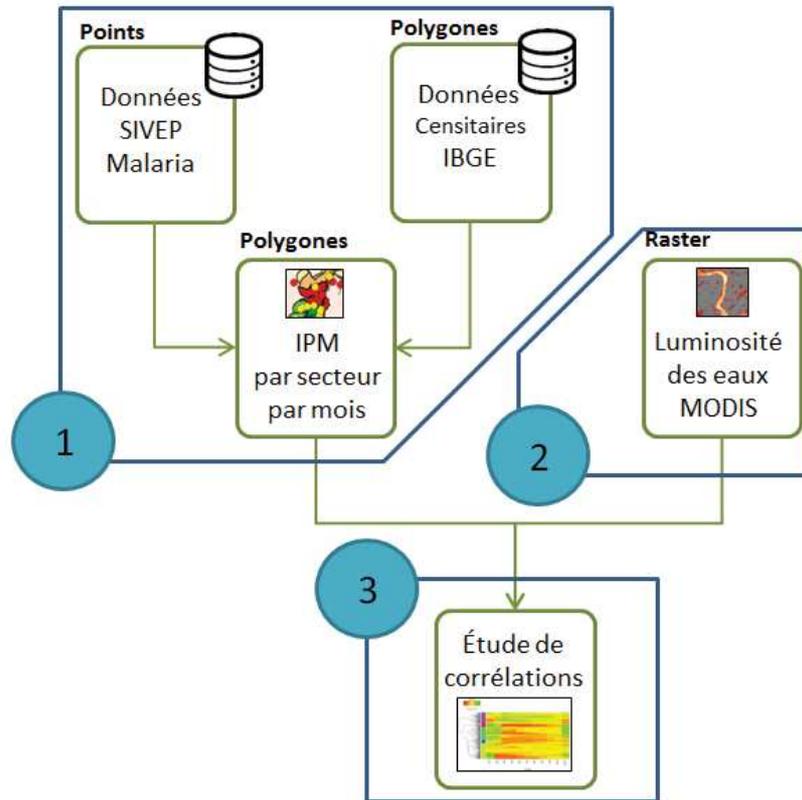


Figure 4-1 Schéma méthodologique des étapes du traitement des données.

4.2.1 Incidence parasitaire

Les données SIVEP-Malaria sont enregistrées par poste de santé, et représentent l'état de santé des populations habitant autour de ces postes. Les données par poste de santé sont plus riches et exactes que les données par lieu de résidence ou par lieu d'infection (voir *Chapitre 2*).

L'Incidence Parasitaire Annuelle (IPA) est l'indice épidémiologique le plus utilisé pour l'étude du paludisme au Brésil. Il s'agit d'une proportion entre le nombre de personnes infectées et le nombre d'habitants; une pondération qui produit une donnée plus stable et significative statistiquement.

Nous avons utilisé l'Incidence Parasitaire Mensuelle (IPM) (voir équation 4.1) pour caractériser les changements saisonniers (Barata 1995).

$$IPM = \frac{\textit{Cas de paludisme du mois}}{\textit{Population du mois}} \times 1000 \quad (4.1)$$

Bien que les données SIVEP-Malaria soient quotidiennes, l'IPM ne peut être calculée que pour des échelles spatiales et temporelles pour lesquelles les données démographiques sont également disponibles. Les données démographiques sont disponibles grâce aux recensements de population réalisés par l'Institut Brésilien de Géographie et Statistiques (IBGE) pour les années 2000, 2007 et 2010, produits selon le découpage territorial spécifique à la logistique de cette institution que sont les secteurs censitaires.

4.2.2 Les secteurs censitaires

Le secteur censitaire est l'unité spatiale élémentaire de recueil de données de l'Institut Brésilien de Géographie et Statistiques (IBGE). La taille du polygone est définie selon le nombre de domiciles qu'un agent est capable de recenser en un mois, soit environ 300 domiciles pour les zones urbaines. On l'utilise ici pour connaître le nombre d'habitants à l'échelle la plus fine afin de calculer l'IPM.

Le découpage de ces secteurs évolue selon les évolutions démographiques. Le nombre de secteurs s'est accru au fil des années comme le montre le Tableau 4-1 pour l'État d'Amazonas.

Tableau 4-1 Nombre de secteurs de recensement par an et par type pour l'État d'Amazonas

Année	2000	2007	2010
Urbain	1942	2124	3603
Rural	1280	1824	2581
Total	3222	3948	6184

Suivant cette évolution, les secteurs de 2000 ont été fractionnés lors des recensements suivants. Il était nécessaire de conserver la même base spatiale sur la profondeur temporelle de notre analyse. Pour cette raison, il a été nécessaire de regrouper les populations des secteurs censitaires de 2007 et de 2010 à l'intérieur des limites des secteurs de 2000.

Les données censitaires sont disponibles sous forme de couches vectorielles de polygones géoréférencés. Premièrement les couches pour les années 2000 et 2007 ont été fusionnées en gardant les attributs « GEOCODE », « Population » et « Surface » des polygones. Le résultat est la division des polygones de 2007 selon les intersections avec les polygones de 2000 pour affecter la population des polygones de 2007 fragmentés sur les contours des polygones 2000. Les règles suivantes ont été appliquées dans cette opération:

- Les polygones de 2007 entièrement contenus dans les contours d'un polygone 2000 ont été sommés;
- Les polygones de 2007 fragmentés sur plusieurs (au moins 2) polygones de 2000 ont été affectés au polygone de 2000 avec lequel ils partageaient la plus grande surface.

Le processus a été ensuite répété entre 2000 et 2010. A la fin de ce traitement nous obtenons les populations pour les 3 années sur une même base de polygones.

La donnée de population des années non recensées a été estimée linéairement entre 2000 et 2007 et aussi entre 2007 et 2010. La population des années entre 2011 et 2014 a été estimée par régression linéaire à partir des données des années de 2007 à 2010. Le résultat est illustré dans le Tableau 4-2.

Tableau 4-2 Echantillon de l'estimation de la population pour les années non recensées par Secteur Censitaire

SC	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
1	348	303	259	214	169	124	80	35	412	789	1166	1543	1920	2297	2674
2	962	1052	1142	1232	1323	1413	1503	1593	1918	2244	2569	2894	3220	3545	3870
3	1211	1068	924	781	638	495	351	208	501	794	1087	1380	1673	1966	2259
4	2027	1748	1469	1190	910	631	352	73	185	298	410	522	635	747	859
5	79	173	268	362	457	551	646	740	2048	3356	4664	5972	7280	8588	9896
6	79	173	268	362	457	551	646	740	2048	3356	4664	5972	7280	8588	9896
7	142	128	114	100	85	71	57	43	845	1646	2448	3250	4051	4853	5655
8	36	158	281	403	526	648	771	893	1031	1168	1306	1444	1581	1719	1857
9	450	921	1392	1863	2334	2805	3276	3747	4189	4632	5074	5516	5959	6401	6843
10	1386	1214	1042	870	698	526	354	182	606	1031	1455	1879	2304	2728	3152
11	135	204	272	341	409	478	546	615	744	873	1002	1131	1260	1389	1518
12	464	563	663	762	862	961	1061	1160	2374	3589	4803	6017	7232	8446	9660

Obs : Les années de recensement sont 2000, 2007 et 2010 (en gris).

4.2.3 Les zones habitées

Les secteurs censitaires de l'État d'Amazonas peuvent avoir une superficie dépassant 40.000 km². Les zones habitées ont été définies selon la méthode du projet Global Human Footprint (Wildlife Conservation Society 2005) (Figure 4-2) de la NASA comme étant celles se trouvant dans un buffer de 15Km autour des axes de communication, soit ici les rivières.

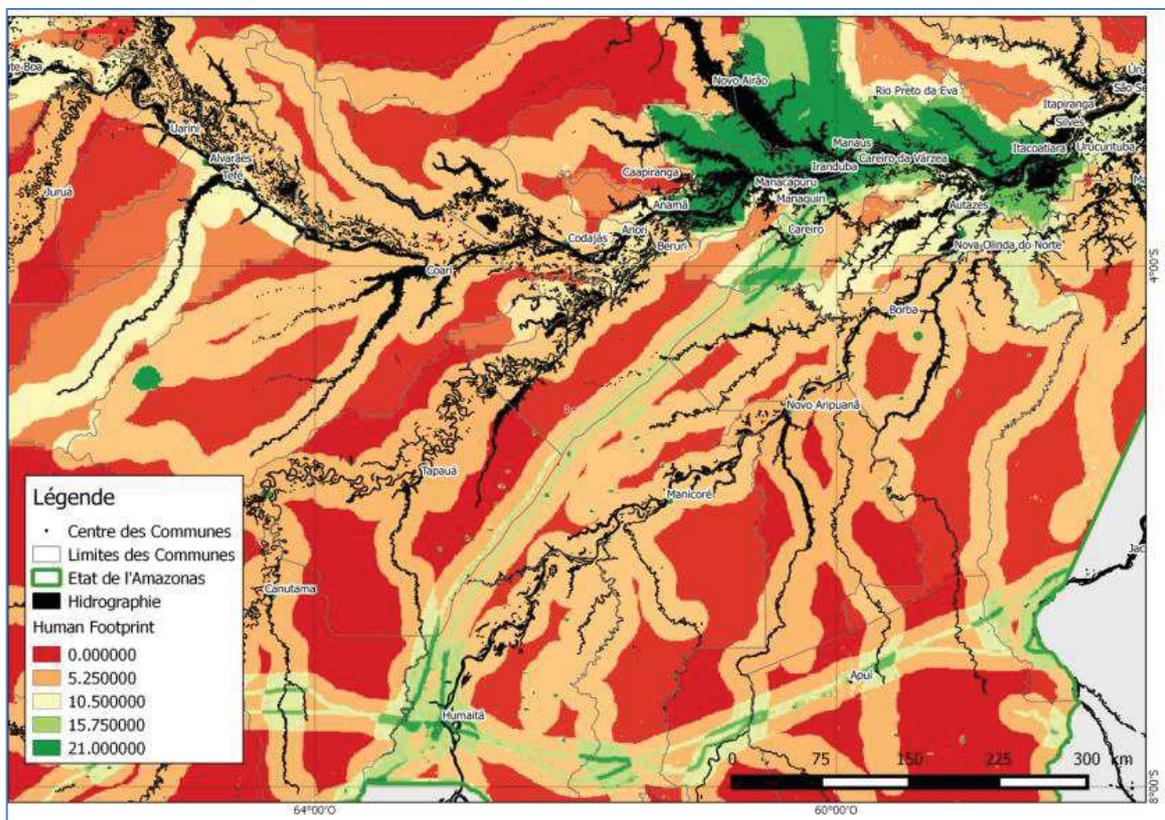


Figure 4-2 Produit du projet Global Human Footprint de la NASA pour la région centrale de l'État d'Amazonas

Le buffer de 15 km a été calculé à partir de la cartographie de la surface maximale des eaux réalisé dans le *Chapitre 3*. De cette manière les eaux présentes dans les vastes étendues non habitées des secteurs censitaires, lointaines des habitations, ne

sont pas comptabilisées. Une grande partie de la surface de l'État d'Amazonas se trouve à moins de 15km d'une rivière, particulièrement le Nord-Ouest caractérisé par la présence de nombreux lacs et petites rivières (Figure 4.3).

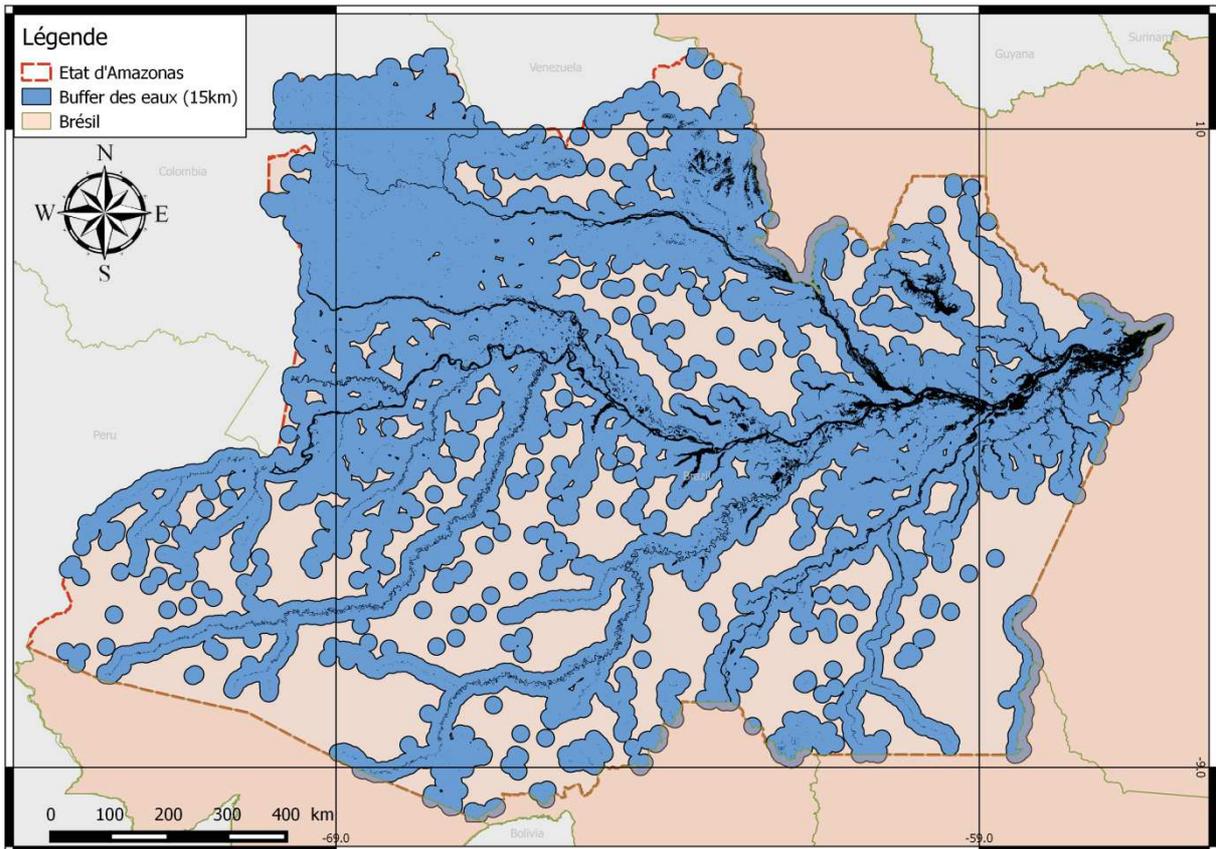


Figure 4-3 Buffer de 15 km des marges des plans d'eau pour l'État d'Amazonas

4.2.4 Les données épidémiologiques du SIVEP-Malaria

Le *Chapitre 2* de cette thèse a mis en évidence que la localisation géographique des notifications de la base de données SIVEP-Malaria la plus cohérente est celle du poste de santé plutôt que celle du lieu de résidence ou du lieu déclaré d'infection du patient. La méthodologie employée ici a suivi les recommandations faites au chapitre 2 afin d'améliorer la cohérence des données SIVEP-Malaria.

Cette mise en cohérence des données a utilisé plusieurs filtres successifs:

1) population d'intérêt. Les notifications réalisées dans l'État d'Amazonas ont été extraites avec la variable « État de notification » (UF_NOTIF).

2) date : notifications entre les années 2003 à 2015 à partir de la variable « *date de notification* » (DT_NOTIF).

On a ainsi obtenu un total de 11.445.939 notifications.

3) résultats positifs de paludisme : élimination des cas suspects à partir de la variable « *résultat d'examen* » (RES_EXAM).

4) unicité : conservation exclusive des notifications issues des premières consultations des patients en utilisant la variable « *consultation de suivi* » (ID_LVC).

Ces notifications ont été ensuite organisées par mois et par poste de santé. Les postes de santé (UNIN_ID) n'ayant pas de coordonnées géographiques ont été exclus car inutilisables.

Voici en résumé l'enchaînement des filtres sur la base de données.

a) UF_NOTIF = 13 (Amazonas)

b) DT_NOTIF < 2003-01-01 & DT_NOTIF < 2015-01-01

- c) RES_EXAM != 1 (Cas positifs) *
- d) ID_LVC = 2 (n'est pas un LVC)
- e) LONGITUDE != "NA" OR LATITUDE != "NA"

* "!=" équivaut à "différent de"

A la fin de cette démarche il est resté 1.411.327 notifications à analyser. Le formatage des données résultantes de ce procédé est illustré dans le Tableau 4.3 qui montre le comptage des cas de paludisme par poste de santé (UNIN_ID) et par mois. Ces données serviront à produire l'indice « Incidence Parasitaire de Paludisme Mensuelle » (IPM).

Tableau 4-3 Echantillon des données résultantes de l'enchaînement des filtres et de la fusion avec les modules de géoréférencement.

LOC_ID	UNIN_ID	2003-01	2003-02	2003-03	...	2014-10	2014-11	2014-12	Long	Lat
13017098	13017071	0	0	0	...	3	1	0	-62,238	-6,276
1301851	13018534	0	0	0	...	1	0	0	-60,083	-3,168
1301851	1301851	33	36	183	...	27	20	9	-60,083	-3,168
13018510	1301853	0	0	4	...	11	13	7	-60,220	-3,214
130185100	13018522	7	39	86	...	28	30	13	-60,181	-3,271
130185108	13018517	0	0	0	...	0	0	0	-60,198	-3,197
130185109	1301852	0	2	17	...	0	0	2	-60,288	-3,167
130185114	13018520	0	5	1	...	7	4	0	-60,316	-3,125
13018514	13018519	3	1	4	...	5	1	0	-60,283	-3,196

Loc_ID est l'identifiant du point SIVEP ; UNIN_ID est l'identifiant du poste de santé.

4.2.5 Croisement et traitement des données

Pour chaque zone habitée de chacun des secteurs censitaires possédant au moins un poste de santé, les indicateurs mensuels de luminosité des eaux ont été croisés

avec l'IPM. Les zones habitées des secteurs censitaires concordant avec la localisation des postes de santé sont illustrées dans la carte de la Figure 4-4.

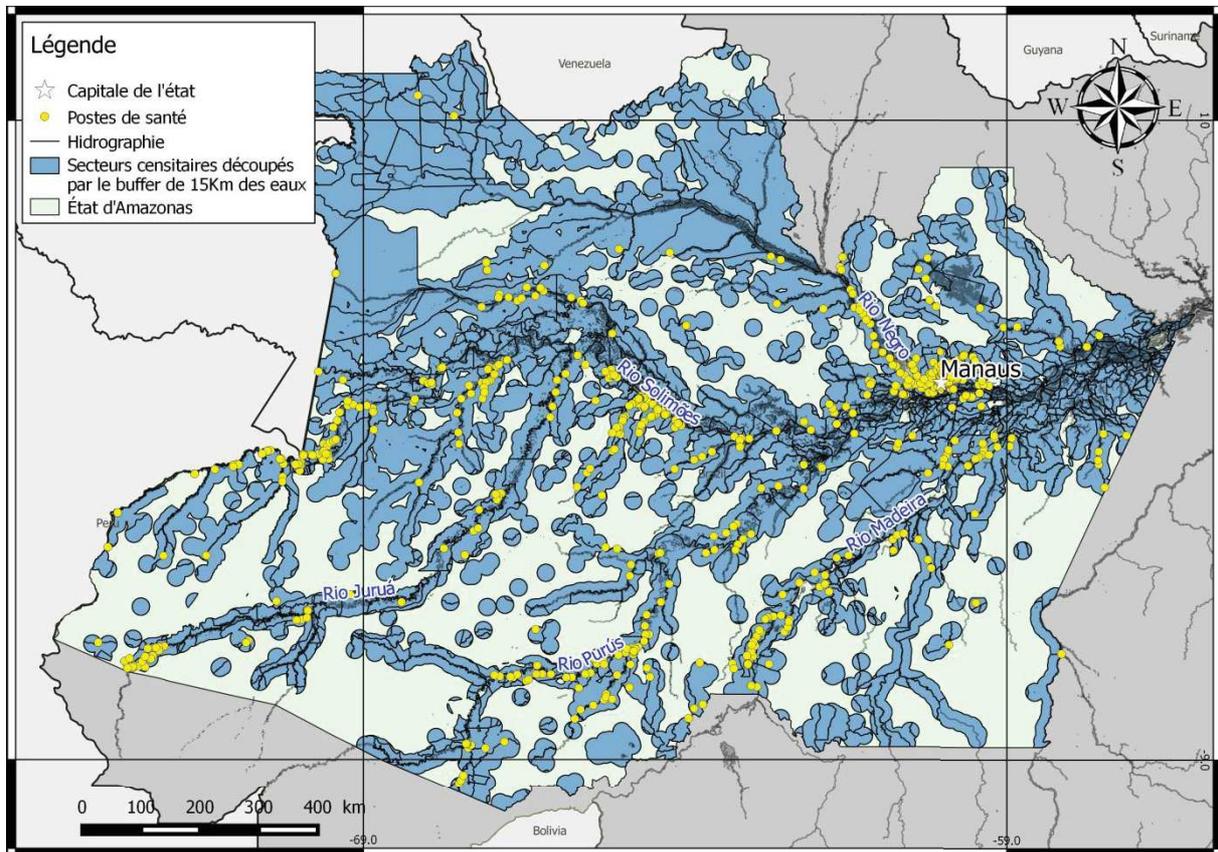


Figure 4-4 Secteurs censitaires découpés par le buffer des eaux et postes de santé se trouvant dessus

Sur les 1130 postes de santé de l'État d'Amazonas, 976 ont enregistré des cas de paludisme sur la période et sont localisés dans les zones habitées telles que définies plus haut.

