



HAL
open science

Analyses spatiales de problèmes de santé publique en Afrique subsaharienne : exemples du VIH/SIDA et de la malnutrition.

Emmanuel Barankanira

► To cite this version:

Emmanuel Barankanira. Analyses spatiales de problèmes de santé publique en Afrique subsaharienne : exemples du VIH/SIDA et de la malnutrition.. Médecine humaine et pathologie. Université Montpellier, 2016. Français. NNT : 2016MONTT048 . tel-01496782

HAL Id: tel-01496782

<https://theses.hal.science/tel-01496782>

Submitted on 27 Mar 2017

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
Sciences Chimiques et Biologiques pour la Santé
(CBS2)**

**Et de l'unité de recherche TransVIHMI, IRD UMI 233,
INSERM U1175, Université de Montpellier**

Spécialité : Biostatistique

Présentée par Emmanuel BARANKANIRA

**Analyses spatiales de problèmes de santé
publique en Afrique subsaharienne :
exemples du VIH/SIDA et de la
malnutrition**

Soutenue le 18 novembre 2016 devant le jury composé de

M. René ECOCHARD, PU-PH, Université de Lyon
M. Ahmadou ALIOUM, PU, Université de Bordeaux
Mme Kirsten BORK, DR, IRD Montpellier
M. Renaud BECQUET, CR, INSERM Bordeaux
M. Christian LAURENT, DR, IRD Montpellier
M. Nicolas MOLINARI, MCU-PH, Université de Montpellier

Président du jury
Rapporteur
Examineur
Examineur
Directeur de thèse
Co-directeur de thèse



Dédicace

À ma chère épouse Sandrine KWIZERA,

À mon frère Tharcisse NIYIBITANGA,

À ma regrettée mère Amélie NZINAHORA,

À mon regretté père Simon NTAMATUNGIRO.

Remerciements

Les travaux de thèse ont été réalisés au sein de l'Unité Mixte Internationale 233 Transitions épidémiologiques, recherches translationnelles appliquées au VIH et aux maladies infectieuses, dénommée UMI 233 TransVIHMI, de l'Institut de Recherche pour le Développement (IRD) de Montpellier. Ils ont été financés par le Gouvernement du Burundi, *via le Ministère de l'Enseignement Supérieur et de la Recherche Scientifique*, et l'École Normale Supérieure (ENS) de Bujumbura.

Au terme de cette thèse, il m'est un agréable plaisir de remercier toutes les personnes sans lesquelles elle n'aurait pas pu avoir lieu. Que les personnes qui ne se retrouveront pas à travers ces lignes se souviennent que « qui cite se trompe ».

À vous, Monsieur le Professeur Eric DELAPORTE, Directeur de l'UMI 233 et responsable de l'équipe Recherche opérationnelle sur le VIH et Maladies Infectieuses, merci de m'avoir accueilli et intégré dans votre unité et de m'avoir facilité les procédures administratives lors de mes inscriptions.

Au Docteur Christian LAURENT, merci d'avoir accepté de diriger cette thèse. Je vous étais une personne inconnue, mais vous avez accepté avec enthousiasme et sans ambage de me lancer dans cette aventure qu'est la thèse. Je vous remercie également d'avoir gardé un bon souvenir du Burundi, mon pays natal, de m'avoir aidé à peaufiner le sujet de thèse, de m'avoir laissé autonome durant toute la durée de la thèse et de m'avoir facilité les démarches administratives chaque fois que de besoin. Merci de m'avoir intégré dans l'équipe « *Recherches clinique et opérationnelle* » et d'avoir créé un environnement favorable et propice pour la réalisation de cette thèse. Je n'oublierai jamais ces courtes mais réconfortantes phrases que vous n'avez cessé de prononcer chaque fois qu'une bonne idée me venait en tête: « *ça roule* » et chaque fois que nous tardions à quitter la cantine pour nous remettre au travail « *La science n'attend pas* ». Merci pour la confiance que vous avez placée en moi tout au long cette thèse. Merci infiniment pour votre accompagnement précieux et pour votre soutien indéfectible.

Au Docteur Nicolas MOLINARI, merci beaucoup d'avoir accepté de me prendre en thèse à la fin du master de Biostatistique et de m'avoir trouvé un directeur de thèse, **Docteur Christian LAURENT**. Ce dernier, pour avoir exploré des thématiques de santé publique en Afrique et plus particulièrement au Burundi, a fait preuve d'un rattachement fort et perpétuel avec mon pays. Merci d'avoir su m'orienter pendant des moments délicats lors de mon admission en thèse. Merci sincèrement d'avoir accepté de co-diriger cette thèse.

Au Docteur Kirsten BORK, un tout grand merci pour avoir su m'orienter vers la thématique liée à malnutrition en Côte d'Ivoire au moment où il me manquait les données des enquêtes nationales de séroprévalence du VIH et celles de surveillance sentinelle du VIH/SIDA au Burundi. Merci infiniment de votre encadrement et des conseils oh combien judicieux que vous n'avez cessé de me prodiguer pour la réalisation de cette thèse.

Au Docteur Philippe MSELATTI, merci pour toutes les discussions que nous avons eues ensemble sur la géopolitique de la Côte d'Ivoire et pour toutes les informations que vous m'avez fournies. Vous avez su me faire profiter de vos expériences en Côte d'Ivoire et au Rwanda. Vous m'avez rappelé ma plus belle chanson apprise au collège « *Donne du rhum à ton homme* » de Georges Moustaki. C'est par vous que j'ai appris que, quoique ce soit en écrivant que viennent les idées, il faut « *savoir arrêter le travail* » pour aller approvisionner le cerveau en glucose. Vous m'avez épaté de la maîtrise des langues parlées au Burundi à savoir le kirundi et le swahili.

A vous, Messieurs les membres du jury, merci d'avoir accepté de lire le manuscrit de cette thèse malgré vos multiples occupations et de participer à son évaluation. Un tout « grand Merci » aux Professeurs **René ECOCHARD** et **Ahmadou ALIOUM** pour avoir accepté d'être rapporteurs de cette thèse et particulièrement au Professeur **René ECOCHARD** pour avoir accepté d'être président du jury. Merci également pour vos remarques oh combien pertinentes afin d'améliorer le manuscrit de cette thèse. **Aux Docteurs Kirsten BORK et Renaud BECQUET**, merci d'avoir accepté de lire cette thèse et d'être examinateurs de cette thèse.

Aux membres de l'UMI 233 : Coralie SIGOUNIOS, Julie PUJET, Cédric CHESNAIS, Sophie SERVAIS, Marie VARLOTEAUX, Laura MARCH, Clotilde COUDERC, Ndeye Fatou NGOM GUEYE, Odette KY ZERBO, Sabrina EYMARD-DUVERNAY, Bruno GRANOULLIAC, Véronique BOYER, Suzanne IZARD, Amandine COURNIL, Michel BOUSSINESQ, Sébastien PION, merci de vos encouragements. Merci à **Coralie SIGOUNIOS** pour vous être toujours préoccupée de la climatisation de mon bureau (fonctionnement de la clim, achat de ventilateur neuf) et de l'achat de mes ordinateurs de bureau.

Tout particulièrement, je remercie **Docteur Clotilde COUDERC** pour le temps qu'elle a consacré à l'organisation des *monthly drinks*, au choix des images de circonstance placées à la fin de chaque annonce envoyée par mail et des beaux endroits tels que *Le Clandestin*, *Les tanins*, chez Amandine COURNIL, *La Barbote*, chez Sabrina EYMARD-DUVERNAY, chez Coralie SIGOUNIOS, au *By Coss Bar*, ..., ce qui a permis de rassembler les membres de notre équipe « *Recherche opérationnelle sur le VIH et les Maladies Infectieuses* » autour d'un verre afin de nous égayer. Merci pour toutes ces photos prises et affichées dans nos services, merci de nous avoir mis dans la logique des apéros déjeunatoires et pour toutes vos propositions diverses afin de changer nos petites sorties mensuelles.

Au Docteur Janvière NDIRAHISHA, Ministre de l'Éducation, de l'Enseignement Supérieur et de la Recherche Scientifique (Burundi), ex-Directeur Général de l'École Normale Supérieure (Bujumbura) ainsi qu'au **Docteur Julien NIMUBONA**, ex-Ministre de l'Enseignement Supérieur et de la Recherche Scientifique, merci de m'avoir permis de poursuivre mes études de troisième cycle.

À mes chers amis Karamoko BEAVOGUI, Boukary Elhadj CAMARA, Claver MPAWENAYO et Serge NDAYIRATA, et à mon très cher garant Lambert BIZABISHAKA, vous avez rendu agréable mon séjour à Montpellier.

Enfin, je remercie vivement **le Gouvernement du Burundi** qui a financé cette formation en France.

Résumé

Cette thèse porte sur l'analyse spatiale de problèmes de santé publique rencontrés en Afrique subsaharienne. Dans une première étude, nous avons analysé les variations spatiales de la prévalence du VIH au Burundi et les déterminants socio-comportementaux chez les personnes âgées de 15 à 49 ans à partir des données de l'Enquête Démographique et de Santé (EDS) de 2010. Dans une deuxième étude, nous avons analysé l'évolution de la répartition spatiale de la prévalence du retard de croissance chez les enfants âgés de moins de 3 ans en Côte d'Ivoire à partir des données des EDS de 1994, 1998-1999 et 2011-2012. Pour ces deux études, nous avons tout d'abord prédit en des points non échantillonnés les niveaux de prévalence du problème étudié, soit par la méthode des noyaux, soit par la méthode du krigeage ordinaire. Nous avons ensuite procédé à la détection des agrégats spatiaux ou clusters par la méthode de Kulldorff puis à l'identification de zones de différents risques par la méthode du partitionnement oblique du territoire. Finalement, nous avons utilisé la modélisation spatiale pour identifier les facteurs associés. Ces techniques d'analyse spatiale présentent de nombreux avantages mais également de nombreuses limites que nous soulignons.

Mots clés : Analyse spatiale, VIH, retard de croissance, Burundi, Côte d'Ivoire.

Summary

This thesis is about the spatial analysis of public health issues in sub-Saharan Africa. In a first study, we analysed the spatial variations of HIV prevalence and the socio-behavioral determinants among people aged 15 to 49 in Burundi using 2010 Demographic and Health Survey (DHS) data. In a second study, we analysed the evolution of the spatial distribution of stunting prevalence among under-three children in Côte d'Ivoire using 1994, 1998-1999 and 2011-2012 DHS data. For both studies, we first predicted the prevalence levels of the public health issue under study in unsampled points by the kernel method or ordinary kriging. We then identified clusters using the Kulldorff's method and zones of different risks using the oblique partitioning method of the study area. Finally, we used spatial modeling to identify associated risk factors. These spatial analysis techniques have not only many advantages but also many limitations that we emphasize.

Key words: Spatial analysis, HIV, stunting, Burundi, Côte d'Ivoire.

Liste des publications et communications associées à la thèse

Article publié

Emmanuel Barankanira, Nicolas Molinari, Théodore Niyongabo and Christian Laurent. Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention. *BMC Public Health*. 2016; 16: 118.

Article soumis

Emmanuel Barankanira, Nicolas Molinari, Philippe Msellati, Christian Laurent, and Kirsten A Bork. Stunting among children under three in Côte d'Ivoire: spatial and temporal variations between 1994 and 2011. *Public Health Nutrition* (article en révision).

Poster

Emmanuel Barankanira, Nicolas Molinari, Théodore Niyongabo, Christian Laurent. Hétérogénéité géographique de la prévalence du VIH au Burundi : des indications utiles pour une prévention efficace de l'infection. Poster L10.7. 7^{ème} Conférence Internationale Francophone sur le VIH et les hépatites AFRAVIH 2014, 27-30 avril 2014, Montpellier, France.

Communication

Emmanuel Barankanira, Nicolas Molinari, Théodore Niyongabo, Christian Laurent, Au-delà des chiffres des Enquêtes Démographiques et de Santé (EDS) : Exemple d'une étude au Burundi, 9^{èmes} Rencontres Nord-Sud IMEA-IRD, 26 décembre 2014, Paris, France.

Liste des abréviations et sigles

ABCMAV	Association burundaise pour la prise en charge des malades vulnérables
ABS	Alliance Burundaise contre le SIDA et pour la promotion de la santé
ACF	Action contre la faim
ACP	Analyse en composantes principales
AG	Assemblée générale
AMS	Assemblée mondiale de la santé
ANSS	Association nationale des séropositifs et des sidéens
AP/SIDA	Appui au programme national de lutte contre le SIDA
APECOS	Association de Prise En Charge des Orphelins du SIDA
APJP	Association des Pères Jésuites du Burundi
APO	Accord Politique de Ouagadougou
ARK	Animation Rurale de Korhogo
ARO	Arbre de régression oblique
ARV	Antirétroviral
AVISI	Association d'Aide aux Victimes du SIDA
BCS	Bureau communal de santé
BDS	Bureau de district sanitaire
BIC	Bayesian Information Criterion
BPS	Bureau de la province sanitaire
CAPES+	Collectif des Associations des Personnes vivant avec le VIH/SIDA
CART	Arbre de classification et de régression
CBS2	Sciences Chimiques et Biologiques pour la Santé
CDS	Centre de santé
CDV	Centres de dépistage volontaire
CE	Comité exécutif
SFPS	Programme Santé Familiale et Prévention du SIDA
CHUK	Centre hospitalo-universitaire de Kamenge
CLS	Comités locaux de lutte contre le SIDA
CMLS	Comités municipaux de lutte contre le SIDA
CNAR	Centre National d'Appareillage et de Réadaptation
CNN	Comité National de Nutrition
COCOLS	Comités communaux de lutte contre le SIDA

CNPK	Conseil neuro-psychiatrique de Kamenge
COSA	Comité de santé
CPLR	Clinique Prince Louis Rwagasore
CRN	Comités Régionaux de Nutrition
CNAR	Centre National d'Appareillage et de Réadaptation
CMLS	Comités municipaux de lutte contre le SIDA
CNC	Centres nutritionnels communautaires
CNLS	Comité national de lutte contre le SIDA
CNN	Comité National de Nutrition
CNO	Centre, Nord et Ouest
CNS	Centres nutritionnels supplémentaires
CNT	Centres nutritionnels thérapeutiques
CPLS	Comité provincial de lutte contre le SIDA
DALYs	Années de vie ajustées sur l'incapacité
DC	Direction de coordination
DRSHP	Directions régionales de la santé et de l'hygiène publique
DR	District de recensement
DSCMP	Direction de la santé communautaire et de la médecine de proximité
EAC	Communauté de l'Afrique de l'Est
ED	European Datum
EDS	Enquête démographique et de santé
ELISA	Enzyme-linked immunosorbent assay
EPSCG	European Petroleum Survey Group
EPVA	Encadrement des personnes vulnérables en Afrique
FAO	Organisation des Nations Unies pour l'Alimentation et l'Agriculture
FARN	Foyer d'apprentissage et de réhabilitation nutritionnelle
FOBULUSI	Forum burundais de lutte contre le VIH/SIDA
FOSA	Formations sanitaires
FVS	Famille pour vaincre le sida
GLIA	Initiative des pays des Grands-Lacs contre le SIDA
GPS	Géo-positionnement par satellite
GSA	Groupe Scientifique d'Appui
HDS	Hôpital de district sanitaire
HAZ	Height-for-age z-score
HMK	Hôpital Militaire de Kamenge

HPRC	Hôpital Prince Régent Charles
HSH	Hommes ayant des rapports sexuels avec d'autres hommes
IC	Intervalle de confiance
IDH	Indice du développement humain
IDW	Pondération par l'inverse de la distance
IGN	Institut national de l'information géographique et forestière
IMC	Indice de masse corporelle
INS	Institut National de Statistique
INSP	Institut National de Santé Publique
IRD	Institut de Recherche pour le Développement
IST	Infections sexuellement transmissibles
ISTEEBU	Institut des statistiques et d'études économiques du Burundi
LISA	Indicateurs locaux d'association spatiale
MCG	Méthode des moindres carrés généralisés
MCMC	Monte Carlo par chaînes de Markov
MCO	Moindres carrés ordinaires
MCP	Moindres carrés pondérés
MHSP	Ministère de la Santé et de l'Hygiène Publique
MICS	Enquête de grappes à Indicateurs Multiples
MMV	Méthode du maximum de vraisemblance
MSLS	Ministère de la Santé et de la Lutte contre le SIDA
MSP	Ministère de la santé publique
MSPLS	Ministère de la santé publique et de la lutte contre le SIDA
MPLS	Ministère à la Présidence chargé de la lutte contre le SIDA
MST	Maladies sexuellement transmissibles
MTE	Modèle numérique d'élevation
MTN	Modèle numérique de terrain
MVR	Maximum de vraisemblance restreinte
NCHS	National Center for Health Statistics
ODD	Objectifs du développement durable
OEV	Orphelins et autres enfants vulnérables
OMD	Objectifs du millénaire pour le développement
OMS	Organisation mondiale de la santé
ONUSIDA	Programme commun des Nations Unies sur le VIH/SIDA
PAM	Programme Alimentaire Mondial
PANLS	Plan d'Action Nationale de Lutte contre le SIDA

PEV	Programme élargi de vaccination
PIEN	Programme intégré d'éducation nutritionnelle
PMA	Paquet minimum d'activités
PMLSAO	Projet multisectoriel de lutte contre le SIDA et d'appui aux orphelins
PMT	Premier plan à Moyen Terme
PND	Plan National pour le Développement Sanitaire
PNLS	Programme national de lutte contre le SIDA
PNN	Programme national de nutrition
PNS	Plan National Stratégique
PNSI/SE	Programme national de santé infantile et de survie de l'enfant
PNSSU	Programme National de la Santé Scolaire et Universitaire
PNUD	Programme de Nations Unies pour le Développement
PRIDE	Projet d'intensification et de décentralisation de la lutte contre le SIDA
PSNLS	Plan stratégique national de lutte contre le SIDA
PTME	Prévention de la transmission mère-enfant
PvVIH	Personnes vivant avec le virus de l'Immunodéficience humaine
QGIS	Quantum Geographic Information System
RBP+	Réseau burundais des personnes vivant avec le VIH
RC	Rapport de cotes
RCA	République Centrafricaine
RDC	République Démocratique du Congo
RGPH	Recensement général de la population et de l'habitat
RN	Route nationale
RNDH	Rapport National sur le Développement Humain
RP	Route poteau
SCR	Système de coordonnées de référence
SEP/CNLS	Secrétariat Exécutif Permanent du CNLS
SFPS	Programme Santé Familiale et Prévention du SIDA
SFPT	Société française de photogrammétrie et télédétection
SIDA	Syndrome de l'Immunodéficience Acquise
SIG	Système de l'information géographique
SNLS	Service national de lutte contre le SIDA
spODT	Spatial Oblique Decision Tree
SUN	Scaling Up Nutrition

SWAA-Burundi	Society for Women Against AIDS in Africa-Burundi
SYM	Service Yezu Mwiza
TR	Taux de réponse
TransVIHMI	Transitions épidémiologiques, recherches translationnelles appliquées au VIH et aux maladies infectieuses
UNGASS	United Nations General Assembly Special Session
UNDP	United Nations Development Programme
UNICEF	Fonds des Nations Unies pour l'enfance
USLS	Unités Sectorielles de lutte contre le SIDA
UMI	Unité Mixte Internationale
UTM	Mercator Transverse Universelle
VIH	Virus de l'Immunodéficience humaine
WGS	Système géodésique mondial
ZD	Zone de dénombrement

Liste des figures

	Page
Figure 1 : Couches vectorielles du Burundi.....	8
Figure 2 : Couche raster du Burundi	13
Figure 3 : Système de coordonnées de référence mondiale	16
Figure 4 : Projection cylindrique directe	17
Figure 5 : Projection cylindrique transverse	17
Figure 6 : Projection conique directe.....	18
Figure 7 : Projection plane polaire directe.....	19
Figure 8 : Spatialisation de l'utilisation des systèmes de projection dans le monde.....	20
Figure 9 : Prévalence du VIH par grappe au Burundi, EDS 2010	22
Figure 10 : Prévalence du VIH par province au Burundi, EDS 2010	23
Figure 11 : Graphe de voisinage entre les provinces du Burundi.....	28
Figure 12 : Graphe de voisinage entre les grappes, Burundi EDS 2010.....	29
Figure 13 : Corrélogramme de la prévalence du retard de croissance en Côte d'Ivoire, 1994 avec comme distance seuil 100 km.....	35
Figure 14 : Corrélogramme de la prévalence du VIH au Burundi, 2010 avec comme distance seuil 80 km	36
Figure 15 : Nuage de points de Moran de la prévalence du retard de croissance, Côte d'Ivoire 1994	45
Figure 16 : Évolution de la prévalence du VIH selon les coordonnées géographiques au Burundi, 2010	53
Figure 17 : Tendence spatiale de la prévalence du VIH au Burundi, 2010.....	54
Figure 18 : Évolution des résidus du modèle de régression linéaire de la prévalence du VIH et des coordonnées géographiques (polynome d'ordre 1).....	55
Figure 19 : Tendence spatiale des résidus du modèle de régression linéaire de la prévalence du VIH et des coordonnées géographiques (polynome d'ordre 1)	56
Figure 20 : Paramètres du variogramme expérimental	58
Figure 21 : Nuée variographique de la prévalence du VIH, Burundi 2010.....	60

Figure 22 : Nuée variographique de la prévalence du VIH avec couples aberrants, Burundi 2010.....	61
Figure 23 : Variogramme expérimental de la prévalence du VIH, Burundi 2010.....	65
Figure 24 : Enveloppes du variogramme expérimental de la prévalence du VIH, Burundi 2010.....	65
Figure 25 : Variogrammes directionnels de la prévalence du VIH, Burundi 2010.....	66
Figure 26 : Estimation par la méthode des noyaux.....	72
Figure 27 : Nombre estimé d'adultes et d'enfants vivant avec le VIH dans le monde en 2015.....	88
Figure 28 : Nombre estimé d'adultes et d'enfants nouvellement infectés par le VIH dans le monde en 2015.....	88
Figure 29 : Nombre estimé d'adultes et d'enfants décédés des causes liées au VIH/SIDA dans le monde en 2015.....	89
Figure 30 : Prévalence du VIH selon le sexe dans la région des Grands Lacs.....	92
Figure 31 : Prévalence du VIH selon le milieu de résidence dans la région des Grands Lacs.....	92
Figure 32 : Évolution du nombre d'adultes et d'enfants vivant avec le VIH dans la région des Grands Lacs, 1990-2015.....	93
Figure 33 : Évolution du nombre d'adultes et d'enfants nouvellement infectés par le VIH dans la région des Grands Lacs, 1990-2015.....	94
Figure 34 : Évolution du nombre de décès pour des causes liées au SIDA dans la région des Grands Lacs, 1990-2015.....	94
Figure 35 : Distribution spatiale de la prévalence du VIH en Afrique en 2015.....	95
Figure 36 : Situation du Burundi en Afrique subsaharienne.....	96
Figure 37 : Répartition spatiale des régions sanitaires et des provinces au Burundi en 2015.....	98
Figure 38 : Organigramme du PNLIS/IST au Burundi.....	108
Figure 39 : Évolution du nombre de personnes sous traitement ARV au Burundi, 2002-2015.....	111
Figure 40 : Évolution du nombre de sites de traitement ARV au Burundi, 2005-2015.....	111
Figure 41 : Évolution du nombre de personnes dépistées au Burundi, 2002-2015.....	112

Figure 42 : Évolution du nombre de sites PTME au Burundi, 2006-2015.....	113
Figure 43 : Évolution du nombre de mères sous PTME au Burundi, 2002-2015.....	113
Figure 44 : Prévalence du VIH selon les régions au Burundi, 2010	118
Figure 45 : Répartition géographique des grappes de l'EDS 2010 du Burundi	120
Figure 46 : Histogramme et boxplot de la prévalence du VIH par grappe	124
Figure 47 : Partitionnement oblique du Burundi en zones de différents risques.....	126
Figure 48 : Modes a posteriori des effets spatiaux.....	134
Figure 49 : Modes a posteriori des paramètres.....	135
Figure 50 : Prévalence du retard de croissance dans le monde, 2011	154
Figure 51 : Prévalence de l'émaciation dans le monde, 2011	155
Figure 52 : Nombre estimé d'enfants de moins de 5 ans souffrant du retard de croissance dans le monde selon les régions des Nations Unies, 2014	156
Figure 53 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant de retard de croissance dans le monde et par région, 1990-2014.....	157
Figure 54 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant de retard de croissance en Afrique subsaharienne, 1990-2014	159
Figure 55 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant du retard de croissance en Afrique subsaharienne par région, 1990-2014	160
Figure 56 : Prévalence du retard de croissance chez les enfants de moins de 5 ans en Afrique de l'Ouest selon le milieu de résidence.....	161
Figure 57 : Prévalence de l'émaciation chez les enfants de moins de 5 ans en Afrique de l'Ouest selon le milieu de résidence.....	162
Figure 58 : Situation de la Côte d'Ivoire en Afrique subsaharienne.....	164
Figure 59 : Répartition spatiale des zones de savane et des zones de forêt en Côte d'Ivoire, 1994	165
Figure 60 : Position de la zone de confiance, flux et zones de concentration de populations déplacées de guerre	167

Figure 61 : Évolution de l'IDH selon les régions, 2002-2011	169
Figure 62 : Répartition géographique des grappes de l'EDS de 1994 selon le lieu de résidence	177
Figure 63 : Répartition géographique des grappes de l'EDS 1998 selon le lieu de résidence	178
Figure 64 : Répartition géographique des grappes de l'EDS 2011 selon le lieu de résidence	179
Figure 65 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 1994	180
Figure 66 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 1998	181
Figure 67 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 2011	182
Figure 68 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 1994	187
Figure 69 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 1998	187
Figure 70 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 2011	188

Liste des tableaux

	Page
Tableau 1 : Table attributaire de la couche vectorielle de polygones (provinces) du Burundi	9
Tableau 2 : Table attributaire de la couche vectorielle de points (grappes) du Burundi	11
Tableau 3 : Table attributaire de la couche vectorielle de polygones (routes) du Burundi.....	12
Tableau 4 : Prévalence du VIH selon le sexe et le milieu de résidence dans la région des Grands Lacs obtenue à partir des données des EDS.....	91
Tableau 5 : Programmes et principaux intervenants de la lutte contre le VIH/SIDA au Burundi.....	103
Tableau 6 : Prévalence du VIH estimée lors des enquêtes nationales	115
Tableau 7 : Prévalence du VIH par province chez les résidents âgés de 15 à 49 ans au Burundi en 2010.....	125
Tableau 8 : Comparaison des risques dans les zones	127
Tableau 9 : Prévalences du retard de croissance et de l'émaciation chez les enfants de moins de 5 ans en Côte d'Ivoire et dans les pays limitrophes.....	160
Tableau 10 : Qualité de l'ajustement des modèles logistiques dans les zones de la Côte d'Ivoire	189

Table des matières

	Page
Dédicace.....	i
Remerciements.....	ii
Résumé.....	v
Summary.....	vi
Liste des publications et communications associées à la thèse.....	vii
Liste des abréviations et sigles.....	viii
Liste des figures.....	xiii
Liste des tableaux.....	xvii
Chapitre 1 : Généralités sur l'analyse spatiale.....	3
1.1. Introduction à l'analyse spatiale.....	4
1.1.1. Définition et système d'information géographique (SIG).....	4
1.1.2. Analyse spatiale et statistique spatiale.....	5
1.1.3. Notion de couches et sources des données.....	7
1.1.4. Projections et systèmes de projection.....	14
1.1.5. Type de données spatiales.....	20
1.2. Définition des poids entre les unités spatiales.....	25
1.2.1. Matrice des distances spatiales.....	25
1.2.2. Graphe de voisinage entre les points.....	28
1.3. Autocorrélation spatiale.....	30
1.3.1. Indices globaux d'autocorrélation spatiale.....	31
1.3.2. Indices locaux d'autocorrélation spatiale.....	42
1.4. Techniques d'interpolation spatiale.....	51
1.4.1. Définition de l'interpolation spatiale, limites et utilisation.....	51
1.4.2. Analyse de la tendance spatiale.....	52
1.4.3. Analyse variographique.....	56
1.4.4. Méthodes d'interpolation spatiale.....	68
1.4.5. Krigeage ordinaire.....	69
1.4.6. Interpolation spatiale par la méthode des noyaux.....	70

1.5.	Modélisation spatiale	73
1.5.1.	Définition d'un modèle spatial	74
1.5.2.	Définition d'un modèle spatial généralisé.....	76
1.5.3.	Les modèles de régression spatiale binomial et de Poisson.....	76
1.6.	Méthodes statistiques d'analyse spatio-temporelle.....	82
Chapitre 2 :	Analyse spatiale du VIH/SIDA au Burundi.....	85
2.1.	Épidémiologie de l'infection par le VIH dans le monde	86
2.2.	Épidémiologie de l'infection par le VIH dans la région des Grands Lacs	90
2.3.	Contexte du Burundi	96
2.3.1.	Situation géographique, administrative et démographique	96
2.3.2.	Organisation du système de santé.....	99
2.3.3.	Historique de la lutte contre le VIH/SIDA	101
2.3.4.	Organisation de la lutte contre le VIH/SIDA	107
2.3.5.	Politique nationale de lutte contre le SIDA.....	109
2.3.6.	Épidémiologie de l'infection par le VIH au Burundi	114
2.4.	Contexte de notre recherche au Burundi	116
2.5.	Objectif de l'étude	118
2.6.	Méthodologie	119
2.6.1.	Sélection de la population d'étude	119
2.6.2.	Données géographiques.....	122
2.6.3.	Analyses statistiques	123
2.6.4.	Explicitation du modèle spatial utilisé dans l'étude	128
2.7.	Article « Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention »	135

Chapitre 3 :	Analyse spatiale et temporelle du retard de croissance en Côte d'Ivoire.....	147
3.1.	Introduction	148
3.2.	Anciennes références et nouvelles normes de croissance de l'OMS	151
3.3.	Situation nutritionnelle dans le monde	152
3.3.1.	Période des 1000 jours, et la tranche d'âge 0-5 ans.....	152
3.3.2.	Sous-alimentation et retard de croissance dans le monde	153
3.4.	Situation nutritionnelle en Afrique subsaharienne.....	158
3.5.	Situation nutritionnelle en Côte d'Ivoire et dans les pays limitrophes	160
3.6.	Contexte de la Côte d'Ivoire.....	163
3.6.1.	Situation géographique, administrative et démographique	163
3.6.2.	Recensements de population en Côte d'Ivoire	165
3.6.3.	Crise ivoirienne et ses conséquences.....	166
3.6.4.	Lutte contre la malnutrition en Côte d'Ivoire.....	170
3.7.	Contexte de notre recherche en Côte d'Ivoire.....	172
3.8.	Objectifs.....	173
3.9.	Méthodologie	173
3.9.1.	Sélection de la population d'étude	173
3.9.2.	Diagrammes des flux des enfants de 0-35 mois	179
3.9.3.	Définition des variables.....	183
3.9.4.	Analyse statistique	185
3.10.	Article « Stunting among children under three years in Côte d'Ivoire: spatial and temporal variations between 1994 and 2011 »	190
	Conclusion générale.....	225
	Références bibliographiques.....	228

Introduction générale

L'analyse spatiale des problèmes de santé publique est encore insuffisamment réalisée, malgré un intérêt croissant au cours des dernières années. À l'inverse, l'analyse spatiale est fréquemment utilisée en économétrie et dans l'estimation des ressources naturelles et minéralogiques. Les méthodes statistiques spatiales, outils de l'épidémiologie spatiale, permettent de décrire, d'expliquer et de modéliser les variations géographiques d'un phénomène d'intérêt, celui-ci étant décrit par des coordonnées géographiques ou toute autre caractéristique permettant de localiser dans l'espace les individus enquêtés.

Contrairement à la statistique classique, la statistique spatiale offre la possibilité de prendre en compte la localisation précise des mesures, d'interpoler des valeurs à des points non observés et de s'affranchir des limites géographiques administratives. Par ailleurs, la statistique spatiale autorise la dépendance spatiale des observations alors que la statistique classique suppose généralement leur indépendance.

Les enquêtes démographiques et de santé (EDS) représentent une formidable source de données sanitaires géolocalisées, malheureusement sous-exploitées. Ce sont des enquêtes internationales réalisées en population générale dans les pays en développement depuis 1985 par la société américaine Inner City Fund (ICF International) en collaboration avec des instituts nationaux de statistiques et/ou des ministères de la santé, et avec l'appui financier de l'agence des États-Unis pour le développement international (USAID). Elles permettent de recueillir de nombreuses données relatives aux caractéristiques des ménages et des résidents, à la fécondité et à la planification familiale, à la santé de la reproduction, à la santé et à l'état nutritionnel des mères et des enfants, à la mortalité des enfants et des adultes, au paludisme et au syndrome de l'immunodéficience acquise (SIDA). Ces enquêtes sont répétées dans le temps selon les besoins et les contextes sociopolitiques des pays. Utilisant une méthode de sondage aléatoire en grappes à deux (ou parfois trois) degrés, elles sont réalisées dans un échantillon représentatif de la population.

Cette thèse porte sur l'analyse spatiale de problèmes de santé publique en Afrique subsaharienne. Plus précisément, nous avons réalisé des analyses statistiques à l'aide de méthodes spatiales de deux problèmes de santé publique majeurs sur ce continent, l'infection par le virus de l'immunodéficience humaine (VIH) et la malnutrition, à partir des données d'EDS. L'objectif de ces travaux était de compléter les informations disponibles pour aider les décideurs de santé publique à définir leurs programmes de lutte correspondants.

La présentation de cette thèse s'articule autour de trois chapitres.

Le premier chapitre synthétise les principaux outils statistiques d'analyse spatiale, ses applications et les techniques d'analyse disponibles.

Le deuxième chapitre est consacré à l'analyse spatiale de l'infection par le VIH au Burundi (mon pays). Les variations spatiales de l'infection par le VIH chez les personnes de 15 à 49 ans ont été étudiées à l'aide des données de l'EDS de 2010, la dernière EDS réalisée au Burundi et la seule disposant simultanément de données biologiques sur le VIH et des coordonnées géographiques des personnes enquêtées.

Le troisième et dernier chapitre est consacré à l'analyse spatiale et temporelle du retard de croissance en Côte d'Ivoire. Nous avons voulu ici compléter l'analyse spatiale par une analyse temporelle et avons étudié les modifications de la distribution spatiale du retard de croissance au cours du temps. Cette étude était justifiée par la guerre civile qui a sévi dans le pays de 2002 à 2011. Les variations spatiales et temporelles du retard de croissance chez les enfants de moins de trois ans ont été étudiées à l'aide des données des EDS réalisées successivement en 1994, 1998 et 2011.

Une conclusion générale clôture cette thèse.

Chapitre 1 : Généralités sur l'analyse spatiale

Ce chapitre s'intéresse à la revue générale de la littérature sur les notions de base permettant d'aborder l'analyse spatiale. Il expose, de manière non exhaustive, les outils et les méthodes à mettre en œuvre afin de répondre à une problématique posée en analyse spatiale, ces méthodes englobant des techniques spatiales exploratoires d'une part et inférentielles d'autre part.

1.1. Introduction à l'analyse spatiale

1.1.1. Définition et système d'information géographique (SIG)

Les données décrites par des coordonnées géographiques sont appelées données géoréférencées, données géospatisées ou données géolocalisées et sont stockées dans des bases données relationnelles. Leur recueil, leur classement, leur stockage, leur gestion et leur géotraitement requièrent la connaissance des outils informatiques adaptés et des techniques d'analyse appropriées. Les structures contenant ces données peuvent être visualisées par des logiciels variés qui prennent en compte un tel type de données. Lors de la visualisation que nous verrons plus loin, elles peuvent être superposées afin d'avoir une idée a priori sur la distribution spatiale du phénomène d'intérêt ou des variables connexes. L'étude de la position de ces structures dans un espace à deux ou à trois dimensions et des relations existant entre elles constitue l'objet de l'analyse spatiale. Les données géographiques, c'est-à-dire les données indexées par des coordonnées géographiques (longitude/latitude), sont collectées, traitées puis traduites en information géographique. Une information géographique est un ensemble de données géolocalisées traduites en information utile afin d'aider les décideurs à prendre des décisions. Une discipline constituée des outils et des techniques permettant ces opérations de recueil, de traitement et de diffusion de l'information géographique est la géomatique, c'est-à-dire l'informatique appliquée à la géographie.

Plusieurs auteurs ont défini le système d'information géographique (SIG) et tous convergent sur le fait que l'information issue des données géolocalisées est stockée sous forme de bases de données géoréférencées au moyen d'un système informatique. C'est notamment la société française de photogrammétrie et

téledétection (SFPT) qui, en 1990, définit un SIG comme « *un système informatique permettant, à partir de diverses sources, de rassembler et organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter des informations localisées géographiquement contribuant notamment à la gestion de l'espace* » [1]. En 2004, Cécile de Joie définit un SIG comme étant « *un ensemble de données, repérées dans l'espace, et d'origines variées, stocké sur support informatique, et structuré de façon à pouvoir en extraire commodément des synthèses utiles à la décision* » [2]. Il apparaît donc qu'un SIG est un outil informatique qui permet de classer, d'ordonner, d'organiser, de stocker, d'interroger, de modifier et de restituer les informations géographiques extraites des bases de données spatialisées. La gestion des bases de données géolocalisées et l'analyse spatiale de ces données sont actuellement possibles grâce aux SIG. Elles permettent de répondre à des questions posées notamment en épidémiologie spatiale.

1.1.2. Analyse spatiale et statistique spatiale

Concrètement, qu'est-ce qu'est l'analyse spatiale ? En quoi diffère-t-elle de la statistique spatiale ? Selon Grasland, l'analyse spatiale et la statistique spatiale se différencient par le fait que « *le concept central en analyse spatiale est celui de l'interaction, tandis qu'en statistique spatiale, c'est l'autocorrélation qui est essentielle* » [3]. L'interaction spatiale et l'autocorrélation spatiale sont deux termes très proches, à une différence que le premier s'intéresse à la relation entre deux endroits et le second à la ressemblance entre eux. En effet, l'interaction spatiale est une notion qui désigne la manière selon laquelle deux endroits tendent à avoir des relations, étant donné que leurs frontières sont poreuses aux mouvements de populations, ce qui facilite les contacts entre les populations. Ces relations peuvent être traduites par exemple par un processus de diffusion spatiale d'un phénomène donné. Cela veut dire qu'un endroit proche d'un endroit ayant par exemple un nombre important de personnes atteintes par une maladie infectieuse a de fortes chances de voir sa population atteinte par cette maladie. L'autocorrélation spatiale, que nous détaillerons plus loin, désigne quant à elle le degré ressemblance ou de dissemblance entre deux endroits, étant donné qu'en général deux endroits qui sont spatialement proches ont plus de chances de se ressembler que deux endroits éloignés. Cependant, cette notion de ressemblance ou de dissemblance entre les

endroits géographiquement proches est à prendre avec précautions. En effet, deux endroits peuvent être caractérisés par les mêmes valeurs d'un phénomène donné, par exemple le paludisme, alors qu'ils sont éloignés dans l'espace (Afrique et Asie par exemple). Actuellement, l'analyse spatiale inclut les statistiques spatiales en tant que branche de la statistique qui s'intéresse à l'analyse des données géolocalisées.

Par ailleurs, l'analyse spatiale est une démarche méthodologique permettant de caractériser un phénomène indexé par des coordonnées géographiques en vue de le décrire, de l'expliquer et de modéliser son comportement dans l'espace et/ou dans le temps, tout cela dans le but d'identifier la tendance à former des structures particulières conduisant ainsi à la formulation des hypothèses et à la prise de décision.

L'interaction spatiale et l'autocorrélation spatiale que nous détaillerons plus loin sont deux étapes importantes et préalables à toute analyse spatiale qui s'intéresse aux similarités d'un phénomène décrit par des coordonnées géographiques. Plutôt que de s'intéresser aux singularités d'un phénomène qui n'ont d'intérêt que lorsque ce phénomène ne peut pas être connu ou n'est pas défini à ces points singuliers, l'analyse spatiale s'intéresse, quant à elle, à mettre en exergue les endroits de l'espace où ce phénomène présente des structures particulières, en liaison ou pas avec les caractéristiques de la population d'étude. Concrètement, l'analyse spatiale permettra de cartographier les zones à haut risque pour une maladie donnée, son évolution spatiale et temporelle, ainsi que de lier la probabilité de survenue de cette maladie aux caractéristiques notamment sociales, démographiques, économiques, environnementales et comportementales de la population d'étude. Utilisée au début pour étudier la distribution spatiale des phénomènes liés aux ressources naturelles et minéralogiques telles que le pétrole et l'or et des phénomènes liés à la gestion des territoires (délimitation des parcelles, traçage des routes), l'analyse spatiale est actuellement utilisée en épidémiologie et dans les sciences sociales [4,5]. Elle est également utilisée dans d'autres domaines tels que l'agronomie, l'environnement et l'océanographie.