La carte de la Figure 4-5 présente les zones d'étude qui sont le croisement entre zones habitées, secteurs censitaires et poste de santé ayant notifié des cas de

paludisme sur la période. L'IPM moyen de chaque zone sur la période de 2003 à 2014 a été porté sur la figure sous la forme d'une légende colorée.

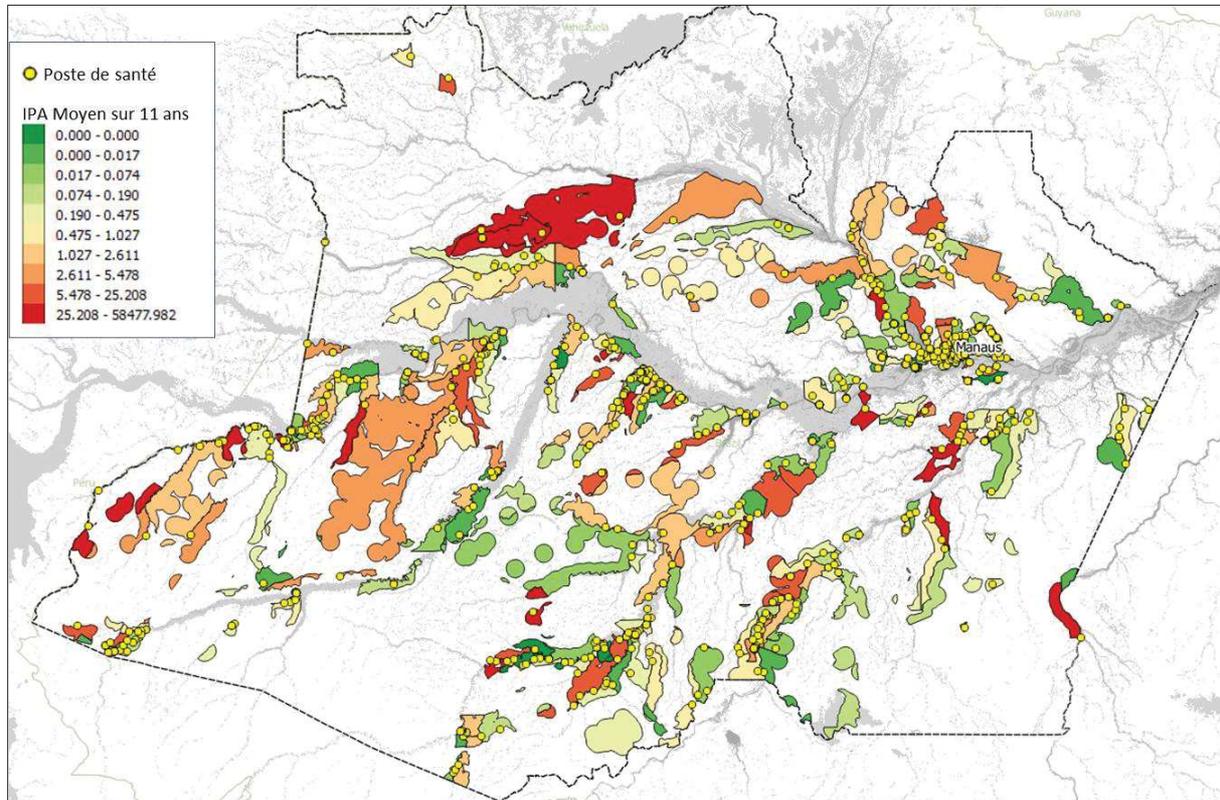


Figure 4-5 IPM moyen des zones d'étude et postes de santé à l'origine de la donnée épidémiologique

Pour utiliser la donnée épidémiologique comme un indicateur du risque d'exposition au paludisme il faut considérer le temps moyen qui s'écoule entre la présence du moustique, le moment de la pique, l'apparition des symptômes et enfin la notification auprès du poste de santé.

En zones tropicales, la phase aquatique des anophèles dure de 1 à 3 semaines et on considère que la longévité moyenne de l'adulte est de 3 à 4 semaines (Carnevale & Robert 2009). L'espace de temps entre la pique et l'apparition des premiers

symptômes (période d'incubation) est de 8 à 12 jours pour le *P. falciparum* et de 13 à 17 jours pour le *P. vivax* (Ávila & Ferreira 2013; Brasil 2014; Hiwat et al. 2011; Girod et al. 2011). On sait aussi, par l'analyse des données SIVEP-Malaria, que le nombre de jours entre l'apparition du symptôme chez le patient et la notification au poste de santé peut atteindre 10 jours (Figure 4-6). Pour ces raisons et afin de réaliser les calculs de corrélations, les données mensuelles de l'IPM ont été décalées d'un mois en arrière, réduisant ainsi la différence de temps entre l'évènement environnemental déclencheur et l'évènement épidémiologique résultant.

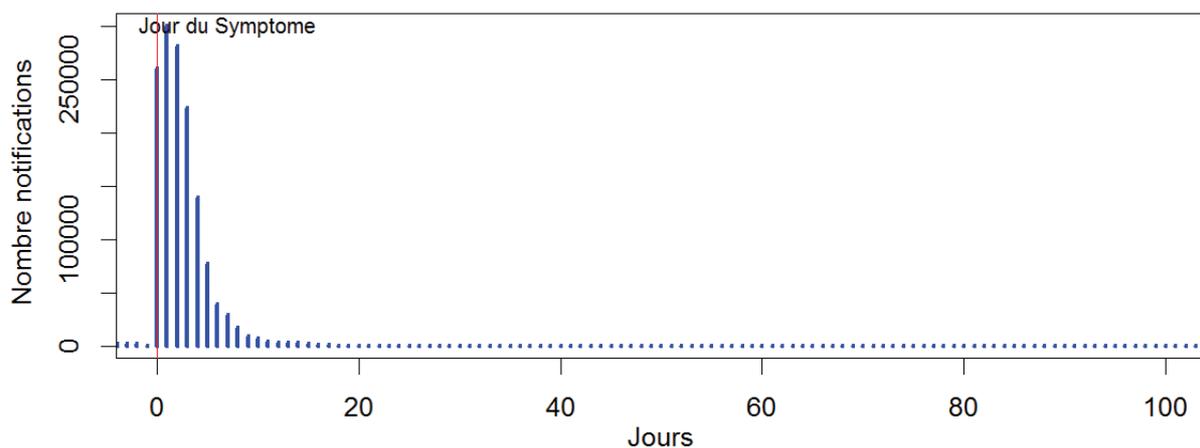


Figure 4-6 Temps entre les symptômes de paludisme et la notification au poste de santé

4.3 Résultats

4.3.1 IPM moyen

La figure 4-7 présente la localisation géographique de l'IPM moyen sur la période 2003-2014. Les grandes zones d'IPM moyen supérieur à 2.6 (Figure 4-7) sont :

Zones 1, 2 et 3 : Sur le Rio Negro, en rive droite, respectivement 1. sur le rio Unini, 2. dans une zone comprise entre le Rio Negro et le Rio Cuiuni (IPM compris entre 2,61

et 5,47) et 3. une zone comprise entre les Rio Uneuixi et Tea , avec un IPM moyen le plus élevé de l'état d'Amazonas (supérieur à 25,20)

Zone 4 : En rive droite du Rio Solimões, au long des fleuves Rio Jandiatuba et Rio Jutai (IPM moyen entre 2,0 et 145,0)

Zone 5 : Au long du Rio Curuçã et du Rio Javari, à Atalaia do Norte (IPM moyen de 4859), et à la frontière avec le Pérou, on trouve des secteurs avec IPM moyen entre 28 et plus de 4000.

Zone 6 et 7 : A l'amont du Rio Purus au sud de Labrea (IPM entre 5 et 25) et à l'aval, aux alentours de la ville de Tapauá. (IPM moyen entre 3,0 et 35,0)

Zone 8 et 9 : Deux zones également d'IPM élevé sur le rio Madeira sur le cours moyen au sud de Manicore et vers Borba, entre le Rio Madeira et son affluent Rio Preto do Igapó-açu (IPM moyen entre 3,0 et 2146,0)

Zones 10 et 11 : Sur le rio Uatuma et son affluent le Rio Pitinga, la zone de la retenue de Balbina et celle de la mine Pitinga

Ces différentes zones sont réparties aussi bien sur des fleuves d'eau blanche que sur des fleuves d'eau noire.

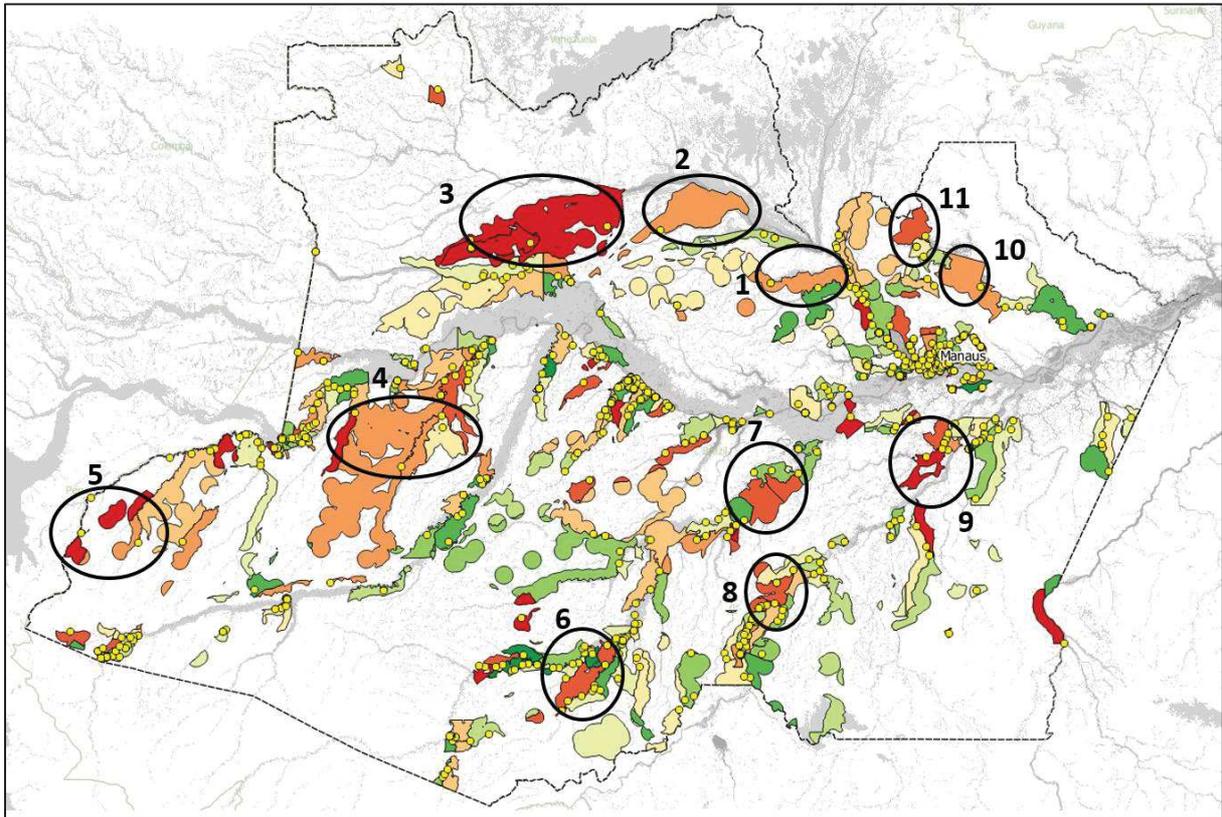


Figure 4-7 Les secteurs à IPM moyen supérieur à 2.6

4.3.2 Corrélation entre IPM et indicateur de luminosité

Le coefficient de corrélation Spearman des signatures temporelles de ces facteurs pour chaque zone d'étude et pour chacune des dix classes de luminosité des eaux est utilisé pour analyser la relation entre la luminosité des eaux et l'IPM.

Une première analyse a été faite sur les similitudes de comportement entre classes de luminosité pour tout l'État d'Amazonas.

Ensuite deux analyses de corrélations ont été effectuées avec deux approches différentes:

- La première approche est la corrélation des IPM et des indicateurs mensuels de luminosité pour chacune des années.

- La deuxième approche est la corrélation entre moyenne mensuelle des IPM et moyenne mensuelle des classes de luminosité de l'eau sur toute la période (moyenne des mois de janvier sur la période, moyenne des mois de février,.....).

4.3.3 Analyse des corrélations entre IPM et classes de luminosité pour tout l'État d'Amazonas

Pour déterminer les similitudes de comportement entre les classes de luminosité des eaux sur l'État d'Amazonas en regard de leur corrélation avec l'IPM, une analyse factorielle en composantes principales a été réalisée avec la production d'un cercle de corrélations entre IPM et réflectance de l'eau (Figure 4-8) permettant d'afficher la distance de corrélation entre les comportements mensuels des classes de luminosités de l'eau de 2003 à 2014.

Les deux premiers axes du nuage de corrélations portent 61,99% de la variance totale, le premier axe totalisant plus de 40% de la variance et le deuxième moins de 20%.

Dans le quadrant des corrélations positives sur les deux axes principaux entre luminosité et IPM se regroupent les classes de luminosité de 7 à 10 (fortes luminosités), indiquant, bien que ces classes soient faiblement représentées en effectif (Figure 3-16) une bonne corrélation avec l'IPM au niveau de l'État.

Dans le quadrant des corrélations positives sur le premier axe et négatives pour le second se regroupent les classes de luminosité 3 à 6 (luminosités faibles à moyennes)

Dans le quadrant des corrélations négatives sur le premier axe et positive sur le second, se regroupent la variable eau totale et la classe de luminosité 1.

Enfin dans le quadrant des corrélations négatives sur les deux axes, on trouve la classe de luminosité 2, avec cependant une signification faible (corrélation faible aplatie sur l'axe).

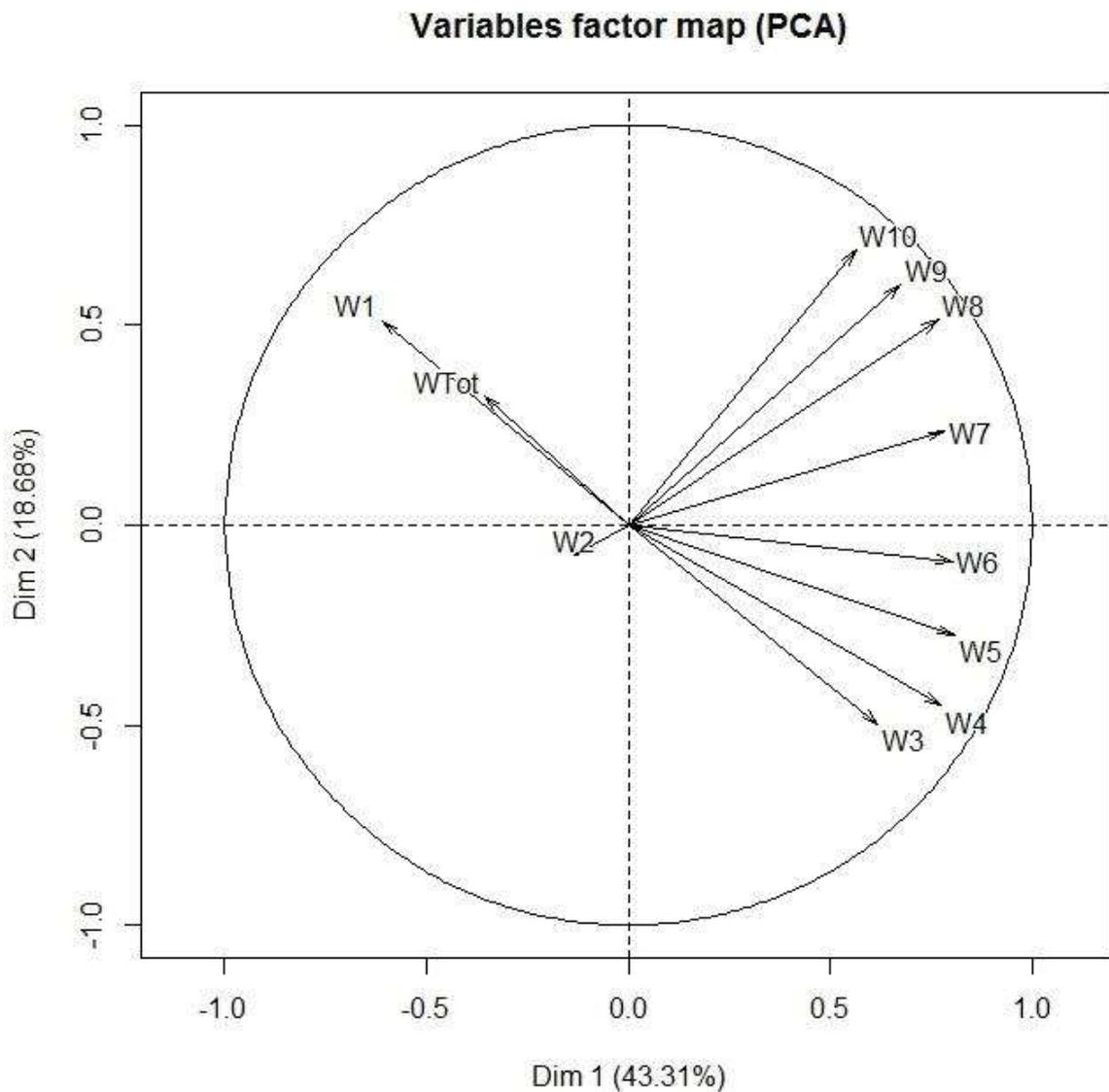


Figure 4-8 Cercle des corrélations entre classes de luminosité des eaux et IPM pour la zone d'étude avec le coefficient de corrélation le plus fort dans l'analyse de corrélation entre 2003 et 2014

Si on considère seulement le premier axe de la variance, on note :

a) L'association de la variable eau totale (WTot), dont on peut faire l'hypothèse qu'elle représente l'eau de crue, à la variable luminosité 1 (W1), permet d'assimiler la classe de luminosité 1 aux eaux claires, diluées par la crue, ces deux variables étant corrélées négativement avec l'IPM sur le premier axe, positivement sur le second ;

b) une opposition entre ces deux variables associées à la crue et toutes les autres variables

Sur le deuxième axe de la variance on note l'opposition entre la variable eau totale et classe de luminosité 1 d'une part et d'autre part les classes de réflectance faible à moyenne, ce qui permet de penser que ces luminosités moyennes à faibles sont plutôt associées à l'étiage.

Pour aller plus loin dans l'analyse et vérifier ces premières inférences, il est nécessaire de caractériser spatialement et temporellement ces corrélations.

1.1.1.1 Corrélation mensuelle pour chacune des années entre 2003 et fin 2014 par zones

Une classification ascendante hiérarchique a été appliquée pour analyser les comportements de corrélation similaires et permettre ensuite de les localiser dans l'espace.

Cette classification a été réalisée en utilisant la fonction « heatmap2 » sous R.