Les données géographiques sont souvent stockées dans des bases de données de différente nature mais, dans certains cas, elles peuvent toutes être stockées dans une même configuration appelée « couche ».

1.1.3. Notion de couches et sources des données

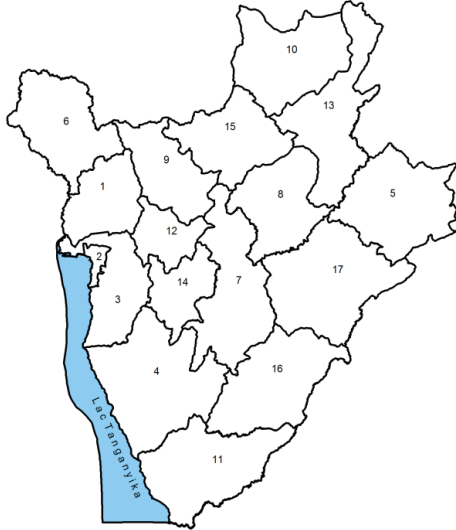
Le type de couche (i.e fond de carte) en présence détermine le type d'opérations à mettre en œuvre pour réaliser une carte. Une couche est un objet qui contient toutes les informations d'une représentation cartographique à réaliser. Elle peut être de type vecteur (couche déformable) ou de type raster (image) (**Figure 1**).

Une couche vectorielle est constituée de vecteurs (lignes, polygones ou points). Elle contient des « entités géographiques » (en lignes) et des attributs (en colonnes), leur intersection étant appelée « champ ».

Ces entités peuvent être repérées soit à l'aide de leurs identifiants (**Figure 1 a**), soit à l'aide de leurs noms de la table d'attributs. Selon que les couches sont constituées de polygones (**Figure 1 b**) ou de points (**Figure 1 c et d**), elles sont appelées couches de type linéaire ou de type ponctuel respectivement. Les termes de « linéarité » et « ponctualité » dont il s'agit ici font référence à la forme géométrique des unités spatiales. C'est ainsi qu'une ligne peut être simple (autoroute à sens unique), double (autoroute à double sens), multiple (route à voies multiples avec par exemple des allées pour piétons), en pointillé ou en hachures (chemin de fer), directionnelle (avec indication du sens de parcours ou du type d'automobile qui y a accès) et d'épaisseur différente (allée, piste, route). De même, les symboles utilisés pour les couches ponctuelles peuvent être simples (étoile, cercle, carré, losange, croix, flèche, triangle), déformés des symboles simples (ellipse, rectangle), sous forme de symboles utilisés dans le traitement des textes ou sous forme d'image (avion, voiture, homme, animal, statue).

Figure 1 : Couches vectorielles du Burundi

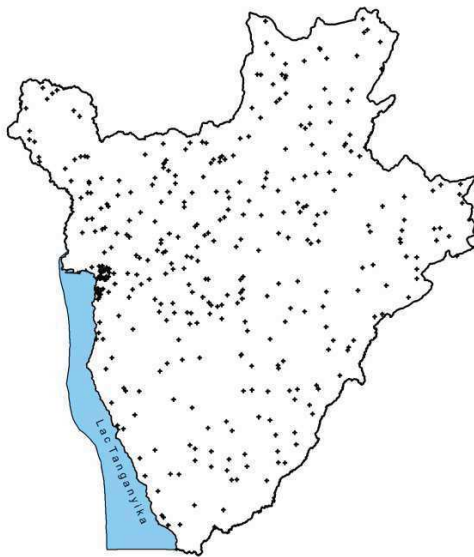
a) Couche de polygones (provinces)



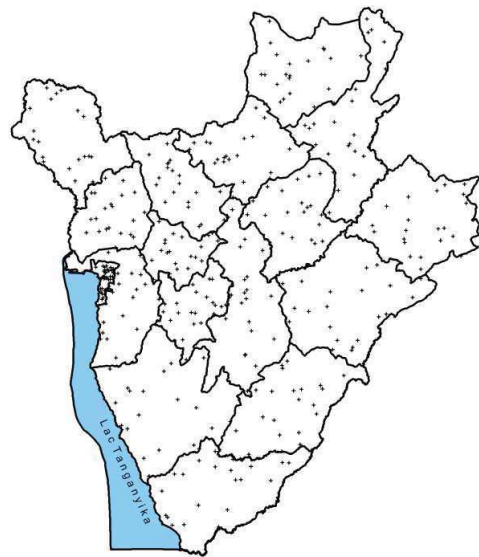
b) Couche de polygones (routes)



c) Couche de points (grappes) non groupés



d) Couche de points (grappes) groupés



La **figure 1 a** illustre la couche vectorielle de polygones représentant les provinces du Burundi (valable jusqu'en 2015), la **figure 1 b** la couche de polygones qui est constituée du réseau routier du Burundi, la **figure 1 c** une couche ponctuelle représentant la localisation géographique des grappes ou zones de dénombrement de l'Enquête Démographique et de Santé (EDS) conduite en 2010 au Burundi, ces zones étant représentées par leurs centres géographiques appelés centroïdes. La

figure 1 d illustre les mêmes grappes groupées dans les provinces du Burundi. Les couches vectorielles sont les plus fréquentes en épidémiologie spatiale tandis que les couches raster sont souvent présentes en télédétection. Les couches vectorielles peuvent contenir des données de type numérique (effectifs, mesures, proportions), de type alphanumérique, les coordonnées géographiques (longitude/latitude), l'identifiant de chaque entité, toutes ces données formant la table attributaire de cette couche (**Tableau 1**).

Tableau 1 : Table attributaire de la couche vectorielle de polygones (provinces) du Burundi

Id_P	Pays	Province	Type	Région	Long_Forme	Surf_Forme
1	Bdi	Bubanza	Province	Ouest	1,529471	0,085629
2	Bdi	Bujumbura-Mairie	Province	Bujumbura	0,857177	0,016407
3	Bdi	Bujumbura-Rural	Province	Ouest	2,590316	0,126920
4	Bdi	Bururi	Province	Sud	2,879989	0,251186
5	Bdi	Cankuzo	Province	Centre-Est	2,254311	0,153507
6	Bdi	Cibitoke	Province	Ouest	2,252415	0,127300
7	Bdi	Gitega	Province	Centre-Est	3,143926	0,154477
8	Bdi	Karuzi	Province	Centre-Est	2,023658	0,114793
9	Bdi	Kayanza	Province	Nord	1,875497	0,096113
10	Bdi	Kirundo	Province	Nord	1,910061	0,144535
11	Bdi	Makamba	Province	Sud	2,968043	0,208227
12	Bdi	Muramvya	Province	Centre-Est	1,403937	0,054206
13	Bdi	Muyinga	Province	Nord	3,161136	0,145819
14	Bdi	Mwaro	Province	Sud	1,752402	0,066804
15	Bdi	Ngozi	Province	Nord	1,988686	0,114714
16	Bdi	Rutana	Province	Sud	2,435968	0,150467
17	Bdi	Ruyigi	Province	Centre-Est	2,545330	0,183063

Bdi : Burundi ; Long : longueur ; Surf : surface

La visualisation de la couche associée à cette table peut faire apparaître soit l'identifiant (**Figure 1 a**), soit le nom de la province. Chaque ligne correspond donc à

un polygone (la province). Cette table stocke également le nom du pays, le nom et le type de chaque entité géographique (province), la longueur et la surface de chaque polygone.

Une couche ne peut contenir que des éléments de même nature (des polygones, des points ou des lignes). Dans le cas d'une couche se présentant sous forme de points, la table attributaire correspondante porte, en plus de l'identifiant de chaque point, des caractéristiques propres à un point comme la longitude, la latitude, l'altitude et le système de projection WGS84 (World Geodetic System 1984) dont nous parlerons plus loin. Le système de projection des couches de points, de polygones ou de polygones est visualisé à travers un logiciel SIG.

Le **tableau 2** montre un extrait des 20 premiers éléments de la table attributaire de la couche des 376 points ou grappes de l'EDS 2010 du Burundi. Ces grappes sont les points d'observation. Il s'agit d'une base géographique contenant également les caractéristiques des grappes telles que la place de résidence (urbain/rural), l'identifiant de la province et la région.

Chaque ligne correspond donc à un point (la grappe). Lors de la fusion, ces deux éléments : l'identifiant de la grappe ou celui de la province (**Figure 1 a**) peuvent servir de lien entre la table attributaire des grappes et la base de données descriptives des individus contenant ces éléments.

Les éléments d'une table attributaire représentant une couche de polygones sont entre autres l'identifiant du tronçon (Id), le nom de la route (Nom), le type (Type) de route (Principale ou Secondaire), sa longueur (Long_Forme) ainsi que d'autres attributs qui ne sont pas présentés dans ce tableau tels que la nature de la route (route/piste), la largeur de la chaussée, le sens de circulation de la voie, les coordonnées et l'altitude du sommet initial et du sommet final de la route.

Tableau 2 : Table attributaire de la couche vectorielle de points (grappes) du Burundi

Id	Longitude	Latitude	Altitude	Datum	Id_P	Région	Résidence
1	29,362186	-3,368500	789	WGS84	2	Bujumbura	Urbain
2	29,963905	-3,816132	1752	WGS84	16	Sud	Rural
3	30,258659	-2,505578	1419	WGS84	10	Nord	Rural
4	30,257027	-3,428142	1610	WGS84	17	Centre-Est	Rural
5	30,310476	-3,758130	1220	WGS84	16	Sud	Rural
6	30,132250	-3,984277	1223	WGS84	16	Sud	Rural
7	29,484858	-3,034616	1655	WGS84	1	Ouest	Rural
8	30,565269	-3,143761	1604	WGS84	8	Centre-Est	Rural
9	30,381834	-3,422767	1605	WGS84	17	Centre-Est	Rural
10	29,671818	-3,211953	1819	WGS84	10	Centre-Est	Rural
11	29,359810	-3,450868	837	WGS84	2	Bujumbura	Urbain
12	29,776478	-3,204809	1871	WGS84	12	Centre-Est	Rural
13	30,087469	-3,701134	1786	WGS84	16	Sud	Rural
14	29,883385	-4,171279	1375	WGS84	11	Sud	Rural
15	29,520090	-3,293669	1714	WGS84	3	Ouest	Rural
16	29,370983	-3,420516	839	WGS84	2	Bujumbura	Urbain
17	29,701815	-2,864700	1647	WGS84	15	Nord	Rural
18	30,423372	-3,165965	1536	WGS84	5	Centre-Est	Rural
19	29,925566	-3,455016	1656	WGS84	7	Centre-Est	Urbain
20	29,718658	-3,555017	1933	WGS84	14	Sud	Rural

Le nom de la route peut être une route nationale (RN) pour le réseau routier principal reliant, pour le cas d'espèces, la province de Bujumbura-Mairie aux axes Nord, Sud, Ouest et Est, ou une route poteau (RP) pour le réseau routier secondaire à trafic intense avec indication du kilométrage. À l'exception des routes non spécifiées, chaque RN ou RP est accompagnée d'un numéro indiquant son importance (**Tableau 3**).

Chaque ligne correspond donc à une route. La géométrie des points à relier pour former ces routes ne peut cependant pas être visualisée dans une table attributaire, ce qui est le cas pour une table attributaire d'une couche de polygones. Cependant, les routes sont données sous forme de tronçons et non sous forme de longueur. Cette couche permet de savoir par exemple combien de provinces sont traversées par une route donnée et avec quelle distance.

Tableau 3 : Table attributaire de la couche vectorielle de polygones (routes) du Burundi

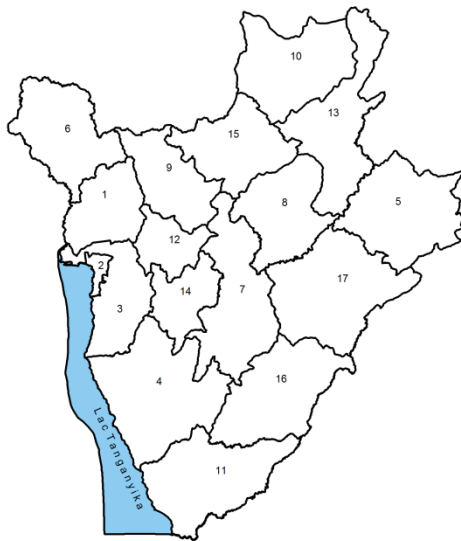
Id	Nom	Type	Long_Forme
22	RN1	Principale	0,123425
62	RN1	Principale	0,060310
68	RN1	Principale	0,058947
81	RN1	Principale	0,351963
83	RN2	Principale	0,535393
28	RN3	Principale	0,136686
38	RN3	Principale	0,151207
42	RN3	Principale	0,164788
45	RN3	Principale	0,209391
49	RN3	Principale	0,171262
2	NULL	Principale	0,036136
4	NULL	Principale	0,064701
54	RN3	Principale	0,176377
210	RP21	Secondaire	0,081037
214	RP21	Secondaire	0,223830
186	RP22	Secondaire	0,169051
188	RP22	Secondaire	0,119914
187	RP23	Secondaire	0,062926
3	NULL	Secondaire	0,069885
5	NULL	Secondaire	0,013188

Les éléments d'une table attributaire permettent de faire des statistiques de base et des opérations mathématiques sur les variables, ou des opérations sur les polygones telles que la fusion, l'intersection et la différenciation des couches, la fusion des polygones et la construction des centroïdes des polygones. D'autres opérations concernent la conversion de la couche vectorielle vers la couche raster et vice-versa.

Par contre, une couche de type raster est composée de pixels, un pixel étant la plus petite partie d'une image (**Figure 2**). Cette couche peut faire apparaître ou non les délimitations des entités géographiques sans pour autant qu'il soit possible d'en extraire les coordonnées géographiques (type 1, type 2).

Figure 2 : Couche raster du Burundi

a) Couche raster de type 1



b) Couche raster de type 2



Les opérations mathématiques faites sur les couches vectorielles ne sont pas les mêmes que celles effectuées sur les couches rasters. Pour ces dernières, les opérations les plus courantes sont le géoréférencement, la conversion vers une couche vectorielle, le découpage, la création des contours, la construction d'une grille d'interpolation, la fusion et le calcul de la pente d'un modèle numérique de terrain (MTN), c'est-à-dire une représentation à trois dimensions de la portion de la surface terrestre à laquelle l'intérêt est porté et qui ne tient pas compte de la hauteur

des objets (maisons, arbres) qui s'y trouvent. C'est le modèle numérique d'élévation (MTE) qui tient compte de l'élévation (hauteur) des objets qui se trouvent sur la surface pour laquelle la pente se veut être calculée.

Les coordonnées géographiques du phénomène d'intérêt et les couches des polygones sur lesquels ce phénomène se veut être cartographié ne sont pas nécessairement dans le même système de projection. C'est le cas des couches des pays téléchargeables gratuitement sur les sites <http://www.diva-gis.org/gdata>, www.maplibrary.org/ et <http://www.gadm.org/country>.

1.1.4. Projections et systèmes de projection

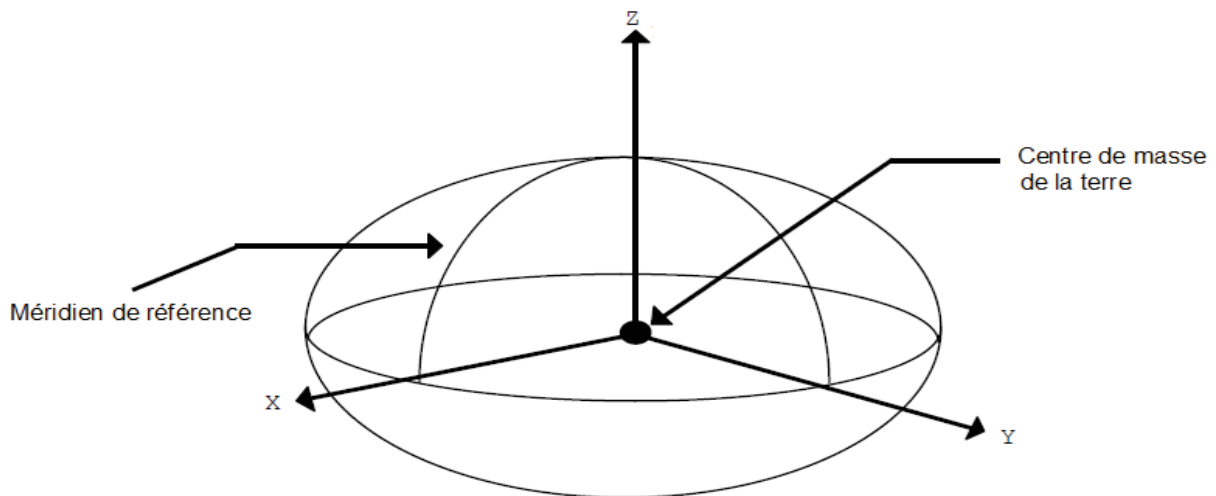
Le début de l'ère spatiale est caractérisée par le lancement du premier missile intercontinental et du premier satellite artificiel de la Terre en 1957 [6]. Depuis lors, des pays comme la Russie et les États-Unis d'Amérique n'ont cessé de fabriquer des outils permettant d'explorer l'espace, de recueillir les coordonnées géographiques des pays et de produire des images satellitaires des portions terrestres et lacustres. Ces portions sont représentées et localisées grâce à des coordonnées qui les caractérisent de manière univoque. Ces coordonnées sont mathématiquement transformées grâce à des systèmes de projection permettant de superposer les couches vectorielles ou les couches raster à l'aide des logiciels SIG tels que QGIS (Quantum Geographic Information System), ArcGIS (une collection de logiciel SIG), GeoDa et R. Un système de projection est un système dans lequel une couche est visualisée. Autrement dit, un système de projection est toute forme de représentation d'un objet à la surface de la Terre, une projection étant cette représentation ou tout simplement la correspondance entre les coordonnées géographiques et les coordonnées planimétriques, qui permet de réduire les déformations lors de la transformation de l'ellipsoïde en un plan. La Terre est une ellipsoïde, étant légèrement aplatie aux pôles. Ainsi donc, pour mieux visualiser un pays du globe terrestre, il convient de le projeter dans un espace à deux dimensions (un plan). Ces projections sont faites de telle sorte que les distorsions, c'est-à-dire les déformations subies lors de la projection, soient moindres. Ces distorsions concernent les surfaces, les angles, les distances, les frontières des pays et les directions. Parmi ces projections, certaines conservent les angles (projections

conformes ou stéréographiques), laissant ainsi le champ aux déformations surfaciques, et d'autres conservent les surfaces (projections équivalentes), laissant ainsi le champ aux déformations des angles. Toutes les projections cartographiques engendrent donc des distorsions selon la position géographique des polygones à projeter et ces distorsions sont plus accentuées soit au niveau des pôles, soit au niveau de l'équateur ou du méridien de Greenwich. La meilleure de ces projections est celle qui présente le moins de distorsions possibles.

La Terre ne pouvant pas être fendue et rabattue sur un plan à deux dimensions afin de visualiser tous les pays sur ce plan, elle ne constitue pas une forme développable au même titre que le cône, le cylindre et le plan lui-même. C'est sur ces trois formes que les points de la surface de la Terre sont projetés, aboutissant ainsi à une correspondance entre les coordonnées géographiques des points de la Terre et ces formes. Cette correspondance porte le nom de projection cartographique et permettra de faire le rendu cartographique.

Le système géodésique mondial de référence de 1984, dénommée WGS84 (World Geodetic System 1984), est un système de coordonnées de référence (SCR) mondiale développée par l'agence nationale géospatiale d'intelligence américaine. Il concerne les coordonnées géographiques (longitude/latitude) exprimées en degrés décimaux ou sexagésimaux. Les coordonnées géographiques exprimées en degrés peuvent être converties en kilomètres ou en mètres soit manuellement, soit en utilisant un système de projection. Plusieurs systèmes de projection correspondent au système de coordonnées WGS84, chacun portant son propre numéro. Par exemple, le système de projection le plus utilisé et qui est associé à ce système de coordonnées est EPSG : 4326 (European Petroleum Survey Group). Le géoïde représentant la forme de la Terre dans le système WGS84 peut être schématisé dans un espace à trois dimensions (**Figure 3**) [7]. Ce système utilise le géopositionnement par satellite (GPS : Global Positioning System).

Figure 3 : Système de coordonnées de référence mondiale



Source : Adapté de « National imagery and mapping agency », 2000

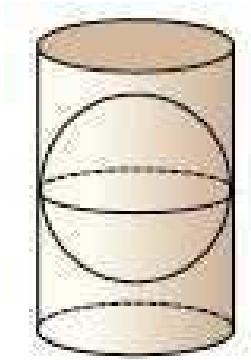
Les trois projections les plus utilisées sont la projection cylindrique, la projection conique et la projection planaire:

La projection cylindrique directe est une projection conforme qui consiste à projeter les points de la surface terrestre sur un cylindre de section elliptique, puis de le rabattre afin d'obtenir une carte du monde sur un plan à deux dimensions (**Figure 4**). Au moment de la projection, le cylindre est tangent (ou sécant) à la Terre avec un axe perpendiculaire au plan de l'équateur (**Figure 4 a**). Après rabattement, les méridiens et les parallèles sont des droites orthogonales (**Figure 4 b**). La projection cylindrique directe entraîne de fortes dispersions aux pôles.

Un autre cas de figure de la projection cylindrique est la projection cylindrique transverse où le cylindre est tangent (ou sécant) à la Terre avec un axe perpendiculaire au plan du méridien d'origine, ici le méridien de Greenwich (**Figure 5**). Comme pour le cas de la projection cylindrique directe, la Terre est projetée sur ce cylindre (**Figure 5 a**). Après rabattement, les méridiens et les parallèles sont des courbes orthogonales (**Figure 5 b**). La projection cylindrique transverse entraîne de fortes dispersions sur le méridien.

Figure 4 : Projection cylindrique directe

a) Représentation cylindrique



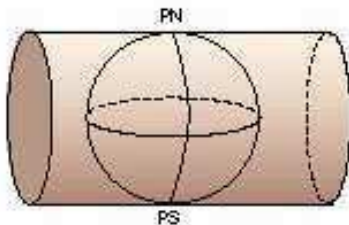
b) Représentation des méridiens et des parallèles



Source : IGN et Éducation Nationale sur l'information géographique

Figure 5 : Projection cylindrique transverse

a) Représentation cylindrique



b) Représentation des méridiens et des parallèles



Source : IGN et Éducation Nationale sur l'information géographique

IGN : Institut national de l'information géographique et forestière

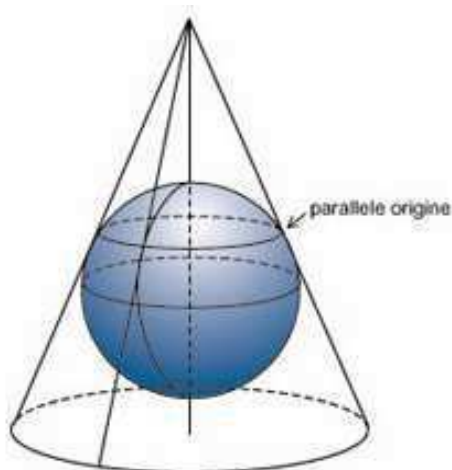
Le cylindre peut aussi être tangent (ou sécant) à la Terre avec un axe oblique à la fois au plan de l'équateur et à celui du méridien d'origine. Le système de projections cylindriques est la Mercator Transverse Universelle (UTM : Universal Transverse Mercator) associée au système géodésique WGS84 et dont le référentiel géodésique est l'ED50 (European Datum 1950). C'est un système de 120 projections cylindriques transverses conformes et de 2 projections correspondant aux fuseaux horaires Nord et Sud [8].

La projection cylindrique directe est appelée projection de Mercator directe et la projection cylindrique transverse la projection de Mercator transverse. Autrement dit, le système de projections cylindriques est le système de Mercator.

La projection conique directe est une projection qui consiste à projeter les points de la surface terrestre sur un cône, puis de le rabattre afin d'obtenir une carte du monde sur un plan à deux dimensions (**Figure 6**). Au moment de la projection, le cône est tangent (ou sécant) à la Terre avec un axe perpendiculaire au plan de la parallèle d'origine (**Figure 6 a**). Après rabattement, les méridiens sont des droites sécantes au pôle et les parallèles des cercles dont le centre est le pôle (**Figure 6 b**). Comme pour le cylindre, le cône peut aussi avoir un axe oblique au plan de l'équateur. La projection conique peut aussi être transverse. Le système de projections coniques est le système de Lambert.

Figure 6 : Projection conique directe

a) Représentation conique



b) Représentation des méridiens et des parallèles

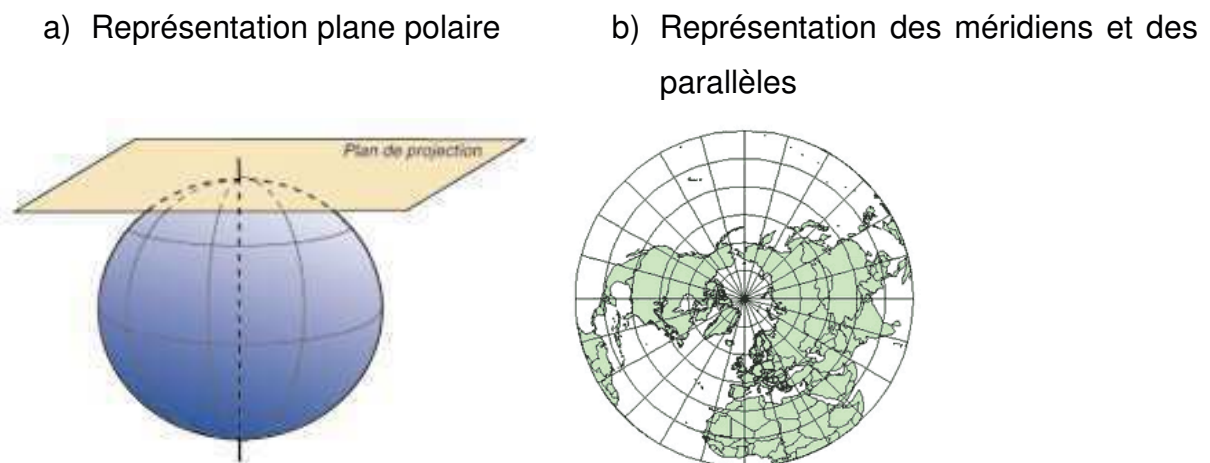


Source : IGN et Éducation Nationale sur l'information géographique

La projection plane polaire directe est une projection qui consiste à projeter les points de la surface terrestre sur un plan tangent au pôle, puis de le rabattre afin d'obtenir une carte du monde sur un plan à deux dimensions (**Figure 7**). Au moment de la projection, le plan est tangent (ou sécant) au pôle (**Figure 7 a**). Après

rabattement, les méridiens sont des droites sécantes au pôle et les parallèles des cercles dont le centre est le pôle (**Figure 7 b**)). Une telle projection est une projection azimutale conforme. Elle peut aussi bien être transverse ou oblique. Le système de projections planes polaires est le système de Lambert. L'utilisation des systèmes de projection dans le monde est hétérogène (**Figure 8**).

Figure 7 : Projection plane polaire directe

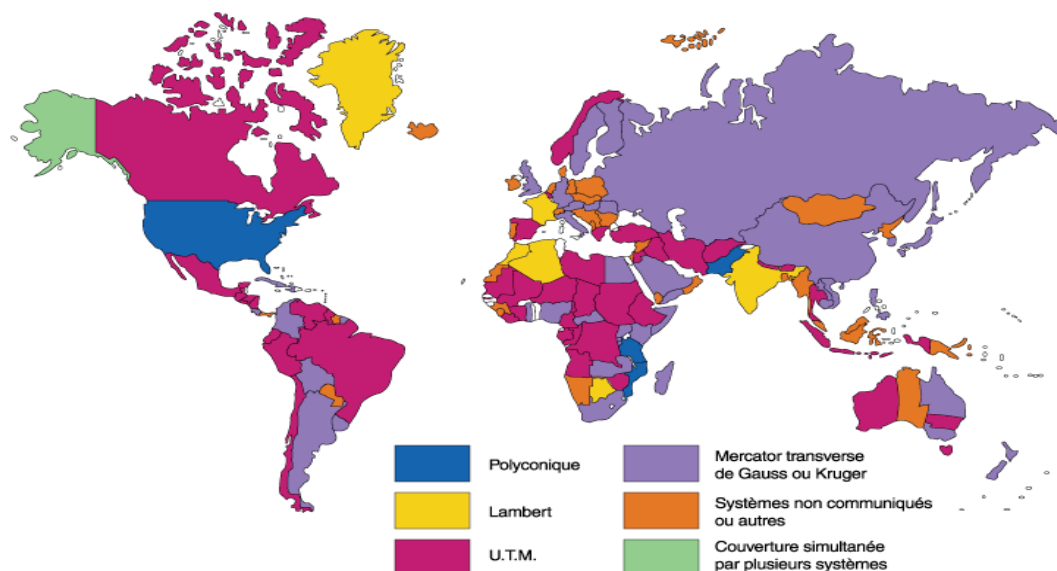


Source : IGN et Éducation Nationale sur l'information géographique

L'utilisation de ces projections conduit à l'existence de plusieurs cartes possibles pour visualiser la Terre. La projection de Mercator, qui privilégie les déformations de surface à la place des déformations des angles, aboutit à une représentation cartographique de la Terre avec des déformations plus accentuées aux hautes latitudes qu'aux basses latitudes et ne traduisant pas ainsi la réalité au point de vue des aires de surface des portions de la Terre situées aux pôles. La carte réalisée sur le maillage des méridiens et des parallèles fausse la surface du Groenland qui devient au moins aussi grande en superficie que l'Afrique alors sa superficie est de loin inférieure à celle de l'Afrique. Puisqu'il n'est pas possible de conserver à la fois les surfaces et les angles, les déformations de la réalité de ce qu'est la Terre doivent avoir lieu. Les projections qui ne conservent ni l'un ni l'autre mais qui peuvent par exemple conserver les distances (projections équidistantes) sont dites aphyllactiques. Les projections qui conservent les surfaces sont entre autres la

projection cylindrique équivalente de Behrmann, la projection conique équivalente d'Albers et la projection azimutale équivalente de Lambert.

Figure 8 : Spatialisation de l'utilisation des systèmes de projection dans le monde



Source : IGN et Éducation Nationale sur l'information géographique

1.1.5. Type de données spatiales

1.1.5.1 Données ponctuelles

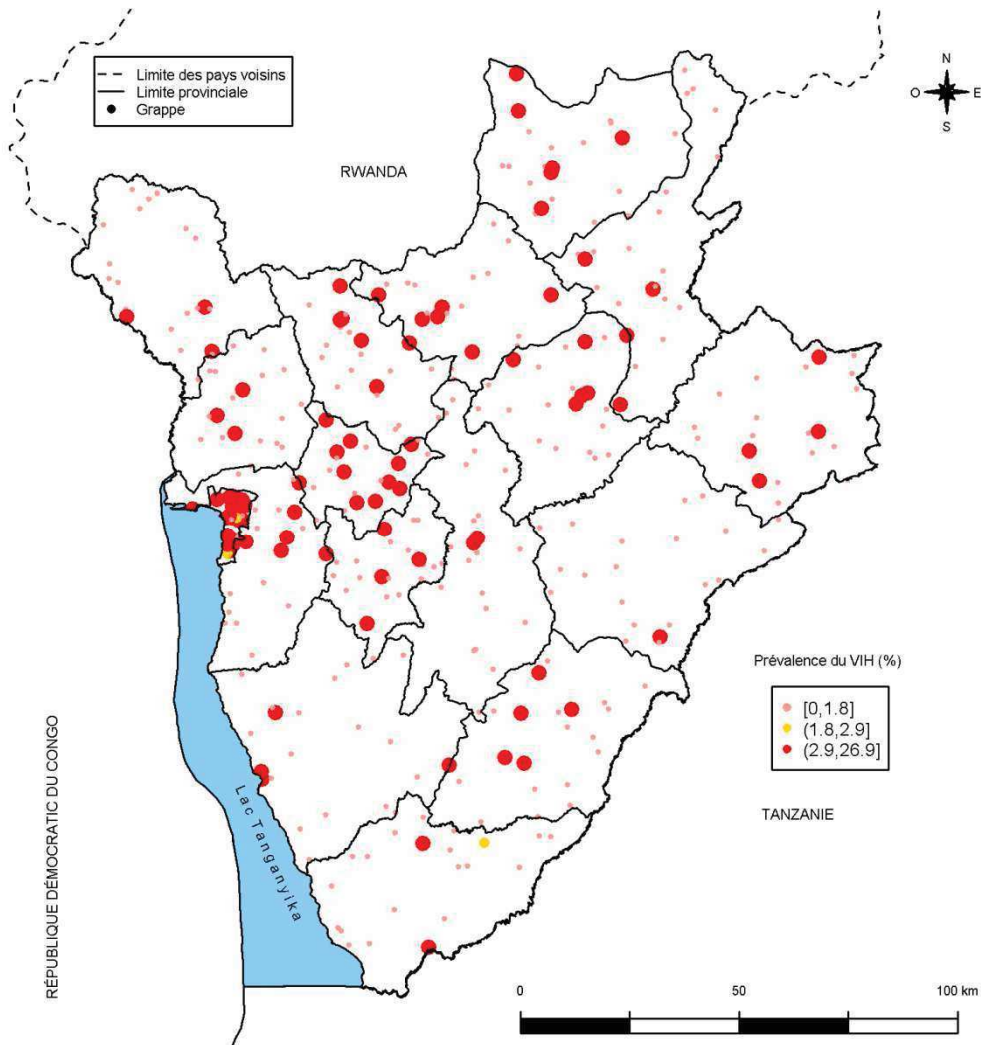
Considérons D le domaine borné représentant la zone d'étude dans laquelle un événement d'intérêt se produit. Les données relatives à cet événement sont dites géolocalisées dans l'espace et l'intérêt principal porte sur son organisation spatiale. Appelé aussi « champ », le domaine D est tel que la variable géographique qui décrit ce phénomène est supposée prendre des valeurs nulles en dehors de ce domaine. Dans ce domaine, les données sont recueillies à différents points de coordonnées géographiques. Ces points d'observation sont le résultat d'un processus spatial ponctuel, c'est-à-dire un processus stochastique au cours duquel le nombre de points et leur répartition spatiale sur un support discret $D \subseteq R^2$ sont aléatoires, ces points étant observés dans une fenêtre $S \subset D$. De telles données sont dites ponctuelles. Les processus de base qui décrivent ce genre de données

sont les processus ponctuels de Poisson, c'est-à-dire les processus au cours desquels le nombre de cas est de loi de Poisson et à accroissements indépendants. Ces points peuvent être les centres géographiques (centroïdes) des lieux de résidence des patients [9], les ménages ou les zones de dénombrement (ZD) lors d'une enquête de population. Une première représentation de la distribution spatiale de la variable d'intérêt se fait en la représentant, à chaque point d'observation, par un symbole dont la taille est proportionnelle à la valeur observée de la variable ou selon la classe d'appartenance de cette valeur afin de détecter la tendance (dérive) du phénomène auquel la variable se rapporte. Une illustration de ce genre de données concerne la prévalence du virus de l'immunodéficience humaine (VIH) par grappe ou ZD obtenue lors d'une enquête conduite au Burundi en 2010 (**Figure 9**).

Chaque couleur correspond à une classe de la prévalence du VIH. Aucune tendance régulière n'est observée dans la distribution spatiale de l'infection par le VIH. Plusieurs raisons expliquent la rareté de ce genre de données : coût élevé de l'enquête, difficultés liés à l'échantillonnage, confidentialité de la localisation géographique des enquêtés.

À vrai dire, il s'agit des données ponctuelles uniquement du fait qu'elles sont rapportées au niveau du centre géométrique de la grappe. Une grappe étant une zone de dénombrement, elle pourrait concerner des données de type agrégé.

Figure 9 : Prévalence du VIH par grappe au Burundi, EDS 2010

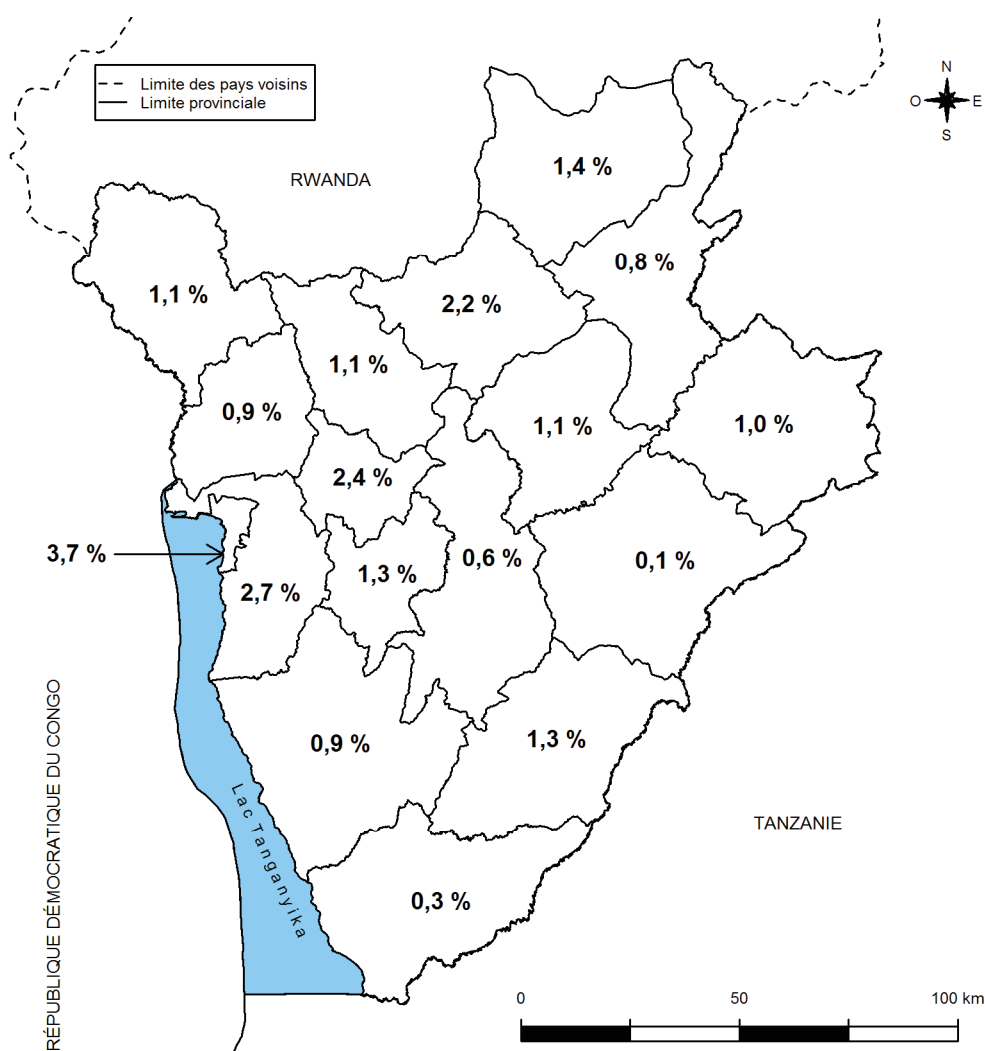


1.1.5.2 Données agrégées

Les données agrégées ou groupées sont un type particulier de données ponctuelles rapportées au niveau d'une variable composite. Cette variable peut par exemple désigner les niveaux du découpage de la zone d'étude en entités géographiques en tant qu'entités administratives contiguës du pays ou une autre caractéristique des enquêtés. La fenêtre S est alors un sous-ensemble discret et fixé. Contrairement aux données ponctuelles, l'intérêt principal porte sur les liaisons entre les caractéristiques mesurées sur les individus dans chaque entité géographique et sur les tendances spatio-temporelles locales et globales. La plupart des données

rencontrées en épidémiologie sont de type agrégé. Une deuxième représentation de la distribution spatiale de la variable d'intérêt consiste à attribuer une valeur moyenne par zone du domaine d'étude et à colorier ces zones selon les valeurs prises par cette variable. C'est le cas par exemple de la prévalence du VIH (**Figure 10**) calculées par province à l'aide des données de l'exemple précédent.