La matrice de distances euclidiennes est donc représentée sous la forme d'une carte de chaleur pour laquelle le résultat de la classification hiérarchique ascendante figure

en ordonnée (Figure 4-9). Les corrélations varient principalement entre -0,3 (vers le rouge) et + 0,4 (vers le vert) pour chacune des 10 classes de luminosité plus une classe eau totale toutes classes confondues (en abscisse). D'une façon générale les corrélations mensuelles sont très faibles entre la luminosité de l'eau et l'IPM. Il semble que la variabilité interannuelle soit trop grande pour pouvoir tirer des conclusions sur ces corrélations. En particulier la diminution spectaculaire du nombre de cas sur la période (figure 2.1) peut rendre ces corrélations compliquées à établir et analyser. La figure

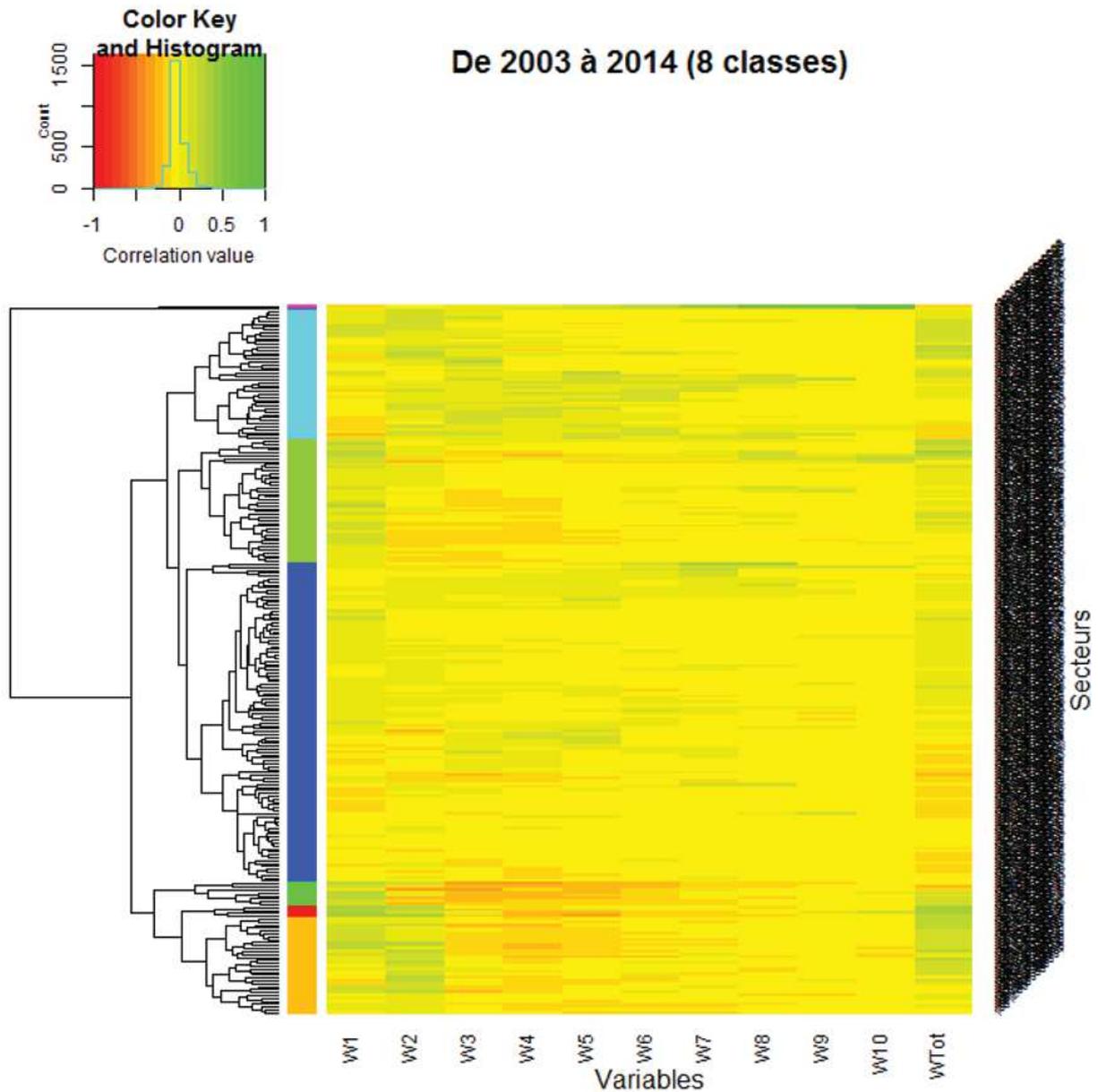


Figure 4-9 Carte de chaleur du regroupement hiérarchique des corrélations entre les classes de réflectance des eaux et l'IPM par zone d'étude. (Obs: les lignes constituent les différentes zones d'étude - secteurs)

4.3.4 Corrélation entre IPM moyen sur la période et moyennes mensuelles des classes de luminosité

La même analyse de corrélation a été faite sur les climatologies plutôt que sur les moyennes mensuelles (moyennes mensuelles sur tous les mêmes mois de la

période). Cette fois les coefficients de corrélations Spearman sont plus robustes et mieux distribués entre 1 et -1.

La carte de chaleur représentant la matrice de distances euclidiennes avec le résultat de la classification hiérarchique ascendante en ordonnée est représentée (Figure 4-10).

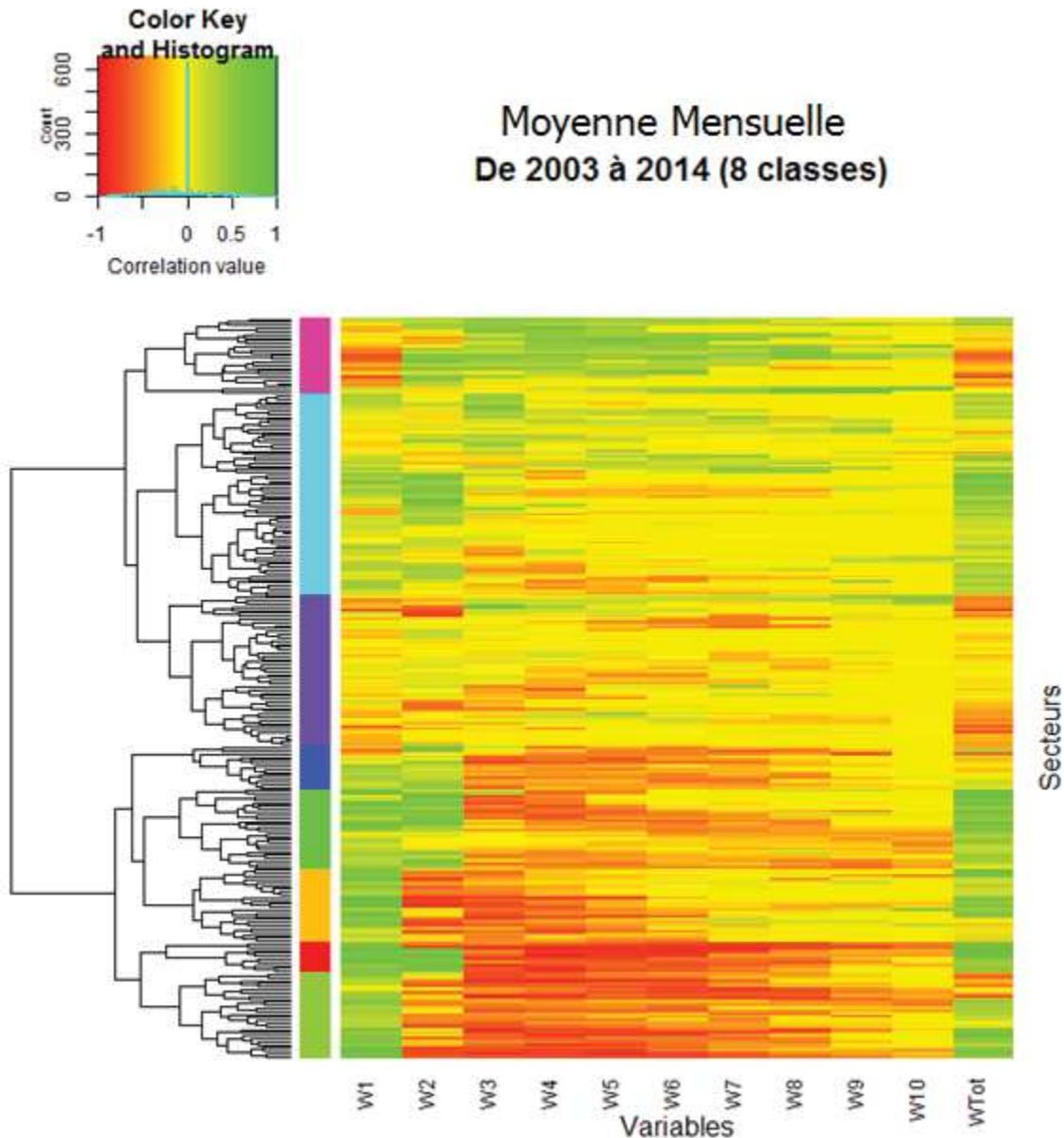


Figure 4-10 Carte de chaleur du regroupement hiérarchique des corrélations entre la moyenne mensuelle des classes de luminosité des eaux et celle de l'IPM par zone d'étude. (Obs: les lignes constituent les différentes zones d'étude)

La classification ascendante hiérarchique, sur les mêmes principes que précédemment, a permis d'identifier 8 classes selon leur similitudes ou différences de corrélations positives ou négatives entre les 10 variables de luminosité + la variable eau totale et l'IPM moyen. On peut distinguer :

Les classes 1 et 4 qui présentent une corrélation positive forte entre IPM et les variables eau totale, 1 et 2, et une corrélation fortement négative pour toutes les autres variables (corrélation négative un peu plus faiblement sur les variables de luminosité élevée pour la classe 4)

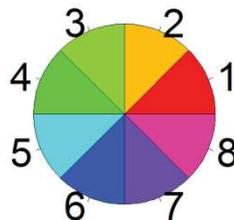


Figure 4-11 Classes obtenues en résultat de la classification hiérarchique

- Les classes 2 et 3 qui présentent une corrélation positive forte entre IMP et la variable 1 ; la classe 2 présente une corrélation négative forte avec les variables 2 à 6 tandis que la classe 3 présente une corrélation négative forte avec toutes les variables 2 à 9.

- La classe 5 présente des corrélations beaucoup plus variables : une dominance de corrélation positive entre IPM et la variable eau totale et également les variables 1 et 2, sans que cette corrélation positive soit constante pour toute la classe, et une distribution variable de corrélations moyennement positives ou négatives sur les autres variables ;

- La classe 6 présente une corrélation faible ou négative entre IPM et la variable eau totale, une corrélation positive avec les variables 1 et 2, une forte corrélation négative avec tous les autres canaux

- La classe 7 présente une anticorrélation de l'IPM avec toutes les variables, parfois faiblement positives pour les variables 2 à 10 mais toujours neutre à négative avec les variables 1 et eau totale.

Enfin la classe 8 présente une corrélation positive forte entre l'IPM et les variables 2 à 5 et une corrélation négative forte avec les variables eau totale et 1.

La localisation géographique des différentes classes est présentée Figure 4-12

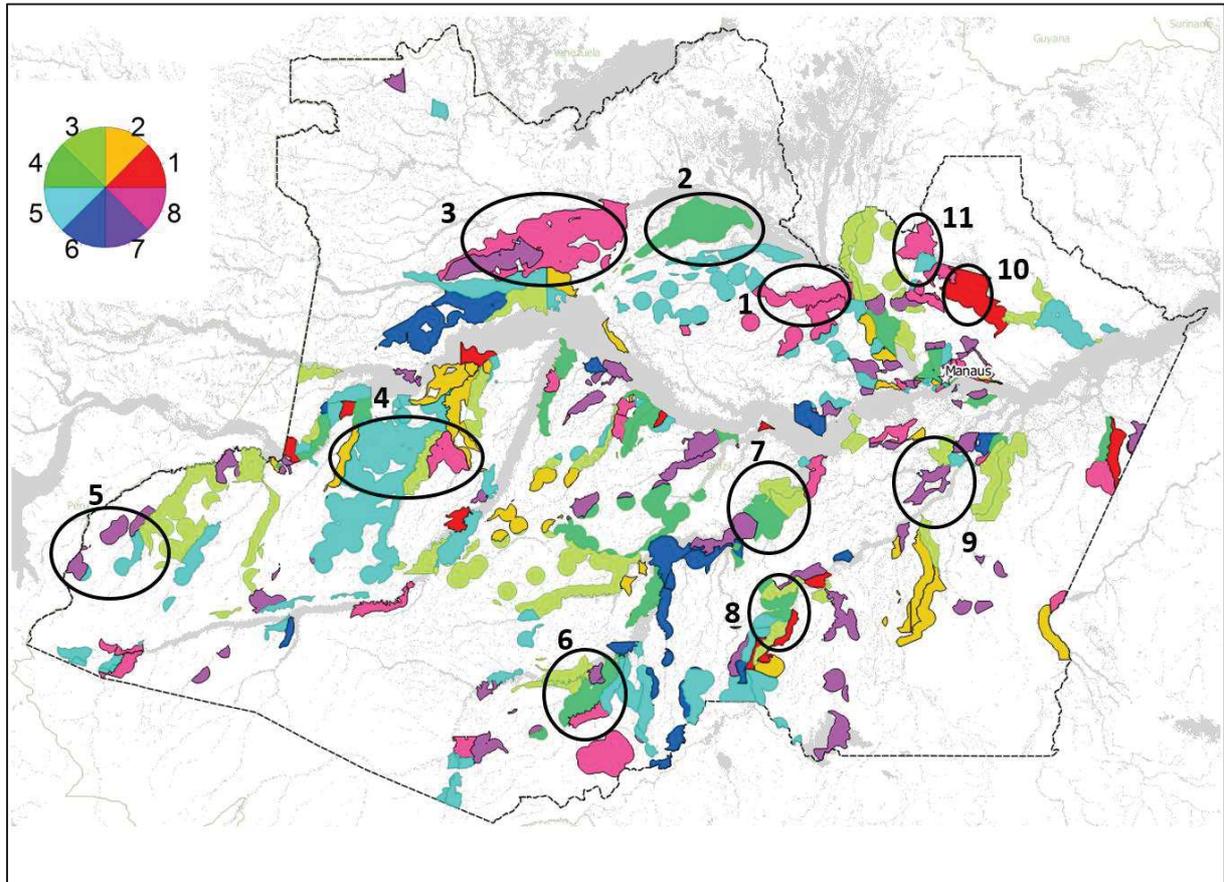


Figure 4-12 Cartographie des classes obtenues par classification hiérarchique des corrélations entre les moyennes mensuelles de l'IPM et celles des classes de luminosité des eaux

Sur cette figure 4-12, on a reporté les zones d'IPM moyen élevé décrites plus haut. Pour analyser chacune des classes de corrélation entre l'IPM et les variables eau totale et luminosité de l'eau, on a analysé

1 : La localisation géographique où ces classes se trouvent représentées lorsqu'elles l'IPM est élevé (Figure 4-12)

2. La temporalité de la variable eau totale et des variables de luminosité associée à celle de l'IPM sur des exemples pris dans chacune des zones considérées.

Les résultats de cette analyse sont les suivants :

- La classe 6 n'est représentée dans aucune des zones d'IPM élevé.
- La classe 1 est représentée avec un IPM élevé sur le Rio Uatuma dans la zone du barrage de Balbina (Figure 4-13)

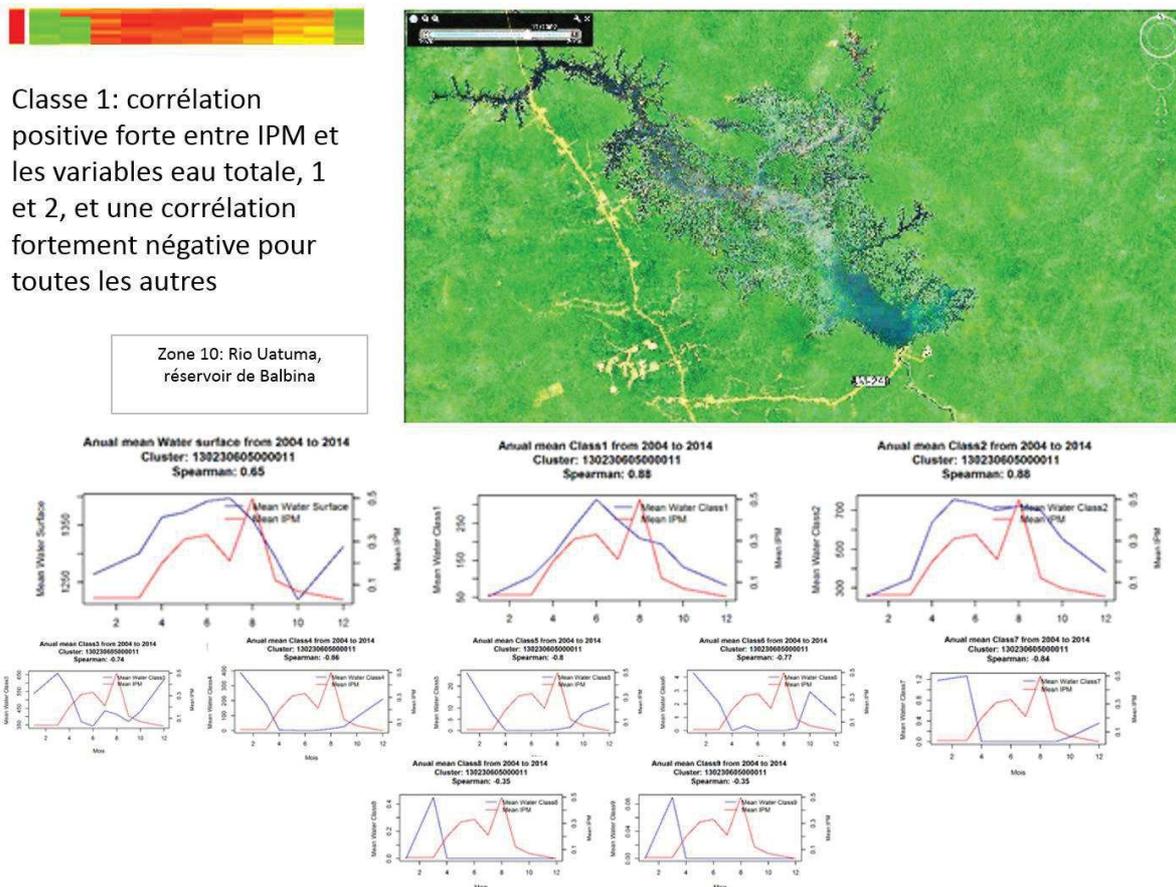


Figure 4-13 : Analyse de la classe de corrélation 1

L'IPM y est bimodal avec un premier pic centré sur la montée de crue en Mai Juin et un second plus important en décrue (Août). Il s'agit d'une zone d'eau noire, peuplée (présence de la BR 174 qui joint Manaus au Vénézuéla) caractérisée par une très grande zone d'inondation sous forêt. L'analyse temporelle confirme bien la bonne corrélation avec la variable eau totale et la corrélation négative avec toutes les autres variables. Les meilleures corrélations temporelles correspondent à la variable 1

(plutôt avec le pic d'IPM de montée de crue) et la variable 2 (sur les 2 pics) (corrélation de 0.88) sur les deux variables.

- L'analyse de la classe 4 montre une localisation géographique de la classe coïncidant avec un IPM élevé pour la zone 6 (amont du Rio Purus au sud de Labrea (IPM entre 5 et 25) et la zone 2 (zone comprise entre le Rio Negro et le Rio Cuiuni - IPM compris entre 2,61 et 5,47); ces zones sont donc caractérisées par une corrélation positive forte entre l'IPM et les variables eau totale, 1 et 2, et une corrélation négative pour toutes les autres variables.

Ces deux zones sont respectivement une zone d'eau blanche (Rio Purus) et une zone d'eau noire (Rio Negro). Le point commun entre les deux zones est qu'elles sont caractérisées par de grandes étendues d'inondation sous forêt.

La variation temporelle de l'IPM moyen et des différentes variables pour ces zones de classe 4 est également présentée figure 4-14. L'IPM comme pour la classe 1 a une distribution bimodale en mai et en juillet correspondant à la montée de crue et à la période qui initie la descente de crue. On note bien le comportement temporel de la variable eau totale qui suit la crue, enfin les corrélations fortes avec les variables d'eau sombre ou diluée par la crue 1 et 2. L'anti-corrélation est nette pour les variables 4 et 5.