Figure 10 : Prévalence du VIH par province au Burundi, EDS 2010



Cela revient donc à faire un résumé statistique de la prévalence du VIH par province et à cartographier ce résumé de façon à montrer la prévalence du VIH à chaque unité géographique du domaine D . Cette cartographie pourrait aussi se faire en coloriant les provinces selon les valeurs ou les classes de la prévalence du VIH. Les données obtenues permettent donc de faire une présentation surfacique de la

prévalence du VIH. L'inconvénient de cette façon de présenter les données est qu'elle surestime (sous-estime) la prévalence du VIH dans certains endroits où cette prévalence est basse (haute). De plus, elle permet de perdre l'information en résumant l'information initialement rapportée au niveau individuel. Contrairement aux données ponctuelles, les données agrégées sont caractérisées par une accessibilité plus ou moins facile dans les systèmes de surveillance de la santé ou dans les programmes chargés de recueillir, de conserver et de distribuer les données de population.

1.1.5.3 Données géostatistiques

En géostatistique, les données se rapportant à la variable d'intérêt sont collectées de manière irrégulière sur la fenêtre S en un nombre fixé de sites s_1, \dots, s_n préalablement choisis lors de l'étape de la sélection de l'échantillon. Cette fenêtre constitue un sous-ensemble continu de $D \subseteq \mathbb{R}^{d;d \geq 2}$. L'intérêt principal de ce genre de données est qu'elles tiennent compte non seulement de l'information sur la localisation géographique (une seule réalisation en un site donné, sauf pour des données répétées) des données mais également du fait que la variable d'intérêt est elle-même considérée comme une variable aléatoire et les valeurs qu'elle est susceptible de prendre en différents sites comme ses réalisations. Elles permettent, en outre, d'intégrer l'information contenue dans d'autres variables afin de comprendre et de modéliser les variations spatiales, et de faire des prédictions spatiales (interpolation) des valeurs de la variable d'intérêt aux sites non échantillonnés et leur incertitude afin de reconstituer la variable d'intérêt en tout point du domaine d'étude [10].

Soit (Ω, F, P) un espace probabilisé où Ω est l'univers des possibles, F une algèbre et P une probabilité. Soit également (E, \mathcal{E}) un espace d'états mesurable. Gaetan et Guyon (2008) définissent un champ ou fonction aléatoire comme une famille $Z = \{Z(s), s \in D\}$ de variables aléatoires définies de (Ω, F, P) vers (E, \mathcal{E}) [11]. Nous noterons par $Z(s)$ le champ aléatoire défini de \mathbb{R}^2 vers E .

Un champ aléatoire est spatialement stationnaire d'ordre 1 s'il reste invariant par translation, c'est-à-dire si son espérance mathématique est constante sur tout le domaine d'étude. Autrement dit, cela veut dire qu'il ne dépend pas de la localisation géographique des points d'observation mais dépend uniquement de la distance séparant les points. Il est dit stationnaire d'ordre 2 si son espérance mathématique est constante et si sa covariance est invariante par translation [12,13] :

$$\begin{cases} E[Z(s)] = \mu = c^{te}; \forall s \in \mathbb{R}^2 \\ C(s+h, s) = Cov[Z(s+h), Z(s)] = C(h); \forall s, h \in \mathbb{R}^2 \end{cases} \quad (1.1)$$

où $C(\bullet)$ est la fonction de covariance ou covariogramme. Cette fonction a comme propriétés la symétrie et la défini-positivité de par la définition de la variance. La covariance caractérise le degré de ressemblance ou la corrélation entre les valeurs prises par la variable aléatoire en deux endroits. La stationnarité d'ordre 2 du champ aléatoire implique que sa variance ne dépend pas de la localisation des points mais dépend uniquement de l'ordre de grandeur et de la direction de h . Dans le cas où la fonction de covariance dépend uniquement de la norme de h et non de sa direction, alors $C(s+h, s) = C(\|h\|)$. Dans ce cas, cette fonction est isotrope au même titre que la variable $Z(s)$. La stationnarité intrinsèque, quant à elle, considère que la variance existe et augmente de façon monotone lorsque la distance entre les paires de points augmente [14]. Les données géostatistiques ont comme clé principale le fait qu'elles sont non indépendantes et que la variable d'intérêt a une réalisation unique. De plus, ces données permettent de caractériser, à courte échelle, le degré de dépendance des données et présentent une organisation ou structure spatiale liée à leur caractère aléatoire.

1.2. Définition des poids entre les unités spatiales

1.2.1. Matrice des distances spatiales

La matrice W des distances spatiales permet de pondérer les observations lors de la vérification de la dépendance spatiale appelée autocorrélation spatiale ou corrélation

de proche en proche. Elle se définit différemment selon qu'il s'agit des points ou des polygones. La matrice des distances euclidiennes entre deux points s_i et s_j du domaine D est une matrice W dont les éléments W_{ij} sont définis par la distance euclidienne $d_{ij} = d(s_i, s_j) = \|s_i - s_j\| = W_{ij}$. Connaissant le nombre n de points, le nombre de valeurs de distances spatiales mutuelles vaut $\frac{n(n-1)}{2}$. La distance euclidienne est une distance à vol d'oiseau ou distance selon la corde entre deux points $s_i(x_i, y_i)$ et $s_j(x_j, y_j)$ dont les coordonnées (x, y) sont connues. Même si la distance euclidienne est la plus utilisée en analyse spatiale, d'autres types de distance peuvent être utilisés. Il s'agit notamment de la distance de Manhattan utilisée dans le cas où il y a des restrictions de déplacements (entre deux voies de circulation par exemple) Est-Ouest et Nord-Sud et définie par $d_{ij} = d(s_i, s_j) = |x_i - x_j| + |y_i - y_j|$ et de la distance de Minkowski définie par $d = d(s_i, s_j) = \left(\sum_k |x_{i,k} - y_{j,k}|^r \right)^{\frac{1}{r}}$ où r est le paramètre indiquant la manière dont la distance varie en fonction de la valeur angulaire entre les points et l'axe de référence (Est-Ouest par exemple) et k la dimension[15]. La distance de Tchebychev donnée par $\max_k |x_{i,k} - y_{j,k}|$ peut également être utilisée. Si d_{ij} est la distance euclidienne entre deux points $s_i(x_i, y_i)$ et $s_j(x_j, y_j)$ se trouvant sur l'ellipsoïde, alors la distance de sur l'ellipsoïde vaut $de = 2R \arcsin\left(\frac{d_{ij}}{2R}\right)$ où $R = 6372 \text{ km}$ est le rayon de la terre.

La matrice des distances définies précédemment est utilisée pour pondérer les observations. Une autre façon de définir les poids spatiaux consiste à utiliser un voisinage global et à considérer que l'influence d'un point sur un autre décroît au fur et à mesure que la distance qui les sépare augmente et de ce fait, les éléments de la matrice des poids s'écrivent :

$$W_{ij} = \frac{1}{d_{ij}^\alpha} \text{ où } \alpha \text{ est un entier positif non nul} \quad (1.2)$$

où le paramètre α contrôle le degré d'influence entre deux points.

La relation (1.2) met plus de poids aux points qui sont proches et ces poids diminuent au fur et à mesure que l'un s'éloigne de l'autre. C'est cela la pondération par l'inverse de la distance (IDW : Inverse Distance Weighting). Une version de cette technique de pondération consiste à utiliser comme poids des distances normalisées :

$$W_{ij} = \frac{d_{ij}^{-\alpha}}{\sum_{k \neq j} d_{ik}^{-\alpha}} \quad (1.3)$$

Pour des polygones, une première façon d'évaluer les interactions spatiales entre eux est de calculer les distances entre leurs centroïdes. Ces distances permettent de connaître, pour chaque polygone, ses voisins au sens de la distance spatiale et au vu de la longueur du segment qui joint les centroïdes (**Figure 11**).

Une deuxième façon est d'utiliser la matrice stochastique de contiguïté qui indique si les zones ont ou n'ont pas de frontière commune [16] :

$$W_{ij} = \begin{cases} 1 & \text{si deux zones ont une frontière commune} \\ 0 & \text{sinon} \end{cases} \quad (1.4)$$

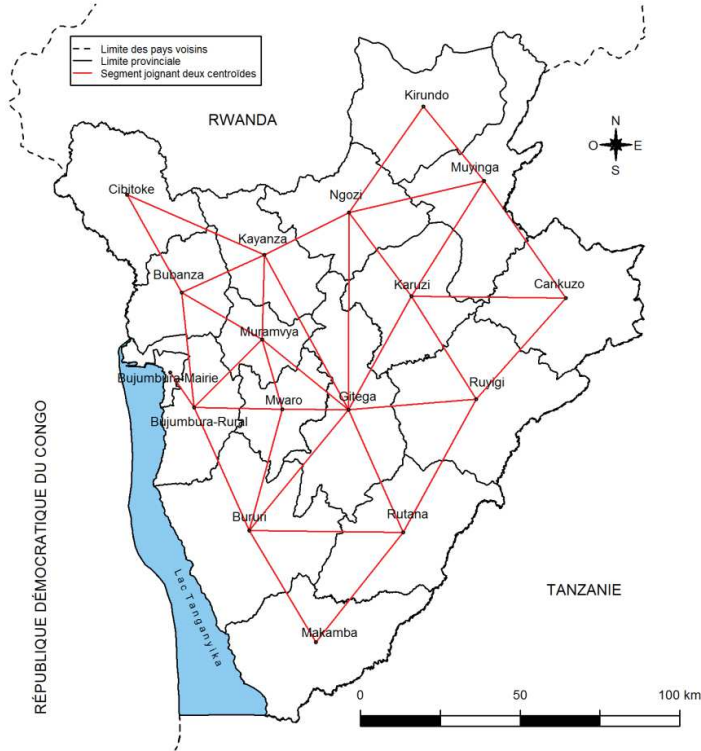
Deux zones sont donc contigües d'ordre 1 si elles ont une frontière commune et d'ordre k s'il faut traverser k frontières pour aller du centroïde de l'une à celui de l'autre.

Une autre façon de définir le degré de similarité entre les unités spatiales est de considérer que les poids entre ces unités diminuent de manière exponentielle au fur et à mesure que les points s'éloignent dans l'espace :

$$W_{ij} = \exp(-\alpha d_{ij}) \quad (1.5)$$

avec α une constante positive dont l'inverse est un paramètre d'échelle spatiale.

Figure 11 : Graphe de voisinage entre les provinces du Burundi



1.2.2. Graphe de voisinage entre les points

Pour définir le nombre de voisins qui sont plus proches d'un point donné dans l'espace au sens d'une métrique (la distance), il importe de fixer la distance seuil, c'est-à-dire la distance au-delà de laquelle deux points ne sont plus voisins. Dans ce cas, les voisins les plus proches se trouvent dans un cercle dont le rayon est égal à cette distance seuil. Si d est la distance seuil, alors les éléments de la matrice des poids basés sur le nombre de plus proches voisins peuvent se définir comme suit :

$$W_{ij} = \begin{cases} 1 & \text{si } 0 \leq d_{ij} \leq d \\ 0 & \text{si } d_{ij} > d \end{cases} \quad (1.6)$$

Plutôt que de fixer la distance radiale définie par la distance seuil, il est possible de considérer un voisinage formé des k voisins les plus proches sans tenir compte de la distance qui les sépare.

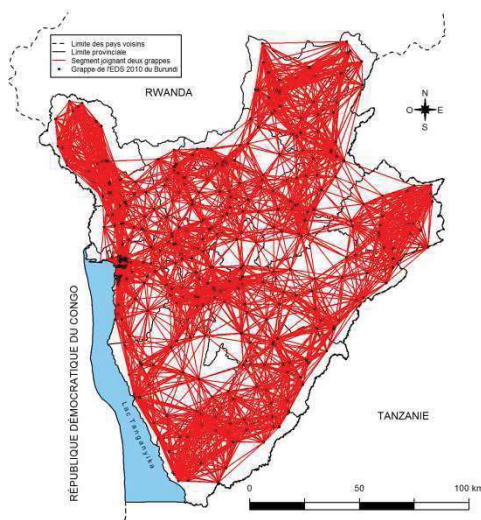
Dans ce cas, en considérant que $N_k(i)$ désigne l'ensemble des k premiers points voisins de i , alors la matrice W des poids est composée de :

$$W_{ij} = \begin{cases} 1 & \text{si } j \in N_k(i) \\ 0 & \text{sinon} \end{cases} \quad (1.7)$$

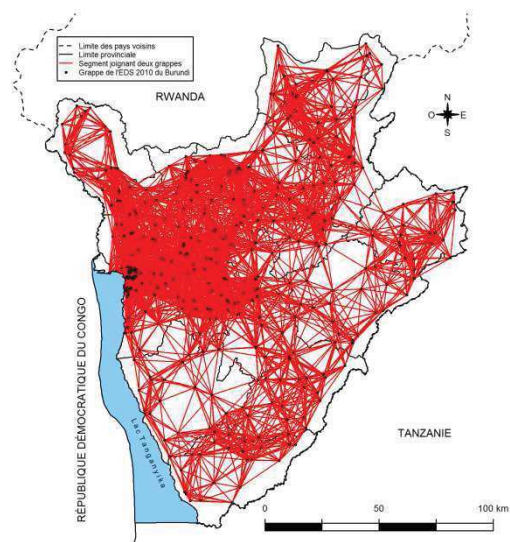
En considérant $k = 20$ grappes voisines proches, la répartition spatiale des grappes de l'EDS 2010 du Burundi (**Figure 12 a**) montre des connexions plus fortes dans certains endroits que dans d'autres. En considérant les grappes qui sont distantes de moins de $d = 25$ km, les connexions plus fortes deviennent plus concentrées dans le Nord-Ouest du pays que dans d'autres endroits (**Figure 12 b**).

Figure 12 : Graphe de voisinage entre les grappes, Burundi EDS 2010

a) 20 voisins proches



b) Voisins distants de moins de 25 km



Le voisinage défini à l'aide d'une distance seuil ou du nombre de voisins sans tenir compte du fait que les voisins sont proches ou distants permet d'évaluer à quelle échelle les valeurs prises par la variable d'intérêt présentent une structure de

dépendance. Cette dépendance spatiale est évaluée lors de l'analyse de l'autocorrélation spatiale.

1.3. Autocorrélation spatiale

L'analyse des données spatialisées passe non seulement par la prise en compte de la dépendance spatiale des données et de l'hétérogénéité spatiale mais également de l'autocorrélation spatiale pour évaluer les effets de l'échelle spatiale [17]. L'autocorrélation spatiale d'une variable, c'est-à-dire la corrélation de proche en proche entre une variable et elle-même dans les différents endroits de l'espace, est une mesure de la dépendance spatiale entre les données. L'idée de base est la loi de Tobler, appelée première loi de Géographie, qui stipule que les objets géographiquement proches ont plus de chances de se ressembler que les objets éloignés. Sous l'hypothèse nulle d'une distribution aléatoire des cas d'une pathologie donnée, l'analyse de l'autocorrélation spatiale permet de détecter le degré de ressemblance ou de dissemblance entre les valeurs prises par une variable d'intérêt.

Dans le cas spatial, l'indépendance des observations est rarement vérifiée. Comme conséquence, les tests statistiques usuels basés sur l'indépendance des observations ne sont plus valables. De plus, lorsqu'il y a présence de l'hétérogénéité spatiale d'un phénomène d'intérêt (par exemple le nombre de sujets séropositifs pour le VIH) traduite par une tendance à l'agrégation spatiale (concentration spatiale) ou à la dispersion spatiale (éparpillement spatial) dans les données, il est important d'en tenir compte dans l'évaluation de l'association entre ce phénomène et les facteurs de risque potentiellement associés à l'échelle de l'individu. L'analyse de l'autocorrélation spatiale va de pair avec la détection des clusters spatiaux. Le terme « cluster » est utilisé dans les EDS pour désigner une « grappe », c'est-à-dire une zone de dénombrement regroupant des ménages. Il sera utilisé ici pour désigner un « agrégat » de cas, c'est-à-dire un regroupement géographiquement délimité de cas, d'une taille et d'une concentration suffisante pour rendre improbable qu'il se soit produit aléatoirement [18]. Autrement dit, il s'agit d'une zone de forte intensité d'un processus donné.

Des techniques spatiales intégrant l'autocorrélation spatiale sont donc utilisées pour faire l'inférence statistique. Les mesures de l'autocorrélation spatiale se classent en deux catégories : les indices globaux et les indices locaux.

1.3.1. Indices globaux d'autocorrélation spatiale

Les indices globaux de l'autocorrélation spatiale, appelés également méthodes globales de détection des agrégats spatiaux ou d'identification de l'hétérogénéité spatiale, permettent de mesurer la tendance à l'agrégation ou à la dispersion globales dans les unités géographiques prises dans leur ensemble, c'est-à-dire dans tout le domaine d'étude. L'analyse de l'autocorrélation spatiale permet de répondre par exemple à la question de savoir si les valeurs de la prévalence du VIH de l'EDS 2010 du Burundi observées dans des grappes géographiquement proches sont similaires à celles observées dans des grappes plus éloignées. S'il en est ainsi, la similarité dans l'espace traduit donc la similarité entre les valeurs de la variable. L'autocorrélation spatiale globale est évaluée à l'aide des indices globaux de l'autocorrélation spatiale et des corrélogrammes (que nous définirons plus loin). Leurs expressions mathématiques diffèrent principalement dans la définition des variables prises en compte pour définir ces indices et dans la définition des poids spatiaux à prendre en compte pour pondérer les observations.

1.3.1.1 Indice global de Moran

L'indice global I_M de Moran, comparable au coefficient de corrélation linéaire de Pearson, est une mesure globale permettant de détecter l'autocorrélation spatiale dans les données observées [11]. Plus utilisé dans le domaine de la recherche en santé publique, il mesure les ressemblances entre voisins [19].

Il s'exprime comme le ratio de la covariance entre les valeurs de la variable observées en deux points de l'espace pondérée par la matrice d'influence des points (matrice des poids), et la variance de cette variable [20] :

$$I_M = \frac{N \sum_{i=1}^N \sum_{j=1}^N W_{ij} (Z_i - \bar{Z})(Z_j - \bar{Z})}{S_0 \sum_{i=1}^N (Z_i - \bar{Z})^2} \quad (1.8)$$

où $S_0 = \sum_{i=1}^N \sum_{j=1}^N W_{ij}$, N est le nombre total de points observés, W_{ij} le poids spatial (éléments de la matrice W) entre les points i et j ($i \neq j$), Z_i la valeur de la variable au point i et \bar{Z} la moyenne de la variable Z . Les éléments de W sont tels que $W_{ij} = 0$. Les valeurs de cet indice donnent donc une idée de son ordre de grandeur : il varie entre -1 (parfaite dispersion) et +1 (parfaite corrélation) dans le cas d'une matrice des poids standardisés. Une autocorrélation spatiale positive corrobore la loi de Tobler et une autocorrélation spatiale négative voudrait dire que les points proches ont tendance à avoir des valeurs plus dissimilaires que celles attendues sous l'hypothèse d'une distribution aléatoire, ce qui traduit le signe que les points éloignés ont tendance à se ressembler et que les points proches ont tendance à se comporter différemment.

Les hypothèses du test de l'indice global de Moran sont :

H_0 : Il y a absence d'autocorrélation spatiale

H_1 : Il y a une autocorrélation spatiale

L'hypothèse nulle H_0 voudrait aussi dire que $I_M = 0$. Le test permettant de vérifier H_0 est basé sur une méthode de rééchantillonnage [21]. Sous H_0 , l'indice global de

Moran standardisé $z_{I_M} = \frac{I_M - E(I_M)}{\text{Var}(I_M)}$ suit une loi normale de moyenne $E(I_M)$ qui ne

dépend que de N et de variance $\text{Var}(I_M)$ qui dépend de N et de W [11]. Les formes

de l'expression de la variance de l'indice global de Moran sont détaillées ailleurs [22,23].

Connaissant le score z_{I_M} pour la statistique de Moran, la p-value s'obtient par :

$$P\text{-value} = P(|z_I| > I) \quad (1.9)$$

La statistique de Moran suppose une distribution spatiale uniforme des cas dans les zones ou aux points observés [24]. De plus, la normalité asymptotique de la distribution des cas ou de la distribution de la proportion des cas est rarement vérifiée [25,26]. Des techniques basées sur les simulations de Monte Carlo par chaînes de Markov (MCMC) peuvent permettre de résoudre le problème en cas de faibles effectifs ou d'absence de normalité [21]. Dans ce cas, après permutation des observations, un ensemble de valeurs de l'indice global de Moran est calculé et la p-value est donnée par le rapport entre le nombre de fois que l'indice de Moran calculé par bootstrap dépasse l'indice de Moran observé et le nombre de réplicats (plus un). Une valeur non significativement différente zéro pour cet indice traduit l'absence d'autocorrélation [23]. Cela veut dire que les cas sont distribués de manière aléatoire dans l'espace d'étude.

Il a été proposé une généralisation par Walter dans le cas d'effectifs regroupés par sites d'observation qui consiste à standardiser les effectifs en soustrayant les effectifs attendus dans chaque site sous l'hypothèse nulle que le risque est constant et en divisant par leurs racines carrées [27] :

$$I_{rc} = \frac{\sum_{i=1}^N \sum_{j=1}^N W_{ij} \frac{Z_i - r n_i}{\sqrt{r n_i}} \frac{Z_j - r n_j}{\sqrt{r n_j}}}{\sum_{i=1}^N \sum_{j=1}^N W_{ij}} \quad (1.10)$$

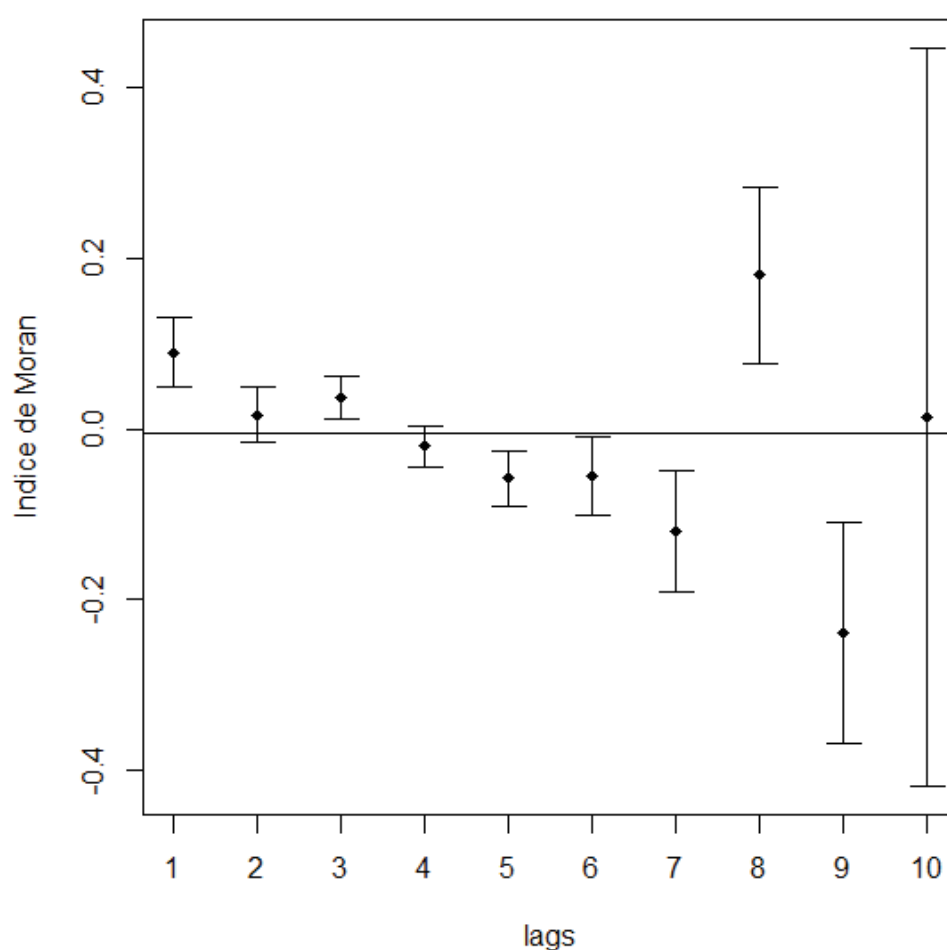
où n_i est la population dans le site i et r le taux d'incidence de la maladie.

Une aide à la recherche des structures spatiales particulières dans les données est le corrélogramme spatial dont l'utilité est de permettre de se rendre compte des distances auxquelles l'autocorrélation spatiale est significative. Autrement dit, le corrélogramme montre la manière dont varie l'autocorrélation spatiale représentée par l'indice global de Moran en fonction des différentes valeurs de la distance seuil ou des distances seuil entre les voisins (classes de distances) et permet d'identifier l'échelle d'organisation de la structure spatiale détectée dans les données [22,28]. Cette distance seuil est la distance au-delà de laquelle deux unités spatiales n'ont plus d'influence l'une sur l'autre. Autrement dit, si la distance entre deux unités spatiales est supérieure à cette distance seuil, alors un poids nul est attribué à ce couple d'unités spatiales, étant donné que l'indice global de Moran fait intervenir la matrice des poids spatiaux dans le calcul. À chaque valeur de distance, un intervalle de confiance de l'indice local est construit et permet de tester l'hypothèse nulle selon laquelle l'espérance mathématique de l'indice local est nulle (**Figure 13**). La distance à laquelle l'autocorrélation spatiale est la plus significative est utilisée comme distance optimale à la recherche des agrégats spatiaux par la méthode de Kulldorff (que nous détaillerons plus loin). Étant donné que la distance entre les grappes varie de 0 à 767,5 km, les distances seuil (50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750) ont été utilisées. Cette figure est une représentation de l'indice global de Moran à une distance seuil maximale de 100 km jusqu'à laquelle l'indice de Moran est calculé, un poids spatial nul étant attribué aux grappes distantes de plus de 100 km. La distance minimale est de 0,01 km.

Ce graphique montre que, à distance minimale, l'indice de Moran vaut 0,12. Cet indice commence à s'annuler ou à changer de signe à partir du voisinage à l'ordre 4. L'autocorrélation spatiale est positive et significative pour les classes de distance faible (égales à 1 et 3 respectivement), négative significative pour les classes de distances fortes aux classes de distances égales 5, 6 et 7, puis redevient positive significative (hautement d'ailleurs) pour la classe de distance égale à 8 et redevient encore négative significative pour la classe de distance égale à 9. Étant donné que l'indice de Moran peut être fragile aux lags plus faibles, la distance seuil à prendre en considération pour la recherche des agrégats spatiaux serait le lag de distance 9, soit 90 km. C'est donc au-delà de 90 km que l'indice de Moran n'est plus significatif.

Cette distance (90 km) pourrait alors servir de portée de l'autocorrélation spatiale positive. Autrement dit, c'est cette distance seuil de 90 km, appelée distance de coupure, qui serait utilisée pour la recherche des agrégats spatiaux et le calcul de l'indice local de Moran. L'intérêt des corrélogrammes est donc qu'ils permettent de déterminer l'échelle optimale pour la recherche des agrégats spatiaux à l'aide des tests locaux.

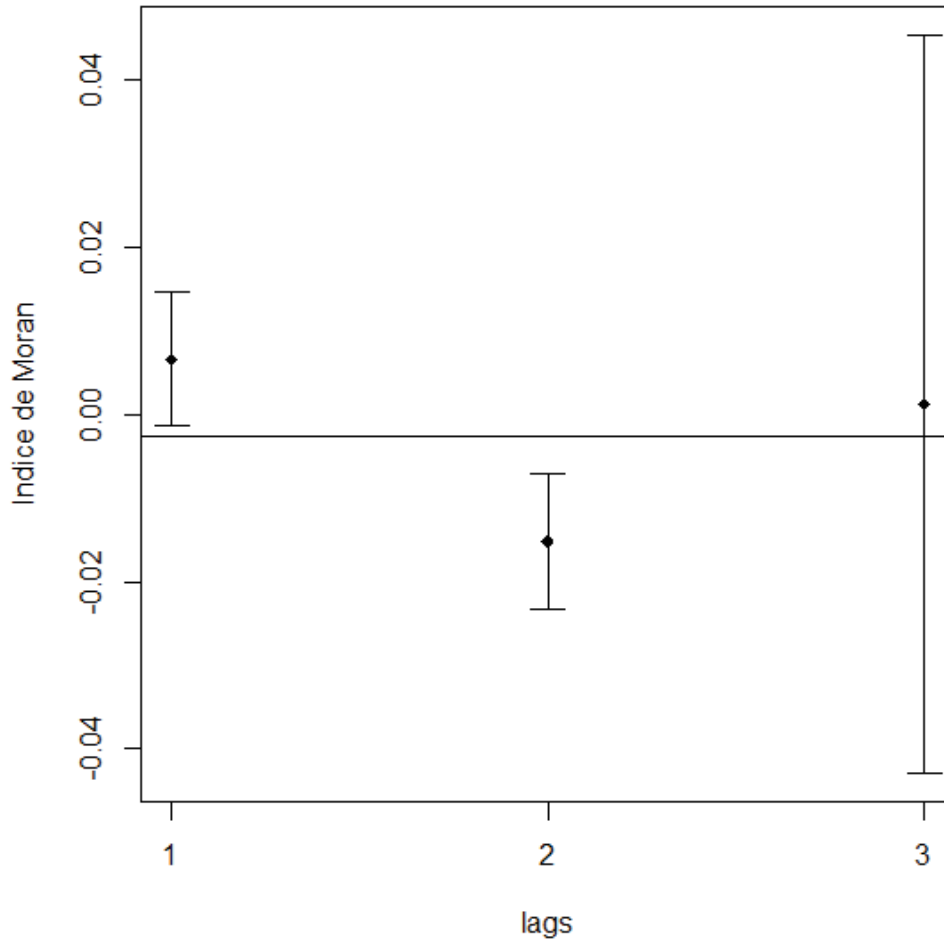
Figure 13 : Corrélogramme de la prévalence du retard de croissance en Côte d'Ivoire, 1994 avec comme distance seuil 100 km



La distance entre les grappes pour les données relatives à la prévalence du VIH au Burundi en 2010 varie entre 0,03 km et 237,4 km. En créant une séquence de valeurs de distances (20, 40, 60, 80, 100, 120 km) à utiliser comme distance seuil dans la définition du voisinage entre les grappes, le corrélogramme basé sur le voisinage de rayon égal à 80 km montre que l'indice de l'autocorrélation spatiale de Moran est significativement différent de zéro au lag numéro 2, ce qui correspond à

une distance de 53,3 km (**Figure 14**). Cette distance maximale seuil sera utilisée lors de l'analyse variographique.

Figure 144 : Corrélogramme de la prévalence du VIH au Burundi, 2010 avec comme distance seuil 80 km



1.3.1.2 Indice global de Geary

L'indice de Geary, comparable au coefficient de Durbin et Watson, est une mesure de la dépendance spatiale qui se présente comme un rapport de variances [29,30] :

$$c = \frac{(N-1) \sum_i \sum_j W_{ij} (Z_i - Z_j)^2}{2S_0 \sum_i (Z_i - \bar{Z})^2} \quad (1.11)$$

Contrairement à l'indice global de Moran, cet indice varie entre 0 (autocorrélation spatiale positive) et 2 (autocorrélation spatiale négative). De plus, son rôle est d'identifier les ressemblances et les dissemblances entre les points voisins. Sous l'hypothèse nulle d'absence d'autocorrélation spatiale ou d'indépendance spatiale, l'indice global de Geary vaut 1 [23]. L'indice de Geary est parasité par de grandes différences entre les valeurs observées en des points plus proches [11]. Autrement dit, il est sensible aux variations locales [10].

L'espérance de cet indice vaut 1 et sa variance dépend, dans une mesure différente de l'indice global de Moran, de N et de W [31]. L'écriture plus complexe des variances des indices de Moran et de Geary a été détaillée par Hubert Jayet, 2001 [32].

Il a été montré que la statistique c suit une loi normale. La différence entre l'indice de Moran et celui de Geary est que le premier est un indice qui s'intéresse à la variabilité entre les zones voisines et que le second s'intéresse à la ressemblance entre les zones voisines. Les indices c de Geary et I_M de Moran sont liés par [33] :

$$c = \frac{N-1}{\sum_i \sum_j W_{ij}} \frac{\sum_i W_{i\cdot} (Z_i - \bar{Z})^2}{\sum_i (Z_i - \bar{Z})^2} - \frac{N-1}{N} I_M \quad (1.12)$$

La plupart des justifications de l'utilisation de l'indice de Moran face à celui de Geary s'appuient sur le fait qu'il est semblable au coefficient linéaire de Pearson, ce qui ne justifie pas la non-utilisation de l'indice de Geary puisque les deux sont liés. L'inconvénient majeur de ces deux indices est qu'ils attribuent le même poids à chaque unité spatiale dans laquelle les points d'observation appartiennent [3].

1.3.1.3 Indice global de Whittemore

L'absence de l'autocorrélation spatiale est un signe d'absence d'agrégation spatiale. Considérons un domaine d'étude divisé en N zones d'échantillonnage contiguës de

telle sorte que les cas d'une pathologie donnée soient placés aux centres de ces zones. Comme conséquence, la distance entre deux cas situés dans une même zone sera nulle. L'indice global de Whittemore s'écrit comme une distance moyenne entre toutes les zones où les cas de maladie ont été observés [34].

Le nombre de cas observés Y_i dans chaque zone i est une variable aléatoire de Poisson :

$$Y_i \sim \text{Poisson}(\lambda) \quad (1.13)$$

Le paramètre de cette loi est proportionnel à la taille ξ_i de chaque zone sous l'hypothèse nulle, λ étant la constante de proportionnalité. Pour obtenir la loi de la variable qui désigne le nombre de cas dans tout le domaine d'étude, il convient de faire la convolution de la loi de Poisson. Sachant que cette loi est stable pour la convolution, la somme de variables aléatoires indépendantes de Poisson est une nouvelle variable de Poisson dont l'espérance est égale à la somme de leurs espérances. La distance totale entre toutes les paires de cas permettant le calcul l'indice de Whittemore s'obtient en calculant la distance entre les paires de zones, puis en multipliant cette distance par le nombre de cas de chaque zone. Soient r la fréquence relative des cas dans chaque zone rapportés au nombre total des cas et W la matrice des distances mutuelles entre les centres des zones où les cas ont été rapportés.

La distance totale vaut $\frac{1}{2} N^2 r' W r$. De fil en aiguille, la distance moyenne (l'indice de Whittemore) s'obtient en divisant la distance totale par le nombre total de valeurs de distances, soit :

$$\delta = \frac{N}{N-1} r' W r \quad (1.14)$$

Les valeurs de l'espérance et de la variance de cet indice sont détaillées ailleurs [34].

1.3.1.4 Indice global de Tango

Contrairement aux indices précédents, l'indice de Tango était initialement destiné à détecter les agrégats temporeux de cas de maladies, notamment infectieuses, lors des études épidémiologiques [35]. L'espace de temps est divisé en k intervalles de temps réguliers d'effectifs Y_1, \dots, Y_k contenant des cas de maladie et dans lesquels la population de taille N était supposée stable au cours du temps. Les observations sont alors les centres des intervalles de temps et la distance temporelle est définie comme étant la valeur absolue de la différence entre ces centres ou peut prendre diverses formes dont l'exponentielle décroissante de cette distance. L'indice d'agrégation de Tango C est alors défini comme une forme quadratique de la matrice W des distances et du vecteur des fréquences $r = \left(\frac{Y_1}{N}, \dots, \frac{Y_k}{N} \right)$ des cas dans chacun des k intervalles de temps :

$$C = r'Wr \quad (1.15)$$

La distribution asymptotique de cet indice a été approximée par une loi normale lorsque le nombre d'intervalles de temps devient important [26]. D'autres auteurs avaient également approuvée que la distribution asymptotique de cet indice était normale [34,36]. Cependant, la distribution de C et de sa version standardisée est asymétrique. Ainsi, il a été recommandé d'utiliser une loi du chi-deux pour approximer sa distribution surtout lorsque la taille de l'échantillon n'est pas grande [26]. Une autre approximation de cet indice transforme celui-ci en une fonction linéaire de la statistique U définie comme une moyenne des distances entre les cas de maladie [37].

Cette version temporelle de l'indice de Tango a ensuite été étendue au cas où les intervalles de temps ne sont pas réguliers et au cas spatial. Dans ce dernier cas, plutôt que de considérer la population totale constituant la population à risque comme dénominateur, les nombres de cas de maladie sont rapportés à la population de chaque unité spatiale. Autrement dit, les cas sont groupés dans des unités

spatiales représentées par leurs centroïdes, la population à risque est aussi supposée stable dans ces unités spatiales [38].

Sous l'hypothèse nulle d'absence d'agrégat spatial ou de risque uniforme dans le domaine d'étude, la statistique de Tango qui compare deux distributions s'écrit comme une somme pondérée des produits des écarts entre la proportion observée de cas dans chaque unité rapportée au nombre total de cas et la proportion attendue, sous l'hypothèse nulle, correspondant à la proportion de la population dans chaque unité spatiale [27] :

$$T = \sum_{i=1}^k \sum_{j=1}^k W_{ij} \left(\frac{Y_i}{Y} - \frac{n_i}{n} \right) \left(\frac{Y_j}{Y} - \frac{n_j}{n} \right) \quad (1.16)$$

où W_{ij} sont les poids spatiaux définis par une relation de la forme (1.4) entre les unités i et j , k le nombre d'unités spatiales, n_i la taille de la population à risque dans l'unité i , Y le nombre total de cas et n le nombre total d'individus à risque. Cet indice est vu comme la somme des produits spatialement pondérés des écarts entre les proportions observée et attendue de cas dans chaque unité spatiale. Le test de sa significativité est basé sur la loi du Chi-deux. En effet, la variable obtenue après la standardisation de la statistique T suit une loi du Chi-deux dont le degré de liberté dépend de la matrice W . Rogerson (1999) a proposé un indice R qui permet de corriger cet indice en le divisant par la racine carrée du produit des proportions observées de la population à risque se trouvant dans les unités i et j [39]. Du fait qu'une valeur élevée de l'indice de Tango (comme pour les autres indices tels que celui de Moran) peut ne pas être un signe d'une forte agrégation mais d'un manque d'ajustement, il a proposé de partitionner l'indice de Tango en une somme de deux indices : un indice de Tango avec $i \neq j$ considéré comme un indice global d'agrégation spatiale (numérateur de l'indice de Moran) et un indice de Tango avec $i = j$ considéré comme un indice de la qualité de l'ajustement représentant la somme des carrés des écarts entre les proportions observées et attendues de cas dans les unités spatiales.

Cela s'écrit [38] :

$$T = \sum_{i=1}^k \sum_{j=1, j \neq i}^k W_{ij} \left(\frac{Y_i - n_i}{Y. - n.} \right) \left(\frac{Y_j - n_j}{Y. - n.} \right) + \sum_{i=1}^k W_{ii} \left(\frac{Y_i - n_i}{Y. - n.} \right)^2 \quad (1.17)$$

1.3.1.5 Indice global d'Oden

La variabilité d'un phénomène au sein des unités spatiales varie en fonction de la taille de la population qui se trouve dans chaque unité spatiale. En effet, des unités spatiales avec des effectifs peu élevés peuvent présenter une variabilité plus grande que les unités avec des effectifs plus élevés [40]. C'est pourquoi Oden (1985) a proposé un indice de détection des clusters corrigé à partir de l'indice global de Moran afin de tenir compte de cette variabilité dans la détection des agrégats spatiaux. Cet indice a la même espérance que l'indice global de Moran. Une description plus détaillée de cet indice et de sa variance peut être trouvée ailleurs [40,41].

La notion de l'autocorrélation spatiale globale est une notion importante dans la mesure où elle permet de mesurer la dépendance spatiale sur tout le domaine d'étude et, par conséquent, de faire la prédiction des valeurs de la variable en des points non échantillonnés. Elle suppose, cependant, que la variable d'intérêt soit décrite par un processus stationnaire, c'est-à-dire que ses moments centrés d'ordre 1 (espérance) et 2 (variance et covariance) existent et ne dépendent que de la distance qui sépare les points d'observation. Cela est rarement vérifié. De plus, l'autocorrélation spatiale globale, qui est supposée être uniforme sur tout le domaine d'étude, peut masquer l'autocorrélation spatiale dans certaines zones. Des indicateurs locaux d'association spatiale (LISA : Local Indicators of Spatial Association) ont été introduits en 1995 pour permettre d'apporter un complément d'information quant à l'existence ou non des agrégats locaux qui pourraient ne pas être identifiés lors des tests sur les indices globaux appelés tests globaux. De ce fait, une méthode globale devrait être accompagnée d'une méthode locale.

1.3.2. Indices locaux d'autocorrélation spatiale

Les indices locaux de l'autocorrélation spatiale, appelés tests locaux d'association spatiale, permettent de localiser les agrégats spatiaux dans des unités spatiales de niveau inférieur à celui du domaine d'étude et ainsi d'affirmer ou d'infirmer si les valeurs d'une variable dans une entité donnée sont similaires à celles des entités avoisinantes. À titre exemplatif, l'étude de l'autocorrélation spatiale locale de l'infection par le VIH au Burundi pourrait permettre de répondre à la question de savoir s'il existe un agrégat local de cas constituant une singularité locale, c'est-à-dire une agglomération de sujets VIH positifs dans une province ou un point chaud, c'est-à-dire un agrégat local de cas agglomérés dans plusieurs provinces avoisinantes.

L'autocorrélation spatiale locale est évaluée à l'aide des indices locaux d'autocorrélation spatiale et du scan spatial. Dans certains cas, le point-source de propagation d'une maladie peut être suspecté. C'est par exemple le cas de la détection des agrégats du nombre de cas d'une épidémie autour d'une source bien spécifiée comme les eaux stagnantes pour le paludisme. Le test de significativité de l'indice permettant de détecter des agrégats de cas avec un point-source prédéfini est qualifié de « focalisé » [41]. Dans la plupart des cas, le point-source n'est pas bien défini : le test de significativité de l'indice est alors qualifié de « non focalisé ». Les tests qui s'intéressent d'une part à la présence des agrégats spatiaux et d'autre part à l'évaluation du risque autour d'un point-source prédéfini respectivement ont été décrits dans la littérature [21]. Nous allons ici nous intéresser aux tests non focalisés.

Les tests locaux permettent, en outre, de détecter et de tester la significativité de 4 types d'endroits : les endroits avec des valeurs élevées (E) de la variable entourés par des endroits avec des valeurs élevées appelés points chauds, les endroits avec des valeurs basses (B) de la variable entourés par des endroits avec des valeurs basses appelés points froids, les endroits avec des valeurs basses (B) de la variable entourés par des endroits avec des valeurs élevées (E) et les endroits avec des valeurs élevées (E) de la variable entourés par des endroits avec des valeurs

basses (B) dans les unités spatiales de la région d'étude. Ces endroits sont visualisés sur la carte en même temps que les valeurs élevées de la variable dans les unités spatiales. Le premier endroit est alors nommé EE (plus-plus ou high-high), le deuxième BB (moins-moins ou low-low), le troisième BE (moins-plus ou low-high) et le quatrième EB (plus-moins ou high-low).

1.3.2.1 Indice local de Moran

L'indice local de Moran est une version locale de l'indice global de Moran. De ce fait, il permet de tester s'il y a une autocorrélation spatiale entre les valeurs observées d'une variable dans chaque entité géographique du domaine d'étude en tenant en compte des entités avoisinantes.

Pour une entité i donnée du domaine d'étude, l'indice local de Moran s'écrit [31] :

$$\rho_i = (Z_i - \bar{Z}) \sum_{j, j \neq i} W_{ij} (Z_j - \bar{Z}) \quad (1.18)$$

Comme pour l'indice global de Moran, la valeur moyenne et la variance de cet indice ont été établies [22]. Sa significativité est évaluée en utilisant la loi normale centrée-réduite.

Les indices local et global de Moran sont liés par :

$$I = \frac{\sum_i \rho_i}{S_0 S^2} \quad (1.19)$$

où S_0 représente le nombre de paires utilisées dans la calcul de l'indice local de Moran et S^2 la variance non réduite de la variable Z_i . Le facteur de proportionnalité

entre ces deux indices est donc $S_0 S^2$ et vaut S_0 dans le cas des variables standardisées.

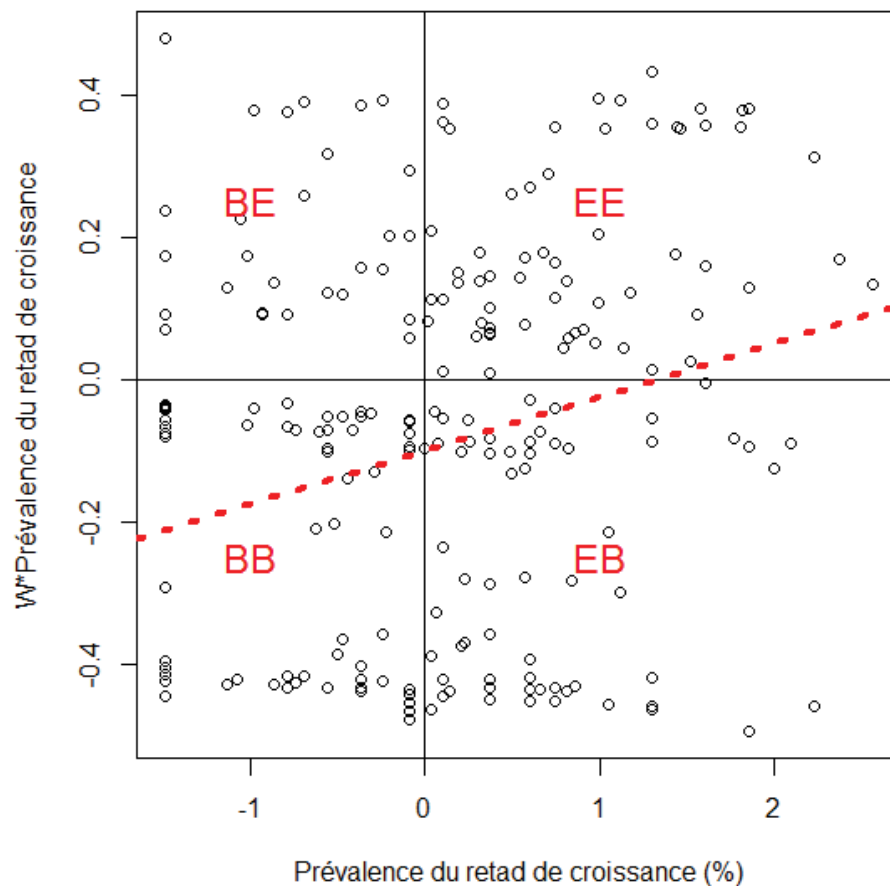
En plus de l'indice global de Moran et de sa version généralisée, le nuage de points de Moran (« Moran scatter plot ») peut être utilisé pour analyser l'autocorrélation spatiale globale. Utilisé conjointement avec l'indice local de Moran, le nuage de points de Moran est un outil essentiel dans l'évaluation des comportements locaux des phénomènes [42]. Pour le construire, il est porté en abscisses les valeurs de la variable quantitative Z d'intérêt, souvent sous sa forme standardisée, et en ordonnée les pas (lags) spatiaux WZ de cette même variable. Appelés « variable spatialement décalée », ces lags sont obtenus en faisant des moyennes spatialement pondérés des valeurs observées Z_j dans les entités géographiquement voisines [43] :

$$lag_i = \frac{\sum_{j=1}^N W_{ij} Z_j}{\sum_{j=1}^N W_{ij}} \quad (1.20)$$

La normalisation de la matrice W des poids spatiaux et le centrage de la variable Z ont pour conséquence que les lags WZ et la variable Z soient de moyennes nulles. La variable WZ est régressée sur la variable Z , ce qui conduit à une droite d'ajustement passant par l'origine des axes et dont le coefficient (ou pente de la droite) correspond à la valeur de l'indice global de Moran. L'intérêt du nuage de Moran est de détecter notamment les valeurs aberrantes et de confirmer la positivité ou la négativité de l'indice global de Moran via le signe de la valeur de la pente de la droite de régression. Une fois la structure spatiale détectée, l'échelle spatiale à laquelle elle est la plus significativité est testée.

Appliqué aux données de prévalence du retard de croissance (que nous définirons plus loin au troisième chapitre) chez les enfants de moins de trois ans en Côte d'Ivoire en 1994 (**Figure 15**), le nuage de points de Moran est ajusté à l'aide d'une droite de régression linéaire dont la pente est égale à l'indice global de Moran.

Figure 15 : Nuage de points de Moran de la prévalence du retard de croissance, Côte d'Ivoire 1994



Les droites horizontale et verticale passant par ces moyennes partagent le nuage de points de Moran en 4 quadrants précédemment dénommés EE (high-high), BB (low-low), BE (low-high) et EB (high-low) qui traduisent 4 types d'association locale. C'est la raison pour laquelle les indices locaux d'autocorrélation spatiale sont appelés indices locaux d'association spatiale.

Comme il s'agit d'une régression linéaire, il est possible de calculer la distance de Mahalanobis associée à chaque observation et de considérer que cette observation est aberrante si cette distance est supérieure au rapport entre le nombre de paramètres (deux ici) multiplié par deux et le nombre d'observations. Les observations aberrantes, qui rendent faible l'association spatiale positive, peuvent être visualisées sur le graphique comme des observations qui s'écartent sensiblement de la distribution des autres observations. Graphiquement, les valeurs basses se trouvant dans un environnement des valeurs élevées (moins-plus) ou les

valeurs élevées se trouvant dans un environnement des valeurs basses (plus-moins) constituent des valeurs localement atypiques.

Le nuage de points de Moran permet de faire à la fois une analyse locale et globale. Le nuage de points de Moran traduit des associations entre les valeurs élevées et basses. Les quadrants EE et BB montrent des associations positives et les quadrants EB et BE des associations négatives. Outre la détection de l'aberrance des points, le nuage de points de Moran permet d'évaluer l'homogénéité du phénomène étudié et de sa dépendance spatiale. En effet, le coefficient R_a^2 de détermination ajusté de la régression linéaire simple de WZ sur Z est faible (7,6 %), ce qui traduit une absence d'homogénéité du processus spatial.

Il est commode de superposer la carte des indices locaux de Moran et le nuage de points de Moran pour détecter, comme souligné plus haut, les valeurs aberrantes et les leviers (mesures qui caractérisent la manière dont une observation s'écarte des autres).

1.3.2.2 Indice local de Geary

Comme pour l'indice local de Moran, pour une entité i donnée du domaine d'étude, l'indice local de Geary s'écrit [31] :

$$\gamma_i = \frac{N \sum_j W_{ij} (Z_i - Z_j)^2}{\sum_i (Z_i - \bar{Z})^2} \quad (1.21)$$

L'indice local de Geary n'est pas normé à l'unité contrairement à l'indice global de Geary [22]. Une valeur faible de cet indice montre des entités avec des valeurs homogènes et une valeur élevée des entités avec des singularités locales. L'indice local de Geary est testé à l'aide de la méthode de ré-échantillonnage.

La relation entre les indices global et local de Geary est telle que :

$$\sum \gamma_i = \frac{2N \sum_i \sum_j W_{ij}}{N-1} c = \frac{2NS_0}{N-1} c \quad (1.22)$$

L'indice local de Geary est moins préféré à l'indice local de Moran et est moins utilisé dans la littérature.

1.3.2.3 Indice local de Getis-Ord

L'indice global de Getis et Ord s'appuie sur la matrice des distances définies pour une distance d bien déterminée séparant deux points d'observation i et j . Getis et Ord ont défini deux statistiques, l'une incluant et l'autre excluant la valeur de la variable au point i [42] :

$$G(d) = \frac{\sum_i \sum_j W_{ij}(d) Z_i Z_j}{\sum_i \sum_j Z_i Z_j} \quad (1.23)$$

$$G_i(d) = \frac{\sum_{j \neq i} W_{ij}(d) Z_j}{\sum_{j \neq i} Z_j} \quad (1.24)$$

Une valeur positive de cet indice montre que les valeurs élevées de la variable sont groupées ensemble et une valeur négative montre que les valeurs basses sont groupées ensemble, ce qui est différent de l'indice local de Moran. Contrairement à ce dernier et à l'indice de Geary, il ne peut montrer ni les points chauds ni les points froids. La significativité de cet indice est évaluée à l'aide d'un test de normalité.

Les résultats des tests locaux sont fortement influencés par le nombre d'entités géographiques formant le découpage du domaine d'étude, l'inégale répartition de la population dans les entités (effets de taille) et l'inégale étendue des entités (effets de surface).

Tous ces indices précédents sont basés sur la notion de distance ou de contiguïté. Une autre technique, qui vise le même objectif mais qui est basée sur le balayage à l'aide d'une fenêtre mobile est le scan spatial pour la détection d'agrégats spatiaux.

1.3.2.4 Le scan spatial

Cette méthode cherche à identifier et localiser les agrégats spatiaux et à tester leur significativité. Une méthode permettant de détecter et de tester les agrégats spatiaux et/ou temporels en se basant sur le scan spatial a été proposée par Kulldorff. Elle est implémentée dans le programme d'analyse des données spatiales et/ou temporelles appelé SaTScan (Software for the spatial, temporal, and space-time scan statistics) [44–49]. L'hypothèse alternative à l'hypothèse nulle d'une distribution aléatoire (ou tout simplement uniforme) des cas dans les entités géographiques du domaine d'étude est la présence d'un excès de cas d'une maladie par rapport à ce qui est attendu sous cette hypothèse nulle, ce qui traduit également la présence de l'autocorrélation spatiale dans les données. Pour détecter les endroits avec un excès de cas, une fenêtre circulaire de recherche balaye graduellement tout le territoire en calculant, à chaque fois, le nombre de cas observé et celui attendu.

Le rayon de la fenêtre et le maximum du nombre de cas en termes de pourcentage maximal de la population à risque dans cette fenêtre doivent être spécifiés. Ce rayon varie entre zéro et une valeur maximale définie par l'utilisateur et le pourcentage de cas dans la fenêtre varie entre 0 % et 50 % de la population à risque. L'utilisateur choisit également le modèle (Bernoulli ou Poisson par exemple) selon le type de données, le type d'analyse (purement spatiale, purement temporelle ou spatio-temporelle par exemple) selon l'objectif de l'étude et le type de cluster à détecter (faible ou haute prévalence par exemple).

Selon le modèle utilisé, la vraisemblance de chaque fenêtre est calculée. Ainsi par exemple, pour le modèle de Bernoulli, la vraisemblance s'écrit :

$$L(c, C) = \left(\frac{c}{n}\right)^c \left(\frac{n-c}{n}\right)^{n-c} \left(\frac{C-c}{N-n}\right)^{C-c} \left(\frac{(N-n)(C-c)}{N-n}\right)^{(N-n)(C-c)} I_{\{c > E(c)\}} \quad (1.25)$$

où C est le nombre total de cas dans tout le domaine d'étude, c le nombre observé de cas dans la fenêtre, $E(c)$ le nombre de cas attendus sous l'hypothèse nulle d'une distribution uniforme, n le nombre total de cas et de témoins dans la fenêtre, N le nombre de cas et de témoins dans le domaine d'étude et I une fonction indicatrice qui, lors de la détection des clusters de haute prévalence d'une pathologie par exemple, vaut 1 s'il y a plus de cas observés qu'attendu dans la fenêtre et 0 sinon. Les valeurs prises par la fonction indicatrice dans la fenêtre s'inversent lors de la détection des clusters de basse prévalence et elles valent 1 lorsqu'il s'agit de détecter les clusters de haute ou de basse prévalence en même temps. La vraisemblance pour une loi de Poisson est décrite ailleurs [50]. Parmi toutes les fenêtres, celle qui a la plus grande vraisemblance constitue le cluster le plus vraisemblable, c'est-à-dire le cluster qui a moins de chances d'être apparu par le simple hasard. Pour chaque cluster détecté, un rapport de vraisemblance est calculé à l'aide des simulations de Monte Carlo ainsi que la valeur de la p-value lui associée. Cette p-value s'obtient en faisant le rapport entre le rang du maximum du rapport de vraisemblances des données et des répliqués (numérateur), et le nombre de répliqués augmenté d'une unité (dénominateur) [51].

Le choix des paramètres (rayon de la fenêtre de recherche, pourcentage maximal de la population à risque) dans SaTScan est délicat [52]. Une valeur trop grande pour le rayon de la fenêtre de recherche conduirait à la détection d'un cluster large de haute prévalence de la pathologie étudiée cachant ainsi les clusters de faible prévalence [53]. Une valeur trop petite conduirait à la non détection d'un cluster significatif [54]. L'obtention d'un cluster large ou étroit de haute prévalence n'est pas toujours le résultat d'un mauvais choix des paramètres.

Les indices locaux de l'autocorrélation spatiale sont un outil qui permet de compléter l'analyse de l'autocorrélation spatiale globale. Une revue des logiciels gratuits (R, SaTScan, GeoDa, CrimeStat) dans lesquels les indices globaux et locaux de l'autocorrélation spatiale sont implémentés a été faite en 2004 par Anselin Luc avec comme champ d'application les données de mortalité liée au cancer aux États-Unis d'Amérique entre 1950 et 1969. Quatre packages de R (spdep, spatstat, spatial, splancs) contenant des fonctions capables de mener à bien l'analyse de l'autocorrélation spatiale ont gardé son attention [55]. À l'issue de ses analyses, il a suggéré que la détection des agrégats spatiaux soit accompagnée du scan spatial de Kulldorff et de l'indice local de Moran.

1.3.2.5 SPODT: Partitionnement spatial oblique

Toutes les techniques précédentes visent globalement à identifier les zones à haut risque d'une maladie donnée. Diverses formes de fenêtre ont été utilisées dans la littérature pour rechercher des endroits avec des excès de cas. C'est la cas notamment des fenêtres elliptiques [56,57]. Une fenêtre variable dont la taille n'a pas besoin d'être choisie a priori a été utilisée pour détecter les agrégats spatiaux [58]. La forme géographique du domaine d'étude peut donner une idée de la forme de la fenêtre, par exemple un domaine de forme ronde ou elliptique. Récemment, une nouvelle technique proposée par Jean Gaudart et ses collègues (2015), basée sur le découpage récursif du domaine d'étude en zones de risques différents à l'aide de l'arbre de régression oblique (ARO), a été mise au point pour mettre en évidence les zones à différents risques par rapport à une zone de référence [59].

Cette technique consiste à découper le domaine d'étude en se basant sur un arbre de classification et de régression (CART : classification and regression tree) [60]. Initialement, CART est une méthode de classification dont le nombre de classes est connu a priori qui permet de construire des arbres de décision afin de classer de nouvelles observations [61]. Adapté au cas spatial, elle découpe le domaine d'étude en zones rectangulaires. Jean Gaudart et ses collègues (2015) ont modifié cette méthode de manière à ce que le domaine d'étude soit découpé en zones de différents risque à l'aide des droites obliques, d'où l'appellation du package spODT

(Spatial Oblique Decision Tree) [59,60,62]. Du fait que cette méthode concerne l'entièreté du domaine d'étude, elle est une méthode globale de détection des clusters. Connaissant les coordonnées géographiques des points d'observation et les valeurs de la variable quantitative d'intérêt en ces points, le domaine d'étude (appelé parent ou racine) est d'abord partitionné en deux zones de telle sorte que la variance inter-zones soit maximale, c'est-à-dire que les deux zones ont des risques les plus différents possibles. Ensuite, chaque zone obtenue est subdivisée en deux sous-zones maximisant le même critère, la variance inter-zones, et ainsi de suite. L'algorithme spODT s'arrête dès qu'il n'est plus possible de faire d'autres subdivisions aboutissant à des zones les plus différentes possibles en termes de risques. Les règles d'arrêt concernent entre autres la taille minimale de la zone-parent (domaine d'étude ou zones créées), le pourcentage maximal de variance expliquée par la coupure ($R^2 < 0,01\%$) et le nombre de nœuds de l'arbre de régression [62]. Une fois le découpage terminé, une valeur du pourcentage de variance expliquée R^2 par la coupure est donnée. Une p-value s'obtient, sous les modèles gaussien, de Poisson, binomial ou uniforme, en comparant cette valeur au R^2 global obtenu à l'aide des simulations de Monte Carlo des ARO. Des fonctions linéaires des coordonnées correspondent aux frontières des zones issues du découpage du domaine d'étude et obtenues de manière récursive de telle sorte que, pour chaque ensemble d'unités spatiales mises dans une zone, ces fonctions maximisent la variance inter-zone. La méthode spODT peut également être qualifiée de locale car elle permet de localiser géographiquement les zones de différents risques et de les comparer à l'aide d'un odds ratio (OR) obtenu à partir du tableau de contingence croisant les individus (cas et témoins) et les zones (zone1, zone2, ...) obtenues à l'issue du partitionnement du domaine d'étude.

1.4. Techniques d'interpolation spatiale

1.4.1. Définition de l'interpolation spatiale, limites et utilisation

L'estimation ou la prédiction spatiale permet d'obtenir les valeurs d'une variable d'intérêt en des points non mesurés en utilisant les valeurs de la variable observées aux points voisins. Lorsque cette estimation spatiale se fait à l'intérieur d'un domaine

d'étude borné, ce qui est souvent le cas, alors il s'agit d'une interpolation spatiale. En fonction du type d'interpolation spatiale choisi, les valeurs interpolées s'accompagnent des mesures de précision, les variances de prédiction. Le krigeage est le meilleur prédicteur linéaire sans biais car la variance de prédiction est minimale, les valeurs prédites s'écrivent comme une combinaison linéaire des valeurs observées.

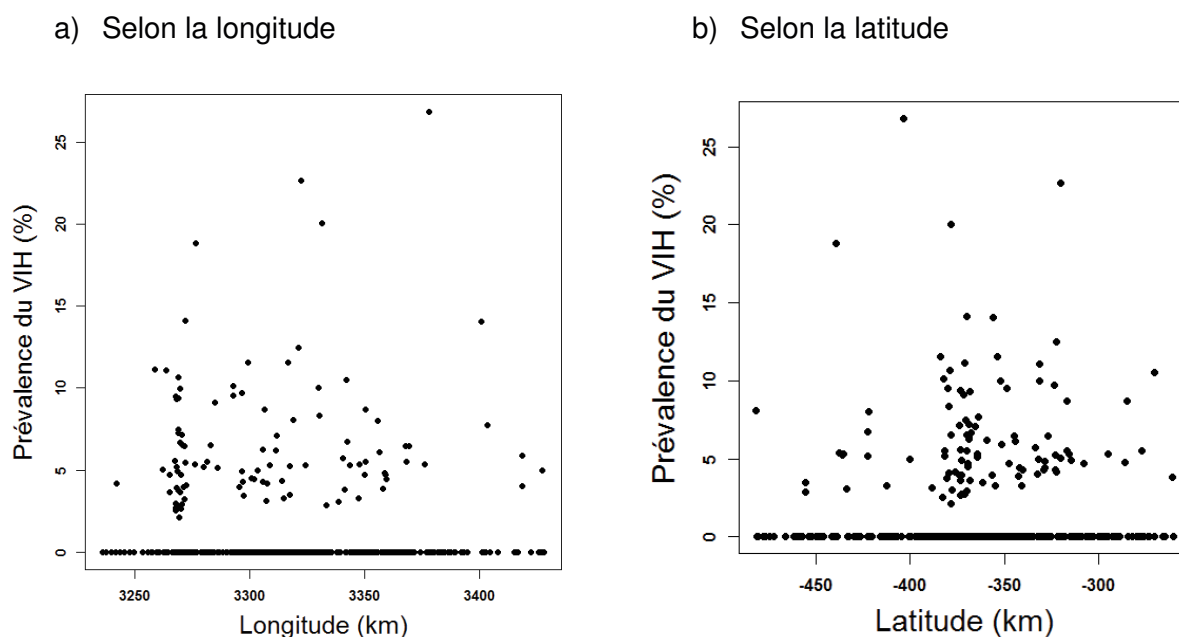
Les résultats du krigeage sont cependant influencés par le choix des distances, de la manière de définir le voisinage (nombre de points voisins, distance seuil à partir de laquelle la notion de voisinage perd le sens) et de la répartition des données dans le domaine d'étude. Cette répartition peut être caractérisée par des observations plus denses dans certains endroits que dans d'autres, un effet d'écran (i.e. l'influence d'une observation est masquée par la présence d'une autre observation qui est intercalée entre elle et l'observation cible), un effet de bord (i.e. les observations situées à la frontière du domaine d'étude ont moins de voisins, ce qui diminue leurs contributions dans l'estimation spatiale dans leurs voisinages).

Une étape préliminaire à l'interpolation spatiale consiste à faire une analyse variographique. Cette analyse est précédée par l'évaluation de la tendance spatiale des valeurs prises par la variable d'intérêt.

1.4.2. Analyse de la tendance spatiale

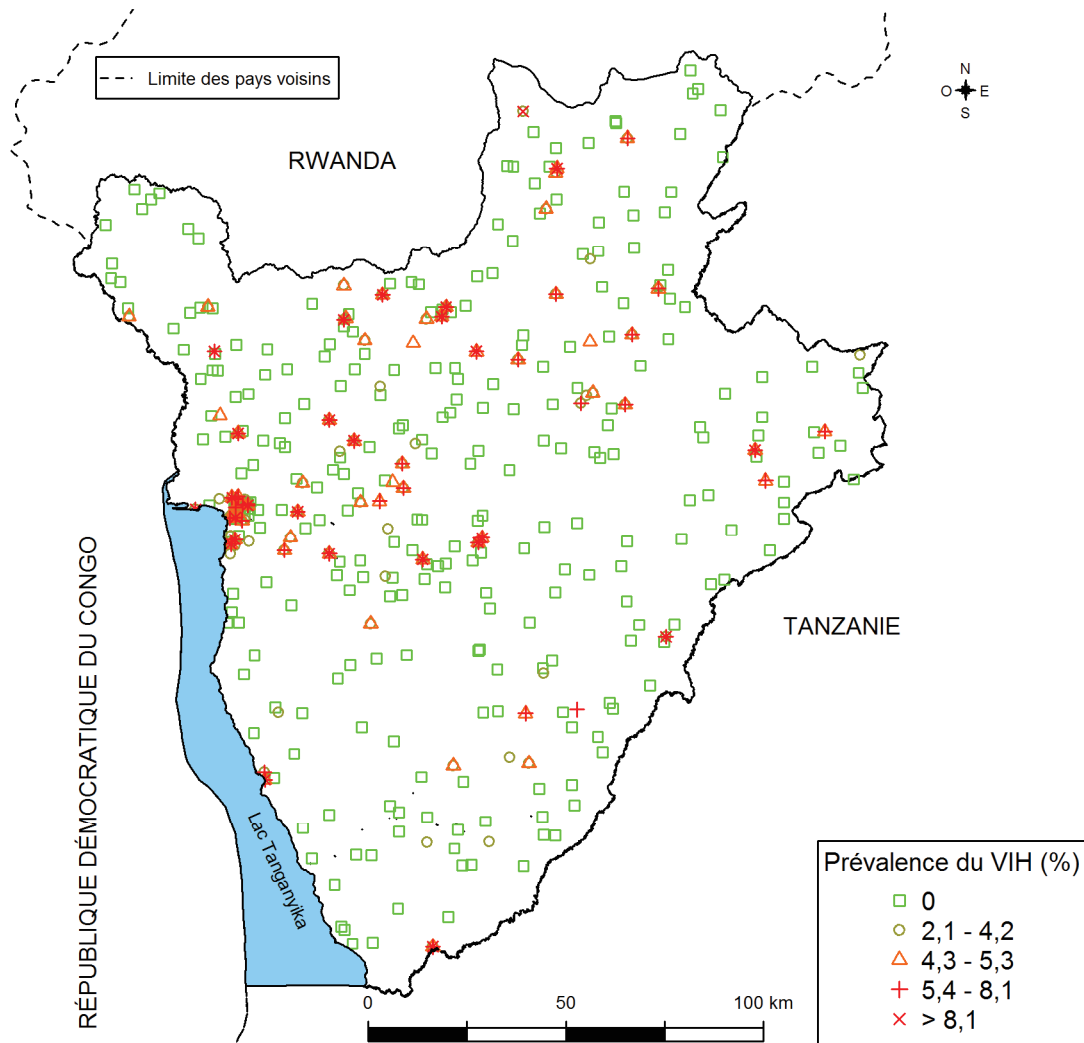
L'analyse de la distribution spatiale d'une variable d'intérêt commence d'abord par la visualisation de ses valeurs selon les valeurs de la longitude et puis de la latitude afin d'évaluer ses variations dans la direction d'un axe (**Figure 16**).

Figure 16 : Évolution de la prévalence du VIH selon les coordonnées géographiques au Burundi, 2010



À travers cette figure, il ressort que la prévalence du VIH ne varie ni dans la direction Est-Ouest, ni dans la direction Nord-Sud. La tendance spatiale dans d'autres directions est évaluée à l'aide de la visualisation de la distribution des valeurs de la variable d'intérêt (divisée en catégories) ou à l'aide de l'analyse variographique (que nous verrons plus loin) en précisant la direction dans laquelle le variogramme est évalué. La distribution spatiale des quartiles de la prévalence du VIH au Burundi est représentée à la **Figure 17**. Elle permet de vérifier si la prévalence du VIH peut être considérée comme constante à travers tout le pays.

Figure 17 : Tendence spatiale de la prévalence du VIH au Burundi, 2010

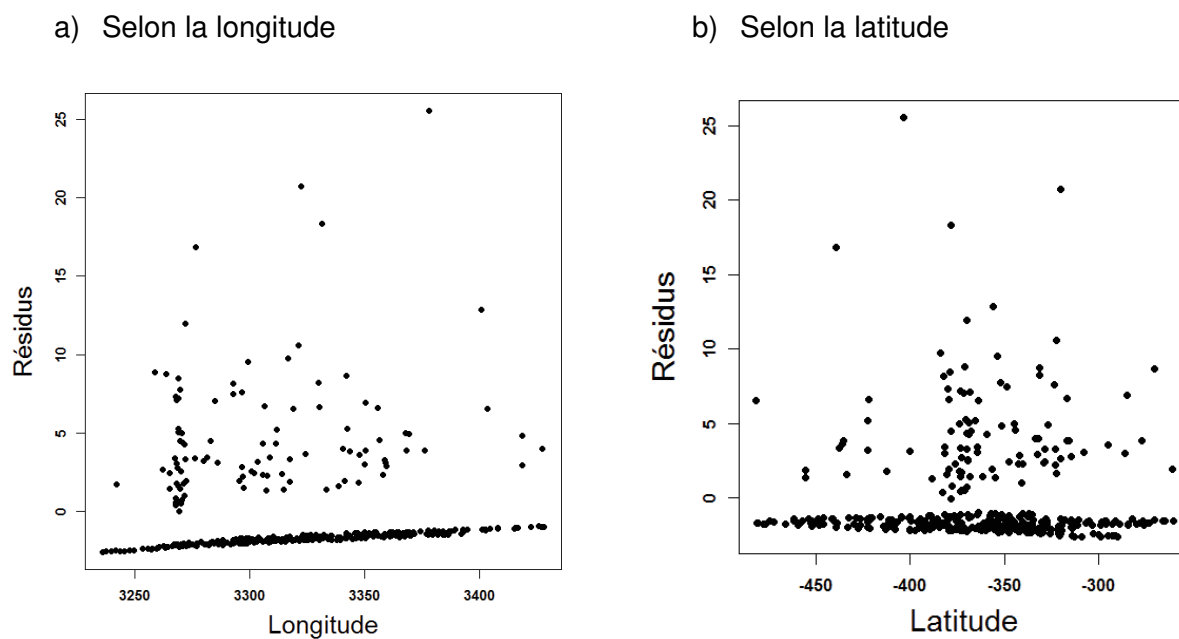


Du fait de la distribution fortement asymétrique de la prévalence du VIH, cette figure est construite en créant la catégorie 1 (la prévalence est nulle) et quatre autres catégories en fonction des valeurs des quartiles de la prévalence non nulle. À travers cette figure, il ressort que l'hypothèse de stationnarité d'ordre 1 ne peut pas être acceptée. Autrement dit, la prévalence du VIH n'est pas constante sur le territoire.

Une alternative consiste à dire que cette prévalence est une fonction polynomiale de premier ordre des coordonnées géographiques. Le but d'étudier le comportement des données selon les coordonnées géographiques est de vérifier si ces coordonnées (ou d'autres variables) seront utilisées comme covariables dans la construction des

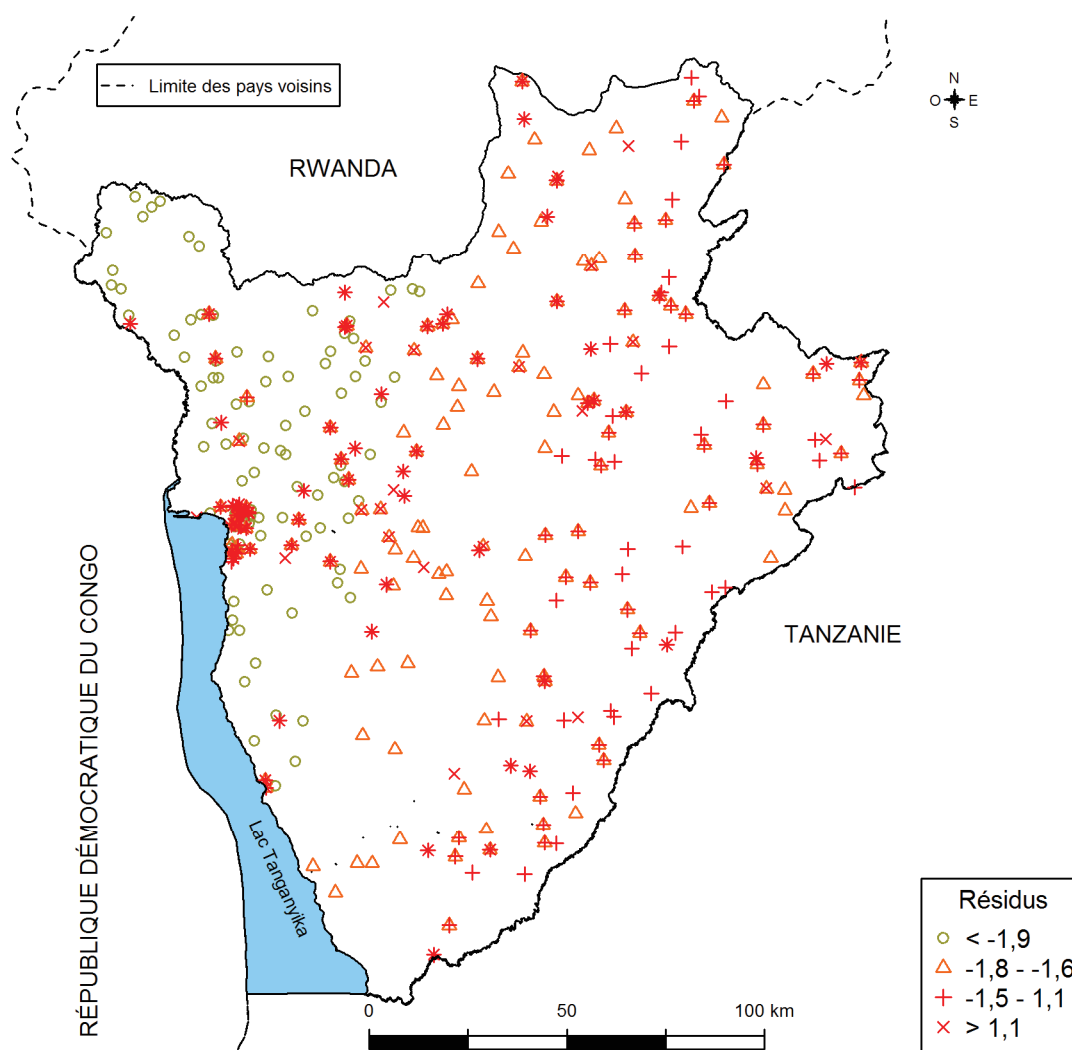
surfaces de tendance. Il revient alors de faire un modèle de régression linéaire des coordonnées géographiques pour l'analyse des résidus. Les résidus de ce modèle sont représentés en fonction des coordonnées géographiques (**Figure 18**).

Figure 18 : Évolution des résidus du modèle de régression linéaire de la prévalence du VIH et des coordonnées géographiques (polynôme d'ordre 1)



Cette figure montre qu'il n'y a pas de tendance spatiale pour la répartition des résidus dans les directions N-S et O-E. Les quartiles de ces résidus sont représentés (**Figure 19**). Nous observons sur cette figure une répartition non homogène des différents quartiles des résidus de la régression d'ordre 1.

Figure 19 : Tendence spatiale des résidus du modèle de régression linéaire de la prévalence du VIH et des coordonnées géographiques (polynôme d'ordre 1)



Une façon complémentaire d'analyser la tendance spatiale est de répéter l'analyse descriptive précédente sur les résidus obtenus avec un polynôme d'ordre supérieur.

1.4.3. Analyse variographique

L'analyse variographique permet de déterminer les poids des voisins lors de l'interpolation spatiale [63]. En tant qu'étape préalable à l'interpolation spatiale, l'analyse variographique (variographie ou analyse structurale) vise à étudier la structure de dépendance spatiale éventuelle à courte échelle des données en

quantifiant cette dépendance spatiale. Cette dépendance caractérise la manière dont les valeurs d'une variable d'intérêt se ressemblent. L'analyse variographique est l'étude du variogramme. Ce dernier, appelé aussi semi-variogramme, est la demi-variance de la différence entre les valeurs prises par une variable aléatoire dans deux sites d'observations s et $s+h$ séparés d'une distance h [64] :

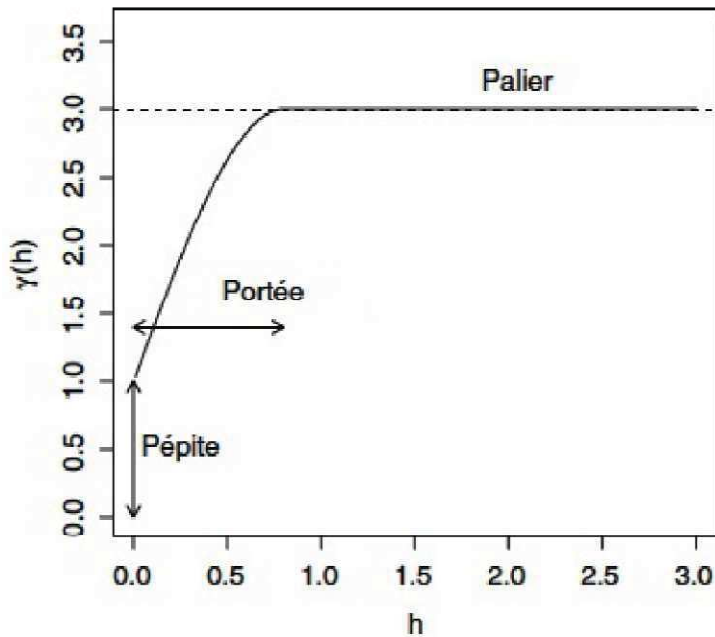
$$\gamma(h) = \frac{1}{2} \text{Var} [Z(s+h) - Z(s)] \quad (1.26)$$

avec $s = s(x, y) \in D$, $h \in \mathbb{R}^2$, x la longitude et y la latitude. Bien que, dans la littérature statistique, le variogramme est anciennement désigné par $2\gamma(h)$, nous conserverons l'appellation de variogramme pour désigné le semi-variogramme.

Le variogramme expérimental ou empirique, c'est-à-dire lié aux données, est une représentation graphique de l'autocorrélation spatiale entre les paires de points où les valeurs de la variable ont été observées.

De ce fait, il caractérise la dissemblance entre les valeurs prises par la variable aléatoire en deux endroits. Son rôle est de contrôler l'interpolation spatiale en évaluant ses paramètres (**Figure 20**) :

- l'effet pépite : valeur de la covariance à une distance nulle ; c'est une discontinuité à l'origine qui peut être due à l'erreur de mesure ;
- la portée : valeur de la distance à partir de laquelle la covariance (la variance) reste constante ; c'est la distance à partir de laquelle les observations ne sont plus dépendantes ; c'est la portée théorique. A la place de la portée théorique, l'on peut aussi s'intéresser à la portée pratique, c'est-à-dire la valeur de la distance pour laquelle la covariance atteint 95% de sa valeur ; les points au-delà de la portée ne sont pas pris en compte lors de l'interpolation ;
- le palier : valeur de la covariance maximale où cette distance est atteinte.

Figure 20 : Paramètres du variogramme expérimental

Source : Gaetan C et al, 2008

L'analyse variographique permettra de sélectionner la distance (portée) à partir de laquelle il n'y a pas de dépendance dans les données ainsi que ses autres paramètres.

L'hypothèse de stationnarité d'ordre 2 suppose que la variance existe et est finie. Dans le cas où cette hypothèse n'est pas vérifiée (cas de variance infinie), elle est substituée par une hypothèse plus faible, l'hypothèse de stationnarité intrinsèque, du champ aléatoire.