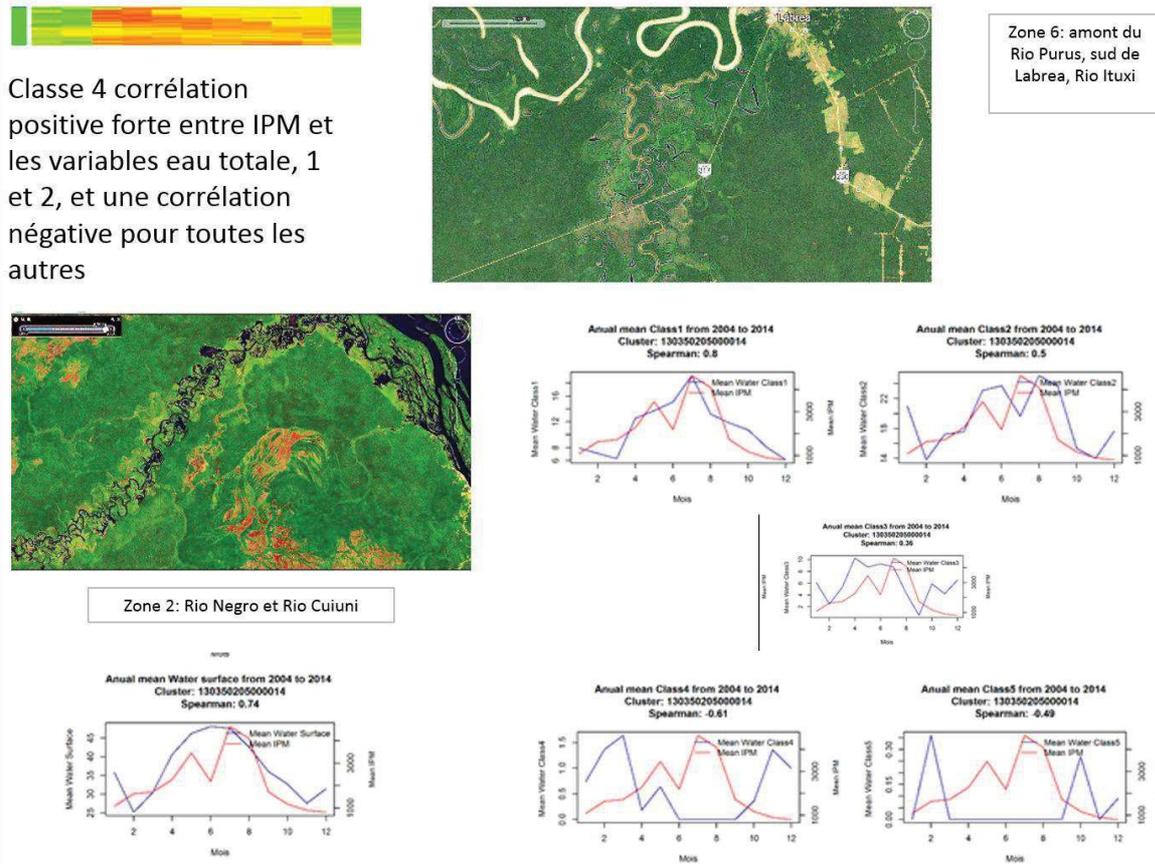


Figure 4-14 : analyse des corrélations pour la classe 4

La classe 3 est représentée dans la zone 7, à l'aval du Rio Purus aux environs de Tapauà (Figure 4-15). Cette zone est une zone d'eau blanche, néanmoins comme on peut le voir sur la copie d'écran Google Earth de la zone, de nombreux affluents sont des rivières d'eau noire.

L'IPM a un comportement temporel particulier ; pratiquement étalé sur toute l'année, avec trois pics, l'un en Avril, l'autre en Juin Juillet et le dernier en Septembre Octobre. Dans l'exemple choisi, l'eau totale et la variable ont pratiquement le même comportement, ce qui n'est pas le cas partout (voir carte de chaleur). Le comportement de la variable 1 est très proche de celui de l'IPM avec des maximums locaux correspondant aux pics d'IPM.

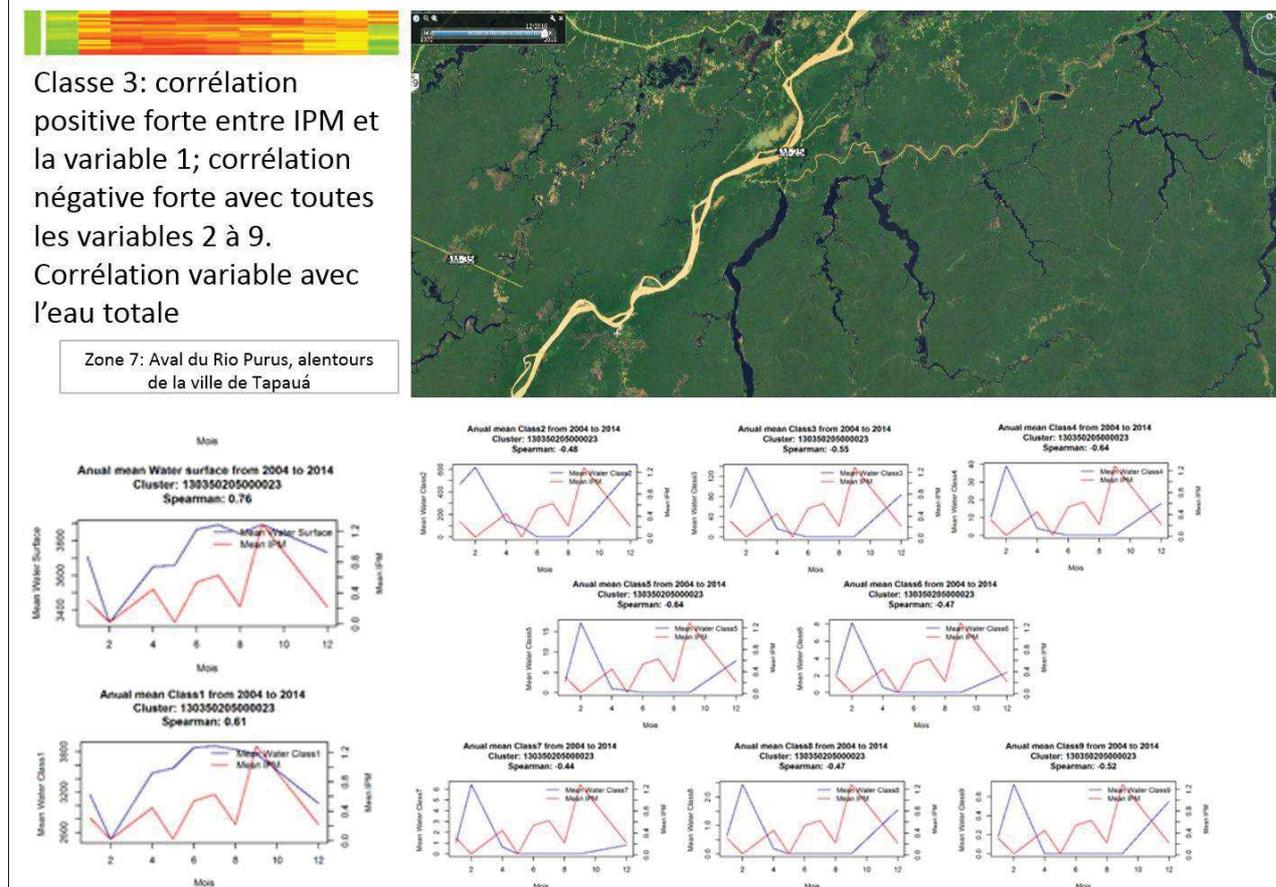


Figure 4-15 : Analyse des corrélations pour la classe 3

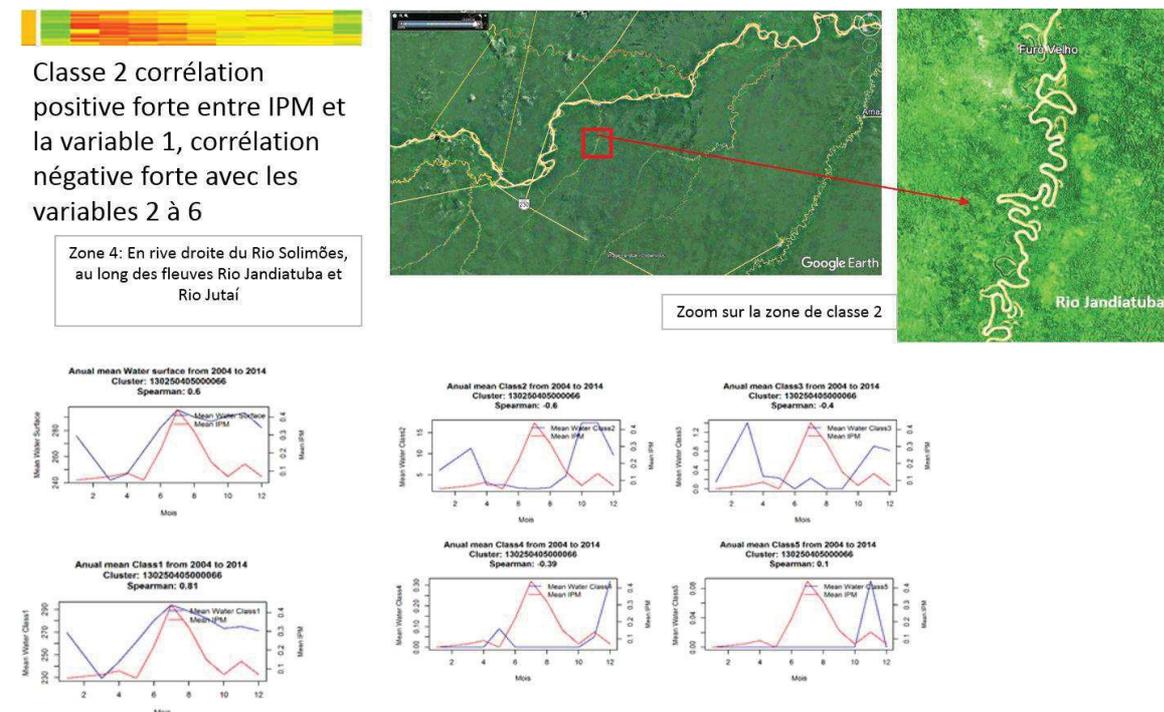


Figure 4-16 : analyse des corrélations pour la classe 2

La classe 2 est représentée dans la zone 4 sur le Rio Jandiutuba, un affluent de rive droite du Rio Solimões. C'est une zone d'eau blanche, une rivière divaguant en nombreux méandres dans une large plaine d'inondation, avec des méandres abandonnés et une eau sans doute relativement stagnante, plutôt sombre, la zone d'inondation étant sous l'eau une bonne partie de l'année comme le montre la courbe d'eau totale : un pic de crue en Juin et une décrue très lente jusqu'en Février. Le comportement de l'IPM est unimodal avec un pic en Juillet à l'amorce de la décrue.

La classe 5 est représentée dans la même vaste zone en rive droite du Rio Solimões, cette fois autour du Rio Jutai (figure 4-17)

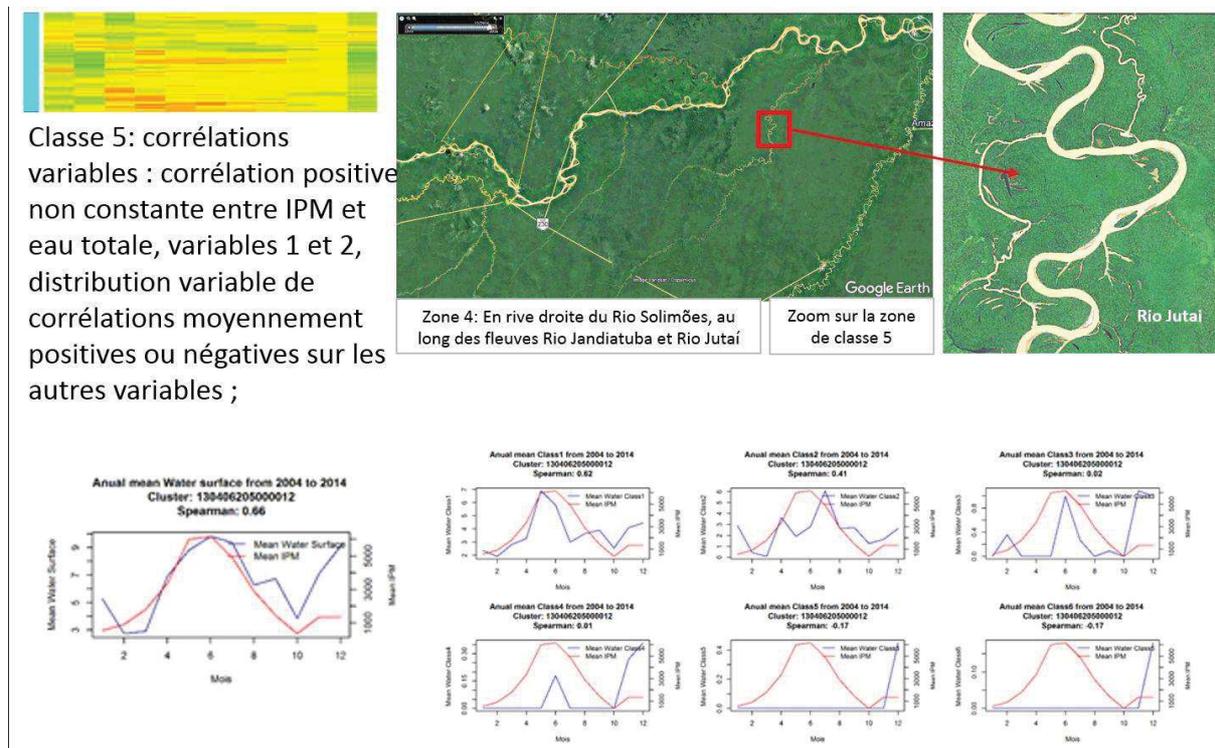


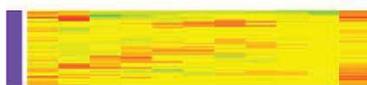
Figure 4-17 : Analyse des corrélations pour la classe 5

L'IPM présente un pic en Juin à base très élargie sur pratiquement toute l'année (de Février à Octobre), particulièrement bien corrélé avec la distribution de l'eau totale. Cette zone est une zone d'eau blanche, assez ressemblante à la précédente, avec

de nombreux bras vifs plutôt que des méandres abandonnés, mais de la même façon une très large plaine d'inondation. La corrélation reste significativement positive pour les variables 2, 3 et 4.

Les deux classes de corrélation suivantes 7 et 8 ont des comportements radicalement différents.

La classe 7 est représentée dans la zone 5, dans la région transfrontalière Brésil Pérou en face de la région de Loreto au Pérou. Le long du fleuve Javari qui marque la frontière, et loin à l'intérieur des terres, se trouvent de grandes zones de marécages, qui semblent deconnectées du fleuve, parfois occupées par une savane.



anticorrélation de l'IPM avec toutes les variables, parfois faiblement positive pour les variables 2 à 10 mais toujours neutre à négative avec les variables 1 et eau totale.



Zone 5: Région de la frontière Brésil Pérou, près de la région de Loreto au Pérou



Zone 9: Le Rio Madeira vers Borba sur l'affluent Rio Preto do Igapó-açu

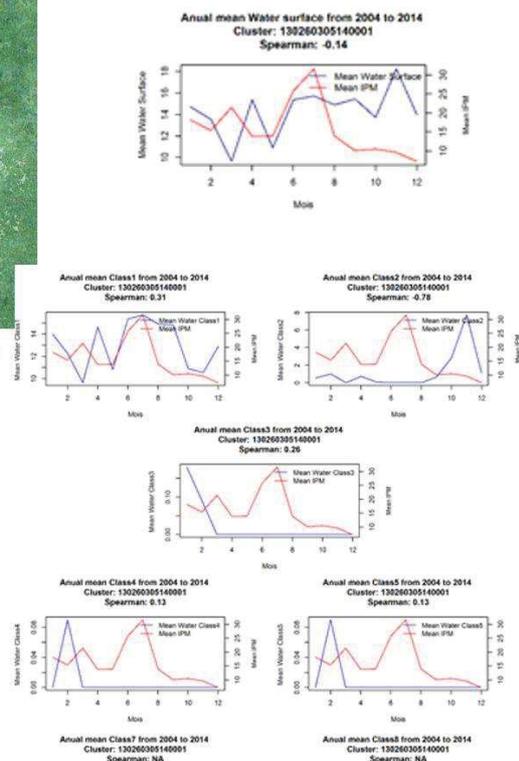


Figure 4 – 18 : Analyse des corrélations pour la classe 7

La zone 0, représentant également la classe 7 se trouve l long du Rio Madeira, un fleuve en eau blanche, mais la zone d'IPM élevé se trouve autour de plusieurs

affluents d'eau noire (Rio Preto do Igapo-açu, et présente un paysage très caractéristique que l'on reverra représenter la classe 8, de savanes aux sols hydromorphes de eau de plateaux, temporairement inondées. L'IPM bimodal à pics en Mars et Juillet présente des corrélations faiblement positives ou négatives avec toutes les variables eau.

La classe 8 est représentée dans plusieurs zones : zone 3 (rive droite du Rio Negro, zone comprise entre les Rio Uneuixi et Tea, avec un IPM moyen le plus élevé de l'état d'Amazonas - supérieur à 25,20), zone 1 (rive droite du Rio Negro, rio Unini) et zone 11 (Mine Pitinga).

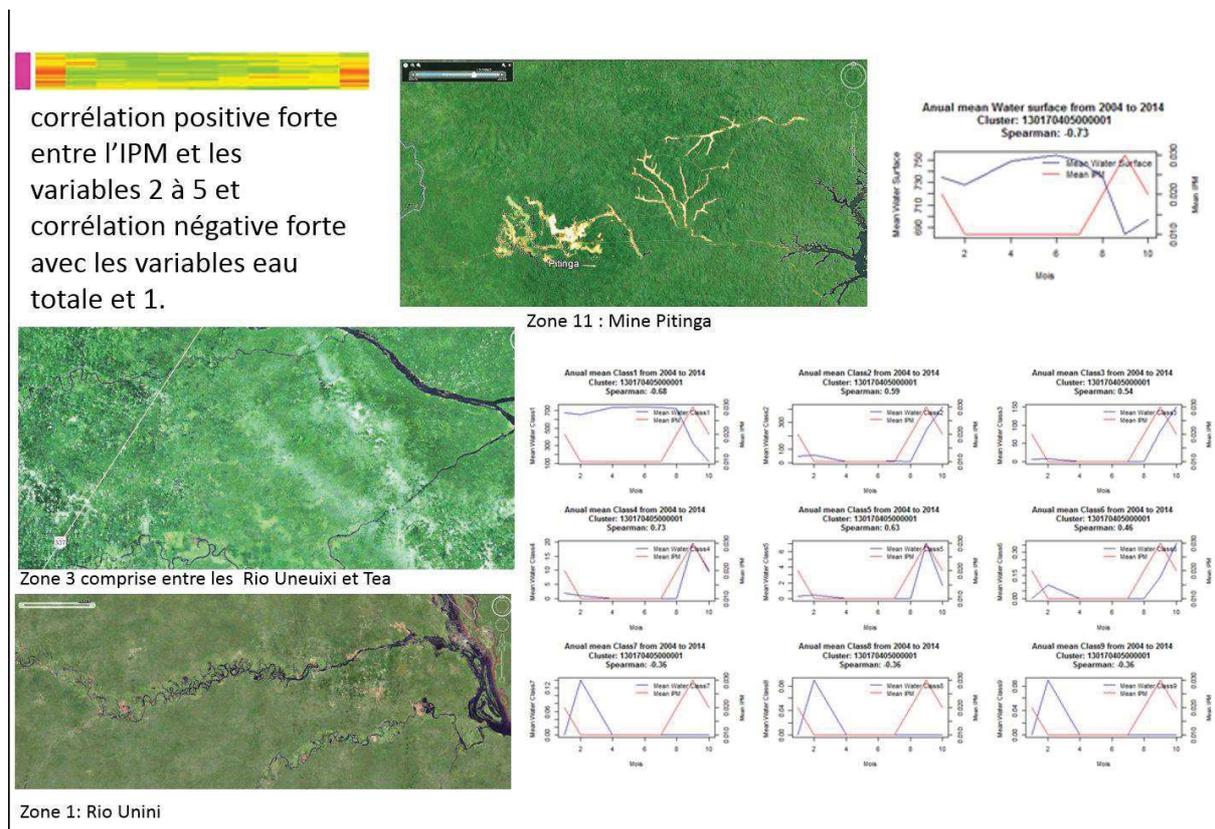


Figure 4-19 : Analyse des corrélations pour la classe 8

Dans ces deux cas du Rio Unini et Rio Uneuixi, le paysage caractéristique est celui de plateaux surbaissés entre des rivières aux bas-fonds toujours inondés. Ces

sommets de plateaux subissent une altération particulière des sols qui transforme les latérites rouges et jaunes initiales en latérites hydromorphes puis en podzols dans des cuvettes sommitales plus ou moins développées, temporairement inondées (Fritsch et al., 2007, Volkoff, 1985 ; Boulet et al., 1993). Lorsque la dégradation est avancée la forêt laisse la place à des savanes.

Le pic d'IPM est en Octobre-Novembre, au plus fort de l'étiage, comme le confirme le graphique représentant l'eau totale. La meilleure corrélation temporelle entre IPM et les variables de luminosité de l'eau se rencontre pour les variables 4 et 5, donc des luminosités moyennes.

4.3.5 Discussion

Une première analyse spatiale des incidences parasitaires mensuelles moyennes sur la période montre que des incidences parasitaires élevées se répartissent indifféremment sur les rives des fleuves d'eau noire ou d'eau blanche.

Au-delà de ce premier constat nos résultats interrogent la classification ancienne des eaux Amazoniennes en eau noire et eau blanche. Comme dans le chapitre 3, l'analyse du comportement temporel des variables de luminosité montre une grande variabilité dans le temps avec pour seul paramètre invariant le fait que les eaux les plus « lumineuses », donc les plus chargées en sédiments sont concentrées dans les périodes d'étiage ; Egalement l'analyse spatiale a montré que les zones à IPM élevé dans un bassin d'eau blanche pouvait se concentrer sur des affluents d'eau noire (figure 4-18) tandis que l'inverse était aussi vrai (figure 4 – 19).