Sous cette hypothèse, l'espérance mathématique et la variance du champ aléatoire sont invariantes par translation :

$$\begin{cases} E[Z(s)] = \mu = c^{te}; \forall s \in \mathbb{R}^2 \\ \frac{1}{2} \text{Var}[Z(s+h) - Z(s)] = \frac{1}{2} E\{[Z(s+h) - Z(s)]^2\} = \gamma(h); \forall s, s+h \in \mathbb{R}^2 \end{cases} \quad (1.27)$$

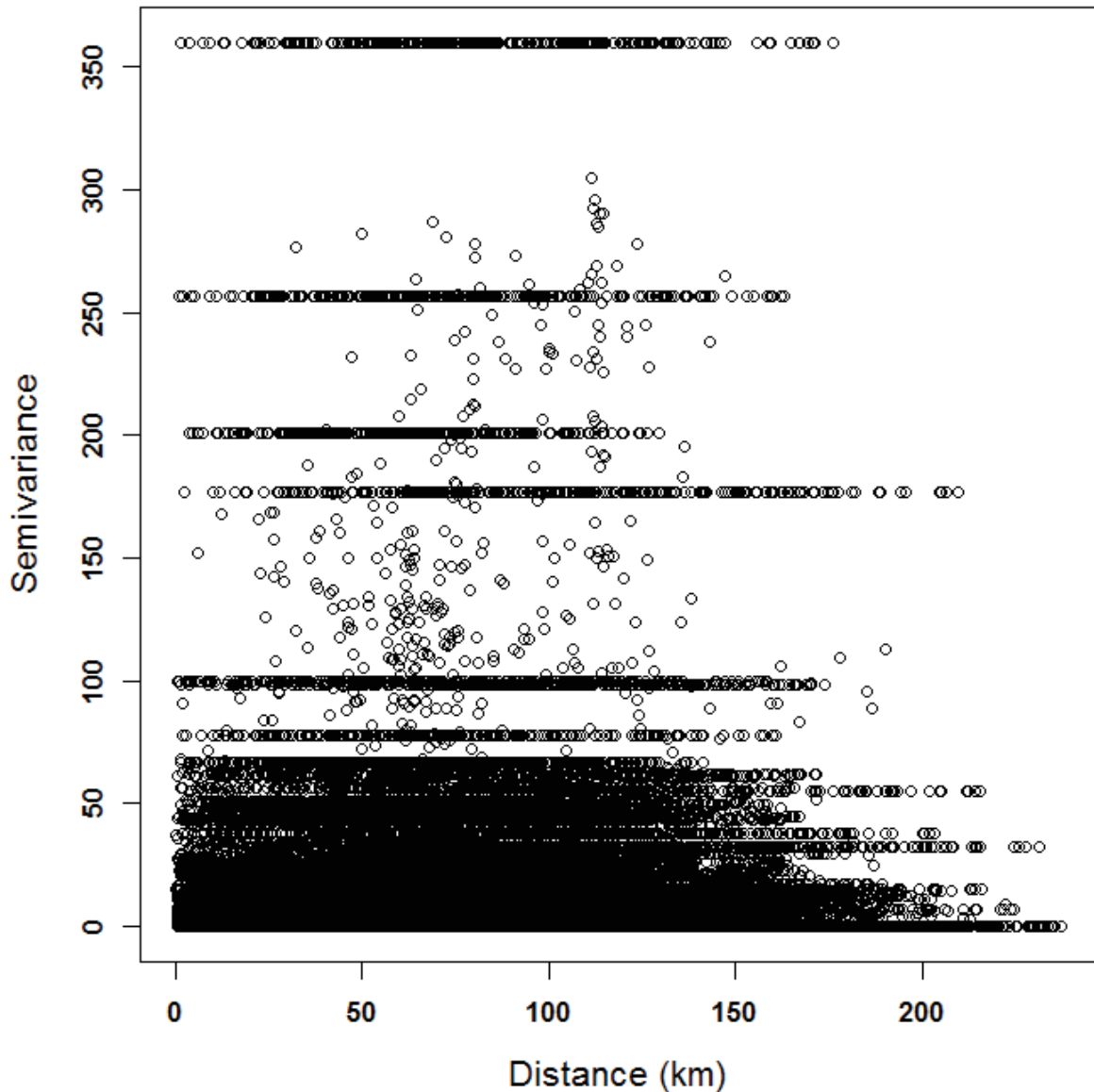
C'est au cours de l'analyse variographique que se fait la vérification de la stationnarité définie plus haut et de l'anisotropie définie comme étant une variation

de la structure spatiale du phénomène d'intérêt selon la direction. En supposant que le variogramme ne varie pas selon les directions (isotropie), la nuée variographique qui désigne le nuage variographique se construit en mettant en abscisse toutes les combinaisons possibles des distances entre les paires formées à partir des n points et en ordonnée les valeurs du variogramme :

$$N_0 = \left\{ \left(\|s_i - s_j\|, \frac{1}{2} [Z(s_i) - Z(s_j)]^2 \right), i, j = 1, \dots, n \text{ et } i \neq j \right\} \quad (1.28)$$

Visuellement, un tel graphique permet de repérer les couples de valeurs aberrantes [11]. Cette aberrance peut avoir pour effet de rendre instable le variogramme expérimental en le faisant varier selon des directions, ce qui traduit la présence d'anisotropies, ou en le faisant dépendre de la position des points, ce qui traduit la non-stationnarité. La nuée variographique, construite à partir des mêmes données de prévalence du VIH au Burundi en 2010, ne montre aucun couple de grappes aberrantes (**Figure 21**).

Figure 21 : Nuée variographique de la prévalence du VIH, Burundi 2010

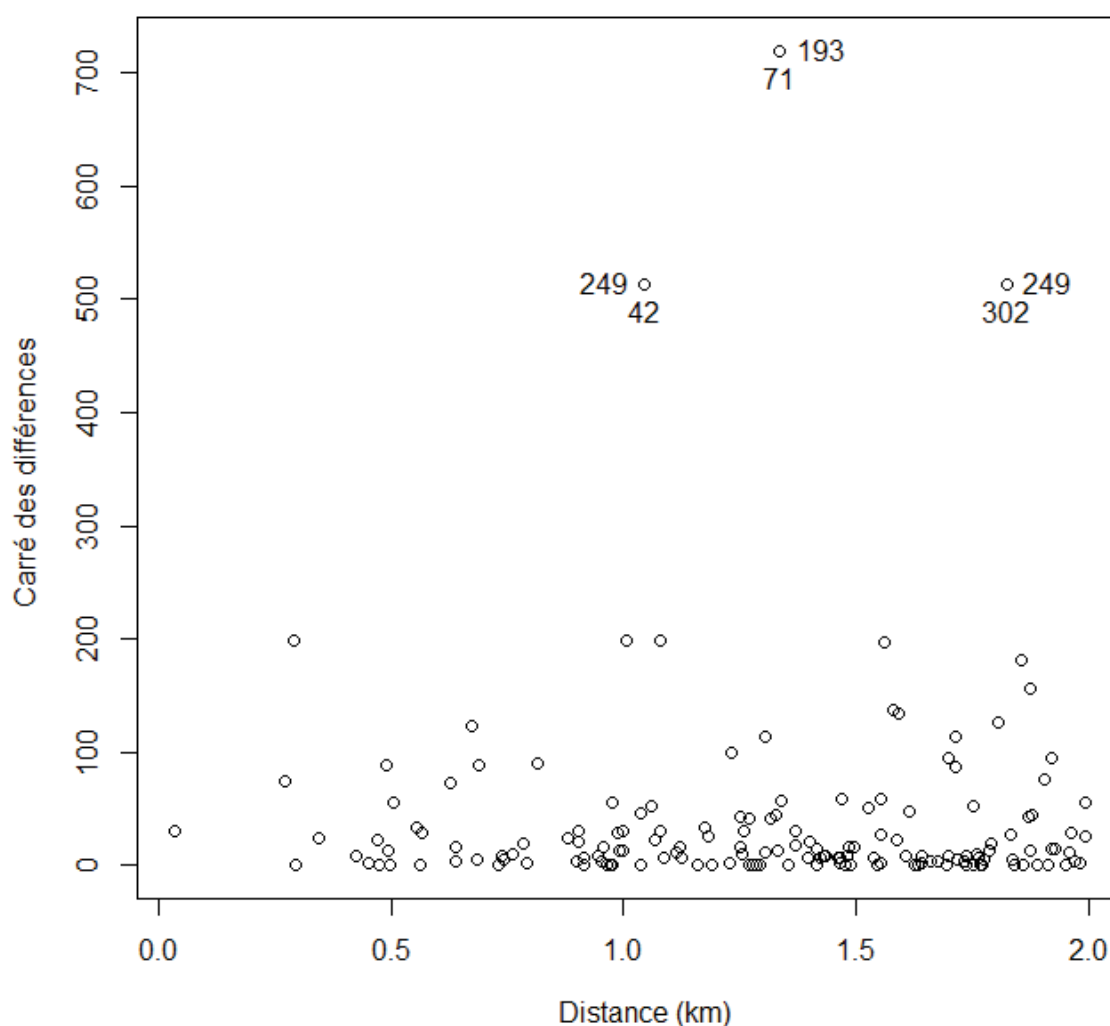


Cette figure ne montre pas de lien linéaire entre les incréments des distances $h_{ij} = |s_i - s_j|$ et la semi-variance qui leur correspond. Après analyse de la nuée variographique, les variogrammes expérimentaux constant et directionnel (angulaire) respectivement sont construits. Cette construction peut se faire de deux manières qui diffèrent par la manière d'estimer l'espérance mathématique du variogramme. Il s'agit de la méthode classique, dite méthode de Mathéron ou méthode des moments, qui estime l'espérance par la moyenne empirique (par classe de distance ou lag) et la méthode robuste contre les valeurs aberrantes, dite méthode de Cressie, qui consiste à élever le carré de l'incrément $Z(s+h) - Z(s)$ à la puissance

0.25 afin de corriger l'asymétrie (absence de normalité) qui apparaît dans la distribution du carré de l'incrément.

Il n'y a pas de critère statistique permettant de se rendre compte des couples aberrants. Leur détection n'est que visuelle, ce qui paraît subjectif. À partir des mêmes données de prévalence du VIH au Burundi en 2010, un exemple de nuée variographique avec des couples aberrants peut s'obtenir en se limitant, à titre illustratif, aux couples de grappes dont la distance ne dépasse pas 2 km (**Figure 22**).

Figure 22 : Nuée variographique de la prévalence du VIH avec couples aberrants, Burundi 2010



Contrairement à la nuée variographique représentée par la figure 16, cette nuée variographique met en ordonnées les carrés des différences entre les valeurs prises par la variable à chaque paire de points et montre que les couples de grappes

(42,249) de la province Ngozi, (71,193) de la province Ruyigi et (302,249) de la province Ngozi sont aberrants. En effet, les grappes n° 42 et 249 sont géographiquement proches (1,0 km) et ont des prévalences du VIH très éloignées (0 % et 22,7 % respectivement). Il en est de même pour les grappes n° 71 et 193 (1,3 km) de prévalences respectives de 26,8 % et 0 %. Les grappes n° 249 et 302 sont géographiquement éloignées (1,8 km) mais ont des prévalences nulles. De même, les grappes n° 302 et 193 sont très éloignées (99,4 km) et ont les mêmes prévalences. Il en est de même pour les grappes n° 249 et 193 distantes de 101,2 km et qui ont des prévalences nulles. Tout cela contredit la loi de Tobler selon laquelle les prévalences du VIH observées dans des sites proches devraient plus se ressembler que les prévalences observées dans des sites éloignés.

La relation (1.26) est valable pour une paire de points. En généralisant cette relation au cas de toutes les paires de points s_i puis en posant $N(h)$ l'ensemble des paires de points séparés d'une distance $h \in \mathbb{R}^2$, le variogramme classique sans dérive est estimé par :

$$\hat{\gamma}_n(h) = \frac{1}{2|N(h)|} \sum_{(s_i, s_j) \in N(h)} [Z(s_i) - Z(s_j)]^2 \quad (1.29)$$

où $|N(h)|$ est le cardinal de $N(h)$.

La variable d'intérêt est supposée continue. Afin de s'assurer de cette continuité dans toutes les directions, il convient de calculer le variogramme directionnel. Le variogramme selon la direction θ est estimé par :

$$\hat{\gamma}_n(h, \theta) = \frac{1}{2|N(h, \theta)|} \sum_{(s_i, s_j) \in N(h, \theta)} [Z(s_i) - Z(s_j)]^2 \quad (1.30)$$

où θ est la valeur de l'angle selon lequel le variogramme est estimé. Ce variogramme n'est valable que lorsque les incréments $Z(s_i) - Z(s_j)$ sont

stationnaires de second ordre. Les angles les plus utilisés sont 0° , 45° , 60° , 90° et 135° avec la possibilité de les personnaliser.

Sous l'hypothèse de stationnarité, le variogramme exprimé par la relation (1.26) est la moitié de l'espérance mathématique du carré de l'incrément $Z(s+h) - Z(s)$. Sous l'hypothèse de normalité, le carré de l'incrément $Z(s+h) - Z(s)$ suit une loi du Chi-deux à un degré de liberté. Du fait que cette loi de distribution est asymétrique, Cressie et Hawkins (1980) ont proposé de transformer l'incrément $Z(s+h) - Z(s)$ en l'élevant au carré, puis en élevant le résultat obtenu à une puissance λ . C'est la valeur $\lambda = 0,25$ qui conduisait à une variable transformée $Y(s) = \left\{ [Z(s+h) - Z(s)]^2 \right\}^{0,25}$ proche de la normalité. La variable transformée s'écrivait alors $Y(s) = |Z(s+h) - Z(s)|^{\frac{1}{2}}$. L'élévation membre à membre à la quatrième puissance conduit à une expression dont le second membre est le carré de l'incrément $Z(s+h) - Z(s)$.

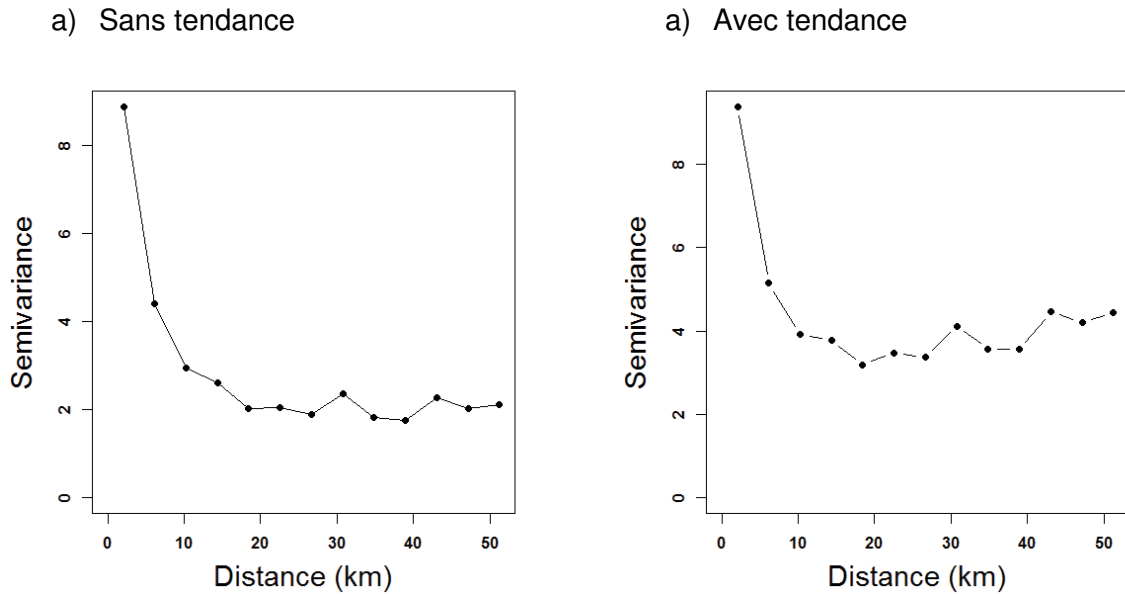
Après quelques transformations, l'espérance mathématique de $\frac{\bar{Y}^4(s)}{2\gamma(h)}$ était donnée par $0,457 + \frac{0,494}{n} + \frac{0,045}{n^2}$ où n est le nombre de paires de points, ce qui a conduit au variogramme robuste [13,65] :

$$\bar{\gamma}_n(h) = \left\{ 0,914 + \frac{0,988}{|N(h)|} \right\}^{-1} \left\{ \frac{1}{|N(h)|} \sum_{(s_i, s_j) \in N(h)} |Z(s_i) - Z(s_j)|^{\frac{1}{2}} \right\}^4 \quad (1.31)$$

Le dénominateur de l'estimateur robuste de Cressie sert à corriger le biais sous la normalité. Son estimateur peut cependant être parasité par la présence d'une valeur aberrante singulière [63]. La robustesse du variogramme a été également étudiée grâce à l'utilisation des quantiles des différences absolues et à l'estimation de ses paramètres et son ajustement par des méthodes bootstrap [63,66].

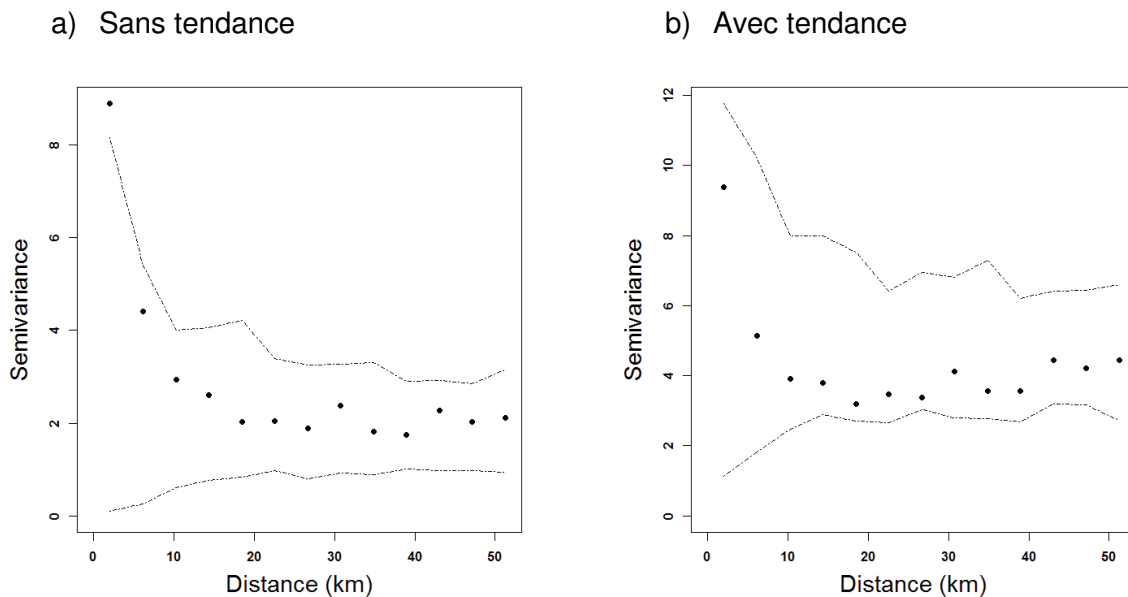
Dans le cas où le champ aléatoire n'est pas stationnaire, c'est-à-dire que son espérance mathématique n'est pas constante sur tout le domaine d'étude, alors il y a tendance. Dans ce cas, il est décomposé en une somme d'une partie déterministe (son espérance) et d'une partie aléatoire. Les variations du variogramme à courte échelle sont alors modélisées par la partie aléatoire qui vaut la différence entre le champ aléatoire et son espérance mathématique. Les variations à grande échelle, quant à elles, peuvent être modélisées par une fonction des coordonnées géographiques (longitude/latitude) et/ou éventuellement d'autres variables quantitatives présentes dans la géodatabase. Lors de l'estimation d'un variogramme, il est donc possible de le considérer avec une tendance ou sans tendance (constante). Les variogrammes sans et avec tendance construits à partir des mêmes données que précédemment à l'aide de la méthode robuste présentent des distributions analogues (**Figure 23**). La distance moyenne est calculée pour chaque classe de distances h (pas ou distance de séparation) qui doit contenir au moins 30 points pour garantir un échantillon de grande taille et la semi-variance est calculée pour cette distance. Le variogramme est construit en mettant en abscisses cette distance et en ordonnées la semi-variance correspondante. À côté du variogramme omnidirectionnel ou du variogramme directionnel, il existe des variogrammes qui tiennent compte de la tendance. Cette tendance est soit définie par les coordonnées géographiques (variogramme avec tendance), soit par une autre variable présente dans les données. Dans tous les cas, comme le variogramme est un outil exploratoire, il permet de comparer les variogrammes sans et avec tendance, les variogrammes sans ou avec direction. Leur inconvénient est qu'il n'y a pas de test statistique connu permettant de choisir le meilleur variogramme. Si la semi-variance est faible pour des faibles distances h , alors cela renforce la présence de l'autocorrélation spatiale vue précédemment ou de dépendance spatiale à courte échelle dans les données. Il permet de détecter les anisotropies éventuelles dans les données en l'évaluant dans plusieurs directions [22].

Figure 23 : Variogramme expérimental de la prévalence du VIH, Burundi 2010



Afin de juger de la significativité (à l'aide d'un graphique) du variogramme expérimental ou de s'assurer de l'existence d'une structure spatiale dans les données, des enveloppes de confiance à 100 % de ces variogrammes sont construits (**Figure 24**).

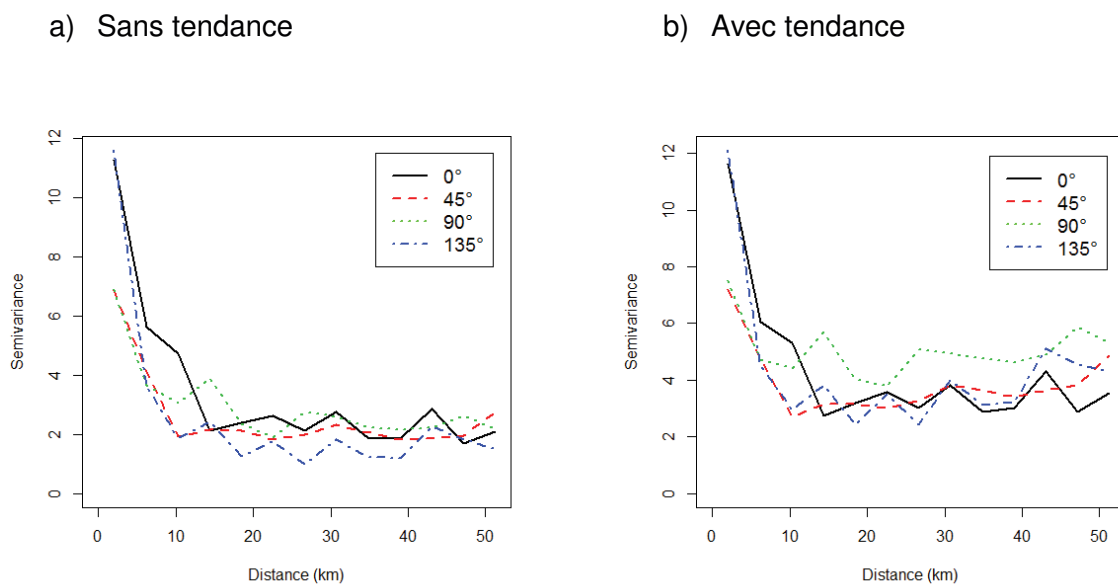
Figure 24 : Enveloppes du variogramme expérimental de la prévalence du VIH, Burundi 2010



Ces enveloppes, situées de part et d'autre des distributions empiriques de la variable d'intérêt, permettent de tester (test de permutation de Monte-Carlo) l'hypothèse nulle d'absence d'une structure spatiale. Comme tous les points sont à l'intérieur des enveloppes, alors pouvons conclure à l'existence d'une structure spatiale dans les données. De plus, afin de s'assurer de l'absence des anisotropies, les variogrammes expérimentaux sont construits selon les directions (**Figure 25**).

Cette figure montre que le variogramme est, à peu près, stationnaire. Cela veut dire que la continuité spatiale de la prévalence du VIH est, à peu près, la même dans toutes les directions, ce qui traduit l'absence des anisotropies.

Figure 25 : Variogrammes directionnels de la prévalence du VIH, Burundi 2010



Une fois que la distance maximale retenue (portée) à partir de laquelle la variance se stabilise, le variogramme avec ou sans tendance, le variogramme omnidirectionnel ou directionnel sont choisis, il convient d'ajuster le variogramme expérimental par un modèle de variogramme théorique soigneusement choisi. Il existe plusieurs modèles théoriques tels que le modèle linéaire, le modèle linéaire exponentiel, le modèle linéaire sphérique et le modèle linéaire de Matérn. L'ajustement se fait par la méthode des moindres carrés ordinaires (MCO), la méthode des moindres carrés généralisés (MCG), la méthode des moindres carrés pondérés (MCP) et la méthode du maximum de vraisemblance (MMV). Le modèle retenu est celui qui minimise la

somme des carrés des résidus ou celui ayant un petit effet de pépite. Il est également d'usage de subdiviser l'échantillon en un échantillon-test et un échantillon d'apprentissage afin d'estimer l'erreur quadratique moyenne comme erreur de prédiction par une méthode de validation croisée. Une méthode avec une erreur de prédiction plus petite est retenue. Cet ajustement permet d'estimer les paramètres du modèle (portée, pépite, palier) qui seront utilisés lors de l'interpolation spatiale. Le choix du modèle dépend de la manière dont les observations doivent être pondérées. Considérons le variogramme $\gamma(h, \lambda)$ où λ est le paramètre qui maximise la somme des carrés des erreurs et Σ la matrice de variances du variogramme estimée à chaque distance h .

Cressie a proposé les valeurs du paramètre qui maximisent la somme des carrés des erreurs pour la MCO, la MCG et la MCP respectivement [67] :

$$\hat{\lambda}_{MCO} = \arg \min \sum_{j=1}^k \{ \hat{\gamma}(h_j) - \gamma(h_j); \lambda \}^2 \quad (1.32)$$

$$\hat{\lambda}_{MCG} = \arg \min \left[\hat{\gamma} - \gamma(\lambda) \right]^t \Sigma^{-1} \left[\hat{\gamma} - \gamma(\lambda) \right] \quad (1.33)$$

$$\hat{\lambda}_{MCP} = \arg \min \sum_{j=1}^k \left\{ \text{Var} \left[\hat{\gamma}(h_j) \right] \right\}^{-1} \{ \hat{\gamma}(h_j) - \gamma(h_j); \lambda \}^2 \quad (1.34)$$

La MCO attribue le même poids à toutes les classes de distances contrairement à la MCP qui met plus de poids aux observations spatialement proches et moins de poids aux observations éloignées. De ce fait, la MCO est moins préférée par rapport à la MCP. La MMV est plus préférée par rapport à la MCP. De façon générale, la minimisation par la MCP peut s'écrire :

$$\hat{\lambda}_{MCP} = \arg \min \sum_{j=1}^k W(h_j) \{ \hat{\gamma}(h_j) - \gamma(h_j); \lambda \}^2 \quad (1.35)$$

où $W(h_j)$ est une fonction de pondération.

1.4.4. Méthodes d'interpolation spatiale

Les techniques d'interpolation spatiale se classent en deux groupes principaux: les méthodes déterministes et les méthodes stochastiques.

Les méthodes déterministes sont des méthodes qui ne font pas intervenir les notions probabilistes dans les prédictions spatiales. Autrement dit, elles considèrent que la variable d'intérêt n'est pas aléatoire et sont simplement basées sur des fonctions mathématiques permettant de pondérer les observations géographiquement voisines ou permettant de lisser les surfaces de tendance. Ces méthodes se subdivisent, à leur tour, en trois groupes : les méthodes barycentriques [12,68,69] dont la méthode de l'inverse des distances pondérées; les méthodes d'interpolation par partitionnement de l'espace [12,68,70] dont les polygones de Thiessen, la triangulation de Delaunay, le diagramme de Voronoï et la mosaïque ou tessellation de Dirichlet et les splines [13,69,71], c'est-à-dire une famille des fonctions régulières de courbure minimale pour lisser une surface. Il peut s'agir des splines d'interpolation (passant par les points d'observation) ou des splines de lissage des discontinuités [72]. Ces méthodes ne fournissent pas de précision autour des estimations et ne tiennent pas compte de la structure spatiale des données en termes de dépendance spatiale contrairement aux méthodes stochastiques.

Basées sur le concept de hasard, les méthodes stochastiques, dites aussi probabilistes, considèrent que la variable d'intérêt est une variable aléatoire et que les valeurs prises par cette dernière sont ses réalisations. Les valeurs de cette variable ne sont connues qu'aux points d'observation, ce qui fait que la connaissance des valeurs de cette variable en des points non échantillonnés se fait par des techniques d'interpolation spatiale. Il existe plusieurs techniques d'interpolation spatiale stochastique telles que le krigeage universel qui considère que l'espérance de la variable est une fonction des coordonnées spatiales, le krigeage avec dérive externe, c'est-à-dire où la tendance est une autre variable présente dans les données [73] incluant des variables auxiliaires, le cokrigeage [74] faisant intervenir des structures déterministes pour les espérances et des fonctions aléatoires, et le krigeage ordinaire qui considère que l'espérance de la variable est une constante μ inconnue. Dans ce dernier cas, la valeur de la variable d'intérêt en

un point quelconque s_i du domaine d'étude peut s'écrire comme $Z(s_i) = \mu + \varepsilon(s_i)$ où $\varepsilon(s_i)$ est une partie aléatoire.

1.4.5. Krigeage ordinaire

Lorsque tous les points du domaine d'étude n'ont pas été mesurés, ce qui est souvent le cas, la connaissance des valeurs de la variable d'intérêt en des points non mesurés se fait par une interpolation spatiale communément appelée krigeage.

Lors du krigeage, la valeur prédite $\hat{Z}(s_0)$ au point donné s_0 s'écrit comme une combinaison linéaire des valeurs aux n points observés s_i , ce qui se traduit par [75]:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad (1.36)$$

où les λ_i sont les poids du krigeage. Le krigeage ordinaire est un interpolateur exact et de variance minimale [13]. Grâce à la stationnarité et à l'absence de biais pour l'interpolateur, le calcul de l'espérance mathématique de l'erreur de prédiction $\hat{Z}(s_0) - Z(s_0)$ conduit à la contrainte $\sum_{i=1}^n \lambda_i = 1$ [76]. La variance vaut $E\left\{\left[\hat{Z}(s_0) - Z(s_0)\right]^2\right\}$. Les valeurs de λ_i qui rendent nul le biais d'estimation et qui minimisent la variance de prédiction s'obtiennent en résolvant le système d'équations ci-après :

$$\begin{cases} E\left[\hat{Z}(s_0) - Z(s_0)\right] = 0 \\ \text{Var}\left[\hat{Z}(s_0) - Z(s_0)\right] = \min\left\{\text{Var}\left[\sum_{i=1}^n \lambda_i Z(s_i) - Z(s_0)\right], \lambda \in \mathbb{R}^n\right\} \end{cases} \quad (1.37)$$

Le lagrangien du système est :

$$L(\lambda_i, K) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}[Z(s_i), Z(s_j)] + \text{Var}[Z(s_0)] - 2 \sum_{i=1}^n \lambda_i \text{Cov}[Z(s_i), Z(s_0)] + \text{Var}(Z(s_0)) + 2K \left(\sum_{i=1}^n \lambda_i - 1 \right) \quad (1.38)$$

où K est le mutiplicateur de Lagrange. En le dérivant respectivement par rapport à λ_i et à K puis en égalant les résultats à 0, il vient :

$$\begin{cases} \sum_{j=1}^n \lambda_j \text{Cov}[Z(s_i), Z(s_j)] + K = \text{Cov}[Z(s_i), Z(s_0)] \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (1.39)$$

ce qui s'écrit à l'aide de la matrice C de covariances des valeurs de la variable aux points de mesure et de la matrice D de covariances des valeurs de la variable aux points de prédiction. Ce système peut s'écrire $C\lambda = D$ et les poids de krigeage s'obtiennent comme suit [77]:

$$\lambda = C^{-1}D \quad (1.40)$$

Le domaine d'étude est transformé en une grille d'interpolation à partir des coordonnées géographiques des points observés. Plus le nombre de points de prédiction est important (grille très fine), plus la carte interpolée devient lisse. Bien que cette méthode lisse moins bien aux frontières à cause des effets de bord, elle permet de mettre en évidence des variations du phénomène étudié cachées par la simple cartographie statistique du phénomène par zone du domaine d'étude.

1.4.6. Interpolation spatiale par la méthode des noyaux

Contrairement au krigeage ordinaire, l'interpolation par la méthode des noyaux s'intéresse à la distribution spatiale du nombre de cas apparaissant dans une portion bien déterminée du domaine d'étude et dont l'aire de surface est connue. Ce nombre

est considéré comme une réalisation d'une variable aléatoire. Toutes les deux sont donc des techniques d'interpolation stochastique et sont sensibles aux effets de bord. Connaissant le nombre de cas d'une pathologie donnée à chaque point d'observation, il est possible de calculer le nombre moyen de cas par unité de surface dans une fenêtre mobile ds englobant le point s et d'aire A .

Cela définit l'intensité de cette pathologie au point s [78] :

$$\lambda(s) = \lim_{ds \rightarrow 0} \left\{ \frac{E[Z^*(ds)]}{A} \right\} \quad (1.41)$$

Z^* étant ici une variable de comptage. Au même titre que les densités de probabilité, les noyaux sont des fonctions intégrables K positives, symétriques et qui sont telles que $\int K(u) du = 1$. L'estimation de la valeur de l'intensité à un point donné du domaine d'étude se fait à l'aide d'une fenêtre tridimensionnelle mobile représentant le noyau. Au cours de cette estimation, le noyau pondère les cas en mettant plus de poids aux cas qui sont proches du point auquel l'intensité se veut être estimée et moins de poids aux cas éloignés à travers une sphère d'influence [79]. Il permet de lisser les discontinuités spatiales.

Cela conduit à une surface d'intensités produite par le lissage [79,80]:

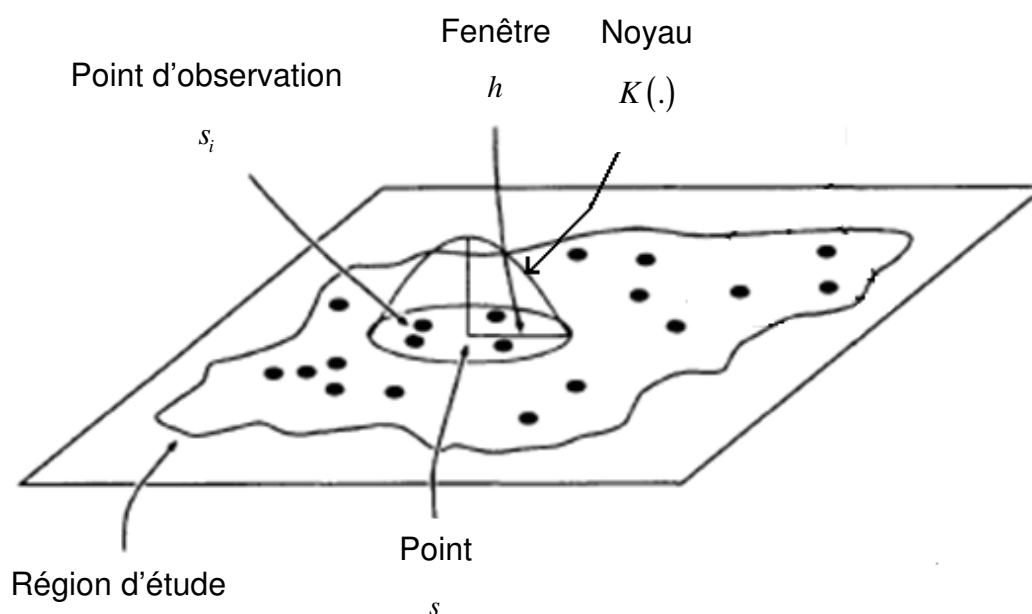
$$\hat{\lambda}_h(s) = \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{s-s_i}{h}\right) \quad (1.42)$$

où h la fenêtre ou rayon d'influence (paramètre de lissage), s_i le point observé, s un point de la grille d'interpolation, et K la forme fonctionnelle du noyau, le point d'observation s_i étant caractérisé par ses coordonnées géographiques. Il existe plusieurs choix de noyaux permettant de faire le lissage de la surface [81,82].

Les formes fonctionnelles des noyaux sont principalement le noyau gaussien, le noyau quartique, le noyau d'Epanechnikov, le noyau triangulaire, le noyau

exponentiel négatif et le noyau uniforme. Du choix de la forme fonctionnelle du noyau et du rayon de la fenêtre de lissage dépend la surface estimée, la fenêtre pouvant être fixe ou adaptative. La fenêtre adaptative permet que le rayon de la fenêtre s'adapte à des endroits avec une forte concentration de cas en devenant plus petit par rapport à celui de la fenêtre dans des endroits avec une faible concentration de cas. Le rayon du cercle varie jusqu'à ce qu'un nombre minimum de points tombant dans le cercle soit atteint. Dans le cas gaussien, une fenêtre est définie par la déviation standard de la distribution normale et dans les autres cas par le rayon du cercle de balayage. Le noyau le plus utilisé est le noyau gaussien tridimensionnel représenté à la **Figure 26**.

Figure 26 : Estimation par la méthode des noyaux



Source: Adapté de Gatrell A.C. et al, 2015

Toutes ces techniques d'analyse spatiale permettent de mettre en évidence les zones à plus haut risque, ce qui permet d'aider les décideurs et chercheurs de santé publique à inventorier les besoins permettant de mettre à jour les programmes d'intervention sanitaire et de générer les hypothèses sur l'étiologie de la pathologie [46]. En vue d'étayer ces hypothèses, l'étape suivante est la recherche du lien entre la variable d'intérêt et les caractéristiques socioéconomiques, démographiques,

comportementales et environnementales des personnes enquêtées à travers des modèles de régression [83].

1.5. Modélisation spatiale

L'épidémiologie spatiale a longtemps été utilisée pour produire des cartes de l'incidence et de prévalence des maladies au niveau des pays dans le but de décrire les maladies et les facteurs associés en santé publique et en épidémiologie [84,85]. Actuellement, les données issues des enquêtes de population et de santé comportent souvent des coordonnées géographiques. Ces données permettent alors de produire des cartes de risque spatial, lesquelles cartes constituent un outil de visualisation par excellence et de transmission de l'information médicale.

La recherche des agrégats spatiaux et l'analyse de l'autocorrélation spatiale visent à évaluer l'hétérogénéité spatiale de la distribution d'un problème de santé publique dans le temps et/ou dans l'espace. Cette recherche épidémiologie spatiale, de nature descriptive, est complétée par l'analyse de l'association entre ce problème de santé et les facteurs de risque démographiques, environnementaux, comportementaux, socio-économiques, génétiques et infectieux potentiels [46]. En d'autres termes, l'évaluation de la distribution spatiale d'une maladie d'intérêt, la détection des structures particulières (agrégats, poches de prévalence) dans cette distribution et l'étude des facteurs de risque qui déterminent ces structures sont autant de questions auxquelles l'épidémiologie spatiale permet de répondre en vue d'aider les décideurs de santé publique à diagnostiquer les facteurs de risque spatial potentiels et à prendre des décisions visant l'orientation des interventions en vue de promouvoir la santé des populations et les chercheurs dans ce domaine à orienter leurs projets de recherche [19].

1.5.1. Définition d'un modèle spatial

Le modèle linéaire spatial s'écrit de la même manière que le modèle linéaire classique [86]:

$$Z = X\beta + \varepsilon \quad (1.43)$$

où $Z = (Z_1, \dots, Z_n)'$ est le vecteur des réalisations de la variable réponse Z respectivement aux n points choisis dans le domaine d'étude, X la matrice de design constituée par une colonne des 1 et les p variables explicatives X^1, \dots, X^p , β le vecteur des paramètres et ε la partie stochastique du modèle. Un tel modèle classique suppose que les valeurs Z_1, \dots, Z_n sont indépendantes, ce qui n'est pas le cas en statistique spatiale. En plus des hypothèses d'indépendance et d'homoskedasticité des erreurs, l'hypothèse de normalité des erreurs considère que $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \sim N(0, \Sigma)$ où $\Sigma = \text{Var}(\varepsilon) = \sigma^2 I_n$ est une matrice diagonale. En statistique spatiale, cette matrice Σ de variances-covariances des erreurs n'est pas diagonale et représente la matrice des interactions spatiales des valeurs de Z ou de ε . La prise en compte de cette dépendance spatiale ou de l'autocorrélation spatiale dans les modèles de régression spatiaux se fait en introduisant dans la relation (1.43) une variable spatialement décalée WZ de la variable réponse et/ou une erreur spatialement décalée $V\varepsilon$.