Le deuxième résultat important concerne l'analyse des corrélations entre IPM moyen sur la période et moyennes mensuelles des classes de luminosité a mis en évidence deux grands groupes très distincts séparés par deux classes intermédiaires:

1) Le premier groupe met en relation les pics saisonniers de l'incidence parasitaire avec la crue (montée de crue, pic de crue ou décrue); ce groupe montre des corrélations positives fortes avec les luminosités caractéristiques des eaux claires diluée par la crue (variable 1) ou avec les variables 1 et 2 et des corrélations négatives fortes avec tous les autres canaux (classes de corrélation 1, 4, 2 et 3).

2) le groupe intermédiaire comprend la classe 5 qui montre des corrélations variables : corrélation positive non constante entre IPM et eau totale, et variables 1 et 2, une distribution variable de corrélations moyennement positives ou négatives sur les autres variables; et la classe 6 qui n'est représentée dans aucune des zones à IPM élevé

3) Le deuxième groupe met en relation l'IPM avec l'étiage et toutes les variables sauf l'eau totale et la variable 1. Dans ce groupe, la classe 8 montre des corrélations positives avec les variables 2 à 9 (la variable 10 n'est presque jamais représentée), et la classe 7 montre des corrélations positives faibles variables sur les variables 2 à 9.

Le comportement temporel de l'IPM est assez variable pour chacun de ces groupes de corrélation. Le bassin Amazonien dans son cours moyen et inférieur a une grande crue unique lissée centrée sur le mois de Juin qui peut se décaler en Juillet certaines années et un étiage prononcé en Novembre. Mais les amonts des bassins peuvent voir une crue décalée dans le temps voir plusieurs pics de crue selon les régimes de pluviosité. Pour analyser les relations de l'IPM avec la crue il est

nécessaire pour chaque exemple de se reporter à la variable eau totale qui représente la variation des surfaces en eau au cours de l'année, et donc un proxy de la crue puisque les surfaces en eau maximales marquent le plus fort de la crue.

Pour le premier groupe, il est important de noter que l'IPM a une distribution temporelle très étalée, mais toujours en phase avec la crue, avec des pics plutôt en montée de crue et tout début de décrue, parfois nettement en relation avec la décrue comme dans le cas du groupe 2.

La localisation géographique des zones de corrélation entre IPM et variables eau pour les IPM élevés montre que les zones du premier groupe sont des zones où les rivières d'eau blanche ou noire divaguent (en méandres ou tresses) dans une très large plaine d'inondation, sous forêt, dans des contextes d'occupation humaine marquée. Ces plaines d'inondation sont en connexion avec le fleuve dans tous les cas.

Pour le deuxième groupe, la classe 7 est caractérisée par un IPM dont le comportement temporel est très étalé avec 2 pics l'un en mars et l'autre en Juillet. Le premier pic plus petit correspond au plus fort de l'étiage et le second à la crue. Dans le cas de la classe 8, le pic d'IPM localisé dans l'année coïncide avec l'étiage.

La localisation géographique de ces deux classes de corrélation permet l'analyse suivante : dans le cas du groupe 7, des zones marécageuses en bordure de fleuve mais aussi assez loin à l'intérieur des terres, souvent en savanes, plus ou moins déconnectées du fleuve (pas de chenaux visibles), sont inondées dans la période de crue, et restent inondées bien après la décrue du fleuve. Dans le cas du groupe 8, ces savanes temporairement inondées sont situées au sommet des plateaux latéritiques surbaissés.

Il faut rappeler cependant que les causes d'erreur ou d'incertitude qui influent sur la corrélation sont nombreuses :

- causalité entre facteurs environnementaux et exposition aux facteurs de transmission du paludisme nécessairement partielle, par rapport aux autres facteurs de risque, et en fonction des actions de santé développées par les pouvoirs publics (campagnes de démoustication par exemple, inégalement réparties dans le temps et dans l'espace.

- interpolation linéaire de la population pour calculer l'IPM

- partie seulement des notifications de la base de données SIVEP Malaria qui sont exploitables

- décalage moyen d'un mois pour faire coïncider les données de cas avec les données d'exposition

4.3.6 Conclusion partielle

Ce travail est une première étude portant sur la relation entre la présence du paludisme et la luminosité de l'eau en utilisant des données satellitaires. C'est aussi un des premiers travaux réalisés à partir des données par poste de santé du SIVEP-Malaria à une aussi grande échelle.

L'objectif de ce travail de thèse était d'évaluer l'impact de la couleur de l'eau, approchée par un indicateur de luminosité de l'eau, sur les risques d'exposition au vecteur du paludisme, au travers de l'analyse des corrélations entre une base de données très riche de notifications de cas de paludisme et une analyse par télédétection, concordante dans le temps et dans l'espace avec la base de données.

Les résultats et l'interprétation qui en a été faite n'établissent pas de lien entre couleur de l'eau et paludisme et mettent même en cause la classification en grandes régions d'eau blanche ou noire. Par contre, ce travail a permis de donner des indications prometteuses sur les relations entre le risque d'exposition au paludisme, et :

1) les saisons clés de l'exposition dans des situations géographiques particulières

- zones forestières associées à des plaines d'inondation étendues, spécialement en période de décrue ou montée de crue, sur des rivières indifféremment d'eau blanche ou d'eau noire. Les zones d'inondation elles-mêmes sont caractérisées par des eaux noires ou claires, quelle que soit la couleur de la rivière qui les alimente.

- zones de savanes périodiquement inondées, en situation de haut de plateau, spécialement en période d'étiage, préférentiellement sur des eaux noires.

Bien que les acteurs publics soient conscients des périodes clés pour le contrôle épidémiologique (Tauil 2003; Silva et al. 2009; Cardoso & Goldenberg 2007; Couto et al. 2001), ces relations entre périodes liées au rythme des inondations et habitats à risque d'exposition au paludisme sont certainement des indications potentiellement utiles pour les pouvoirs publics et les systèmes de santé.

Essai de synthèse et considérations finales

La finalité de cette recherche était d'explorer les implications de la variation de la couleur de l'eau dans le comportement écologique du paludisme. Nous savons que la saisonnalité de l'incidence de paludisme est corrélée à la saisonnalité de l'*Anophèles darlingi* en Amazonie (Tadei et al. 2009) et que la saisonnalité de l'*Anophèles darlingi* est corrélée à la saisonnalité de la hauteur de l'eau des fleuves (Wolfarth et al. 2013). L'étude de l'implication de la couleur des eaux sur l'*Anophèles* est à ses débuts (Tadei et al. 2016) et ne disposait pour l'Amazonie, avant cette thèse, d'aucune analyse spatiale.

En Amazonie brésilienne, les scénarios de transmission du paludisme sont divers étant donné la diversité d'écosystèmes où le vecteur est retrouvé. Trois composantes doivent être présentes pour la transmission du paludisme, la population à risque, l'agent pathologique et la présence du vecteur. De cette manière les analyses qui sont valables pour une région ne le sont pas forcément pour une autre région mais aident à comprendre de quelle manière les différents facteurs de risque influencent le comportement épidémiologique du paludisme (Costa et al. 2010).

La relation entre le paludisme et l'homme est particulière. Les zones urbaines sont les moins affectées par la présence de moustiques mais l'*Anophèles darlingi* est aussi rarement retrouvé dans la forêt en Amazonie (Lourenço-de-Oliveira & Luz

Essai de synthèse et considérations finales

1996). Sa progression et prolifération comme espèce dominante en régions endémiques est le fruit de l'action de colonisation humaine. Les grandes épidémies de paludisme en Amazonie brésilienne ont été déterminées par des facteurs tels que l'occupation désordonnée des périphéries des grandes villes, l'exploitation forestière, l'élevage, les viviers, les établissements de réforme agraire ou l'extraction minière (Katsuragawa et al. 2008; Deane et al. 1948; Vieira et al. 2014; Lapouble et al. 2015).

Ainsi nous avons établi une méthode qui vise à comparer la relation du paludisme avec la couleur de l'eau dans différentes réalités régionales de l'État d'Amazonas en observant avec un maximum de détail spatial le lien étroit que le paludisme possède avec les populations à risque dans les limites de la disponibilité des données.

L'originalité de cette thèse réside principalement dans le fait d'étudier un facteur de risque de paludisme encore mal connu mais aussi dans la méthode proposée qui aborde le problème à grande échelle avec la préoccupation de la production de connaissances au niveau local. Les résultats ont permis de caractériser la problématique sur une grande portion du territoire de l'État d'Amazonas.

Les résultats obtenus dans cette recherche sont ici listés dans l'ordre de leur apparition dans les chapitres de ce travail.

En premier, l'analyse critique et la réflexion méthodologique sur l'usage scientifique de la base de données de vigilance sanitaire du SIVEP-Malaria nous a permis de comprendre quelles sont les limites de l'utilisation de cette source de données. L'étude du résultat du nettoyage des données du SIVEP-Malaria montre que les différentes variables qui composent les notifications de cas de paludisme ne sont pas produites avec les mêmes critères d'exactitude. Ainsi toutes les variables ne peuvent

Essai de synthèse et considérations finales

pas être exploitées de la même manière selon la période d'étude et les communes étudiées. Nous avons même trouvé des différences de qualité de donnée entre les zones urbaines et rurales d'une même commune. Nous proposons une utilisation attentive des variables disponibles et concluons que c'est une source de données très importante et en constante amélioration pour le contrôle et le suivi des épidémies mais aussi qu'elle peut servir comme donnée pour la recherche en géographie de la santé si les sources d'incertitudes sont connues et contrôlées.

Suite à l'analyse critique de la base de données SIVEP-Malaria, nous faisons une brève critique de l'indice épidémiologique appelé Incidence Épidémiologique Annuel (IPA) qui est le plus utilisé au Brésil pour comparer le risque d'infection de paludisme entre les communes de la région endémique. Nous concluons qu'il est risqué de comparer par cet indice des lieux qui ont des accès aux services de santé inégaux et qui par conséquent enregistrent de manière inégale les cas de paludisme des populations.

La partie du traitement des données environnementales de cette thèse commence par la production des données mensuelles de la couleur des eaux des fleuves et lacs de l'État d'Amazonas, une région avec une grande concentration de nuages tout au long de l'année qui empêche l'acquisition d'images de qualité en grand nombre. Pour ce faire nous avons produit une chaîne de traitement de données satellitaires du capteur MODIS qui télécharge des serveurs de la NASA les produits MODIS complémentaires à l'analyse et les croise pour produire les bandes spectrales nécessaires pour réaliser les indices utilisés dans ce travail sur huit tuiles MODIS et sur 11 ans. Nous avons observé que le pas de temps mensuel est adapté pour cette recherche et est ce qui peut se faire de mieux pour une étude régulière avec un minimum d'interférence de nuages sur cette région d'études.

Essai de synthèse et considérations finales

La production de l'indice de luminosité de l'eau, produit à partir des bandes du visible des données MODIS et dont le contraste est mis en évidence par la technique de la composante « luminosité » permet de déterminer la transparence ou la luminosité du spectre visible de l'eau donnant une indication sur la couleur de l'eau. L'étude de la dynamique de l'indice de luminosité sur 11 ans a produit une cartographie de la variance saisonnière des eaux montrant la spatialisation du changement de couleur des eaux entre les saisons. Cela a permis aussi de caractériser mensuellement la luminosité des différents fleuves de l'État d'Amazonas. Les résultats montrent une bien plus grande variabilité spatiale et temporelle que le laisse supposer la classification des grands fleuves en fonction d'une couleur de l'eau, supposée constante.

L'étude comparative entre l'IPM et la luminosité des eaux par secteur censitaire de l'IBGE est le mieux qu'on peut faire pour approcher une échelle d'analyse locale sur une étendue régionale avec les données disponibles. La signature temporelle de la saisonnalité des changements de la luminosité des eaux corrélée à la signature temporelle de l'IPM nous a permis d'observer quelles zones d'études et pour quelles temporalité de la crue, pour les IPM moyens significatifs, affichent une plus grande corrélation entre la couleur de l'eau et le paludisme. Les principales conclusions sont les suivantes :

Deux grands groupes de corrélations entre l'IPM et les variables eau totale et variables caractérisant la luminosité peuvent être distingués :

- l'un montrant des corrélations liées à la crue, montée de crue ou tout début de décrue en zone forestière dans de grandes zones d'inondation liées à la crue du

Essai de synthèse et considérations finales

fleuve, indifféremment pour des fleuves d'eau blanche ou noire, mais toujours en zone habitée.

- le deuxième groupe caractérisant des zones de transition forêt - savane périodiquement inondées, en situation de haut de plateau ou zones marécageuses étendues, spécialement en période d'étiage, caractéristiques des zones d'eau noire pour le groupe 8 ou sans relation avec la couleur de l'eau parce que sans relation nette avec le fleuve.

La chaîne de traitement automatisée sur des données à production en continue, comme le sont les données MODIS et celles du SIVEP-Malaria, permet la réalisation d'un produit d'analyse doté d'une mise à jour automatique pour un accompagnement de l'évolution de la situation, ce qui est un atout pour la recherche et pour le contrôle sanitaire.

Les données utilisées pour cette recherche sont toutes sujettes à erreur et incertitudes et doivent faire l'objet de précautions particulières pour leur emploi.

Les données épidémiologiques ont été prétraitées au mieux et les données résultantes ne sont pas exhaustives du profil épidémiologique local de l'État d'Amazonas. La responsabilité de la production des données SIVEP-Malaria repose sur les municipalités et chaque commune donne une priorité plus ou moins grande aux programmes de santé publique selon les années, ce qui produit une donnée non standardisée. Par exemple, une campagne de démoustication très abondante une certaine année suivie d'une année sans investissements dans ce sens est un facteur important d'incertitudes.

Le traitement des secteurs censitaires pour produire des zones d'études munies des informations de population sur 11 ans à partir de la forme des secteurs censitaires

Essai de synthèse et considérations finales

plus anciens permet la standardisation de l'information et la sélection de ces secteurs dans les limites du buffer des eaux permet de focaliser l'échantillonnage des pixels autour des zones habitées. Malgré la mise en œuvre de cette méthode, beaucoup de zones d'études sont encore trop grandes pour établir un lien direct entre les pixels d'eau et la population habitant le secteur censitaire. Certaines zones d'études ont une relation postes de santé/nombre de pixels plus représentatifs que d'autres, notamment les petites zones, dont les zones urbaines.

L'indice de luminosité utilisé ici est limité dans la détection de la couleur de l'eau. En effet il produit un contraste qui permet une subdivision par seuillage de différentes nuances de luminosité de l'eau entre la plus réfléchissante et la plus absorbante mais les eaux noires ne sont pas les seules eaux absorbantes. Les eaux claires ont un comportement spectral proche de celui des eaux noires car elles absorbent les radiations du spectre visible. L'État d'Amazonas ne possède pas de fleuve d'eau claire comme l'exemple du Tapajós, affluent du fleuve Amazonas dans l'État du Para. C'est la raison pour laquelle l'indice de luminosité a été conçu de cette façon. On observe qu'en période de crue l'eau blanche est moins réfléchissante car elle se dilue et perd de sa densité optique. Cette extrême variabilité temporelle des eaux dites « blanches » n'avait pas été anticipée.

Le seuillage de la classification de l'indice de luminosité des eaux par intervalles égaux en 10 classes est basé sur les valeurs minimum et maximum de luminosité identifiées dans les traitements des eaux. Les 5 premières classes, les moins réfléchissantes, ont concentré la plupart des occurrences de pixels et peuvent dans le futur être plus détaillées pour produire des informations plus précises sur les qualités d'eau prédominantes.

Essai de synthèse et considérations finales

Différentes questions d'échelle se posent dans ce travail. En plus de la relation entre la taille des zones d'étude et la présence de population et postes de santé dans la zone il y a la question de la résolution spatiale des pixels des produits MODIS entre 250 et 500 mètres. Certains habitats des gîtes larvaires sont bien sur indétectables à cette échelle. Le moustique du genre *Anophèles* et particulièrement l'espèce *Anophèles darlingi* possède une capacité d'adaptation importante. L'étude étant basée sur l'exploitation d'une base de données de cas, fait qu'elle est relativement indépendante en termes d'échelle, ce qui ne serait pas le cas si on s'était intéressé sur les facteurs influençant les niches écologiques des vecteurs.

Les résultats ne sont pas du tout décisifs sur le rôle de la couleur de l'eau comme facteur environnemental influant sur les incidences épidémiologiques et indiquent plutôt une conclusion contraire. Cependant, la mise en relation des périodes clé du cycle hydrologique avec une typologie des localisations sujettes à l'exposition au risque d'infection ou de transmission du paludisme est un résultat important.

La systématisation de toutes les étapes des traitements réalisés dans cette thèse en langage de programmation R permet de reproduire les résultats mais aussi de les adapter pour réaliser les traitements nécessaires aux études qui suivront sur la grande quantité de données qu'exige une étude sur plusieurs années sur une grande surface.

Références

- Amanajás, J.C. et al., 2011. Associação entre incidência de malária autóctone e precipitação no Estado do Amapá. In sbmet, ed. *Simposio Internacional de Climatologia*.
- Angelo, J.R., 2015. *Modelagem espacial dinâmica dos determinantes sociais e ambientais da malária e simulação de cenários 2020 para município*. Instituto Nacional de Pesquisas Espaciais.
- Arst, H., 2003. *Optical Properties and Remote Sensing of Multicomponental Water Bodies* Springer., Chichester: Praxis.
- Assis, M. de & Gurgel, H. da C., 2008. Precipitação pluviométrica e a dinâmica da malária na bacia do rio Purus. In *Congresso Brasileiro de Meteorologia*. INPE-15488-PRE/10216.
- Assis, M.C. et al., 2008. A malária e a dinâmica ambiental na bacia do rio Purus. *Geografia Ensino e Pesquisa*, 12, pp.5034–5049.
- Ávila, S.L. & Ferreira, A.W., 2013. *Diagnóstico laboratorial das principais doenças infecciosas e auto-imunes* 3rd ed., Rio de Janeiro: Guanabara Koogan.
- Ayana, E.K. et al., 2015. Assessing the potential of MODIS/Terra version 5 images to improve near shore lake bathymetric surveys. *International Journal of Applied Earth Observation and Geoinformation*, 36, pp.13–21.
- Barata, R. de C.B., 1995. Malária no Brasil: panorama epidemiológico na última década. *Cadernos de Saúde Pública*, 11(1), pp.128–136.

Références

- de Barros, F.S.M. et al., 2007. Parity and age composition for *Anopheles darlingi* root (Diptera: Culicidae) and *Anopheles albitarsis* Lynch-Arribálzaga (Diptera: Culicidae) of the northern Amazon Basin, Brazil. *Journal of vector ecology: journal of the Society for Vector Ecology*, 32(1), pp.54–68.
- Barros, F.S.M. de & Honório, N.A., 2007. Man biting rate seasonal variation of malaria vectors in Roraima, Brazil. *Memórias do Instituto Oswaldo Cruz*, 102(3), pp.299–302.
- Bartsch, A. et al., 2012. Detection of open water dynamics with ENVISAT ASAR in support of land surface modelling at high latitudes. *Biogeosciences*, 9, pp.703–714.
- Basurko, C. et al., 2011. Influence of climate and river level on the incidence of malaria in Cacao, French Guiana. *Malaria journal*, 10, p.26.
- Bayat, S. et al., 2001. Les bases de données épidémiologiques. *Les Cahiers du numérique*, Vol. 2(2), pp.155–176.
- Becker, B., 2013. *A urbe amazônida: a floresta e a cidade* Garamond Universitária, ed., Rio de Janeiro.
- Berry, P.A.M. et al., 2005. Global inland water monitoring from multi-mission altimetry. *Geophysical Research Letters*, 32(16), p.L16401.
- Bitencourt-Pereira, M.D., 1992. *Parametros limnologicos passíveis de serem monitorados através de dados espectrais obtidos por sensores remotos orbitais*. Universidade Federal de São Carlos.
- Boulet, R. et al., 1993. Géochimie des paysages: le rôle des couvertures

Références

- pédologiques. Coll. Acad. Sci. Cadas, Sédimentologie et géochimie de la surface, Paris, 55-76.
- Brasil, Ministério da Saúde, S. de V. em S., 2015. DATASUS. *Departamento de Informática do SUS*. Available at: <http://www2.datasus.gov.br/DATASUS/> [Accessed February 9, 2015].
- Brasil, Ministério da Saúde, S. de V. em S., 2014. *Portaria Nº 1.271, de 6 de junho de 2014*, Brasília.
- Brasil, Tribunal de Contas da Uniao, 2003. *Relatório de Monitoramento. Programa Nacional de Controle da Malária – PNCM*, Brasília.
- Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica., 2016. Brasil registra menor número de casos de malária nos últimos 35 anos. Available at: <http://portalsaude.saude.gov.br/index.php/o-ministerio/principal/secretarias/svs/noticias-svs/23370-brasil-registra-menor-numero-de-casos-de-malaria-nos-ultimos-35-anos> [Accessed May 4, 2016].
- Brasil, M. da S.S. de V. em S., 2014. *Guia de vigilância em saúde* 1ª edição. E. MS, ed., Brasília: Ministério da Saúde.
- Brasil, M. da S.S. de V. em S., 2006. *Portaria Nº 5, de 21 de fevereiro de 2006*,
- Braz, R.M., 2013. *Monitoramento da incidência da malária na Amazônia Brasileira utilizando algoritmo automatizado*.
- Braz, R.M.R., Andreozzi, V.V.L. & Kale, P.L.P., 2006. Detecção precoce de epidemias de malária no Brasil: uma proposta de automação. *Epidemiologia e*

Références

- Serviços de Saúde*, 15(2), pp.21–33.
- Van den Broeck, J. et al., 2005. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS medicine*, 2(10), p.e267.
- Cabral, A.C. et al., 2010. Increasing incidence of malaria in the Negro River basin, Brazilian Amazon. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(8), pp.556–62.
- Cardoso, R.F. & Goldenberg, P., 2007. Malária no Estado do Amapá, Brasil, de 1970 a 2003: trajetória e controle. *Cadernos de Saúde Pública*, 23(6), pp.1339–1348.
- Carnevale, P. & Robert, V., 2009. *Les anophèles: biologie, transmission du Plasmodium et lutte antivectorielle* IRD Éditio. Institut de Recherches pour le Développement, ed., Marseille.
- Carroll, M.L. et al., 2009. A new global raster water mask at 250 m resolution. *International Journal of Digital Earth*, 2(4), pp.291–308.
- CONASS, C.N. de S. de S., 2011. *Coleção para Entender a Gestão do SUS 2011* 20th ed. CONASS, ed., Brasília.
- Confalonieri, U.E.C., 2003. Variabilidade climática, vulnerabilidade social e saúde no Brasil. *Terra Livre*, 1(20), pp.193–204.
- Confalonieri, U.E.C., Margonari, C. & Quintão, A.F., 2014. Environmental change and the dynamics of parasitic diseases in the Amazon. *Acta Tropica*, 129, pp.33–41.
- Consoli, R.A.G.B. & Lourenço De Oliveira, R., 1994. *Principais mosquitos de importância sanitária no Brasil* Editora FIOCRUZ, ed., Rio de Janeiro.
- Costa, K.M. de M. et al., 2010. Malária em Cruzeiro do Sul (Amazônia Ocidental

Références

- brasileira): análise da série histórica de 1998 a 2008. *Revista Panamericana de Salud Pública*, 28(5), pp.353–360.
- Coutinho, P.E.G., 2015. *Análise do padrão de incidência da malária e sua relação com fatores climáticos e hidrológicos em escala sub-regional e local na bacia do Rio Negro*. Instituto Nacional de Pesquisas da Amazônia.
- Couto, Á.A. et al., 2001. Controle da transmissão da malária em área de garimpo no Estado do Amapá com participação da iniciativa privada. *Cadernos de Saúde Pública*, 17(4), pp.897–907.
- DAAC, N.L., 2013. *MODIS Land Products Quality Assurance Tutorial: Part-3 How to interpret and use MODIS QA information in the Land Surface Reflectance product suite* N. L. DAAC, ed.,
- Danson, F.M. & Plummer, S.E., 1995. *Advances in environmental remote sensing*, Wiley.
- Deane, L.M., 1988. Malaria studies and control in Brazil. *The American journal of tropical medicine and hygiene*, 38(2), pp.223–30.
- Deane, L.M., Causey, O.R. & Deane, M.P., 1948. Notas sobre a distribuição e a biologia dos anofelinos das regiões Nordeste e Amazônica do Brasil. *Revista do Serviço Especial*, (1), pp.827–965.
- Diegues, A.C.S., 2001. *O mito moderno da natureza intocada* HUCITEC, ed.,
- Diegues, A.C.S., 2002. *Povos e águas: inventário de águas úmidas brasileiras* 2a ed. NUPAUB, ed., São Paulo.
- Dunne, T. et al., 1998. Exchanges of sediment between the flood plain and channel

Références

- of the Amazon River in Brazil. *Geological Society of America Bulletin*, 110(4), pp.450–467.
- Fajardo, J.D.V., Souza, L.A.G. de & Alfaia, S.S., 2009. Características químicas de solos de várzeas sob diferentes sistemas de uso da terra, na calha dos rios baixo Solimões e médio Amazonas. *Acta Amazonica*, 39(4), pp.731–740.
- Feng, M. et al., 2015. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic–spectral classification algorithm. *International Journal of Digital Earth*, 8947(July), pp.1–21.
- Fernandes, F.B., Lopes, R.G.C. & Mendes Filho, S.P. de M., 2010. Malária grave em gestantes. *Revista Brasileira de Ginecologia e Obstetrícia*, 32(12), pp.579–583.
- Ferreira, M.U. & Silva-Nunes, M. Da, 2010. Evidence-based public health and prospects for malaria control in Brazil. *Journal of Infection in Developing Countries*, 4(9), pp.533–545.
- Feyisa, G.L. et al., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, 140, pp.23–35.
- Forattini, O.P., 1962. *Entomologia médica. 1o volume: Parte geral, Díptera, Anophelini* F. De & H. e S. Pública, eds., São Paulo.
- Furch, K., 1984. Water chemistry of the Amazon Basin: the distribution of chemical elements among freshwaters. In *The Amazon: Limnology and landscape ecology of a mighty tropical river and its basin*. Dordrecht: Dr W. Junk Publishers, pp. 167–169.

Références

- Fritsch, E. et al., 2007. Mobilité des métaux (Fe et Al) dans des podzols hydromorphes développés sur formations latéritiques du haut bassin amazonien.
- Gaillardet, J. et al., 1997. Chemical and physical denudation in the Amazon River Basin. *Chemical Geology*, 142(3–4), pp.141–173.
- Giglioli, G., 1938. Malaria in British Guiana; part III. Breeding habits of *An. darlingi* natural factors which limit the distribution of this species and of malaria. *Agricultural journal of British Guiana*, (9), pp.197–206.
- Gil, L.H.S. et al., 2003. Seasonal Malaria Transmission and Variation of Anopheline Density in Two Distinct Endemic Areas in Brazilian Amazônia. *Journal of Medical Entomology*, 40(5), pp.636–641.
- Girod, R. et al., 2011. Unravelling the relationships between *Anopheles darlingi* (Diptera: Culicidae) densities, environmental factors and malaria incidence: understanding the variable patterns of malarial transmission in French Guiana (South America). *Annals of tropical medicine and parasitology*, 105(2), pp.107–22.
- Gomes, A.P. et al., 2011. Malária grave por *Plasmodium falciparum*. *Revista Brasileira de Terapia Intensiva*, 23(3), pp.358–369.
- Goulding, M. et al., 2003. *The Smithsonian Atlas of the Amazon*, Washington and London: Smithsonian Institution press.
- Gualtieri, C. et al., 2015. Hydrodynamics and mixing at the confluence of Rio Negro and Rio Solimões. In *Andean and Amazonian rivers: deciphering the global change from the hydroclimatic variability in the critical zone*, 6th Observation Service HYBAM Scientific Meeting.

Références

- Guerschman, J.P. et al., 2011. *MODIS-based standing water detection for flood and large reservoir mapping: algorithm development and applications for the Australian continent*, Canberra.
- Gurgel, H., 2006. *Paludisme et dynamiques environnementales dans l'état de Roraima au Brésil*. Université de Nanterre.
- Hahn, M.B. et al., 2014. Conservation efforts and malaria in the Brazilian Amazon. *The American journal of tropical medicine and hygiene*, 90(4), pp.591–4.
- Herbreteau, V. et al., 2007. Thirty years of use and improvement of remote sensing, applied to epidemiology: From early promises to lasting frustration. *Health & Place*, 13(2), pp.400–403.
- Hiwat, H. et al., 2011. Ecology of *Anopheles darlingi* Root with respect to vector importance: a review. *Parasites & Vectors*, 4(1), p.177.
- Hu, Y. et al., 2015. Monitoring Spatial and Temporal Dynamics of Flood Regimes and Their Relation to Wetland Landscape Patterns in Dongting Lake from MODIS Time-Series Imagery. *Remote Sensing*, 7(6), pp.7494–7520.
- Ianniruberto, M. et al., 2015. River bed morphology and stratigraphy at the confluence of Rio Negro and Rio Solimões. In *Andean and Amazonian rivers: deciphering the global change from the hydroclimatic variability in the critical zone - 6th Observation Service HYBAM Scientific Meeting*.
- Jensen, J.R., 2007. *Remote Sensing of the Environment: An Earth Resource Perspective* Second Edi., Upper Saddle River: Pearson Prentice Hall.
- de Jonge, E. & van der Loo, M., 2013. *An introduction to data cleaning with R*, The

Références

Hague: Statistics Netherlands.

Junk, W.J., 1983. As águas da Região Amazonica. In Brasiliense, ed. *Amazonia: desenvolvimento, integração e ecologia*. Sao Paulo, pp. 45–100.

Justice, C.O. et al., 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, 36(4), pp.1228–1249.

Kar, N.P. et al., 2014. A review of malaria transmission dynamics in forest ecosystems. *Parasites & vectors*, 7, p.265.

Katsuragawa, T.H. et al., 2008. Endemic and epidemic diseases in Amazonia: malaria and other emerging diseases in riverine areas of the Madeira river. A school case. *Estudos Avançados*, 22(64), pp.111–141.

Katsuragawa, T.H. et al., 2009. Malária e aspectos hematológicos em moradores da área de influência dos futuros reservatórios das hidrelétricas de Santo Antônio e Jirau, Rondônia, Brasil. *Cadernos de Saúde Pública*, 25(7), pp.1486–1492.

Katsuragawa, T.H.T. et al., 2010. The dynamics of transmission and spatial distribution of malaria in riverside areas of Porto Velho, Rondônia, in the Amazon region of Brazil. *PLoS One*, 5(2), p.e9245.

Kffuri, C., 2014. Etnobotânica de plantas antimaláricas em comunidades indígenas da região do Alto Rio Negro–Amazonas–Brasil.

Kilham, N.E. & Roberts, D., 2011. Amazon River time series of surface sediment concentration from MODIS. *International Journal of Remote Sensing*, 32(10), pp.2659–2679.

Références

- Konhauser, K.O., Fyfe, W.S. & Kronberg, B.I., 1994. Multi-element chemistry of some Amazonian waters and soils. *Chemical Geology*, 111(1–4), pp.155–175.
- Kriegler, F. et al., 1969. Preprocessing Transformations and Their Effects on Multispectral Recognition. *Proceedings of the Sixth International Symposium on Remote Sensing of Environment (University of Michigan, USA)*, 2, pp.97–131.
- Kutser, T. et al., 2005. Mapping lake CDOM by satellite remote sensing. *Remote Sensing of Environment*, 94(4), pp.535–540.
- Land Cover CCI, 2016. *Product User Guide* UCL-Geomatics, ed.,
- Land Processes DAAC, 2011. *MODIS Reprojection Tool User's Manual* Release 4., USGS Earth Resources Observation and Science (EROS) Center.
- Lapouble, O.M.M., Santelli, A.C.F. e S. & Muniz-Junqueira, M.I., 2015. Epidemiological situation of malaria in the brazilian amazon region, 2003 to 2012. *Revista Panamericana de Salud Pública*, 38(4), pp.300–306.
- Lathrop, R.G., Lillesand, T.M. & Yandell, B.S., 1991. Testing the utility of simple multi-date Thematic Mapper calibration algorithms for monitoring turbid inland waters. *International Journal of Remote Sensing*, 12(10), pp.2045–2063.
- Liu, R. & Liu, Y., 2013. Generation of new cloud masks from MODIS land surface reflectance products. *Remote Sensing of Environment*, 133, pp.21–37.
- Lourenço-de-Oliveira, R. & Luz, S.L., 1996. Simian malaria at two sites in the Brazilian Amazon - II: Vertical distribution and frequency of anopheline species inside and outside the forest. *Memórias do Instituto Oswaldo Cruz*, 91(6), pp.687–694.

Références

- Luo, Y., Trishchenko, A. & Khlopenkov, K., 2008. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment*, 112(12), pp.4167–4185.
- Lyzenga, D.R., 1981. Remote sensing of bottom reflectance and water attenuation parameters in shallow water using aircraft and Landsat data. *International Journal of Remote Sensing*, 2(1), pp.71–82.
- Magris, M. et al., 2007. Vector bionomics and malaria transmission in the Upper Orinoco River, Southern Venezuela. *Memórias do Instituto Oswaldo Cruz*, 102(3), pp.303–312.
- Martinelli, P., 2014. *Amazônia - O povo das águas* 2a ed. T. Virgem, ed.,
- Martinez, J.M. et al., 2009. Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. *Catena*, 79(3), pp.257–264.
- Martinez, J.M. et al., 2007. Surface water quality monitoring in large rivers with MODIS data application to the amazon basin. In *2007 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 4566–4569.
- Martinis, S. et al., 2013. A Multi-Scale Flood Monitoring System Based on Fully Automatic MODIS and TerraSAR-X Processing Chains. *Remote Sensing*, 5(11), pp.5598–5619.
- McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), pp.1425–1432.

Références

- Meade, R.H. et al., 1985. Storage and remobilization of suspended sediment in the lower Amazon river of Brazil. *Science (New York, N.Y.)*, 228(4698), pp.488–90.
- Melfi, A.J., Forti, M.C. & Amorim, P.R.N., 2011. Hidroquímica das águas de drenagem de uma pequena bacia hidrográfica no nordeste da Amazônia (Estado do Amapá, Brasil): efeitos da sazonalidade. *Geochimica Brasiliensis*, 11(3), pp.325–340.
- Melo, W.F. de & Pinheiro, M. da C.N., 2012. Malaria in women in reproductive age: analysis of clinical-epidemiological aspects of the region Itaituba 2005 to 2007. *Revista paraense de medicina*, 26(1).
- Moquet, J.-S. et al., 2016. Amazon River dissolved load: temporal dynamics and annual budget from the Andes to the ocean. *Environmental science and pollution research international*, 23(12), pp.11405–11429.
- Motta, P. de N., 2015. *Variabilidade na incidência da malária no Estado de Rondônia, na Amazônia Brasileira e circulação atmosférica de larga escala: uma contribuição para a adaptação*. Universidade de Trás-os-Montes e Alto Douro.
- Mourao, F.F.R. et al., 2014. A vigilância da malária na Amazônia Brasileira. *Biota Amazônia*, 4(2), pp.161–168.
- Naimi, B., 2011. ModisDownload: an R function to download, mosaic, and reproject the MODIS images. Available at: <http://r-forge.r-project.org/projects/modis>.
- Nath, R.K. & Deb, S.K., 2010. Water-Body Area Extraction from High Resolution Satellite Images-An Introduction, Review, and Comparison. *International Journal of Image Processing (IJIP)*, 3(6), pp.353–372.

Références

- Naziano Filizola, Jean-Loup Guyot, Hella Wittmann, J.-M.M. and E. de O., 2011. *Sediment Transport in Aquatic Environments* A. Manning, ed., InTech.
- Novo, E., Steffen, C. & Braga, C., 1991. Results of a laboratory experiment relating spectral reflectance to total suspended solids. *Remote Sensing of Environment*, 36(1), pp.67–72.
- Novo, E.M.M., Hansom, J.D. & Curran, P.J., 1989. The effect of sediment type on the relationship between reflectance and suspended sediment concentration. *Remote Sensing*, 10(7), pp.1283–1289.
- Oliveira, A.A. de & Dalcy, D.C., 2001. *Florestas do Rio Negro* A. Oliveira, A. De, & D. C. Daly, eds., São Paulo: Companhia das Letras: UNIP.
- Oliveira-Ferreira, J. et al., 2010. Malaria in Brazil: an overview. *Malaria journal*, 9(1), p.115.
- Olson, S.H. et al., 2009. Links between climate, malaria, and wetlands in the Amazon Basin. *Emerging infectious diseases*, 15(4), pp.659–662.
- Overton, I.C., 2005. Modelling floodplain inundation on a regulated river: Integrating GIS, remote sensing and hydrological models. *River Research and Applications*, 21(9), pp.991–1001.
- Palmer, S.C.J., Kutser, T. & Hunter, P.D., 2015. Remote sensing of inland waters: Challenges, progress and future directions. *Remote Sensing of Environment*, 157, pp.1–8.
- Pantoja, G.M.T. & Andrade, R.F. de, 2012. Impactos sócioambientais decorrentes dos projetos hidrelétricos na bacia do Rio Araguari: do aumento populacional a

Références

- disseminação da malária. *Planeta Amazônia: Revista Internacional de Direito Ambiental e Políticas Públicas*, (4), pp.61–74.
- Parise, E., 2009. Malária grave em Palmas, Estado do Tocantins: relato de caso. *Rev Soc Bras Med Trop*.
- Pompêo, M. et al., 2015. *Ecologia de reservatórios e interfaces* Instituto de Biociências da Universidade de São Paulo, ed., São Paulo.
- Powell, S.J., Letcher, R.A. & Croke, B.F.W., 2008. Modelling floodplain inundation for environmental flows: Gwydir wetlands, Australia. *Ecological Modelling*, 211(3–4), pp.350–362.
- Queiroz, M.M.A. et al., 2009. Hidroquímica do rio Solimões na região entre Manacapuru e Alvarães: Amazonas - Brasil. *Acta Amazonica*, 39(4), pp.943–952.
- R Development Core Team, R., 2011. *R: A Language and Environment for Statistical Computing*,
- Rahm, E. & Do, H.-H., 2000. Data Cleaning: Problems and Current Approaches. *IEEE Bulletin of the Technical Committee on Data Engineering*, Vol 23(No. 4), p.11.
- Richardson, L.L. & LeDrew, E.D., 2006. *Remote Sensing of Aquatic Coastal Ecosystem Processes: Science and Management Applications*, Springer Science & Business Media.
- Ritchie, J., Zimba, P.V. & Everitt, J., 2003. Remote Sensing Techniques to Assess Water Quality. *American Society for Photogrammetry Engineering and Remote*

Références

- Sensing*, (69), pp.695–704.
- Rouse, J. et al., 1973. Monitoring Vegetation Systems in the Great Okains with ERTS. *Third Earth Resources Technology Satellite-1 Symposium*, 1, pp.325–333.
- Rouse, J.W. et al., 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. *Progress Report RSC 1978-1*, p.112.
- Ryerson, R.A. & Rencz, A.N., 1999. *REMOTE SENSING for the Earth Sciences* 3rd ed. A. N. Rencz, ed., New York: Wiley.
- Samanta, A. et al., 2012. Why Is Remote Sensing of Amazon Forest Greenness So Challenging? *Earth Interactions*, 16(7), pp.1–14.
- Sampaio, V.V.S. et al., 2015. Malaria in the State of Amazonas: a typical Brazilian tropical disease influenced by waves of economic development. *Revista da Sociedade Brasileira de Medicina Tropical*, 48 Suppl 1, pp.4–11.
- Sánchez, L.S.H. et al., 2015. Spatial-temporal variation of dissolved inorganic material in the Amazon basin. *Acta Amazonica*, 45(2), pp.175–186.
- Santelli, A., Ribeiro, I. & Daher, A., 2012. Efeito da combinação de dose fixa de artesunato e mefloquina na transmissão da malária em comunidades da bacia do rio Amazonas. *Malar J.*
- Santos, J.M.M. dos, Contel, E.P.B. & Kerr, W.E., 1981. Biologia de Anofelinos Amazônicos. 1- Ciclo biológico, postura e estádios larvais de *Anopheles darlingi* Root 1926 (Diptera: Culicidae) da rodovia Manaus - Boa Vista. *Acta Amazônica*, 11(4), pp.789–797.

Références

- Santos, T.B., Gurgel, H.C. & Angelis, C.F., 2007. Bacia do Rio Purus: Potencialidades para exportação da malária. In UFPR, ed. *III Simpósio Nacional de Geografia da Saúde - I Fórum Internacional de Geografia da Saúde*. Curitiba.
- Saraiva, M. das G.G. et al., 2009. Expansão urbana e distribuição espacial da malária no município de Manaus, Estado do Amazonas. *Revista da Sociedade Brasileira de Medicina Tropical*, 42(5), pp.515–522.
- Schroeder, W., Csiszar, I. & Morisette, J., 2008. Quantifying the impact of cloud obscuration on remote sensing of active fires in the Brazilian Amazon. *Remote Sensing of Environment*, 112(2), pp.456–470.
- Sharma, R. et al., 2015. Developing Superfine Water Index (SWI) for Global Water Cover Mapping Using MODIS Data. *Remote Sensing*, 7(10), pp.13807–13841.
- Silva, A.R. da et al., 2009. Controle da malária no Estado do Maranhão. *Revista da Sociedade Brasileira de Medicina Tropical*, 42(3), pp.318–324.
- Sioli, H., 1968. Hydrochemistry and Geology in the Brazilian Amazon Region. *Amazoniana*, 3, pp.267–277.
- Sioli, H. & Klinge, H., 1962. Solos, tipos de vegetação e águas na Amazônia. *Boletim do Museu Paraense Emilio Goeldi*.
- Stallard, R.F. & Edmond, J.M., 1983. Geochemistry of the Amazon: 2. The influence of geology and weathering environment on the dissolved load. *Journal of Geophysical Research: Oceans*, 88, pp.9671–9688.
- Stefani, A. et al., 2013. Land cover, land use and malaria in the Amazon: a systematic literature review of studies using remotely sensed data. *Malaria*

Références

journal, 12(1), p.192.