Cela s'écrit :

$$Z = \rho WZ + X\beta + \varepsilon \text{ avec } \varepsilon = \lambda V\varepsilon + u \quad (1.44)$$

où W et V sont les matrices de voisinage définies respectivement pour la variable réponse et l'erreur du modèle de régression et, ρ et λ les coefficients des modèles autorégressifs et $u = (u_1, \dots, u_n)' \sim N(0, \sigma^2 I_n)$ l'erreur. Trois cas de figure peuvent se présenter : une indépendance spatiale ($\rho = \lambda = 0$), une dépendance spatiale captée par la variable réponse ($\lambda = 0$) et une dépendance spatiale captée par les erreurs

($\rho = 0$), ce qui se résume par les modèles pour chaque unité spatiale i du domaine d'étude :

$$Z_i = (X\beta)_i + u_i \quad (1.45)$$

$$Z_i = \rho(WZ)_i + (X\beta)_i + u_i \quad (1.46)$$

$$Z_i = (X\beta)_i + \lambda(V\varepsilon)_i + u_i \quad (1.47)$$

Les modèles spatiaux définis par les relations (1.46) et (1.47) montrent que, pour chaque unité spatiale i , les fluctuations de la variable réponse Z dépendent non seulement de celles des variables explicatives X_1, \dots, X_p mais également des valeurs de la variable réponse (ou de l'erreur) dans les unités avoisinantes. Comme conséquence, une fluctuation de la variable réponse dans une unité spatiale donnée entraîne des fluctuations de la même variable dans les unités spatiales avoisinantes. Étant donné que l'utilisation de la méthode des moindres carrés ordinaires biaise l'estimation des paramètres dans les modèles de régression spatiale et que les tests usuels de Student et de Fisher permettant de faire l'inférence sur les paramètres ne sont plus valides, le vecteur des paramètres β des modèles spatiaux définis par les relations (1.46) et (1.47) est estimé par la méthode du maximum de vraisemblance [87,88]. Le test du multiplicateur de Lagrange permet de tester la nullité simultanée des paramètres et l'hypothèse nulle selon laquelle l'autocorrélation spatiale non expliquée (provenant de la variable réponse ou de l'erreur) est nulle. Ces modèles permettent également d'estimer les indices de l'autocorrélation spatiale globale calculées respectivement sur la variable réponse et les résidus (estimations) du modèle de régression non spatiale en tenant compte de la localisation géographique des unités spatiales d'analyse. La nullité de ces indices est testée soit à l'aide du test t de Student basée sur la normalité asymptotique, soit à l'aide du test du rapport de vraisemblance. L'ajustement de chaque modèle est évalué à l'aide du critère de l'information d'Akaike (AIC) en comparant cet AIC à celui qui serait obtenu du modèle ne faisant pas intervenir la matrice de voisinage, ce qui permet d'évaluer l'intérêt de l'utilisation d'un modèle spatial par rapport à un modèle non spatial. L'inconvénient de ces méthodes est que les résultats sont fortement influencés par le choix de la matrice W .

Les associations entre les problèmes de santé et les facteurs de risque sont souvent évaluées à travers des modèles spatiaux généralisés.

1.5.2. Définition d'un modèle spatial généralisé

Les données spatialisées présentent un caractère particulier du fait que, non seulement elles sont indexées des coordonnées géographiques mais également elles ne sont pas ordonnées et sont souvent corrélées surtout pour de faibles distances entre les points d'observation. Faire fi à la spatialité des données dans un modèle de régression spatiale revient à s'intéresser aux indicateurs globaux qui ne représentent pas les variations locales éventuelles [89]. Les modèles spatiaux tiennent compte de la localisation et de la structure de dépendance spatiale des données. Les variables présentes dans la base de données sont souvent utilisées à leur état brut. Cependant, des fonctions mathématiques peuvent les lier à la variable réponse. Un modèle spatial, dit généralisé, construit à partir de la variable réponse, des variables explicatives et des fonctions de lien entre ces dernières et la variable réponse tout en tenant compte de leur caractère spatialisé est un processus analytique par lequel les données spatialisées sont analysées et visualisées afin d'en extraire une information utile à la prise de décision. Parmi les modèles spatiaux généralisés figurent les modèles binomial et de Poisson qui sont des modèles décrits par une fonction exponentielle. Sous les critères de convergence, le modèle spatial binomial converge vers le modèle spatial de Poisson.

1.5.3. Les modèles de régression spatiale binomial et de Poisson

1.5.3.1 Approche bayésienne

L'approche bayésienne consiste à estimer un paramètre θ en partant de l'information *a priori* sur les valeurs possibles de ce paramètre, puis en mettant à jour cette information à travers la vraisemblance (lors de l'expérience) afin d'obtenir la distribution *a posteriori* de θ . Elle permet d'obtenir une meilleure précision que la statistique classique.

Partant de la formule de Bayès, la probabilité conditionnelle de θ sachant les observations z , appelée probabilité *a posteriori* de θ , est donnée par [90] :

$$p(\theta|z) = \frac{p(z|\theta)p(\theta)}{p(z)} \propto L(\theta; z)p(\theta) \quad (1.48)$$

avec

$$p(z) = \int p(z|\theta)p(\theta)d\theta \quad (1.49)$$

où $L(\theta; z) = p(z|\theta)$ est la fonction de vraisemblance. Pour une loi binomiale par exemple, la vraisemblance s'écrit :

$$L(\theta; z) = p(z|\theta) \propto \theta^k (1-\theta)^{n-k} \quad (1.50)$$

où n est le nombre d'essais. La relation (1.48) montre que la probabilité *a posteriori* de θ est proportionnelle au produit de la vraisemblance et de la probabilité *a priori* de θ . Un *a priori* sur θ est dit non informatif si son effet sur la distribution *a posteriori* de θ est minimale et conjugué si $p(\theta)$ et $p(\theta|z)$ sont de même loi.

1.5.3.2 Estimation et inférence bayésiennes

En utilisant des simulations de Monte-Carlo par chaînes de Markov (MCMC : Markov Chain Monte Carlo) intégrant la distribution *a priori* du paramètre θ , sa distribution *a posteriori* permet de calculer la moyenne, la médiane ou le mode *a posteriori* de θ (estimation ponctuelle du paramètre) après vérification de la convergence. Cette convergence sera atteinte en fonction du nombre de chaînes, du nombre d'itérations, de la loi *a priori* et des valeurs initiales du paramètre. L'inférence bayésienne est faite alors en construisant un intervalle de crédibilité à $(1-\alpha)$ % et en vérifiant si cet intervalle contient ou non la valeur du paramètre sous l'hypothèse nulle.

Dans les modèles de régression logistique, la vraisemblance est donnée par :

$$L(\pi; z) = \prod_{i=1}^n \pi(x_i)^{z_i} [1 - \pi(x_i)]^{1-z_i} \quad (1.51)$$

où l'hyperparamètre $\theta = \pi = p(z_i = 1)$ est la probabilité qu'un individu i subisse l'événement d'intérêt qui s'écrit comme une fonction des covariables et x_i la valeur de la variable indépendante x . Si aucune information n'est connue sur la distribution des paramètres du modèle, un a priori non-informatif peut être mis sur les paramètres. La distribution *a posteriori* des paramètres peut être représentée graphiquement sous forme d'histogrammes. Une revue complète de la méthode bayésienne peut être trouvée ailleurs [90–92].

1.5.3.3 Cas concret

Lorsque les données sont recueillies à l'échelle de l'individu considéré comme unité d'analyse et que la variable réponse est binaire, un modèle de régression spatiale binomiale avec le logit comme fonction de lien convient pour analyser les relations avec les covariables et la réponse géospatialisées [93,94]. Considérons Z_{is} les valeurs de la variable binaire d'intérêt Z pour un individu i de l'entité géographique s avec une probabilité inconnue π_{is} que l'événement d'intérêt se réalise et, $X = (X_1, \dots, X_p)$ la matrice des variables explicatives.

La variable Z suit une loi de Bernoulli de paramètre π_{is} :

$$Z_{is} \sim \text{Bernoulli}(\pi_{is}) \quad (1.52)$$

Les valeurs de la variable Z sont dans $\{0,1\}$ et les probabilités $\pi_{is} \in [0,1]$ sont modélisées à l'aide de la régression logistique spatiale.

Si, par contre, les données sont groupées dans les entités géographiques s considérées comme des unités spatiales d'analyse, une analyse écologique (modèle spatial binomial ou de Poisson) permet d'évaluer la relation entre la variable réponse transformée (fréquences ou proportions) et les covariables. Dans le cas d'un modèle spatial binomial ou de Poisson, la variable Z qui désigne le nombre d'individus ayant subi l'événement d'intérêt suit une loi binomiale de paramètres n et π_s représentant respectivement le nombre d'individus se trouvant dans l'entité s et la proportion d'individus ayant subi l'événement d'intérêt dans la même entité ou une loi de Poisson de paramètre $\lambda_s = n\pi_s$:

$$Z_s \sim Bin(n, \pi_s) \quad (1.53)$$

$$Z_s \sim Poisson(\lambda_s) \quad (1.54)$$

Les valeurs de la variable Z sont des nombres naturels et les probabilités π_s sont modélisées à l'aide de la régression spatiale binomiale ou de Poisson.

La distribution de la variable réponse Z utilisée dans les relations (1.48) et (1.50) est de loi exponentielle et ces modèles font donc partie des modèles généralisés additifs (GAM) au même titre que le modèle de Cox [95,96]. Ces modèles, utilisés pour des analyses univariées, font appel à des fonctions de lissage des variables explicatives continues. L'introduction des effets aléatoires spécifiques aux individus dans les modèles GAM a conduit aux modèles mixtes additifs généralisés (GAMM) utilisant des fonctions non paramétriques des covariables et des fonctions des effets aléatoires [97].

Pour des données géolocalisées et groupées dans des unités spatiales s , l'introduction d'un effet spatial de ces unités dans les modèles GAMM a conduit aux modèles géo-additifs [93] :

$$\eta = U\gamma + \sum_{j=1}^p f_j(X_j) + f_{spat}(U_s) \quad (1.55)$$

où η est le prédicteur linéaire, U la matrice des effets fixes de nature qualitative relatifs aux n individus, γ le vecteur des paramètres correspondant à ces effets, $f_j(X_j) = X_j\beta_j = (f_j(x_{1j}), \dots, f_j(x_{nj}))$ des fonctions non-paramétriques des effets linéaires X_1, \dots, X_p , $f_{spat}(U_s)$ un effet spatial spécifique à l'unité spatiale s et β_j les paramètres correspondant à ces variables [94,96].

L'introduction des effets aléatoires dans les modèles géo-additifs généralisés conduit aux modèles mixtes géo-additifs généralisés [97,98]. Ces modèles font partie des modèles additifs structurés de régression [99–101]. Les effets fixes rentrent dans la définition de l'espérance du modèle et les effets aléatoires dans sa variance. $f_{spat}(U_s)$ peut se décomposer en une somme d'un effet aléatoire spatial (structuré) et d'un effet aléatoire non spatial (non structuré) dû au fait que, dans le modèle, il y a un effet fixe, ce qui s'écrit [102,103] :

$$f_{spat}(U_s) = f_{structuré}(U_s) + f_{non-structuré}(U_s) \quad (1.56)$$

La décomposition de l'effet aléatoire a été décrite par Besag, York et Mollié [83,104].

L'espérance mathématique du prédicteur linéaire s'écrit alors :

$$\eta = U\gamma + X_1\beta_1 + \dots + X_p\beta_p + f_{spat}(U_s) \quad (1.57)$$

Le modèle de la relation (1.51) peut s'écrire :

$$\eta = U\gamma + X_1\beta_1 + \dots + X_p\beta_p + f_{structuré}(U_s) + f_{non-structuré}(U_s) \quad (1.58)$$

La variable réponse Z_{is} binaire est telle que la probabilité $\pi_{is} = p(Z_{is} = 1)$ est inconnue mais pouvant être estimée par $p_{is} = E(Z_{is})$. La somme des modalités des variables explicatives qualitatives U vaut q et le nombre de variables explicatives continues X vaut p .

Le modèle binomial spatial pour lequel la variable réponse Z_{is} est binaire peut s'écrire [105–107] :

$$\log it(p_{is}) = \gamma_0 + \gamma_1 U_{i1} + \dots + \gamma_q U_{iq} + f_1(X_{i1}) + \dots + f_p(X_{ip}) + f_{spat}(U_s) \quad (1.59)$$

où γ_0 est l'intercept, $\gamma = (\gamma_0, \dots, \gamma_q)$ le vecteur des paramètres correspondant aux q modalités des variables qualitatives pour un individu i donné appartenant à l'unité spatiale U_s , f_1, \dots, f_p des fonctions de lissage non linéaires des variables continues et $f_{spat}(U_s)$ l'effet spatial des entités géographiques U_s .

Le modèle spatial de Poisson s'écrit :

$$\eta_s = \beta_0 + \beta_1 X_{s1} + \dots + \beta_p X_{sp} + f_1(X_{s1}^*) + \dots + f_m(X_{sm}^*) + f_{spat}(U_s) \quad (1.60)$$

où η_s est le nombre d'individus ayant subi l'événement d'intérêt dans l'unité spatiale U_s , β_0 l'intercept, $\beta = (\beta_0, \dots, \beta_p)$ le vecteur des paramètres correspondant aux p variables continues considérées comme linéaires et X^* la matrice des m variables continues considérées comme non linéaires.

Que le modèle spatial soit binomial ou de Poisson, l'estimation et l'inférence des paramètres se fait par des techniques de simulations MCMC dont la mise en pratique peut être faite sous le logiciel BayesX [108,109]. Avec ce logiciel, l'effet spatial $f_{spat}(s)$ est modélisé par un champ aléatoire gaussien de Markov pour des données groupées [110]. Pour des données non groupées, l'effet spatial $f_{spat}(s)$ est modélisé par un champ aléatoire gaussien et les variables explicatives continues par des marches aléatoires comme lois a priori et des fonctions p-splines bidimensionnelles, c'est-à-dire des fonctions mathématiques définies par morceaux et permettant le lissage [111].

L'estimation et l'inférence des paramètres peut se faire soit à l'aide des méthodes basée sur des techniques de simulations de Monte-Carlo par chaînes de Markov (MCMC pour *Markov Chain Monte Carlo*) avec une inférence entièrement bayésienne [99–101,103,112,112] ou la méthode basée sur des modèles mixtes avec une inférence empirique bayésienne [97,99,113,114].

Un a priori sur les fonctions f_j et donc sur les paramètres β_j est telle que :

$$p(\beta_j | \tau_j^2) \propto \exp\left(-\frac{1}{2\tau_j^2} \beta_j' K_j \beta_j\right) \quad (1.61)$$

avec τ_j^2 la variance du paramètre β_j et K_j la matrice de pénalité. Aux variances inconnues τ_j^2 sont assignées des lois a priori Gamma-inverse. Dans le cas de l'estimation bayésienne empirique dans des modèles mixtes, les variances τ_j^2 sont des constantes inconnues et sont estimées à partir de la vraisemblance marginale. L'estimation des paramètres se fait par maximum de vraisemblance restreinte. La même démarche pour le modèle de Poisson peut être effectuée mais dans le cas des données groupées se rapportant au nombre de cas observés.

1.6. Méthodes statistiques d'analyse spatio-temporelle

L'étude de l'évolution spatio-temporelle d'un phénomène d'intérêt requiert la connaissance de sa distribution dans le temps d'une part et dans l'espace d'autre part. L'analyse spatio-temporelle peut être vue comme une analyse spatiale intégrant la notion de temporalité ou bien une analyse temporelle intégrant la notion de spatialité. Autrement dit, il s'agit de montrer que les événements sont proches aussi bien dans l'espace que dans le temps. Lorsque les dates d'apparition des événements (rechute d'une maladie, survenue d'une maladie, survenue d'un effet indésirable due à la consommation d'un produit par exemple) sont connues, il peut être intéressant de détecter des clusters temporeux du phénomène représenté par une série chronologique. Si cette série est indexée par des coordonnées

géographiques et des entités géographiques d'appartenance, un exemple de test statistique permettant d'évaluer la proximité des événements est le test de Knox [115]. Ce test fait intervenir la spatialité et la temporalité des données. Autrement dit, il tient compte de l'interaction espace-temps afin d'identifier les clusters spatio-temporaux. La statistique du test de Knox est la somme des paires d'événements dont les distances spatiale et temporelle ne dépassent pas des seuils spatial δ et temporel τ bien définis :

$$X = \sum_{i=1}^n \sum_{j=1}^n a_{ij}^s a_{ij}^t \quad (1.62)$$

où n est le nombre d'événements, a_{ij}^s une indicatrice qui montre si la distance spatiale d_{ij}^s entre les événements i et j est inférieure à la distance spatiale seuil δ et a_{ij}^t une indicatrice qui montre si la distance temporelle d_{ij}^t est inférieure à la distance temporelle seuil τ . Une version modifiée de ce test a été décrite ailleurs comme étant la moitié de la différence entre la statistique de Knox et son espérance mathématique sous l'hypothèse d'une distribution spatio-temporelle aléatoire des événements [116].

La distance entre les événements est nulle lorsqu'ils se produisent au même moment ou à un même lieu. La statistique de Knox a également été modifiée en ajoutant aux distances spatiales et temporelles une constante c pour éviter que ces distances soient nulles, puis en élevant à la puissance p chaque facteur. Cela a conduit au test de Mantel dont la statistique de test est de la forme [117] :

$$X = \sum_{i=1}^n \sum_{j=1}^n (d_{ij}^s + c)^p (d_{ij}^t + c)^p \quad (1.63)$$

Les statistiques (1.62) et (1.63) sont basées sur les distances spatiales et temporelles seuil. Jacquez a proposé une statistique basée sur la notion de distance au plus proche voisin [118]. L'hypothèse nulle est que la probabilité d'observer deux

plus proches voisins dans l'espace est indépendante de la probabilité d'observer deux plus proches voisins dans le temps.

La statistique de Jacquez est ainsi égale au nombre d'événements qui tombent dans un jeu de k plus proches voisins dans le temps et dans l'espace :

$$J_k = \sum_{i=1}^n \sum_{j=1}^n a_{ijk}^s a_{ijk}^t \quad (1.60)$$

où a_{ijk}^s est une indicatrice qui montre si l'événement j est parmi les k événements géographiquement plus proches voisins de i et a_{ijk}^t une indicatrice qui montre si l'événement j est parmi les k événements temporellement plus proches voisins de i . Du fait de l'absence d'indépendance entre les J_k , Jacquez a proposé une statistique $\Delta J_k = J_k - J_{k-1}$ qui mesure l'interaction espace-temps pour capturer l'interaction non mesurée pour les $(k-1)$ plus proches voisins.

La statistique du scan spatial par la méthode de Kulldorff permet de détecter des clusters purement spatiaux, des clusters purement temporeux sous les modèles de probabilité de Bernoulli ou de Poisson ainsi que des clusters spatio-temporeux [50]. Pour ces derniers, au lieu d'une fenêtre circulaire comme pour le cas purement spatial ou purement temporel, un cylindre de base circulaire et de la hauteur égale à la période de temps au cours de laquelle les clusters se doivent être identifiés balaye tout le domaine d'étude pour détecter des endroits avec des excès de cas dans l'espace et dans le temps. Dans chaque cluster détecté, le nombre de cas observé est comparé à celui attendu sous l'hypothèse d'indépendance entre les positions spatiales et temporelles ou d'une distribution aléatoire des cas dans l'espace et dans le temps, ce qui traduit une absence d'interaction spatiale et temporelle.

Pour l'approche de Kulldorff qui est la plus utilisée, l'inférence est faite à l'aide des simulations de Monte Carlo. Ces méthodes d'analyse spatio-temporelle sont l'objet de nombreuses recherches ; nous n'avons pas vocation dans ce travail à en présenter une liste exhaustive.

Chapitre 2 : Analyse spatiale du VIH/SIDA au Burundi

2.1. Épidémiologie de l'infection par le VIH dans le monde

Depuis que les premiers cas de Syndrome de l'Immunodéficience Acquise (SIDA) ont été diagnostiqués aux États-Unis en 1980, l'infection par le Virus de l'Immunodéficience Humaine (VIH) n'a cessé de se propager dans tous les pays. À l'heure actuelle, elle constitue un problème de santé publique majeur. En 2015, le programme commun des Nations Unies sur le VIH/SIDA (ONUSIDA) estimait à 36,7 millions le nombre de personnes vivant avec le VIH (PvVIH) dans le monde [intervalle de confiance : IC 95 % : 34,0-39,8 millions] parmi lesquels 34,9 millions [IC 95 % : 32,4-37,9 millions] d'adultes et 1,8 millions d'enfants [1,5-2,0 millions] (**Figure 27**). Le rapport souligne également que 19 millions [17,7-20,5 millions] de PvVIH résidaient en Afrique de l'Est et Afrique australe et que 6,5 millions [5,3-7,8 millions] de PvVIH résidaient en Afrique de l'Ouest et Afrique centrale, soit près de 70 % des PvVIH dans le monde qui résidaient en Afrique subsaharienne [117]. Le fardeau de la maladie est notamment mesuré en termes d'années de vie ajustées sur l'incapacité (DALYs : Disability-adjusted life year). Le VIH/SIDA est une des principales causes de mortalité et de morbidité dans le monde : il était la 7^{ème} cause d'années de vie perdues en bonne santé dans le monde (91,907 millions de DALYS sur un total de 2743,857 millions, soit 3,4 %) et la 2^{ème} cause d'années de vie perdues en bonne santé après les infections respiratoires basses en Afrique en 2012 (68,109 millions de DALYS sur 690,532 millions, soit 9,9%).

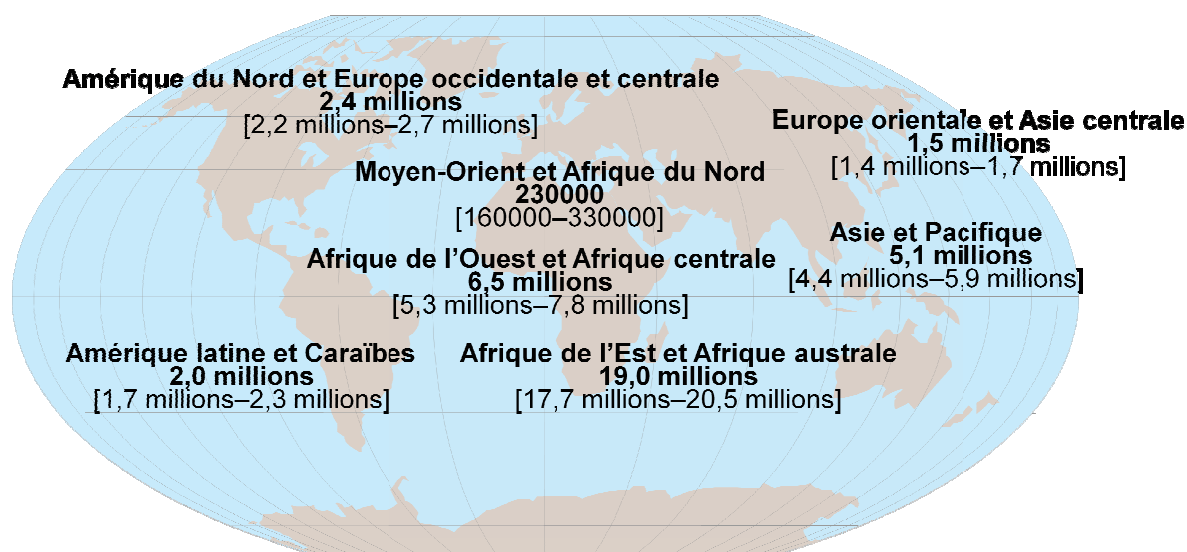
Alors que l'Afrique subsaharienne n'héberge que 13 % de la population mondiale, elle continue ainsi à porter le plus lourd fardeau de l'infection par le VIH. Afin de contribuer à réduire ce fardeau dans le monde en général et en Afrique subsaharienne en particulier, l'ONUSIDA s'est fixé un objectif pour 2020 intitulé « 90-90-90 » : 90 % des PvVIH devront connaître leur statut VIH, parmi elles 90 % devront être sous traitement antirétroviral (ARV) et parmi ces dernières 90 % devront avoir une charge virale indétectable, c'est-à-dire une charge virale inférieure à 1000 copies/ml [118]. Malgré les efforts des gouvernements africains et la collaboration avec les pays du nord, l'Afrique subsaharienne est encore loin de cet objectif. En effet, dans la population adulte de 15 ans et plus, 45 % [39 %-62 %] seulement des PvVIH connaissaient leur statut VIH, parmi elles 86 % étaient sous traitement ARV

et parmi ces dernières 76 % [53 %-89 %] avaient une charge virale indétectable en 2013 [119].

Dans le but d'éradiquer l'épidémie du VIH, l'ONUSIDA avait fixé en 2015 pour objectif de réduire de 90 % le nombre de nouvelles infections, de 90 % la stigmatisation et la discrimination liées à la maladie et de 90% le nombre de décès liés au SIDA d'ici 2030. Les progrès dans la lutte contre le VIH/SIDA se sont traduits par une réduction du nombre de nouvelles infections : le nombre total d'adultes et d'enfants nouvellement infectés était ainsi estimé à 3,2 millions [2,9 millions-3,5 millions] en 2000 contre 2,1 millions [1,8 millions-2,4 millions] en 2015, soit une réduction de 34 %. En Afrique subsaharienne, le nombre d'adultes et d'enfants nouvellement infectés était estimé à 2,3 millions [2,2 millions -2,4 millions] en 2000 contre 1,4 millions [1,2-1,5 millions] en 2014 (**Figure 28**), soit une diminution de 39 % [120]. Parallèlement, le nombre de décès liés au SIDA était estimé à 1,2 millions [1,0 millions-1,5 millions] en 2000 contre 790 000 [940000-1,3 millions] en 2015 (**Figure 29**), soit une diminution de 34 %.

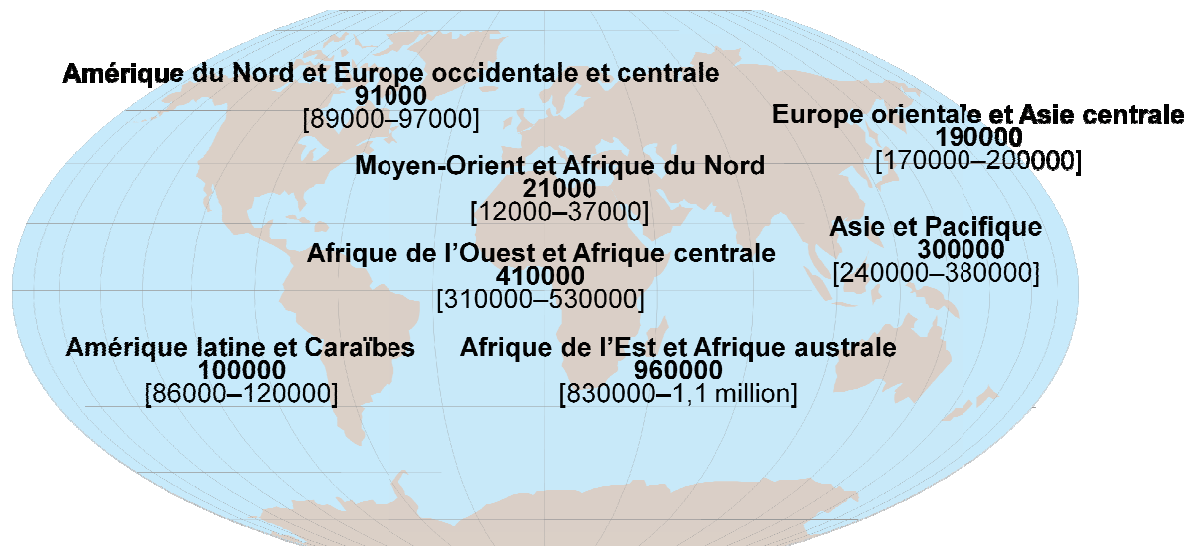
L'augmentation du nombre de PvVIH connaissant leur statut, du nombre de personnes séropositives mises sous traitement ARV et du nombre de personnes sous traitement ARV ayant une charge virale indétectable vise l'atteinte du troisième objectif du développement durable (ODD) qui consiste à « permettre à tous de vivre en bonne santé et à promouvoir le bien-être de tous à tout âge », notamment en sa troisième cible qui consiste à « mettre fin à l'épidémie de SIDA, à la tuberculose, au paludisme et aux maladies tropicales négligées et combattre l'hépatite, les maladies transmises par l'eau et autres maladies transmissibles d'ici 2030 » [121].

Figure 27 : Nombre estimé d'adultes et d'enfants vivant avec le VIH dans le monde en 2015



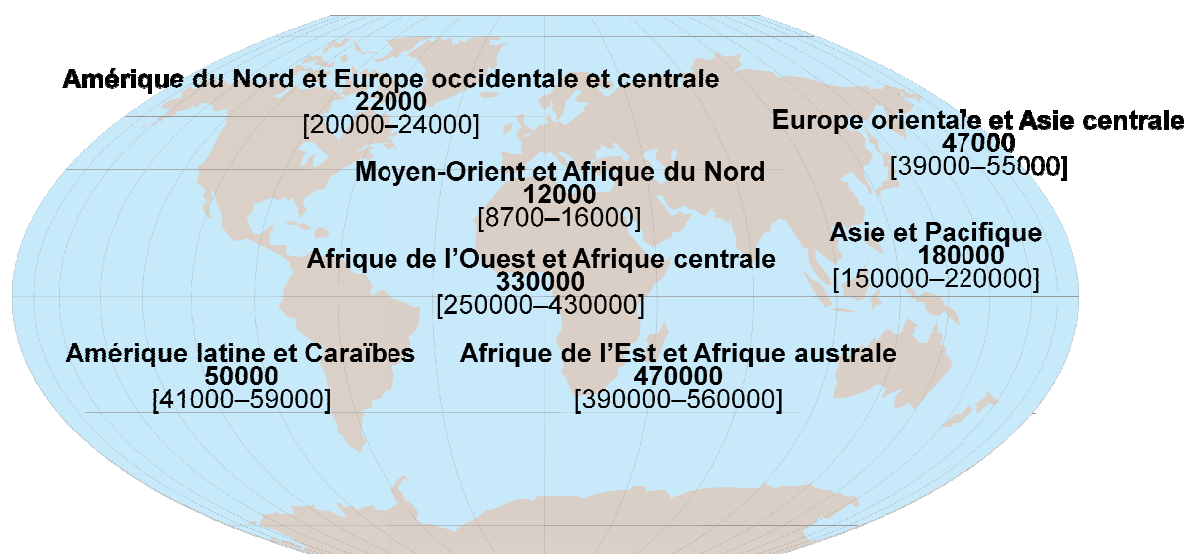
Total: 36,7 millions [34,0 millions–39,8 millions]

Figure 28 : Nombre estimé d'adultes et d'enfants nouvellement infectés par le VIH dans le monde en 2015



Total: 2,1 millions [1,8 millions–2,4 millions]

Figure 29 : Nombre estimé d'adultes et d'enfants décédés des causes liées au VIH/SIDA dans le monde en 2015



Total: 1,1 millions [940 000 – 1,3 million]

Source : Journée mondiale SIDA - Rapport ONUSIDA 2016, Diapositives clés sur les données épidémiologiques

D'autres progrès en matière de lutte contre le SIDA ont été réalisés ces dernières années :

- le nombre de PvVIH bénéficiant d'un traitement ARV était estimé à 800 000 personnes en 2003 contre 15,8 millions de personnes en 2015, soit une augmentation de 95 % ;
- le nombre de décès pour des causes liées au SIDA était estimé à 1,6 millions de cas en 2000 et à 2,0 millions de cas en 2005, soit une diminution respective de 31 % et de 45 % par rapport à l'année 2015.

Les progrès réalisés sont aussi à l'échelle de la région des Grands Lacs située en Afrique de l'Est. L'Afrique de l'Est et l'Afrique australe étaient les régions les plus touchées par l'infection par le VIH dans le monde en 2011. En effet, 1,2 million [1,1 million-1,3 million] de personnes y ont nouvellement été infectées par le VIH et 800 000 personnes sont décédées des causes liées au SIDA en 2011 [122]. Le mode le plus fréquent de transmission du VIH est la transmission hétérosexuelle.

2.2. Épidémiologie de l'infection par le VIH dans la région des Grands Lacs

Les pays de la région des Grands Lacs désignent le Burundi, le Rwanda, la République Démocratique du Congo (RDC), anciennement appelée « Zaïre », la République Unie de Tanzanie (ou tout simplement la Tanzanie), l'Ouganda et le Kenya. Ces pays sont membres de la communauté des pays de l'Afrique de l'Est (EAC pour East African Community) et de l'initiative des pays des Grand-Lacs contre le SIDA (GLIA – Great Lakes Initiative on HIV/AIDS). Dans le cadre de cette initiative, les données de population et de santé sont mises à jours par la conduite d'enquêtes nationales.

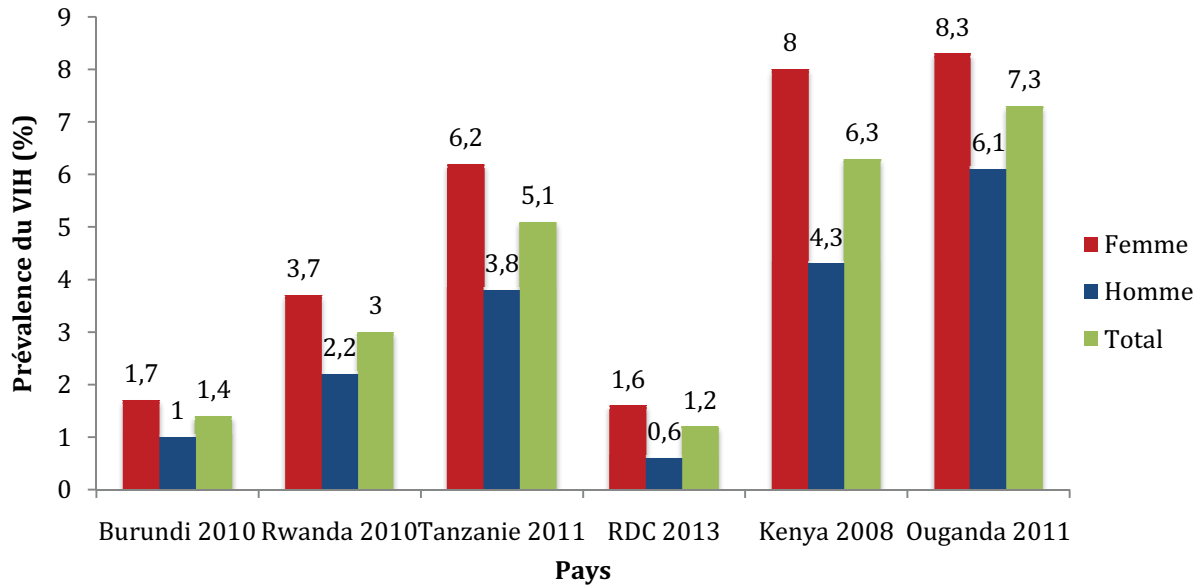
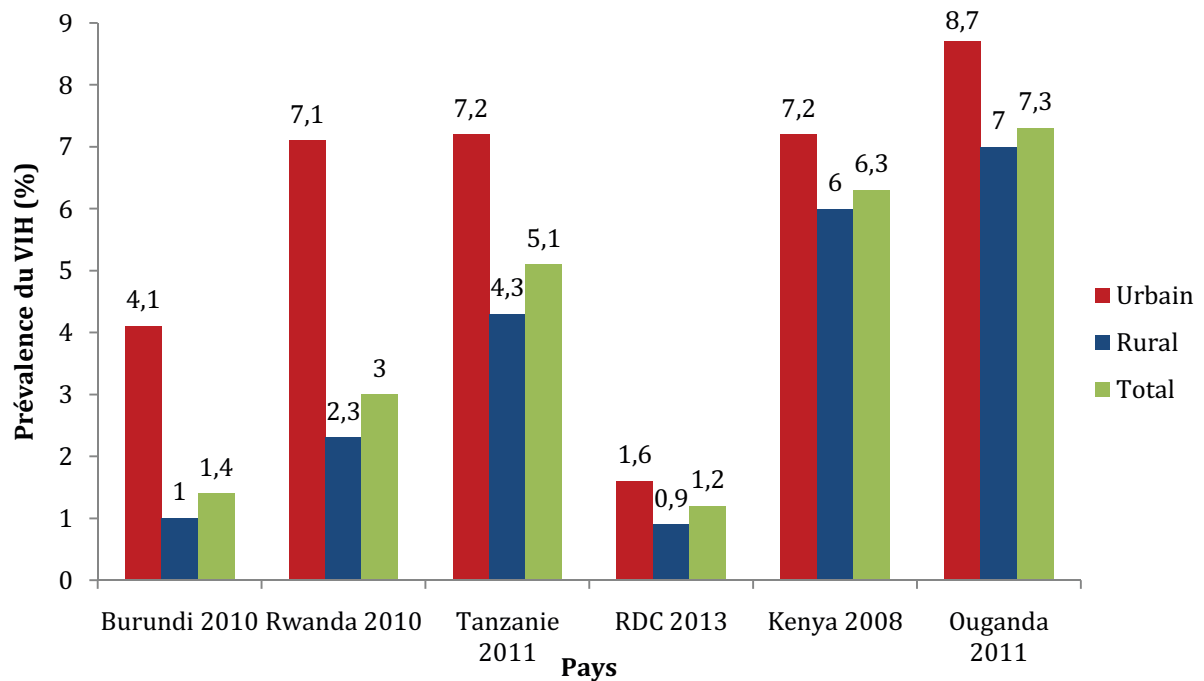
Une enquête démographique et de santé (EDS) conduite au Rwanda en 2010 a montré une prévalence nationale de 3,0 % chez les femmes et les hommes âgés de 15 à 49 ans avec des disparités selon le sexe et le milieu de résidence (urbain/rural) (**Tableau 4**). Dans ce pays, la prévalence du VIH était estimée à 2,9 % [2,6 %-3,2 %] selon l'ONUSIDA en 2015 [123,124]. Une enquête sur les indicateurs du SIDA (EIS) a été conduite en Tanzanie en 2011 et a montré une prévalence nationale de 5,0 % chez les femmes et les hommes âgés de 15 à 49 ans avec une prévalence du VIH de 6,2 % chez les femmes, 3,8 % chez les hommes, 7,2 % en milieu urbain et 4,3 % en milieu rural [125]. La prévalence du VIH était estimée dans ce pays à 4,7 % [4,2 %-5,3 %] chez les femmes et les hommes de 15-49 ans par l'ONUSIDA en 2015 [126]. L'EDS de 2013 réalisée en RDC a montré une prévalence globale du VIH chez les femmes et les hommes de 15 à 49 ans inférieure à celle des deux pays précédents : 1,2 % avec une prévalence du VIH presque trois fois plus élevée chez les femmes que chez les hommes (1,6 % versus 0,6 %) [127]. Les estimations de l'ONUSIDA spécifiques à ce pays et faites en 2015 ont porté cette prévalence à 0,8 % [0,7 %-1,1 %] [128]. Au Kenya, le rapport de l'EDS de 2008 a rapporté une prévalence nationale du VIH de 6,3 % chez les femmes et les hommes de 15-49 ans avec une prévalence aussi prononcée en milieu urbain qu'en milieu rural (7,2 % versus 6,0 %) mais environ deux fois plus élevée chez les femmes que chez les hommes (8,0 % versus 4,3 %) [129]. Les estimations de l'ONUSIDA de 2015 portaient cette prévalence du VIH à 5,9 % [4,9 %-7,0 %] dans la population de même tranche d'âge [130]. Il n'y a aucune EDS conduite en Ouganda avec une

composante VIH. Selon les estimations de l'ONUSIDA de 2015, la prévalence du VIH en Ouganda est estimée à 7,1 % [6,6 %-7,7 %] dans la population âgée de 15 à 49 ans [131].

Tableau 4 : Prévalence du VIH selon le sexe et le milieu de résidence dans la région des Grands Lacs obtenue à partir des données des EDS

Pays	Année	Global	Sexe		Milieu de résidence	
			Femmes	Hommes	Urbain	Rural
Rwanda	2010	3,0 %	3,7 %	2,2 %	7,1 %	2,3 %
Tanzanie	2011	5,1 %	6,2 %	3,8 %	7,2 %	3,8 %
RDC	2013	1,2 %	1,6 %	0,6 %	1,6 %	0,9 %
Kenya	2008	6,3 %	8,0 %	4,3 %	7,2 %	6,0 %
Burundi	2010	1,4 %	1,7 %	1,0 %	4,1 %	1,0 %

En comparaison des autres pays de la région des Grands Lacs, le Burundi affiche une prévalence du VIH plus faible, à l'exception de la RDC. En effet, la deuxième et plus récente EDS réalisée en 2010 a montré une prévalence du VIH de 1,4 % en population générale des femmes et des hommes de 15-49 ans [132]. La comparaison des prévalences du VIH dans les pays de la région des Grands Lacs est illustrée par les **Figures 30 et 31**.