Suárez-Mutis, M.C. & Coura, J.R., 2007. Mudanças no padrão epidemiológico da malária em área rural do médio Rio Negro, Amazônia brasileira: análise retrospectiva. *Cadernos de Saúde Pública*, 23(4), pp.795–804.

Tadei, W.P. et al., 2016. Adaptive processes, control measures, genetic background, and resilience of malaria vectors and environmental changes in the Amazon region. *Hydrobiologia*, pp.1–18.

Tadei, W.P. et al., 1988. Biologia de anofelinos amazônicos: XII. Ocorrência de espécies de Anopheles, dinâmica da transmissão e controle da malária na zona urbana de Ariquemes (Rondônia). *Revista do Instituto de Medicina Tropical de São Paulo*, 30(3), pp.221–251.

Tadei, W.P., Thatcher, B.D., et al., 1998. Ecologic observations on anopheline vectors of malaria in the Brazilian Amazon. *The American journal of tropical medicine and hygiene*, 59(2), pp.325–35.

Tadei, W.P. et al., 2009. Malária, meio ambiente e mudanças climáticas: A dinâmica ambiental e a transmissão na Amazônia. In *Anais da 61ª Reunião Anual da SBPC*. Manaus-AM.

Tauil, P.L., 2003. Avaliação de uma nova estratégia de controle da malária na Amazônia brasileira. *Revista do Instituto de Medicina Tropical de São Paulo*, 45(6), pp.306–306.

Ticehurst, C., Guerschman, J. & Yun, C., 2014. The Strengths and Limitations in Using the Daily MODIS Open Water Likelihood Algorithm for Identifying Flood Events. *Remote Sensing*, 6(12), pp.11791–11809.

Références

- Ticehurst, C.J. et al., 2013. Using MODIS for mapping flood events for use in hydrological and hydrodynamic models: Experiences so far. In *20th International Congress on Modelling and Simulation*. Adelaide, pp. 1721–1727.
- UCSF, G.H.G., 2014. *Surveillance Systems to Facilitate Malaria Elimination*, UCSF.
- Vermote, E.F., Kotchenova, S.Y. & Ray, J.P., 2011. MODIS Surface Reflectance User's Guide.
- Vermote, E.F. & Vermeulen, 1999. Atmospheric correction algorithm: spectral reflectances (MOD09). *ATBD version*, 4(April), pp.1–107.
- Verpoorter, C. et al., 2014. A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18), pp.6396–6402.
- Vieira, G. de D. et al., 2014. Insecticide-treated bed nets in Rondônia, Brazil: evaluation of their impact on malaria control. *Revista do Instituto de Medicina Tropical de São Paulo*, 56(6), pp.493–7.
- Vittor, A.Y. et al., 2009. Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*. *The American journal of tropical medicine and hygiene*, 81(1), pp.5–12.
- Volkoff, B. 1985. Organisations régionales de la couverture pédologique du Brésil. Chronologie des différenciations. Cah. Orstom, sér. Pédologie, 21, pp. 225-236.
- Walker, I., 1987. The biology of streams as part of Amazonian forest ecology. *Experientia*, 43(3), pp.279–287.
- Wiefels, A. et al., 2016. Accuracy of the malaria epidemiological surveillance system

Références

- data in the state of Amazonas. *Acta Amazônica*, 46(4), pp.383–390.
- Wiefels, A. & Silveira, M., 2015. Malaria dynamic in Belo Monte's dam impact. *International Association for Impact Assessment Conference*.
- Wildlife Conservation Society, C. for I.E.S.I.N.-C.-C., 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Footprint Dataset (Geographic).
- Williams, J. & Pinto, J., 2012. *Manual de Entomologia da Malária Para Técnicos de Entomologia e Controlo de Vetores (Nível Básico)* USAID. RTI international, ed.,
- Wolfarth, B., Filizola, N. & Tadei, W.P., 2012. Análise epidemiológica da malária e as relações com variáveis ambientais em quatro municípios do Amazonas, Brasil. *selper brasil*.
- Wolfarth, B.R. et al., 2013. Epidemiological analysis of malaria and its relationships with hydrological variables in four municipalities of the State of Amazonas, Brazil. *Hydrological Sciences Journal*, 58(7), pp.1495–1504.
- World Health Organization, 2012. *Disease Surveillance for Malaria Control. An operational manual*, Geneva: WHO Library Cataloguing-in-Publication Data.
- World Health Organization, 2002. *Protocole d'évaluation des systèmes nationaux de surveillance et de riposte concernant les maladies transmissibles*, Gene: World Health Organization.
- World Health Organization, W., 2015. *World Malaria Report 2015*, Geneva.
- Zhang, Q. et al., 2007. Automatic extraction of water body based on EOS/MODIS remotely sensed imagery. In T. Zhang et al., eds. *International Symposium on Multispectral Image Processing and Pattern Recognition*. International Society

Références

for Optics and Photonics, pp. 678642-678642–7.

Annexe 1

Article publié pendant la thèse:

Wiefels A., Wolfarth-Couto B., Filizola N., Durieux L., Mangeas M., 2016. Accuracy of the malaria epidemiological surveillance system data in the state of Amazonas. *Acta Amazônica*, 46(4), pp.383–390.

Accuracy of the malaria epidemiological surveillance system data in the state of Amazonas

Alexandre WIEFELS¹, Bruna Wolfarth-COUTO², Naziano FILIZOLA³, Laurent DURIEUX^{1,4}, Morgan MANGEAS^{1,4}

1 Unité Mixte de Recherche Espace pour le Développement, Montpellier, France.

2 Instituto Nacional de Pesquisas da Amazônia, Manaus, Brasil.

3 Universidade Federal do Amazonas, Laboratório de Potamologia Amazônica, Manaus, Brasil.

4 Institut de Recherche pour le Développement, Montpellier, France.

* Corresponding author: alexandre.wiefels@teledetection.fr.

ABSTRACT

The Epidemiological Surveillance System for Malaria (SIVEP-Malaria) is the Brazilian governmental program that registers all information about compulsory reporting of detected cases of malaria by all medical units and medical practitioners. The objective of this study is to point out the main sources of errors in the SIVEP-Malaria database by applying a data cleaning method to assist researchers about the best way to use it and to report the problems to authorities. The aim of this study was to assess the quality of the data collected by the surveillance system and its accuracy. The SIVEP-Malaria data base used was for the state of Amazonas, Brazil, with data collected from 2003 to 2014. A data cleaning method was applied to the database to detect and remove erroneous records. It was observed that the collecting procedure of the database is not homogeneous among the municipalities and over the years. Some of the variables had different data collection periods, missing data, outliers and inconsistencies. Variables depending on the health agents showed a good quality but those that rely on patients were often inaccurate. We showed that a punctilious preprocessing is needed to produce statistically correct data from the SIVEP-Malaria data base. Fine spatial scale and multi-temporal analysis are of particular concern due to the local concentration of uncertainties and the data collecting seasonality observed. This assessment should help to enhance the quality of studies and the monitoring of the use of the SIVEP database.

KEYWORDS: Erroneous data, Database, Health surveillance.

Exatidão dos dados do sistema de vigilância epidemiológica da malária no estado do Amazonas

RESUMO

O Sistema de Vigilância Epidemiológica de Malária (SIVEP-Malária) é um programa governamental brasileiro que arquiva automaticamente todas as informações sobre casos de malária registrados em todas as unidades de saúde e consultórios médicos. O objetivo deste estudo foi avaliar a qualidade dos dados coletados pelo sistema de vigilância e sua precisão. Foram utilizados os dados do SIVEP-Malária para o estado do Amazonas, Brasil, de 2003 a 2014. Um método de limpeza de dados foi aplicado para detectar e remover registros errôneos. Observamos que a coleta de dados não é homogênea entre os municípios e ao longo dos anos. Algumas variáveis tinham diferentes padrões de coleta, falta de dados, dados discrepantes e inconsistências. Dados que dependem do agente de saúde possuem boa qualidade mas aqueles que dependem dos pacientes são frequentemente imprecisos. Mostramos que um pre-processamento meticuloso é necessário para produzir dados estatisticamente corretos a partir do SIVEP-Malária. Análises em escala espacial detalhada ou multi-temporais são particularmente afetadas devido à concentração local de incertezas e a sazonalidade observada na coleta de dados. Esta avaliação deve auxiliar a melhorar os estudos e monitoramentos que fazem uso dos dados do SIVEP.

PALAVRAS-CHAVE: Dados errôneos, Base de dados, Vigilância sanitária.

INTRODUCTION

The Epidemiological Surveillance System for Malaria (SIVEP-Malaria) is the Brazilian governmental program that registers all information about compulsory reporting of detected cases of malaria by all medical units and medical practitioners. It collects and provides information on the number and distribution of malaria cases and deaths in a territory. It is a critical tool for implementation of governmental malaria control programs (Deane 1988; Ferreira and Silva Nunes 2010; Silva *et al.* 2010). It is needed to determine which areas or population groups are most affected by malaria, so that resources on prevention and control can be targeted to the relevant populations (Martens *et al.* 1999; Almeida *et al.* 2010; Tauil and Daniel-Ribeiro 1998). Northern Brazil is an endemic malaria region (Barata 1995, Oliveira-Ferreira *et al.* 2010) and a growing number of researches make use of the Brazilian epidemiological surveillance system of malaria (SIVEP-Malaria) database.

In 1999, the Ministry of Health launched the Plan for Intensification of Malaria Control Actions [*Plano de Intensificação das Ações de Controle de Malária* (PIACM)] as a response to the World Health Organization (WHO) Roll Back Malaria Program (Sampaio *et al.* 2015). In 2003, in partnership with states and municipalities of the Amazon region, the Brazilian Ministry of Health's Secretariat of Health Surveillance (SVS) created the National Program for Prevention and Control of Malaria [*Programa Nacional de Prevenção e Controle da Malária* (PNCM)] for the prevention and control of the disease, continuing the progress made with the PIACM (Pina-Costa *et al.* 2014). The Malaria's Epidemiological Surveillance System (SIVEP-Malaria) is the main tool used by PNCM to improve the production of information about malaria. From 2003 to 2004, the municipal participation in the SIVEP-Malaria program increased from 60% to 100% in the Amazon region where 99% of the malaria cases in Brazil are concentrated (Brasil 2003; Braz *et al.* 2006). In Brazil, patients suspected of having malaria are compulsorily reported to the Ministry of Health through SIVEP-Malaria (Brasil 2014). The reporting form includes the medical unit data, the patient's personal and clinical data, and also the georeferenced data, generating a total of 43 variables. Thus, the SIVEP-Malaria database allows the spatial and temporal monitoring of epidemics but also serves to assess the coverage of diagnosis and treatment (Oliveira-Ferreira *et al.* 2010). However, research find that the SIVEP information is not utilised well by local managers (Costa *et al.* 2010; Hermes *et al.* 2013; Peiter *et al.* 2013). There are a variety of errors spread over the data, mainly due to mistyping or lack of precision on declarations of patients. In 2013 the Ministry of Health decentralized the managing of data acquisition to

the municipal administration, which has generated different data collection methods.

Our hypothesis is that the raw SIVEP-Malaria data is exploitable for epidemiological studies with scientific credibility after preprocessing it. The objective of this study is to point out the main sources of errors in the SIVEP-Malaria database by applying a data cleaning method to assist researchers about the best way to use it and to report the problems to authorities.

MATERIALS AND METHODS

The usual analysis approach is to begin with descriptive analyses to explore the data (Schoenbach and Rosamond 2000). This type of prior analysis allows one to identify and intercept errors in order to obtain a valid and consistent data set that is technically correct, thus ready for statistical analysis. The filtering mechanism to obtain this data quality is called data cleaning and it is executed in four steps: detecting missing values, outliers, inconsistencies and finally making the corrections (Broeck *et al.* 2005; Jonge and van der Loo 2013; Hang-Hai and Erhard 2000).

The SIVEP-Malaria data is available on database files kept by the Health System Informatics Department (DATASUS) (Brasil 2015). It contains administrative, personal and laboratorial variables as shown in Table 1. The data banks do not include names or other information that might allow identification of the subjects.

The data cleaning was performed over a selection from the database which included all notified and all infected patients (variables UF_NOTIF & UF_INFEC) from the state of Amazonas. This selection allows consideration, at the same time, of the cases registered in the Amazonas health centers and those infected in the Amazonas territory but registered abroad responding to different research objectives. From this selection, a new filter picked only the positive cases (variable RES_EXAM). Positive cases ensure that one can handle the complete record of all variables, such as clinical ones. Both steps resulted at first in 12,286,851 reported cases and then in 1,801,229 positive cases from 2003 to 2014.

RESULTS

The data collection for each patient was discontinuous over the time and the variables (Table 2). Some variables from the malaria notification form are systematically filled in, while some others are missing. Finding out where the missing values are is the first step in choosing which variable to use in research for a specific study area.

However, the availability of data is not constant over time. From 2003 to 2014, SIVEP-Malaria improved as a surveillance system. The systematicity of the data production

Table 1. SIVEP-Malaria variables code names and definition

Variable	Definition	Variable	Definition	Variable	Definition	Variable	Definition
COD_NOTI	Notification number	DT_NASCI	Birth date	MUN_RESI	Municipality of residence	LOC_INFE	Locality of infection
DT_NOTIF	Notification date	ID_PACIE	Patient age	LOC_RESI	Locality of residence	DT_EXAME	Examination date
TIPO_LAM	Active/passive notification	ID_DIMEA	Age writing format	DT_SINTO	First symptoms date	EXAME	Examination method
UF_NOTIF	State of notification	SEXO	Gender	DT_TRATA	Date of treatment	RES_EXAM	Examination results
MUN_NOTI	Municipality of notification	GESTANTE	Pregnancy length	VIVAX	Patient is under Vivax treatment	QTD_CRUZ	Parasitaemia
COD_UNIN	Health unit of notification	NIV_ESCO	Schooling level	FALCIPARUM	Patient is under Falciparum treatment	QTD_PARA	Parasites by mm ³
COD_AGEN	Health agent code	RACA	Ethnic group	ID_LVC	Follow-up consultation	HEMOPARASI	Hemoparasites
SEM_NOTI	Notification week	COD_OCUP	Employment	PAIS_INF	Country of infection	EXAMINADOR	Examiner code
DT_DIGIT	Date of digitalization	PAIS_RES	Country of residence	UF_INFEC	State of infection	ESQUEMA	Treatment schedule
DT_ENVLO	Data entering into National database date	UF_RESID	State of residence	MUN_INFE	Municipality of infection	SINTOMAS	Symptoms

■ Administrative data

■ Personal data

■ Laboratorial data

Table 2. Data availability per variable from 2003 to 2014 for the state of Amazonas

Variable	Data Availability (%)	Variable	Data Availability (%)	Variable	Data Availability (%)
COD_NOTI	100	PAIS_INF	99,90	GESTANTE1	79.50
DT_NOTIF	100	QTD_CRUZ	99,80	NIV_ESCO_1	78.60
DT_ENVLO	100	MUN_RESI	99,80	ESQUEMA_1	75.30
TIPO_LAM	100	UF_RESID	99,30	DT_DIGIT	71.40
UF_NOTIF	100	UF_INFEC	99,30	DT_NASCI	70.90
MUN_NOTI	100	MUN_INFE	99,30	RACA	14.10
COD_UNIN	100	LOC_INFE	97,60	FALCIPARUM	14.10
SINTOMAS	100	LOC_RESI	96,60	VIVAX	14.10
RES_EXAM	100	GESTANTE	93,50	EXAME	14.10
ID_PACIE	99,90	DT_SINTO	93,30	EXAMINADOR	14.10
PAIS_RES	99,90	DT_TRATA	90,50	HEMOPARASI	14.10
ID_LVC	99,90	ESQUEMA	89,40	QTD_PARA	11.90
ID_DIMEA	99,90	COD_AGEN	88,80	SEM_NOTI	0
SEXO	99,90	NIV_ESCO	88,60		
DT_EXAME	99,90	COD_OCUP	88,00		

Variables names are described in Table 1

could be assessed when plotted over time. The entered data values were counted for each variable per day, based on the notification date at the health unit. The proportion between available data, for each variable, and the total of notification collected was plotted (Figure 1).

The results were classified into five *Groups* from *A* to *E*. *Group E* contained the six variables whose values were nearly or always present, and for which a plot was unnecessary. *Group E* variables were: the municipality of infection, the state of residence, the state of infection, the parasitaemia, the municipality of residence, and the country of infection.

Variables of the *Group A* began to be recorded in 2011, and were systematically registered in 2013. By contrast, *Group B*'s three variables developed a systematicity since 2003 but stopped being collected when SIVEP-Malaria's register form changed in 2010. They were transposed to *Group C* with the new method of classification. *Group C* assembled the variables with the most expected behavior, gaining in systematicity and continuous filling over time. *Group D* contained variables that presented a seasonal trend, with a stationary mean value over time. These variables had an annual seasonality. The birth date and the positive malaria cases had a Spearman correlation of -0.65 (Figure 2). It suggested that repeatedly,

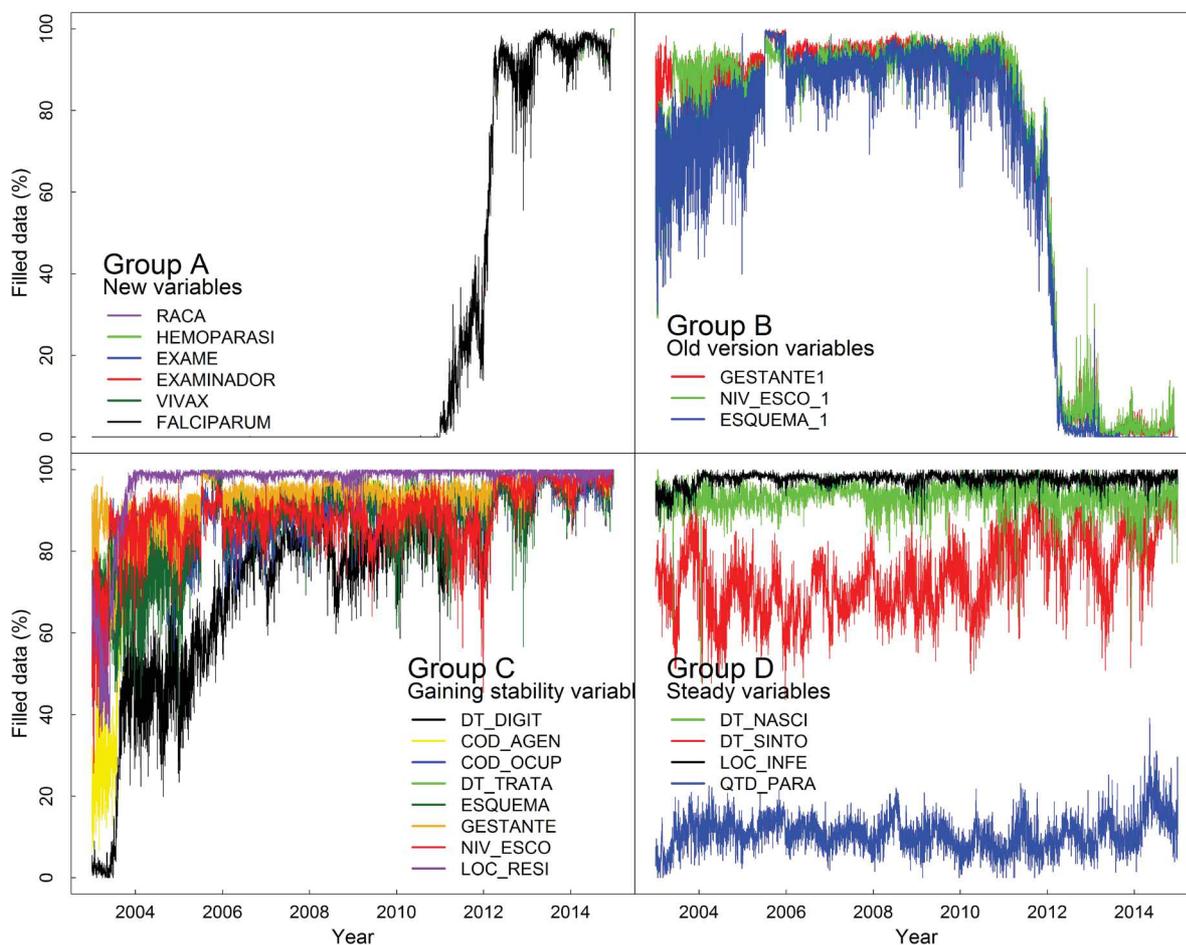


Figure 1. Classification of the data availability by variable over the years. This figure is in color in the electronic version.

at the epidemiological peak, the health agents made a more efficient data collection, decreasing the amount of missing birth dates on notifications. This periodicity was observed for all the others variables in different proportions but it was more visible on the birth date, particularly because of its stability over time. This observation suggested that the epidemiological cycle affects the notification of other variables.

Most parts of SIVEP-Malaria data are qualitative. Each variable was analyzed separately, comparing the data provided to the expected interval of values to identify unrealistic outliers. There were seven categories representing dates, among which the following present outliers: the birth date - expected to be in the interval between “1900-01-01” and “2014-12-31” - presented 19 occurrences from the year of “1006” to “1900” and 22 from “2030” to “2091”; the symptom date - expected to be between “2000-01-01” and “2014-12-31” - presented 148 occurrences from the year of “1002” to “1999” with a peak around “1200”; and the treatment date - expected to be

between “2003-01-01” and “2014-12-31” – which registered 796 occurrences from the year of “2015” to “9620”.

The exam results (with values from 1 to 11) and the parasitaemia (from 1 to 6) showed 17 times the value “0”, which was not an option, in 2003 for the municipalities of Coari and Manaus. The registration code, with positive values, showed 18 outliers occurrences (16 x “0”, “-1264” and “-682”) from 2003 to 2007. The epidemiological week variable was always absent except for one case in 2003. The municipality of residence has only one unrealistic outlier named “539924” which is not a municipality code. There were 20,242 registered health agents in the SIVEP-Malaria database for Amazonas state. Each one is represented by a code number. From those, 15,692 were responsible for notifications between 2003 and 2014. Another 302,236 agents appearing as notifiers were not registered.

Some values are logically or biologically impossible (i.e. gender=male AND pregnant=yes). The following most

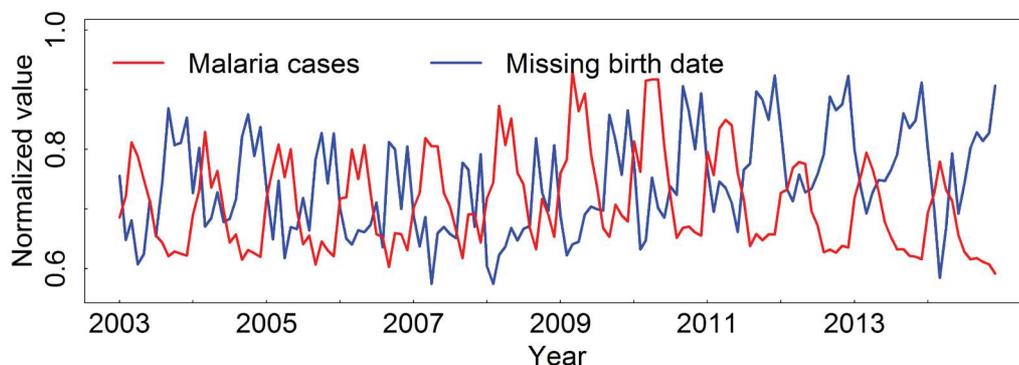


Figure 2. Relation between birth date (DT_NASCI) data filling and malaria cycle. This figure is in color in the electronic version.

frequently used variables were analyzed, relating pairs to check the coherence of values.

The follow-up consultation was related to the active/passive notification. The patient's follow-up consultation occurs weeks after the confirmation of the malaria infection (World Health Organization 2012). It is a second consultation to certify the healing and is registered as a new notification, which is why it is essential to distinguish it from the first consultation to avoid duplicates. Two variables - the detection type, called TIPO_LAM (1-Passive, 2-Active and 3- Follow-up consultation) and the follow-up consultation, called ID_LVC (1-Yes, 2-No) certify the follow-up consultation. There were 201 inconsistent notifications checked as a follow-up consultation notification (TIPO_LAM=3) but also as its opposite (ID_LVC=2).

The age writing format was related to the patient age. Both variables work together. The age writing format shows whether the patient age value is registered in days, months or years. There were 299 values showing patients aged between 101 and 1004.

The patient age was related to the pregnancy length. There were seven patients less than one year old showing as pregnant between the 1st and 3rd trimester of pregnancy, 118 notifications of 10-year old patients with ignored gestational age, and 38 pregnant but over 55 years old.

The exam, the notification and the symptoms dates were related. Between the malaria infection and the cure, the order of events was not always symptom, then notification, and finally exam. When comparing the dates of these events, some behaviors were unexpected. For not-followed-up consultation notifications (1,530,499 notifications) there were 348,292 patients notifying the same day they reported the malaria symptoms and 1,151,271 patients who delayed notifying it until one week later. On the other hand, 13,597 persons

notified before having symptoms, which is not logical and 6,290 persons notified 60 days after the symptoms.

There were 1,395,019 (91.15%) clinical exams carried out on the same day as the notification and 9,276 (0.6%) realized one week afterwards. However, there were 60 unexpected clinical exams performed before the notification.

All those cases were concentrated mostly in the municipalities of Manaus and Coari around 2003.

The SIVEP-Malaria georeferencing system is organized by localities and the municipalities are responsible for updating this information. One locality can accommodate a health unit, a patient's residence or his declared infection site. The localities module provides the georeferenced points. However, only 1.06% of all 525,795 of SIVEP-Malaria's localities were actually georeferenced, of which 28.6% were on the Amazonas state, the most covered one.

Data cleaning product

In order to maximize the capacity of analysis of SIVEP-Malaria, a data cleaning was made covering Amazonas state from 2003 to 2014. The goal was to identify the fully filled notifications with georeferenced health units, residences and infection localities. The data cleaning started excluding the 10 variables having the larger amount of missing values (Table 2) and the three ones that were replaced (Group B). The data entry included 1,801,229 malaria positive notifications. At first, 49.8% of notifications were excluded for having missing values. Next, less than 0.01% were excluded for having outliers. Then, 0.02% were removed for being inconsistent. Finally, the georeferencing of localities was analyzed. There were 2.03% of notifications without georeferenced health units, 37.59% of notifications without georeferenced patient residences and 44.82% of notifications without georeferenced infection localities.

The data cleaning product is summarized in Figure 3 as an erroneous data concentration map by municipality. The missing data, outliers and inconsistencies were counted for each notification and the result was matched with its respective notification municipality geocode to quantify the erroneous data occurrence by municipality. The total of invalid data was divided by the total of notifications for all years and multiplied by 100 returning the so called “erroneous data index”. Values varied from 4.47 for the municipality of São Gabriel da Cachoeira to 153.93 for Novo Airão.

After the data clearance and the subsequent selection of notifications based on georeferenced localities, 49.17% of data entry was exploitable for a study based on the health units, 31.33% for a study based on the patient’s residence and 27.70% for one based on the infection site.

Impact on research

Using the Google Scholar, the Malaria Journal, and the Web of Science internet search engines, and giving “SIVEP” and “Malaria” as keywords, we have counted a total of 349 publications: 280 articles, 14 books and 18 master/PhDs from

2003 to 2014. The number of publications gradually grows every year. The most used variables from SIVEP-Malaria were those used to elaborate the Annual Parasite Incidence (API) employed by the National Plan to Malaria Control (PNCM) to set transmission risk and priority areas for disease control. It is a reference index, often used at the municipal scale, and composed of four SIVEP-Malaria variables: exam results, municipality of residence, notification date and follow-up consultation. Those were also on the top of the list of used variables when used separately. The next most used variables are the residence locality, the municipality of notification, gender, age, the infection locality, the pregnancy length, active/passive notification, the treatment date and the symptom date.

DISCUSSION

Medical datasets often have missing and erroneous values due to a bad record or human error (Bai *et al.* 2015). There was a major difference of data availability between the 43 variables that makes it impossible to compare all data for every patient. Missing data formed the first step of the data cleaning,

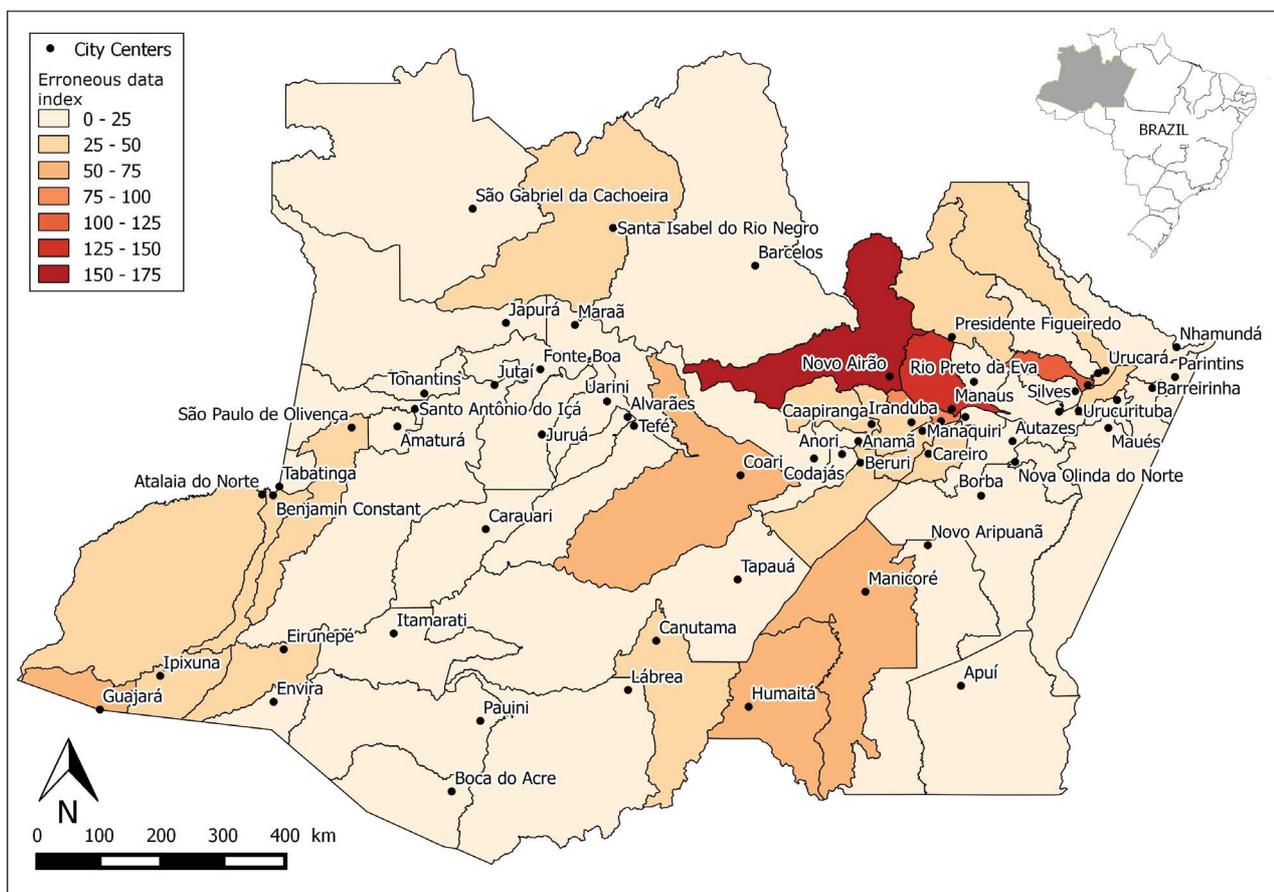


Figure 3. Erroneous data by Amazonas state municipalities from 2003 to 2014. This figure is in color in the electronic version.

discarding 49.8% of patients' notifications if the research demanded having all the variables. The variables from *Group A* are often responsible for this. Usually a personalized data cleaning has to be made for each study discarding undesirable variables before discarding missing data notifications. It is recommended to make a descriptive analysis and exploration to identify for each variable its availability for specific region and period.

Using the raw SIVEP-Malaria data rather than available processed products allows multiple approaches of the epidemiological reality and the method to be used has to consider the spatial and temporal scale to be analyzed. Local scale studies are often impossible, as for the municipalities of Itapiranga, Urucurituba, Nhamundá, Nova Olinda do Norte, Barreirinha, Boa Vista do Ramos and Silves, which accumulates zero georeferenced localities. The lack of fine geographical information explains why researches are made on a municipal scale. The georeferenced SIVEP localities points are missing for most parts of the territory and the provided population values are outdated. This disables a demographic dependent approach like the Annual Parasite Incidence (API) (Ray and Beljaev 1984) on a smaller scale than the municipal, forcing SIVEP-Malaria to be combined with the municipal demographic data of the Brazilian Institute of Geography and Statistics (IBGE).

Multi-temporal studies may contain errors if they exclude notifications with missing data from some erroneous variable, such as the new variables from *Group A*. This would lead to underestimation of the real amount of malaria cases. A study based on the ethnics is only possible after 2011. Using the data from *Group B* is also a mistake since they have been replaced, though they still have some insertions after 2011 that are possibly complementary to variables from *Group C*.

Compulsory reporting produces a great amount of data. SIVEP-Malaria gets data from 805 municipalities in nine states. Invalid data may occur, but we observe improvement in terms of data production and quality. It is possible to mark exactly where and when invalid data occurs and some are concentrated in certain municipalities or localities and also in specific time periods. Manaus and Coari are the two municipalities highlighted by having more invalid data. The earliest years of the system are clearly a transition phase to an operational system. The variables that are more constant in providing technically correct data are the variables of exclusive responsibility of the medical unit/agent. In second place is the information provided by the patient.

SIVEP-Malaria is in continuous development. It will only be fully functional when all variables are systematically filled and when the complementary modules, like the list of georeferenced locations, health agents or examiners, are updated by municipalities. The outliers and the

inconsistencies are not very important statistically when looking at the whole region of the Amazonas state, but can be a serious threat to a local study for specific periods, since the accumulation of erroneous data can be precise in space and time. Erroneous data may come from the patient declaration, the health agent reporting or from mistyping while uploading to the database. Visualizing the data behavior through space and time allows one to learn much about the database and the collecting system, and enables one to adjust the methodology of research to avoid unexpected situations. SIVEP-Malaria still lacks strategic data, such as geographical, and the fact that it has passed through major changes in its young lifetime does not allow its full temporal capacity to be used for some variables. However, it has a great capacity to provide information on the distribution and trends in malaria, especially at the municipal scale or using the georeferenced medical unit for the local scale. Although, the system is designed so as to, eventually give a good data quality to the residential and the infection site localities. It is already a great data source for scientific research, even though it has been created for monitoring and administration.

CONCLUSIONS

The results of the SIVEP-Malaria data base data cleaning for the state of Amazonas show that some variables are best to be used than others. The major limitation in using variables is the absence of data in certain places, and certain periods, which could affect comparisons between places and dates. Plotting data availability by variable over the years allowed a global view of the most robust SIVEP-Malaria variables. We detected several errors, as unrealistic outliers, and incoherent data. The source of the detected erroneous data can be identified, and most part originates from the municipalities of Coari and Manaus, mainly from the early dates of the SIVEP-Malaria. The errors can also be influenced by the epidemic fluctuations. The number of studies that uses SIVEP data increases every year. A best knowledge of the structure of this data base is critical. The results shows that it is highly recommend to realize the applied pretreatment method before performing statistics with the data.

ACKNOWLEDGEMENT

This research was partially supported by the Ciência sem Fronteiras (CsF), Brazilian Scholarship Program, the National Institute of Science and Technology Innovation for Neglected Diseases (INCT-IDN) and the Observatory of Climate and Health from Oswaldo Cruz Foundation in Rio de Janeiro, Brazil and the Unité Mixte de Recherche Espace pour le Développement (UMR ESPACE-DEV) from Montpellier, France.

REFERENCES

- Almeida, L.B.; Barbosa M.G.V.; Espinosa, F.E.M. 2010. Malária em mulheres de idade de 10 a 49 anos, segundo o SIVEP- Malária, Manaus, Amazonas, 2003-2006. *Revista da Sociedade Brasileira de Medicina Tropical*, 43: 304-308.
- Bai, E.W.; Johnson, H.; Xu, W.; Jacob, M. 2015. A preliminary study on cleaning up erroneous data and filling in missing values in a medical record. *IFAC-PapersOnLine*, 48(20): 493-498.
- Barata, R.C.B. 1995. Malaria in Brazil: Trends in the last ten years. *Cadernos de Saúde Pública*, 11: 128-136.
- Brasil, 2003. Relatório de monitoramento. Programa Nacional de Controle da Malária. *Tribunal de Contas da União*, Brasília, 27p.
- Brasil, 2014. Portaria nº 1.271, de 6 de junho de 2014. Brasília.
- Brasil, 2015. Departamento de informática do Sistema Único de Saúde do Brasil. (<http://www2.datasus.gov.br/DATASUS/>). Accessed on 09/02/2015.
- Braz, R.; Andreozzi, V.; Kale P. 2006. Detecção precoce de epidemias de malária no Brasil: Uma proposta de automação. *Epidemiologia e Serviços de Saúde*, 15: 21-33.
- Broeck, J. van der.; Cunningham, S.A.; Eeckels, R.; Herbst, K. 2005. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Medicine* 2: e267.
- Costa, K.M.M.; Almeida W.A.F.; Magalhães, I.B.; Montoya, R.; Moura, M.S.; Lacerda, M.V.G. 2010. Malária em Cruzeiro do Sul (Amazônia Ocidental brasileira): Análise da série histórica de 1998 a 2008. *Revista Panamericana de Salud Pública*, 28: 353-60.
- Deane, L. M. 1988. Malaria studies and control in Brazil. *The American journal of tropical medicine and hygiene*, 38: 223-30.
- Ferreira, M. U.; Silva-Nunes, M. 2010. Evidence-based public health and prospects for malaria control in Brazil. *Journal of Infection in Developing Countries*, 4: 533-545.
- Hang-Hai, D.; Erhard, R. 2000. Data Cleaning: Problems & Current Approaches. *IEEE bulletin of the technical committee on Data Engineering*, 23.
- Hermes, S.C.N.M.; Nunes, V.L.B.; Dorval, M.E.C.; Brillhante, A.F. 2013. Aspectos epidemiológicos da malária humana no município de aripuanã, estado de Mato Grosso, Brasil, 2005 a 2010. *Hygeia*, 9: 42-51.
- Jonge, E.D.; van der Loo, M. 2013. An introduction to data cleaning with R. Statistics Netherlands, *The Hague*, 53p.
- Martens, P.; Kovats, R.S.; Nijhof, S.; Vries, P.; Livermore, M.T.J.; Bradley, D.J.; Cox, J.; McMichael, A.J. 1999. Climate change and future populations at risk of malaria. *Global Environmental Change*, 9: 89-107.
- Oliveira-Ferreira, J.; Lacerda, M.V.G.; Brasil, P.; Ladislau, J.L.B.; Tauil, P.L.; Ribeiro, C.T.D. 2010. Malaria in Brazil: An overview. *Malaria Journal*, 9: 115
- Peiter, P.C.; Franco, V.C.; Gracie, R.; Xavier, D.R.; Mutis, M.C.S. 2013. Situação da malária na tríplice fronteira entre Brasil, Colômbia e Peru. *Cadernos de Saúde Pública*, 29: 2497-2512.
- Pina-Costa, A.; Brasil, P.; Santi, S.; Araujo, M. P.; Suárez-Mutis, M. C.; Santelli, A. C. F. S. 2014. Malaria in Brazil: what happens outside the Amazonian endemic region. *Memórias do Instituto Oswaldo Cruz*, 109: 618-633.
- Ray, A. P.; Beljaev, A. E. 1984. Epidemiological surveillance: a tool for assessment of malaria and its control. *The Journal of communicable diseases*, 16: 197-207.
- Sampaio, V.S.; Siqueira, A.M.; Alecrim, M.G.C.; Mourão, M.P.G.; Marchesini, P.B.; Albuquerque, B.C.; et al. 2015. Malaria in the State of Amazonas: A typical Brazilian tropical disease influenced by waves of economic development. *Revista da Sociedade Brasileira de Medicina Tropical*, 48: 4-11.
- Schoenbach, V.J.; Rosamond, W.D (eds). 2000. Understanding the fundamentals of epidemiology: An evolving text. Chapel Hill, NC: University of North Carolina at Chapel Hill. *Fall 2000 Edition*, 481p.
- Silva, N. S.; Silva-Nunes, M.; Malafronte, R.S.; Menezes, M.J.; D'Arcadia, R.R.; Komatsu, N.T.; Scopel K. K. G.; Braga, E. M.; Cavasini, C. E.; Cordeiro, J. A.; Ferreira, M.U. 2010. Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104: 343-50.
- Tauil, P.L.; Daniel-Ribeiro, C. 1998. Some aspects of epidemiology and control of malaria in Brazil. *Research and Reviews in Parasitology*, 58:163-167.
- World Health Organization. 2012. Disease surveillance for malaria control. An operational manual. *WHO Press*, Geneva, 84p.

Recebido em 08/02/2016
Aceito em 13/05/2016