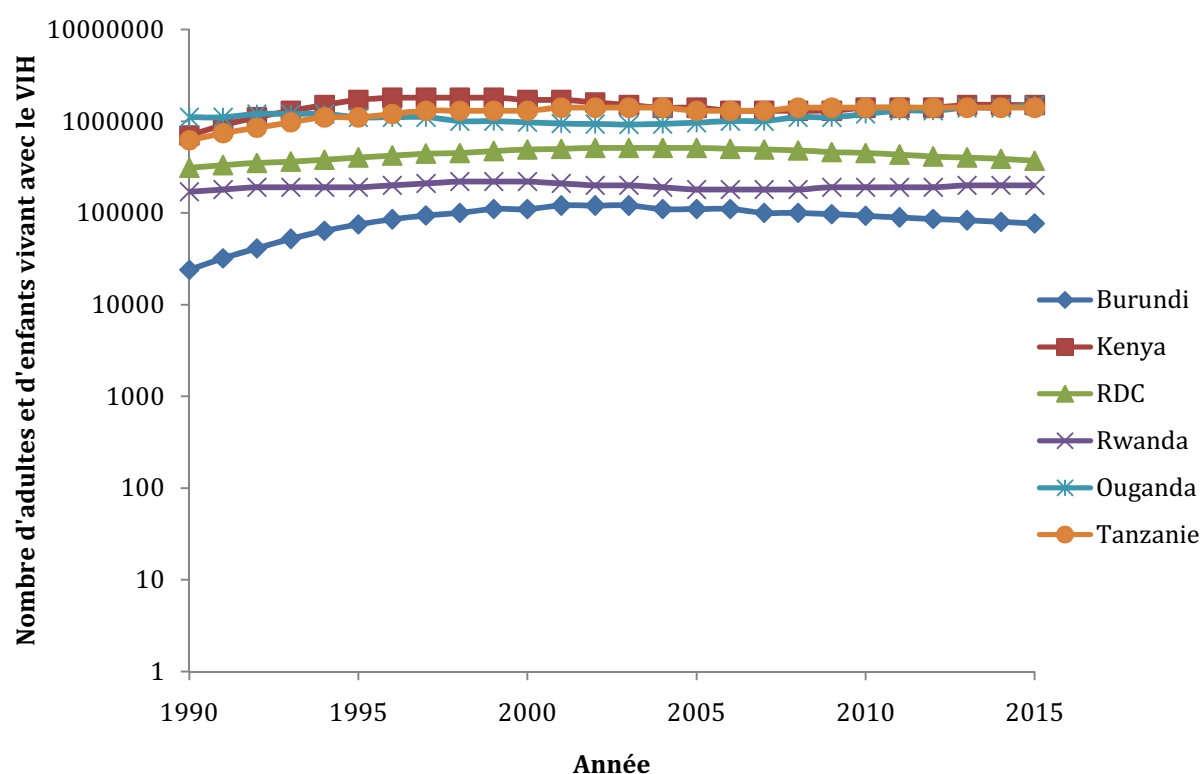
Figure 30 : Prévalence du VIH selon le sexe dans la région des Grands Lacs**Figure 31** : Prévalence du VIH selon le milieu de résidence dans la région des Grands Lacs

Source : À partir des rapports des EDS

Au début des années 1990, le nombre de PvVIH était constamment élevé en Ouganda par rapport à celui des autres pays de la région des Grands Lacs au moment où ce nombre croissait au Kenya. De 1990 à 2015, ce nombre est resté plus bas au Burundi par rapport à celui des autres pays.

Selon les estimations de l'ONUSIDA, le Burundi avait un nombre d'adultes et d'enfants vivant avec le VIH plus faible que celui des pays voisins durant cette période (**Figures 32**). Le nombre de PvVIH a montré une croissance rapide entre 1990 et 1993 en Tanzanie et au Kenya. À partir de 1994, ce nombre est resté élevé au Kenya jusqu'en 2003. Il est resté plus élevé au Kenya que dans les pays voisins de la région des Grands Lacs de 2013 à 2014.

Figure 32 : Évolution du nombre d'adultes et d'enfants vivant avec le VIH dans la région des Grands Lacs, 1990-2015



Toujours selon l'ONUSIDA, le nombre de nouvelles infections par le VIH est resté particulièrement plus élevé au Kenya jusqu'en 1997, en Tanzanie de 1998 à 2005, puis en Ouganda à partir de 2008. Le Rwanda a montré une faible diminution en matière du nombre de nouvelles infections par le VIH (**Figure 33**). S'agissant des décès liés au SIDA, l'Ouganda présentait un nombre élevé de cas jusqu'en 1997, puis le Kenya jusqu'en 2005 et la Tanzanie de 2009 à 2014. Le nombre de décès liés au SIDA était plus faible au Burundi que dans les autres pays de la région des Grands Lacs jusqu'en 2006. Ce nombre était quasiment le même au Burundi et au Rwanda à partir de l'année 2007 (**Figure 34**).

Figure 33 : Évolution du nombre d'adultes et d'enfants nouvellement infectés par le VIH dans la région des Grands Lacs, 1990-2015

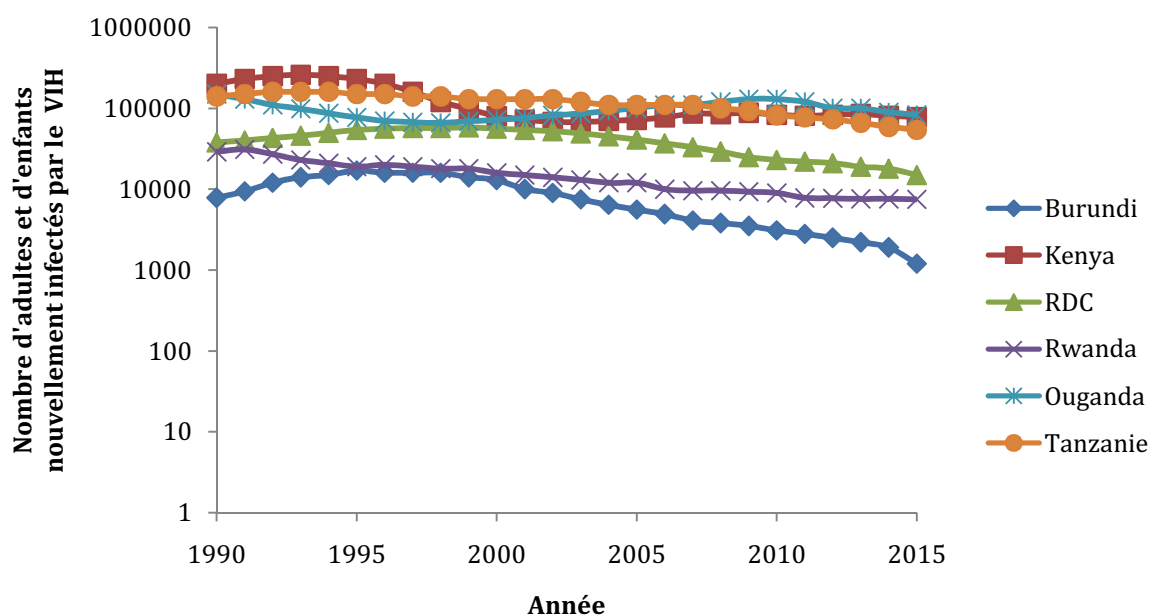
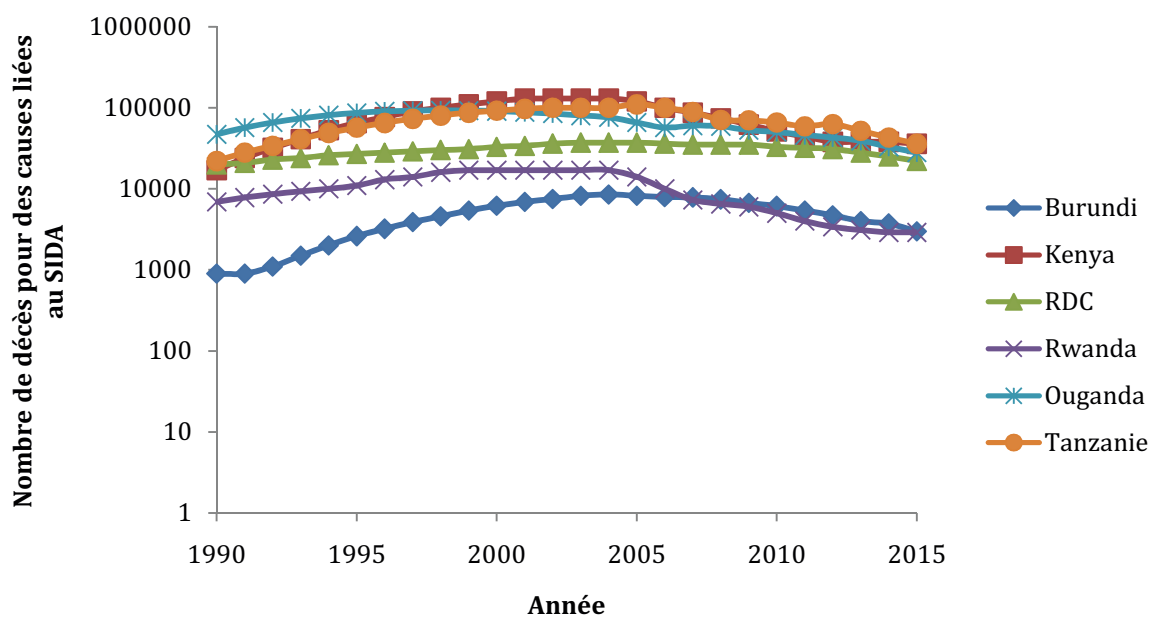


Figure 34 : Évolution du nombre de décès pour des causes liées au SIDA dans la région des Grands Lacs, 1990-2015



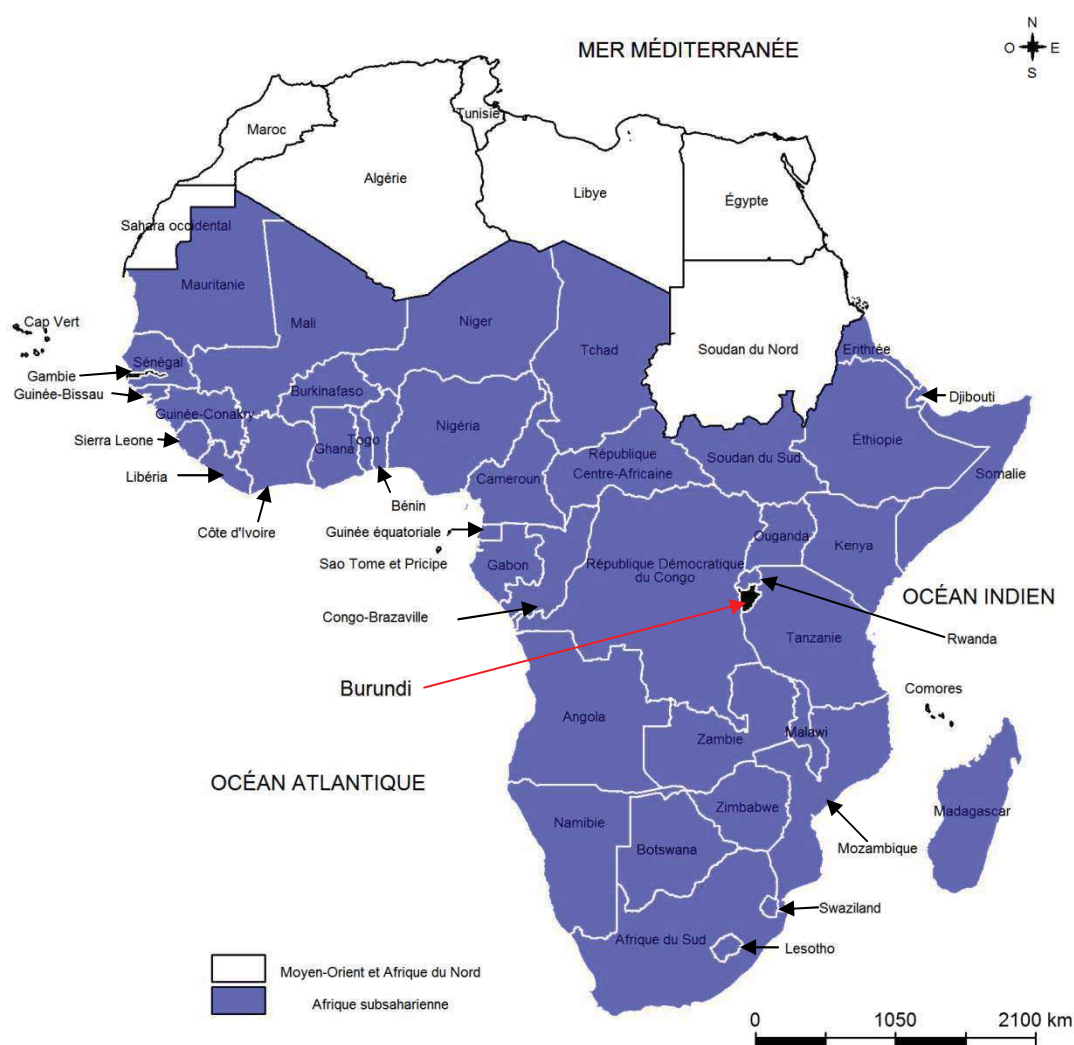
Source : Réalisées à partir des estimations de l'ONUSIDA, 1990-2015

2.3. Contexte du Burundi

2.3.1. Situation géographique, administrative et démographique

Le Burundi, un pays enclavé de l'Afrique subsaharienne (**Figure 36**), est situé à la jonction entre l'Afrique centrale (zone équatoriale) et l'Afrique de l'Est. La capitale du Burundi est Bujumbura et la superficie du Burundi est de 27834 km².

Figure 36 : Situation du Burundi en Afrique subsaharienne



Géographiquement, le Burundi se situe entre 2°20 et 4°27 de latitude sud et entre 28°50 et 30°53 de longitude est. Son altitude varie entre 775 m (Lac Tanganyika) et 2670 m (Crête Congo-Nil). Il est situé à 2200 km de l'Océan Atlantique (à l'Ouest), à

1200 km de l'Océan Indien (à l'Est) et à 3800 km de la mer Méditerranée (au Nord) [133]. Il est limité au nord par le Rwanda, à l'ouest par la RDC, et au sud et à l'est par la Tanzanie (**Figure 36**).

En plus du lac Tanganyika qui occupe une superficie de 2634 km², le Burundi dispose aussi des lacs internes au nord du pays qui font partie du bassin du Nil. Ces lacs sont Rwero (102 km²), Cohoha (78,50 km²), Kanzigiri (7,50 km²), Rwihinda (3,40 km²) et Gacimirinda (2,50 km²), occupant ainsi une superficie totale de 193,90 km².

La température oscille autour de 20°C dans les plateaux centraux, autour de 16° C dans les montagnes et autour de 23° C à Bujumbura. Le Burundi comprend deux saisons : une longue saison sèche mêlée d'une courte période de pluie et une longue saison des pluies mêlée d'une courte période sèche. Le Burundi occupe la 169^{ème} place sur 177 pays avec un indicateur de développement humain (IDH) de 0,339 en 2015 [134].

Administrativement, le Burundi est subdivisé en 5 régions sanitaires qui ont constitué un niveau de stratification pour l'EDS de 2010 (Bujumbura-Mairie, l'Ouest, le Nord, le Sud et le Centre-Est). Le pays est actuellement composé de 18 provinces (au lieu de 17 en 2015) réparties au sein de ces régions, ces dernières ayant aussi actuellement changé (**Figure 37**).

Figure 37 : Répartition spatiale des régions sanitaires et des provinces au Burundi en 2015



Ainsi, l'Ouest comprend les provinces de Cibitoke, Bubanza et Bujumbura-Rural. Le Nord comprend Ngozi, Kayanza, Kirundo et Muyinga. Le Centre-Est comprend Muramvya, Gitega, Ruyigi, Karuzi et Cankuzo. Le Sud comprend Mwaro, Rutana, Rumonge, Bururi et Makamba. La nouvelle province de Rumonge est issue de la fusion de deux communes provenant de l'ancienne province de Bujumbura-Rural et qui sont Muhuta et Bugarama et de trois communes provenant de l'ancienne province de Bururi et qui sont Rumonge, Buyengero et Burambi, et cela pour des raisons de décentralisation administrative.

La région de Bujumbura-Mairie (ou tout simplement Bujumbura) est à la fois la capitale du Burundi et une province complètement urbaine issue de la séparation de l'ancienne province de Bujumbura en deux provinces : Bujumbura-mairie et Bujumbura-Rural. Chaque province est, à son tour, divisée en communes (il y en a 129 au total) et chaque commune en zones (il y en a 2911 au total). Chaque zone regroupe plusieurs collines (il y en a 2847 au total), l'entité administrative la plus périphérique précédant le ménage.

La population burundaise est essentiellement rurale (90 %) et vit principalement de l'agriculture. Actuellement, la population burundaise est d'environ 10,7 millions d'habitants avec une densité de 386 habitants/km², une proportion d'hommes de 49,5 %, un nombre moyen de 6 enfants par femme et un taux de croissance démographique annuelle de 3,3 % [135]. La population burundaise est jeune. En effet, 45,6 % des femmes et des hommes sont âgés de 0-14 ans, et les 65 ans et plus ne représentent que 2,5 % de la population. Avec l'extension des villes, la population urbaine représente 12,1 % avec un taux annuel d'urbanisation de 5,7 %.

Secoué par des guerres civiles successives (1965, 1972, 1988, 1993-2003, 2016), le Burundi est un pays à statistique déficiente.

2.3.2. Organisation du système de santé

Le Ministère de la Santé Publique et de la lutte contre le SIDA (MSPLS) constitue l'administration centrale et joue le rôle de régulateur et de normalisateur [136]. Issu de la fusion entre le Ministère de la Santé Publique (MSP) et le Ministère à la Présidence chargé de la lutte contre le SIDA (MPLS) en 2009, le MSPLS a pour mission de définir la politique sanitaire nationale, sectorielle et des normes structurelles et sanitaires, de faire la planification stratégique, la coordination du secteur de la santé, la gestion et l'affectation stratégiques des ressources, et d'appuyer les niveaux intermédiaire et périphérique. Le Burundi est le deuxième pays à avoir créé un ministère ayant la lutte contre le VIH dans son intitulé après la Côte d'Ivoire [137]. Spécifiquement, le niveau central est constitué d'un cabinet du ministre, d'une inspection générale de la santé qui lui est rattachée, de deux directions générales, des institutions personnalisées, de six départements de santé,

de 9 programmes de santé et des organes de direction spécifiques à chacun de ces programmes [136]. Pour faciliter le suivi et l'évaluation de ces programmes, il a été créé un secrétariat permanent (comme dans tous les autres ministères), une direction générale de la planification et du suivi-évaluation et un programme national intégré de lutte contre le VIH/SIDA.

Le niveau intermédiaire comprend les 18 Bureaux Provinciaux de Santé (BPS) qui se trouvent dans chaque province. Une telle déconcentration attribue aux BPS le rôle d'assurer la coordination de l'action sanitaire et de lutter contre le SIDA au niveau de chaque province. Les BPS fournissent également un appui matériel et logistique aux districts sanitaires établis dans les provinces.

Le niveau périphérique est le niveau des unités opérationnelles du système de soins [136]. Il est constitué de 45 districts sanitaires dont 36 ont un hôpital de district sanitaire (HDS) appelé hôpital de première référence. Chaque district sanitaire regroupe deux ou trois communes et a environ 10 à 15 Centres de Santé (CDS). Une population d'environ 150000 habitants y fait des recours aux soins. Chaque district dispose d'un Bureau de District Sanitaire (BDS). Les BDS collaborent étroitement avec les BPS. Il existe également des Bureaux Communaux de Santé (BCS) dans chaque commune. Le niveau périphérique a pour rôle de coordonner l'action sanitaire de proximité, de veiller au bon fonctionnement des CDS, des HDS et des structures des associations ayant pour but de promouvoir la santé des populations et de lutter contre le SIDA. Étant proche de la population, il permet de relayer des messages de sensibilisation, de prendre en charge les malades et d'en assurer le suivi. De plus, c'est à ce niveau que les communautés et les confessions religieuses participent à l'action sanitaire via les comités de santé (COSA).

Au même titre que l'administration des structures de soins, le niveau de l'administration des soins est hiérarchisé : niveau de base, niveau de première référence et niveau de référence nationale [136].

Le niveau de base est constitué des hôpitaux de première référence. Ces hôpitaux offrent des services d'hospitalisation, d'urgences aux patients transférés par les CDS

ou gravement malades et des services de consultation externe sur demande des CDS, ces derniers offrant un paquet minimum d'activités (PMA) à la population résidant dans une aire géographique bien déterminée [138]. Les CDS et hôpitaux publics offrent des soins gratuits depuis 2006 pour les femmes enceintes et les enfants âgés de moins de 5 ans. Les CDS, dont le nombre s'élève actuellement à 863, sont des lieux de consultation de soins les plus accessibles par la population. Parmi les 863 CDS, 636 ont une activité VIH. Sur 64 hôpitaux que compte le pays, 37 sont des hôpitaux publics, 14 des hôpitaux confessionnels et 13 des hôpitaux privés [139].

Les 5 régions sanitaires du Burundi ont chacune, à l'exception des régions de Bujumbura-Mairie et de l'Ouest, un hôpital régional appelé hôpital de deuxième référence. Ces hôpitaux de deuxième référence sont situés dans les provinces de Gitega (Hôpital de Gitega) pour la région du Centre-Est, de Ngozi (Hôpital de Ngozi) pour la région du Nord et de Bururi (Hôpital de Bururi) pour la région du Sud. Ils offrent des services de soins plus spécialisés que ceux des hôpitaux de première référence.

Le niveau de référence nationale est constitué des hôpitaux plus spécialisés appelés hôpitaux de référence nationale. Ces hôpitaux ont l'avantage d'offrir des soins et examens non fournis par les hôpitaux de première et de deuxième référence. Il s'agit principalement du Centre Hospitalo-Universitaire de Kamenge (CHUK), de l'Hôpital Prince Régent Charles (HPRC), de la Clinique Prince Louis Rwagasore (CPLR), de l'Hôpital Militaire de Kamenge (HMK) et du Centre Neuro-Psychiatrique de Kamenge (CNPK) tous situés à Bujumbura [140]. Cette liste est complétée par le Centre National de prise en charge de la tuberculose, anciennement appelé « Senatorium de Kibumbu » dans la province de Mwaro et le Centre National d'Appareillage et de Réadaptation (CNAR) de Gitega.

2.3.3. Historique de la lutte contre le VIH/SIDA

Depuis la description des premiers cas de SIDA au Burundi en 1983, l'infection par le VIH n'a cessé de se propager dans le pays, affectant les jeunes et les adultes des milieux urbains et ruraux du pays, ce qui a contribué à la paupérisation des familles

et à la baisse de l'économie du pays. Les principaux modes de propagation du VIH au Burundi sont la transmission par voie sexuelle et la transmission du VIH de la mère à l'enfant. Au cours de l'année 1983, l'infection par le VIH touchait uniquement les zones urbaines. Pour contrecarrer cette propagation qui prenait une allure inquiétante, le gouvernement a mis en place, via le Ministère chargé de la lutte contre le SIDA, des campagnes de sensibilisation afin de prévenir l'infection et des programmes de prise en charge des personnes infectées. C'est dans ce cadre qu'il a été créé, pour la première fois, un service national de lutte contre le SIDA et les maladies sexuellement transmissibles (MST) en 1986, un service qui deviendra le Programme National de Lutte contre le SIDA et les MST (PNLS/MST) en 1988 (**Tableau 5**).

Tableau 5 : Programmes et principaux intervenants de la lutte contre le VIH/SIDA au Burundi

Année	Programmes/intervenants
1983	Déclaration des premiers cas de SIDA
1986	Service national de lutte contre le SIDA et les Maladies Sexuellement Transmissibles (SNLS/MST)
	Plan de lutte à courte terme (1987-1988)
1988	Programme National de Lutte contre le SIDA et les MST (PNLS/MST)
	Premier plan à Moyen Terme de lutte contre le SIDA 1988-1992 (PMT1)
1990	Association d'Aide aux Victimes du SIDA (AVISI)
1992	Society for Women Against AIDS in Africa-Burundi (SWAA-Burundi)
	Famille pour Vaincre le Sida (FVS)
	La Nouvelle Espérance pour les malades du SIDA
	Accord-Burundi (the African Center for the constructive resolution of disputes - Burundi)
1993	Association de Prise En Charge des Orphelins du SIDA (APECOS)
	Association Nationale de Soutien aux Séropositifs et Sidéens (ANSS)
	Deuxième plan à Moyen Terme de lutte contre le SIDA 1993-1997 (PMT1)
1998	Premier Plan Stratégique National de Lutte contre le SIDA (1999-2003)
1999	Alliance Burundaise contre le SIDA et pour la promotion de la santé (ABS)
	Association d'aide aux mères célibataires (NTURENGAHA)
	Association pour l'encadrement des personnes vulnérables en Afrique (EPVA) est une ONG
2000	Projet Multisectoriel de Lutte contre le SIDA et d'appui aux Orphelins (PMLSAO)

Tableau 5 : Programmes et principaux intervenants de la lutte contre le VIH/SIDA au Burundi (suite)

Année	Programmes/intervenants
2002	Appui au Programme National de Lutte contre le Sida » (AP/SIDA) dont la couverture s'étend au début sur quatre provinces à savoir RUYIGI, GITEGA, MUYINGA et BUJUMBURA Mairie
	Projet d'Appui aux organisations locales dans la lutte contre le SIDA et dans la promotion de la santé communautaire
	Deuxième Plan Stratégique National de Lutte contre le SIDA (2002-2006)
	Mise en place d'un cadre institutionnel conforme aux trois principes de l'ONUSIDA
2006	Association burundaise pour la prise en charge des malades vulnérables (ABCMAV)
	Troisième Plan Stratégique National de Lutte contre le SIDA (2007-2011)
2007	Deuxième Projet Multisectoriel de lutte contre le SIDA 2007-2011
	Projet d'intensification et de décentralisation de la lutte contre le SIDA (PRIDE)
2008	Service Yezu Mwiza (SYM)
2009	Forum burundais de lutte contre le VIH/SIDA (FOBULUSI)
2010	Réseau Burundais des Personnes vivant avec le VIH (RBP+)
	Collectif des Associations des Personnes vivant avec le VIH/SIDA (CAPES+)
2014	Quatrième Plan Stratégique National de Lutte contre le SIDA (2014-2017)

A partir de l'année 1988, des conflits sociopolitiques ont commencé à surgir dans les communes du nord du pays (Ntega-Marangara), conduisant ainsi à des situations de pauvreté, à la détérioration des services publics et des services de santé, à la dégradation de l'état sanitaire et aux migrations internes favorisant la diffusion du VIH/SIDA. Durant ces années, l'estimation de la prévalence du VIH au niveau

national se faisait à l'aide des données de surveillance sentinelle de la prévalence du VIH chez les femmes enceintes.

À partir de l'année 1989, pour freiner la propagation du VIH, des associations de lutte contre le SIDA qui prônaient la lutte contre la stigmatisation et la discrimination liées au SIDA ont commencé à voir le jour. Ces associations de lutte contre le SIDA ont joué un rôle particulièrement prépondérant dans la lutte contre le SIDA. Par exemple, l'Association Nationale de Soutien aux Séropositifs et Sidéens (ANSS) a créé un centre dénommé « La maison de la joie » pour accueillir les enfants orphelins pour des causes liées au SIDA, séropositifs ou non. Depuis sa création, elle a contribué et continue à maintenir les PvVIH en vie par leur suivi et leur prise en charge médicale en leur facilitant l'accès aux ARV. Actuellement, elle s'intéresse aussi à la prise en charge des hommes ayant des rapports sexuels avec d'autres hommes (HSH). Elle a également créé un centre dénommé « TURIHO » qui veut dire « Nous sommes encore en vie » qui fait la promotion de la prévention de la transmission du VIH, tout en contribuant à l'amélioration du bien-être des personnes déjà atteintes par le VIH. La branche burundaise de l'organisation panafricaine de femmes face au VIH/SIDA (Society for Women Against AIDS in Africa), dénommée SWAA-Burundi, a contribué et contribue actuellement à prévenir les infections par le VIH principalement chez les filles et les femmes, à prendre en charge médicalement et psychologiquement les PvVIH vivant à Bujumbura et dans ses antennes (Gitega, Ngozi, Kayanza, Muyinga, Muramvya, Ruyigi, Burundi) et à protéger les enfants de la transmission du VIH de la mère à l'enfant lors de la PTME. De plus, elle offre un service de dépistage et encourage la population à l'adoption de comportements à faible risque d'infection par le VIH. Comme pour l'ANSS, elle a inscrit dans ses activités un volet HSH. L'association pour l'encadrement des personnes vulnérables en Afrique (EPVA) est contribue à assurer la prise en charge des PvVIH, des sidéens et des orphelins de guerre et du SIDA. La Famille pour Vaincre le Sida (FVS) prend en charge les orphelins et autres enfants vulnérables (OEV) et plus particulièrement les orphelins du SIDA.

En 1999, des organisations de la société civile qui se proposent de lutter contre le VIH et de promouvoir la santé des populations se sont coalisées au sein de l'alliance

burundaise contre le SIDA et pour la promotion de la santé (ABS). Elles se sont impliquées dans l'élaboration du plan stratégique national de lutte contre le SIDA et les MST. L'association d'aide aux mères célibataires (NTURENGAHA) a été créée par des bénévoles et des anciens membres de la SWAA-Burundi la même année. C'est au cours de cette année que le projet multisectoriel de lutte contre le SIDA et d'appui aux Orphelins (PMLSAO) a été mis en place. Deux ans plus tard, l'appui au programme national de lutte contre le SIDA (AP/SIDA) et le projet d'Appui aux organisations locales dans la lutte contre le SIDA et dans la promotion de la santé communautaire ont été mis en place, de même que le cadre institutionnel conforme aux trois principes « un-un-un » de l'ONUSIDA. Le premier principe consiste en « un cadre d'action contre le SIDA, qui offre une base pour la coordination des activités de tous les partenaires » qui se décline en un plan national stratégique, le deuxième en « une instance nationale de coordination de la lutte contre le SIDA, à représentation large, multisectorielle » et le dernier en « un système de suivi et d'évaluation à l'échelon pays » [141].

D'autres organisations de lutte contre le SIDA furent progressivement créées. C'est notamment :

- le Réseau Burundais des PvVIH, dénommée (RBP+), fut créé en 2010 mais, à notre connaissance, il n'administre pas des soins aux PvVIH contrairement aux autres associations de lutte contre le SIDA comme l'ANSS, la FVS et la SWAA-Burundi ;
- le Service Yezu Mwiza (SYM), pour dire « Service Bon Jésus », est un service agréé en 2008 qui exerce ses activités sous le label de l'Association des Pères Jésuites du Burundi (APJP) fondée en 1964 et appartenant au Réseau Jésuite Africain de lutte contre le SIDA créé en 2002. Il s'occupe des PvVIH en leur procurant des antirétroviraux et des soins et propose des tests de dépistage pour le VIH et la tuberculose. Il fait également de la sensibilisation auprès des jeunes, de la prévention, de la promotion de la santé sans le VIH/SIDA, de la prise en charge médicale et psychosociale et combat les violences faites aux femmes;
- le Collectif des Associations des PvVIH (CAPES+) créé en 2010 ;

- l'Association burundaise pour la prise en charge des malades vulnérables (ABCMAV) créée en 2006 et qui est membre de l'ABS regroupant 433 associations;
- le forum burundais de lutte contre le VIH/SIDA (FOBULUSI) créé en 2009 ;
- le CED Caritas Burundi, créé en 1962, est une organisation non gouvernementale chrétienne.

2.3.4. Organisation de la lutte contre le VIH/SIDA

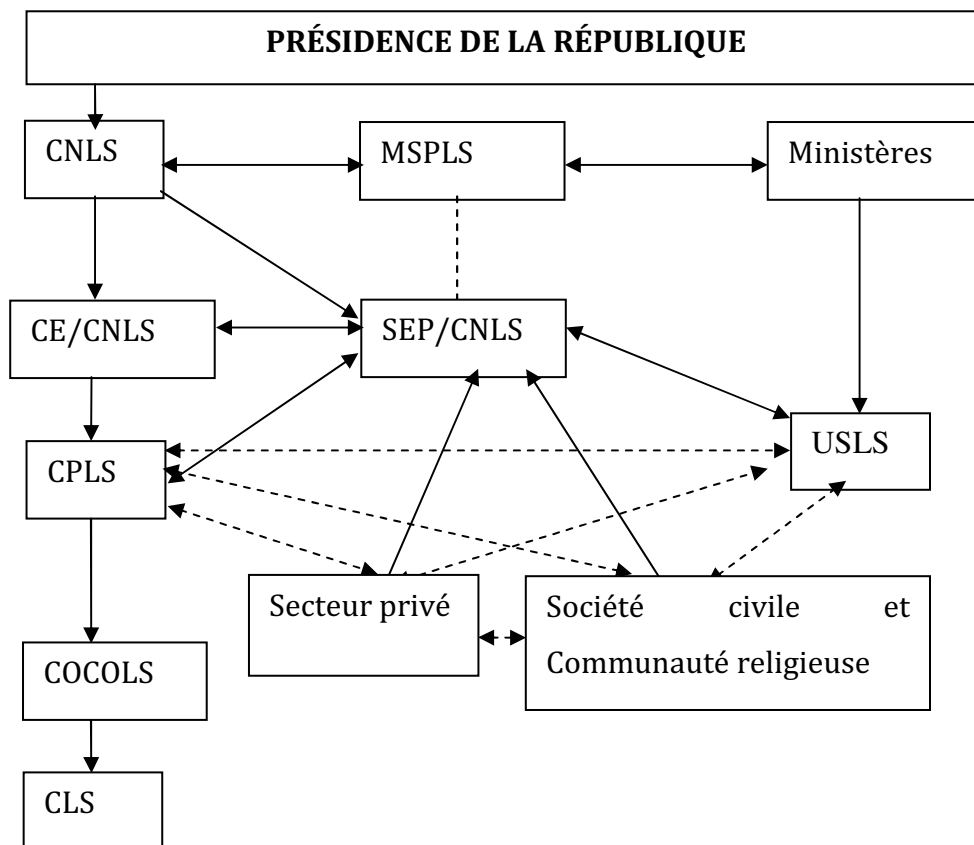
Le gouvernement burundais a créé un volet sur la lutte contre le VIH au niveau de la présidence de la République afin de contribuer à freiner la vitesse de propagation de l'infection par le VIH au Burundi. Le Président de la République est lui-même le Président du Conseil National de Lutte contre le SIDA (CNLS), le seul organe national de coordination de l'action de lutte contre le SIDA et dont la création est conforme avec les « trois principes » de l'ONUSIDA. Ce « cadre institutionnel », le CNLS, se réfère au plan stratégique national comme « seul plan d'action » de lutte contre le SIDA et « un plan de suivi et évaluation » de tous les programmes de lutte contre le SIDA est mis en place.

Le CNLS a un comité exécutif (CE/CNLS), une assemblée générale (AG/CNLS) et un Secrétariat Exécutif Permanent (SEP/CNLS). Le SEP/CNLS et le Programme National de Lutte contre le SIDA et les MST (PNLS/MST) relèvent du MSPLS. Pour alléger les responsabilités et étendre les niveaux d'exécution des tâches, le CNLS a créé des Unités Sectorielles de lutte contre le SIDA (USLS) dans tous les ministères, des comités provinciaux de lutte contre le SIDA (CPLS), des comités communaux de lutte contre le SIDA (COCOLS), des comités locaux de lutte contre le SIDA (CLS) sur les collines et des comités municipaux de lutte contre le SIDA (CMLS) dans la municipalité de Bujumbura (**Figure 38**).

Au niveau des infrastructures, 167 centres de dépistage volontaire (CDV), 40 sites de prise en charge avec traitement ARV, 27 sites du programme de lutte contre la transmission du VIH de la mère à l'enfant (PTME), 129 centres de jeunes ont été créés afin d'offrir à la population des services VIH de proximité.

Les associations de lutte contre le SIDA collaborent entre elles, avec les CPLS et avec les USLS et contribuent dans l'élaboration du Plan d'Action Nationale de Lutte contre le SIDA (PANLS). Le PNLIS inclut la lutte contre les infections sexuellement transmissibles (IST) et s'étend de la présidence de la République aux comités locaux de lutte contre le SIDA (CLS) (**Figure 38**). La décentralisation de leurs programmes de lutte contre le SIDA a été concrétisée par l'établissement des antennes dans les provinces afin de prévenir, au mieux, le SIDA et de prendre en charge les personnes déjà atteintes et les personnes affectées par le VIH.

Figure 38 : Organigramme du PNLIS/IST au Burundi



Source : Adapté à partir de MPLS, 2006 (PSN 2007-2011)

- PNLS : Programme National de Lutte contre le SIDA
 IST : Infections sexuellement transmissibles
 MSPLS : Ministère de la Santé Publique et de la Lutte contre le SIDA
 MPLS : Ministère à la Présidence chargée de la lutte contre le SIDA

IST	: Infections sexuellement transmissibles
SEP	: Secrétariat Exécutif Permanent
CNLS	: Conseil National de Lutte contre le SIDA
CPLS	: Comités Provinciaux de Lutte contre le SIDA
USLS	: Unités Sectorielles de Lutte contre le SIDA
COCOLS	: Comités Communaux de Lutte contre le SIDA
CLS	: Comités Locaux de Lutte contre le SIDA

La flèche aller (tiret continu) montre un lien hiérarchique, la flèche aller-retour ((tiret continu) la coordination et l'appui à la mise en œuvre du PANLS, la flèche aller (trait discontinu) la tutelle et la flèche aller-retour (trait discontinu) la collaboration dans la mise en œuvre du PANLS [137].

2.3.5. Politique nationale de lutte contre le SIDA

La politique nationale de lutte contre le SIDA s'appuie sur des plans stratégiques nationaux (PSN) à court et moyen termes définis successivement pour les années 1999-2003, 2002-2006, 2007-2011, 2012-2016 et 2014-2017 (le PSN 2012-2016 ayant été revu). Ces plans stratégiques définissent les axes stratégiques à suivre dans la lutte contre le SIDA et les progrès réalisés sont évalués après chaque année. Cette évaluation concerne les programmes définis dans les PSN et les PANLS. Ces programmes relatifs à l'infection par le VIH sont définis dans la Politique Nationale Sanitaire (PNS 2005-2015, 2016-2026) et cette politique est mise en œuvre à travers les Plans Nationaux pour le Développement Sanitaire (PNDS 2006-2010, 2011-2015, 2016-2020).

Les axes stratégiques de lutte contre le SIDA pour le PSN 2014-2017 sont la réduction des nouvelles infections par le VIH et des IST chez les jeunes (scolarisés ou non) et chez les femmes (particulièrement les femmes vulnérables) ; la prise en charge globale des PvVIH, des enfants infectés par le VIH, des personnes affectés ; la santé de la reproduction ; les programmes d'intervention de proximité auprès des groupes à haut risque de contamination par le VIH (les HSH, les travailleuses de sexe, les détenus, les jeunes, les hommes en uniforme, les déplacés, les usagers de drogues et les rapatriés en provenance des pays avec une forte prévalence du VIH

comme la Tanzanie) ; la réduction de l'impact socio-économique du VIH et du SIDA ; le renforcement de la coordination de la réponse nationale et du suivi-évaluation [142,143]. Comme pour les précédents PSN, le PSN 2014-2017 s'inscrit dans le cadre du respect des trois principes de l'ONUSIDA (un seul plan stratégique national, un organe national de coordination et un seul plan national de suivi-évaluation) [144].

Le Burundi a fait siennes toutes les initiatives tant internationales que régionales de lutte contre le SIDA telles que la déclaration d'engagement sur le VIH/SIDA émise lors de la session spéciale de l'Assemblée Générale des Nations Unies sur le VIH/SIDA (UNGASS : United Nations General Assembly Special Session) de 2001 à laquelle 189 États membres avaient participé, et l'accès universel à la prévention, aux soins et aux traitements [136,145]. Dans le but d'atteindre les objectifs définis dans les PSN et les Objectifs du Millénaire pour le Développement (OMD) définis en 2000, le Burundi a toujours inscrit au premier plan la lutte contre le SIDA et les maladies transmissibles. En 2002, le Burundi a défini un vaste programme de lutte contre le SIDA à travers le PSN 2002-2006. Ce programme visait l'atteinte des objectifs et engagements définis lors de la session spéciale de l'Assemblée Générale des Nations Unies sur le VIH/SIDA et celle des OMD. Ces objectifs visent les actions de prévention, de dépistage, de prise en charge et d'appui psychosocial en vue d'un accès universel au traitement et de la création en 2030 d'un monde avec Zéros nouvelles infections, Zéros décès liés au SIDA et Zéro discrimination [143].

Les progrès réalisés par le Burundi en matière de lutte contre le SIDA ont été illustrés notamment par une augmentation du nombre de personnes sous traitement ARV de 2002 à 2015 (**Figure 39**), du nombre de sites de traitement ARV de 2006 à 2015 (**Figure 40**) et du nombre de personnes dépistées de 2003 à 2015 (**Figure 41**).

Figure 39 : Évolution du nombre de personnes sous traitement ARV au Burundi, 2002-2015

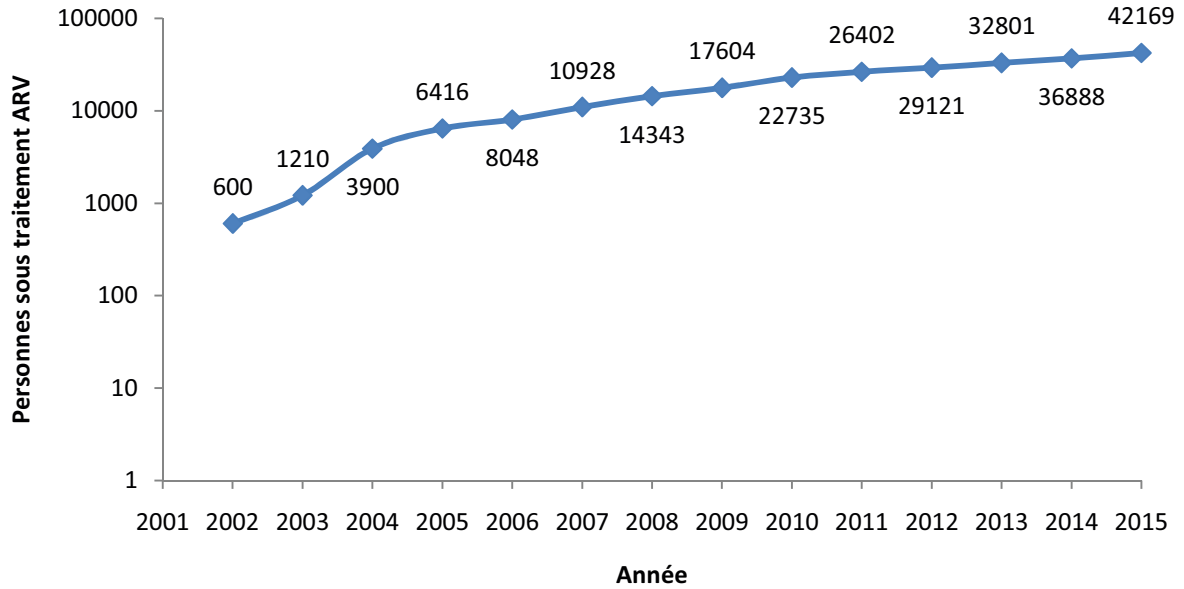


Figure 40 : Évolution du nombre de sites de traitement ARV au Burundi, 2005-2015

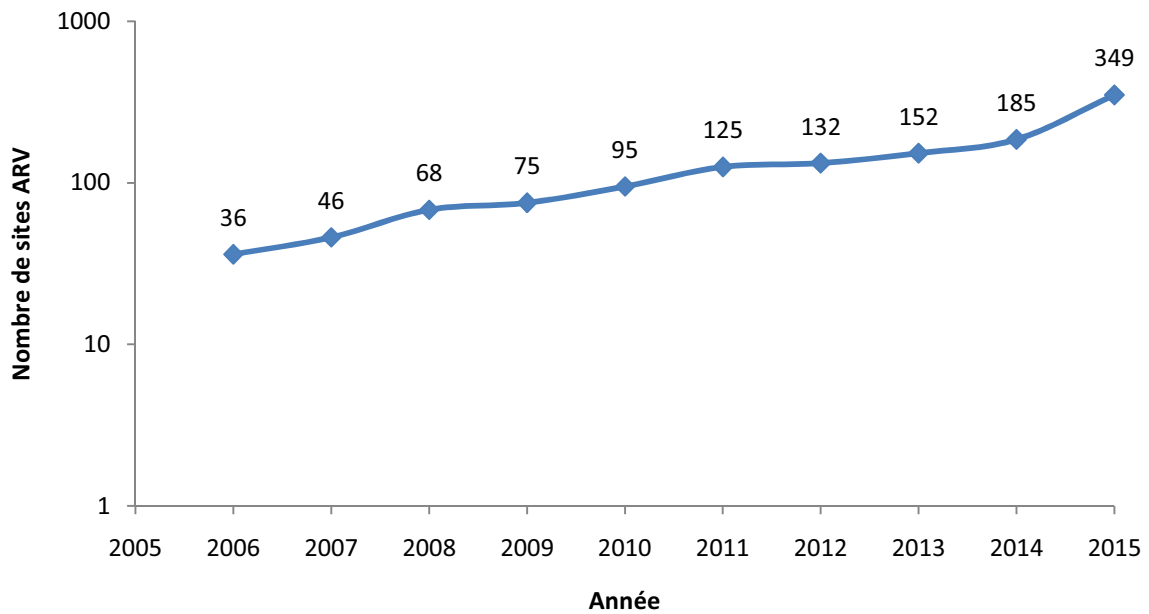
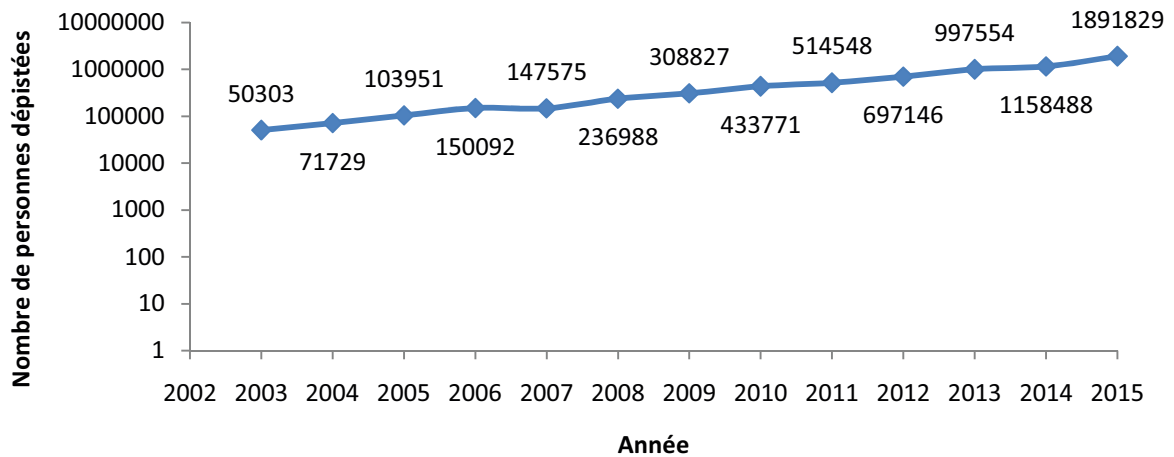
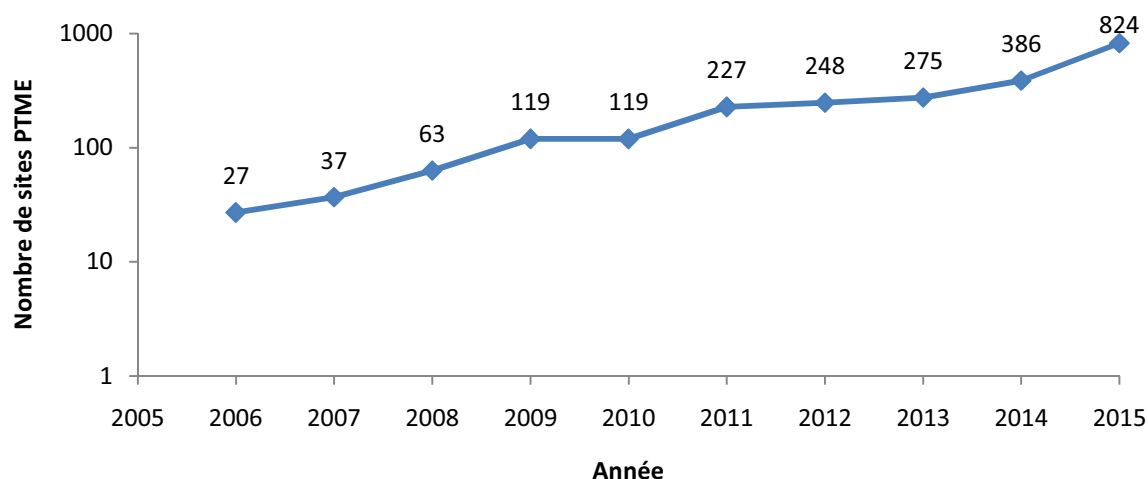
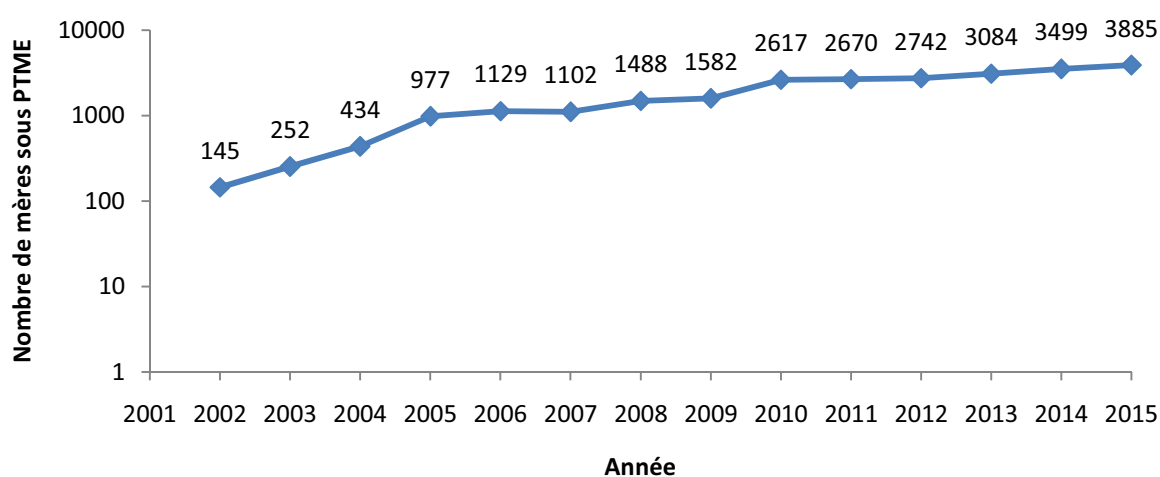


Figure 41 : Évolution du nombre de personnes dépistées au Burundi, 2002-2015

Source : Adapté à partir des données du PNLIS/IST, 2016

À côté des enquêtes nationales et internationales de séroprévalence, les données de prévalence du VIH sont obtenues à travers les sites de surveillance sentinelle chez les femmes enceintes suivies dans le cadre de la PTME. Les données de la PTME permettent de faire des projections de la prévalence du VIH en population générale et ces projections ont donné une prévalence de 1,1 % en 2014 et de 1,0 % en 2015.

De 2006 à 2015, le nombre de sites PTME est passé de 27 à 824 (**Figure 42**) et, de 2002 à 2015, le nombre de mères sous PTME est passé de 145 à 3885 (**Figure 43**).

Figure 42 : Évolution du nombre de sites PTME au Burundi, 2006-2015**Figure 43** : Évolution du nombre de mères sous PTME au Burundi, 2002-2015

Source : Adapté à partir des données du PNLS/IST, 2016

Même si le nombre de PvVIH mises sous traitement ARV a augmenté d'année en année au Burundi, passant par exemple de 1210 en 2003 à 8234 en 2006 et à 18129 en 2009, la couverture thérapeutique reste faible. À la fin de l'année 2013, le nombre de PvVIH recevant un traitement ARV était de 31643, ce qui représente environ la moitié (47,8 %) de la population nécessiteuse. Le pourcentage de PvVIH qui avaient accès au traitement ARV en 2014 reste faible (43,4 %), soit 36888 personnes (2440 enfants et 34448 adultes) sur un nombre total d'adultes et d'enfants éligibles au traitement ARV estimé à 84915.

En 2014, 3499 mères séropositives enceintes ont reçu un traitement ARV durant les 12 derniers mois et le nombre de mères séropositives enceintes qui ont accouché durant les 12 derniers mois était estimé à 4838, ce qui représente un taux de couverture de 72,3 %. Le nombre d'enfants nés des mères séropositives ayant reçu un traitement ARV durant les 6 dernières semaines était de 2304, ce qui représente 47,6 % du nombre total de mères séropositives enceintes qui ont accouché durant les 12 derniers mois [143].

En 2015, 1891829 personnes ont été dépistées pour le VIH. Parmi elles, 1860234 (99,0 %) ont été informées de leur statut VIH. Les résultats des tests VIH ont montré que 29628 personnes avaient un résultat positif, ce qui a conduit à une prévalence du VIH de 1,6 %. Le nombre de PvVIH était estimé à 76541 (44760 femmes et 31781 hommes) en 2015. Au cours de cette année, 42169 PvVIH (27868 femmes et 14301 hommes) ont reçu le traitement ARV, ce qui représente une couverture thérapeutique de 55,1 % (62,3 % chez les femmes, 45,0 % chez les hommes, 60,1 % chez les personnes âgées de 15 ans et plus). De plus, sur 4138 échantillons testés, 85 % avaient une charge virale indétectable.

Alors que le Burundi s'était fixé d'atteindre en 2015 le sixième OMD « Combattre le VIH/SIDA, le paludisme et d'autres maladies » en sa cible A « D'ici 2015, avoir enrayé la propagation du VIH/SIDA et avoir commencé à inverser sa tendance la tendance actuelle », le chemin qui lui restait à parcourir avant la fin de l'année 2015 était encore long. Malgré tous les efforts fournis par le gouvernement et les associations dans la lutte contre le SIDA au Burundi, un long chemin reste à parcourir pour arriver à un accès universel.

2.3.6. Épidémiologie de l'infection par le VIH au Burundi

La première enquête d'envergure nationale de séroprévalence du VIH a été réalisée en 1989-1990 auprès des adultes de 14-44 ans et a montré une prévalence de 0,6 % en zones rurales, 14,7 % en zones semi-urbaines et 15,2 % en zones urbaines.

La deuxième enquête nationale de séroprévalence du VIH a eu lieu en 2002 auprès de 5569 personnes âgées de 12 ans et plus. Les résultats de cette enquête montrent une prévalence globale du VIH de 3,2 % [134] avec une forte

hétérogénéité géographique : une prévalence de 9,5 % en milieu urbain, de 10,4 % en milieu semi-urbain et de 2,5 % en milieu rural.

Une troisième enquête nationale de séroprévalence du VIH a été réalisée en 2007 auprès de 18000 personnes âgées de 18 mois et plus. Elle était combinée à une enquête de surveillance des comportements chez les femmes et les hommes âgés de 15 à 49 ans dans cette même population. Cette enquête a montré une prévalence globale du VIH de 3,0 %. Cette prévalence variait, comme pour les enquêtes de 1989-90 et 2002, selon la place de résidence : 4,6 % en milieu urbain, 4,4 % en milieu semi-urbain et 2,8 % en milieu rural [142].

Une première EDS a été réalisée par le Département de la population en 1987 mais ne comportait ni les coordonnées géographiques, ni la composante VIH (cette dernière a été introduite dans les EDS en 2001). Une deuxième EDS a été réalisée entre le 29 août 2010 et le 30 janvier 2011 par l'Institut de Statistiques et des Études Économiques du Burundi (ISTEEBU) en collaboration avec l'Institut National de Santé Publique (INSP) et l'IRC Macro International auprès de 8088 personnes âgées de 15-59 ans. La prévalence du VIH était estimée à 1,4 % chez 8088 femmes et hommes âgés de 15-49 ans (**Tableau 6**) [133]. Comme pour les enquêtes nationales précédentes, la prévalence du VIH reste plus élevée en milieu urbain (4,1 %) qu'en milieu rural (1,0 %). Ces différences observées globalement restent valables pour les femmes (6,1 % en milieu urbain contre 1,2 % en milieu rural) et les hommes (2,4 % en milieu urbain contre 0,7 % en milieu rural).

Tableau 6 : Prévalence du VIH estimée lors des enquêtes nationales

Année	Global	Sexe		Milieu de résidence		
		Femmes	Hommes	Urbain	Semi-urbain	Rural
1989	-	-	-	15,2 %	14,7 %	0,6 %
2002	3,2 %	2,9 %	2,8 %	9,5 %	10,4 %	2,5 %
2007	3,0 %	-	-	4,6 %	4,4 %	2,8 %
2010	1,4 %	1,7 %	1,0 %	4,1 %	-	1,0 %

À côté des enquêtes nationales et internationales de séroprévalence, les données de prévalence du VIH sont obtenues à travers les sites de surveillance sentinelle

chez les femmes enceintes suivies dans le cadre de la PTME. Les données de la PTME permettent de faire des projections de la prévalence du VIH en population générale et ces projections ont donné une prévalence de 1,0 % [0,8 %-1,3 %] chez les adultes de 15-49 ans [146].

2.4. Contexte de notre recherche au Burundi

Des disparités spatiales de prévalence du VIH ont été rapportées dans la littérature, y compris à un niveau local. Une étude conduite en Afrique du Sud en 2009 a concerné 12221 participants âgés de 15 à 49 ans pour les femmes et de 15 à 54 ans pour les hommes. Cette étude a utilisé deux techniques d'analyse spatiale (méthode à noyaux et scan spatial) pour détecter des variations géographiques de prévalence du VIH et les agrégats ou clusters de basse et de haute prévalence du VIH [147]. Elle a montré trois clusters non disjoints de haute prévalence du VIH dans des endroits de haute prévalence du VIH et trois clusters de faible prévalence du VIH ailleurs. Une étude réalisée en Ouganda en 2014 a utilisé les données de l'EDS 2011 et a concerné uniquement un échantillon de 7518 personnes âgées de 15 à 24 ans. Entre autres résultats, cette étude a identifié un seul cluster de haute prévalence du VIH au centre de l'Ouganda à l'aide de la statistique du scan spatial de Kulldorff. En revanche, peu d'études, à notre connaissance, se sont intéressées à la distribution spatiale de la prévalence du VIH à une échelle plus fine que la région au Burundi.

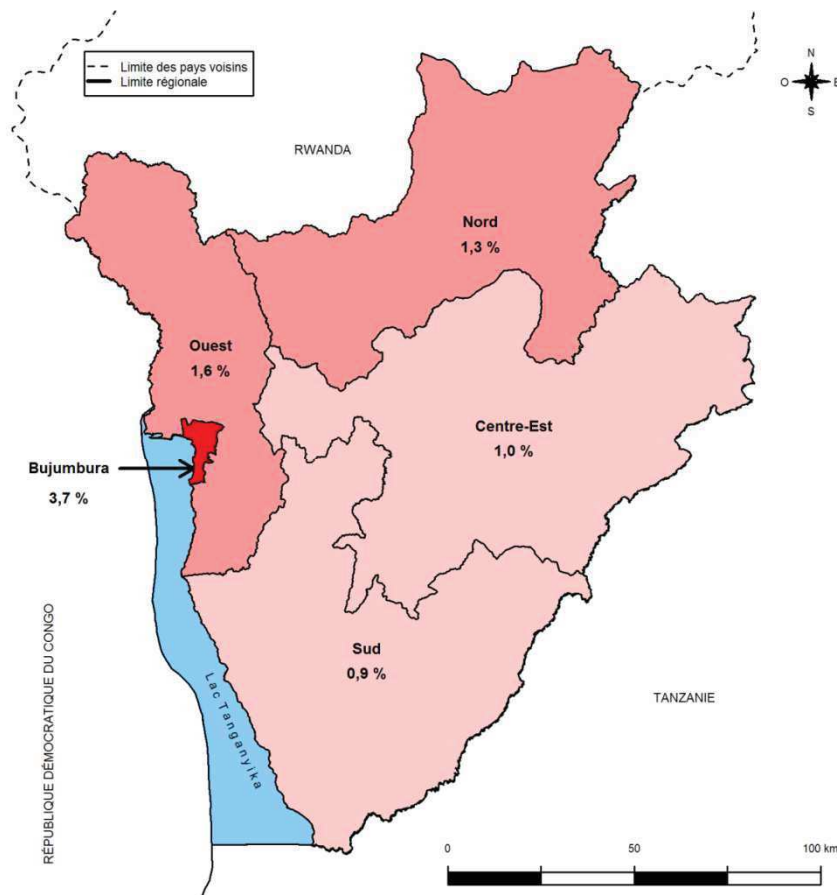
Une d'entre elles, faite à partir des données des EDS conduites dans 20 pays de l'Afrique subsaharienne incluant le Burundi, a détecté un cluster de haute prévalence à Bujumbura [148]. Ce cluster contenait un nombre de cas égal à 12,7 % de la population du cluster. Pour la détection des clusters, une fenêtre de recherche de 100 km de rayon a été utilisée et aucun cluster de faible prévalence du VIH dans la population de 15-59 ans n'a été détectée même en fixant la taille de la fenêtre de recherche à 50 % maximum de la population à risque ($p=0,069$). Les auteurs de cette étude n'ont pas utilisé les poids d'échantillonnage de l'EDS car ils s'intéressaient à la corrélation entre la prévalence nationale du VIH et la force de

l'agrégation spatiale. Cette dernière, exprimée en termes de risque relatif (RR) de l'infection par le VIH dans le cluster versus à l'extérieur du cluster, tendait à être plus élevée dans les pays ayant une faible prévalence nationale du VIH. Par exemple, au Burundi (2010), ce RR vaut 2.65 avec une prévalence nationale du VIH de 1,9 % alors qu'il est de 1,28 au Lesotho (2009), un pays où la prévalence nationale du VIH est estimée à 22,2 %.

Une autre étude qui s'est intéressée à l'analyse spatiale des données de prévalence du VIH au Burundi avait pour objectif d'estimer, au deuxième niveau administratif, c'est-à-dire à l'échelle des 17 provinces, la prévalence du VIH chez les personnes âgées de 15 à 49 ans et le nombre de PvVIH à partir des données de l'EDS 2010 [149]. La prévalence du VIH chez les personnes âgées de 15 à 49 ans et le nombre de PvVIH ont été estimés à l'aide de la méthode à noyau avec une fenêtre adaptative implémentée dans le module `prevR` de R [149,150]. La surface de prévalence produite avec cette méthode est égale au rapport entre la surface d'intensité des cas (définie comme étant le nombre de cas par unité de surface) et la surface d'intensité de la population à risque (définie comme étant la population à risque par unité de surface). La prévalence du VIH chez les personnes âgées de 50 ans et plus a été estimée à l'aide du rapport de prévalence obtenu à partir des estimations de l'ONUSIDA [151].

Au Burundi, les résultats de la dernière EDS conduite en 2010 chez les femmes de 15-49 ans et les hommes de 15-59 ans ont montré que la prévalence du VIH était plus élevée à Bujumbura (3,7 %), la capitale, que dans les autres régions où elle était estimée à 1,6 % dans la région de l'Ouest, 1,3 % dans la région du Nord, 1,0 % dans la région du Centre-Est et 0,9 % dans la région du Sud (**Figure 44**).

Figure 44 : Prévalence du VIH selon les régions au Burundi, 2010



Source : Adapté à partir de l'EDSB-II, 2010

Notre première hypothèse de recherche était qu'il existe des disparités intra-régionales de prévalence du VIH au Burundi. Notre deuxième hypothèse était que le risque d'infection par le VIH variait d'une province à une autre.

2.5. Objectif de l'étude

L'objectif de cette étude était d'évaluer, d'une part, l'hétérogénéité spatiale de l'infection par le VIH et, d'autre part, l'association entre les principales caractéristiques sociales et comportementales des résidents et l'infection par le VIH en tenant compte de cette hétérogénéité spatiale chez les femmes et les hommes âgés de 15-49 ans à l'aide des données de l'EDS 2010 du Burundi. Cette étude

visait à aider les décideurs de santé publique à allouer efficacement les ressources afin de combattre le SIDA au Burundi, en leur apportant des données complémentaires aux données déjà disponibles.

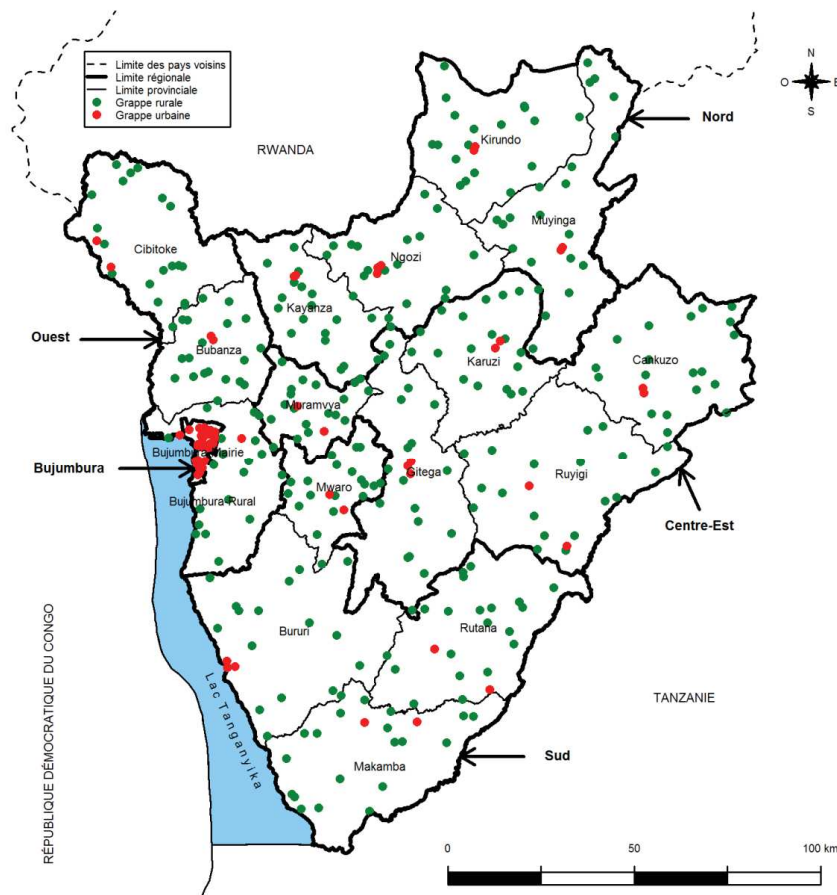
2.6. Méthodologie

2.6.1. Sélection de la population d'étude

Les données utilisées pour cette étude proviennent de la deuxième EDS conduite entre le 29 août 2010 et le 30 janvier 2011 par l'Institut de Statistiques et d'Études Economiques du Burundi (ISTEEBU), l'Institut National de Santé Publique (INSP) et ICF Macro International [132]. Ces données nous ont été fournies par MEASURE DHS, un programme des Nations Unies chargé de mener des EDS dans les pays en développement [152]. L'EDS de 2010 est basée sur un sondage aléatoire en grappes, stratifiée à deux degrés. La stratification se fait d'une part au niveau de la province ($n=17$) et d'autre part au niveau du milieu de résidence (rural versus urbain), soit un total de 33 strates puisque Bujumbura est complètement urbaine. L'échantillon fourni par cette enquête est un échantillon représentatif de la population au niveau national et au niveau régional et, pour certains indicateurs, au niveau provincial. Les régions (sanitaires) sont des regroupements de provinces. Contrairement aux régions, les provinces sont à la fois des entités administratives et des entités sanitaires. Seule Bujumbura-Mairie est en même temps une région et une province.

Au premier degré, 376 grappes (301 rurales et 75 urbaines) ou zones de dénombrement (ZD), soit 5,0 % des grappes du pays, ont été aléatoirement tirées avec une probabilité proportionnelle à leur taille (nombre de ménages) dans une liste de 8104 ZD (7402 rurales et 702 urbaines). Cette liste a été établie à l'occasion du deuxième RGPH de 2008 lors duquel 1695252 ménages (1538634 ruraux et 156618 urbains) ont été identifiés. Une grappe est une agglomération de ménages créée pour les besoins du recensement et peut être un village ou un ensemble de villages. Ces grappes (urbaines et rurales) sont représentées sur la **Figure 45**.

Figure 45 : Répartition géographique des grappes de l'EDS 2010 du Burundi



Source : A partir des données de l'EDSB-II, 2010

Au deuxième degré, un échantillon de 9024 ménages (7224 ruraux et 1800 urbains) a été systématiquement sélectionné dans la liste des 376 grappes avec une probabilité égale, soit 24 ménages par grappe. Que ce soit au premier ou au deuxième degré, les grappes ou les ménages ont des probabilités d'être tirés au sort et ces probabilités sont utilisées pour corriger la surestimation ou la sous-estimation de l'échantillon afin d'assurer sa représentativité. Ainsi, si P_{1hi} est la probabilité de tirer la grappe i au premier degré dans la strate h et P_{2hi} la probabilité de tirer la grappe i au deuxième degré dans la strate h , alors la probabilité de sélection du ménage de la grappe i dans la strate h vaut $P_{hi} = P_{1hi} \times P_{2hi}$. Le poids

d'échantillonnage associé à chaque ménage vaut alors l'inverse de la probabilité de sélection du ménage de la grappe i dans la strate h , soit $W_{hi} = \frac{1}{P_{hi}}$.

Les strates avec de grands effectifs seront sous-échantillonnées, d'où des faibles poids d'échantillonnage leur seront attribués et les strates avec des faibles effectifs seront sur-échantillonnées, d'où des poids d'échantillonnage forts leur seront attribués. Il convient alors de corriger ces poids d'échantillonnage en les ajustant pour le taux de réponse TR des ménages, ce qui donne $W_{hi}^* = \frac{W_{hi}}{TR}$.

Comme ces poids d'échantillonnage sont exprimés en termes de milliards dans les bases de données de l'EDS, il convient de les diviser par un million avant de les utiliser et de vérifier si leur somme vaut alors le nombre d'individus enquêtés.

Trois questionnaires ont été utilisés pour recueillir les données : un questionnaire « Ménages », un questionnaire « Femmes » et un questionnaire « Hommes ».

Sur 8677 ménages identifiés, 8596 ont été effectivement enquêtés, soit un taux de réponse de 99 %. Pour l'enquête individuelle, sur 9737 femmes âgées de 15-49 ans et 4592 hommes âgés de 15-59 ans éligibles, 9389 femmes et 4280 hommes ont été respectivement enquêtés, soit un taux de réponse respectif de 96 % et 93 %. L'enquête-hommes a été menée dans 50% des ménages sélectionnés pour l'enquête-femmes.

Les personnes éligibles au test VIH étaient celles qui résidaient ou qui avaient passé la nuit précédant l'enquête dans le ménage et qui ont accepté d'être testées pour le VIH. Ce test s'est fait dans les ménages sélectionnés pour l'enquête-hommes. Une note informative concernant l'obtention des conseils et des tests volontaires gratuits auprès des services VIH a été remise aux personnes éligibles au test. Les personnes âgées de 18 ans et plus ainsi que les représentants légaux des mineurs ont fourni leur consentement écrit pour participer à l'enquête. En outre, les mineurs de moins de 18 ans ont fourni leur assentiment (expression de l'accord du mineur

n'ayant pas la même valeur juridique que le consentement des majeurs). Le comité national d'éthique pour la protection des êtres humains du Burundi ainsi que le comité d'éthique de l'ICF Macro International ont approuvé ce protocole.

Le test VIH s'est fait en deux étapes : un premier test ELISA (enzyme-linked immunosorbent assay) a été utilisé pour tester la séropositivité des échantillons de sang et les échantillons qui se sont avérés positifs ainsi que 10 % des échantillons négatifs (pour le contrôle-qualité) ont été soumis à un second test ELISA. Ensuite, tous les échantillons ayant des résultats discordants ont été testés pour confirmation à l'aide d'un test Inno-Lia HI I/II score. Sur un total de 9503 personnes (4911 femmes et 4592 hommes) éligibles pour le test VIH, 90,4 % ont été interviewées. Parmi les 8588 personnes (4533 femmes de 15-49 ans et 4055 hommes de 15-59 ans) pour lesquelles le résultat du test VIH est connu (positif, négatif ou indéterminé), nous avons exclu 501 hommes âgés de 50-59 ans ainsi qu'une femme dont le statut VIH était indéterminé. Notre échantillon est donc composé de 8086 personnes (4532 femmes et 3554 hommes) âgées de 15-49 ans observées dans 3816 ménages issus des 376 grappes.

2.6.2. Données géographiques

Les données géographiques indiquent notamment la position (longitude/latitude) de la grappe dans laquelle les individus ont été interviewés, le milieu de résidence (urbain/rural, région), l'altitude (en m) de la grappe et le type de projection (WGS84). Les contours du pays, les entités administratives de premier niveau (régions) dans lesquelles les estimations de l'EDS sont valables sont obtenues en fusionnant les entités administratives de deuxième niveau (provinces) selon leur appartenance dans les régions. Hormis les données géographiques décrivant les coordonnées géographiques des grappes qui sont accessibles sur le site de MEASURE DHS, les autres données géographiques sont gratuitement accessibles sur le site <http://www.diva-gis.org/gdata> ou <http://www.maplibrary.org/>.

Pour les EDS réalisées après 2008, les grappes sont déplacées de façon à ce qu'elles restent dans les entités administratives de deuxième niveau, c'est-à-dire dans les entités géographiques directement inférieures aux régions (les provinces

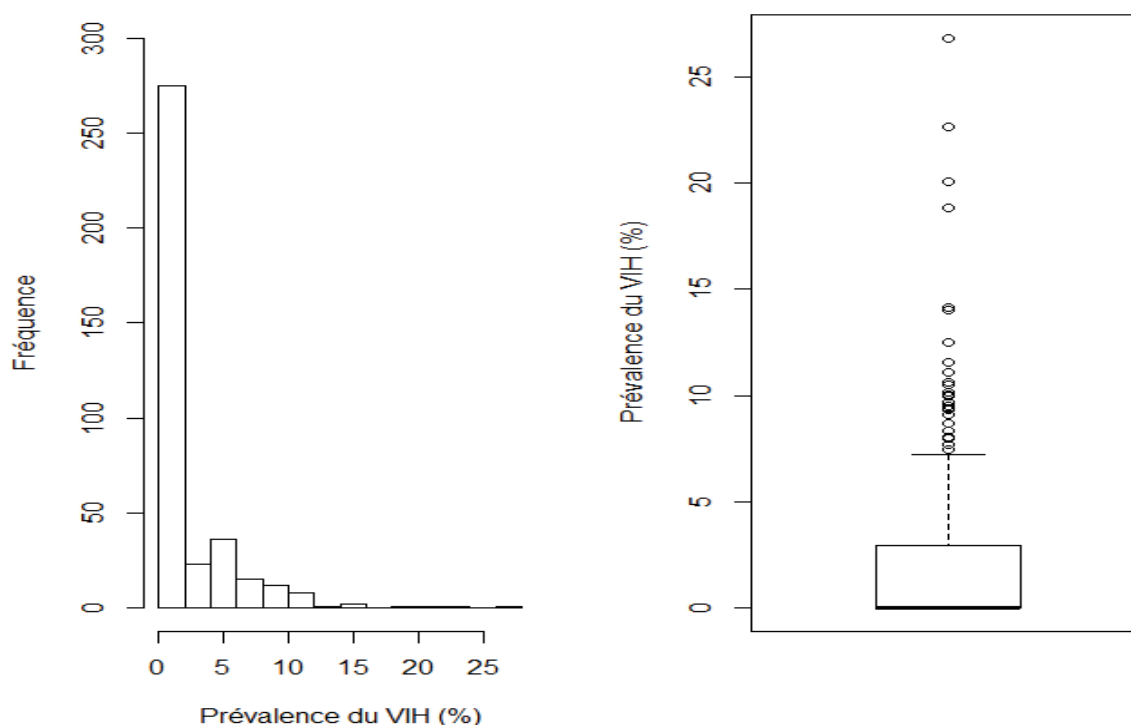
pour le Burundi) [153]. Au Burundi, le codage des niveaux administratifs dans le domaine sanitaire est 0 (Burundi), 1 (Régions), 2 (Provinces), 3 (Districts). Pour l'EDS de 2010 du Burundi, les grappes ont été déplacées de façon qu'elles restent dans les entités administratives (au niveau sanitaire) de premier niveau, c'est-à-dire à l'intérieur des régions. Le déplacement des grappes, surtout celles qui se trouvent tout près des limites régionales, peut donc biaiser les résultats puisqu'elles peuvent se retrouver dans une région avoisinante. Les grappes et les contours du pays appartiennent à des systèmes de projection différents. Les données géographiques externes aux bases de données de l'EDS et utilisées pour cette étude ont été mises dans le même système de projection WGS84 que celui des données géographiques de l'EDS.

La prise des coordonnées géographiques à l'aide d'un Système de Positionnement Global (GPS, Global Positioning System) lors du recueil des données des EDS n'est devenue standard que dans les années 2000. Les coordonnées géographiques (longitude et latitude) sont mesurées au centre des grappes.

2.6.3. Analyses statistiques

Nous avons ré-analysé les données de l'EDS 2010 du Burundi. Nous avons d'abord calculé le nombre de cas d'infection par le VIH et la prévalence de celle-ci globalement puis selon les régions/provinces et selon les caractéristiques sociales et comportementales des résidents âgés de 15-49 ans. Nous avons ensuite cartographié cette prévalence selon les provinces afin de nous rendre compte de l'hétérogénéité spatiale de la distribution de l'infection par le VIH dans les provinces.

Nous avons d'abord calculé la prévalence du VIH par grappe avec le logiciel Stata, version 11 [154]. Cette prévalence était comprise entre 0,0 % (275 grappes sur 376 ont une prévalence nulle) et 26,8 % (**Figure 46**).

Figure 46 : Histogramme et boxplot de la prévalence du VIH par grappe

Source : A partir des données de l'EDSB-II, 2010

Nous avons ensuite calculé la prévalence du VIH par province afin de vérifier l'hétérogénéité spatiale (**Tableau 7**).

Nous avons ensuite fait une analyse de l'autocorrélation spatiale à l'aide du test global de Moran. Les poids spatiaux étaient donnés par les inverses des distances entre les grappes. L'auto-corrélation spatiale étant significative (Moran's $I = 0,01$; $p = 0,021$), nous avons ensuite interpolé les données de prévalence du VIH à l'aide d'une grille régulière de 1000×1000 cellules (856000 points) et de la méthode des noyaux [80,150]. L'interpolation spatiale et la cartographie ont été effectués sous le logiciel R version 3.2.1 [155]. L'interpolation spatiale a été réalisée à l'aide du noyau gaussien décrit dans le module `prevR` de R [150,156]. Le noyau gaussien a été choisi car il n'impose pas que les limites géographiques des zones (provinces ici) dans lesquelles la variable à interpoler est définie soient statiques et produit des estimations de prévalence qui sont robustes aux effets de bruit et sensibles aux fluctuations locales [147].

Tableau 7 : Prévalence du VIH par province chez les résidents âgés de 15 à 49 ans au Burundi en 2010

Région	Province	VIH+	VIH-	N	%	IC 95%
Centre-Est	Ruyigi	0	361	362	0,1	0,0-0,4
	Gitega	3	587	590	0,6	0,0-1,4
	Karuzi	5	434	439	1,1	0,2-2,0
	Muramvya	8	305	313	2,4	1,0-3,8
	Cankuzo	2	207	209	1,0	0,0-1,9
Ouest	Bubanza	3	376	379	0,9	0,0-1,9
	Cibitoke	5	484	489	1,1	0,0-2,2
	Bujumbura-Rural	13	479	493	2,7	0,9-4,6
Nord	Muyinga	5	606	610	0,8	0,0-1,8
	Kayanza	6	554	560	1,1	0,2-2,0
	Kirundo	8	540	547	1,4	0,2-2,5
	Ngozi	13	599	613	2,2	0,5-3,8
Sud	Makamba	2	473	474	0,3	0,0-0,7
	Bururi	6	648	654	0,9	0,0-1,8
	Mwaro	4	285	288	1,3	0,0-2,7
	Rutana	4	330	334	1,3	0,2-2,4
Bujumbura	Bujumbura-Mairie	27	704	731	3,7	2,5-4,9
Burundi		114	7972	8086	1,4	1,1-1,7

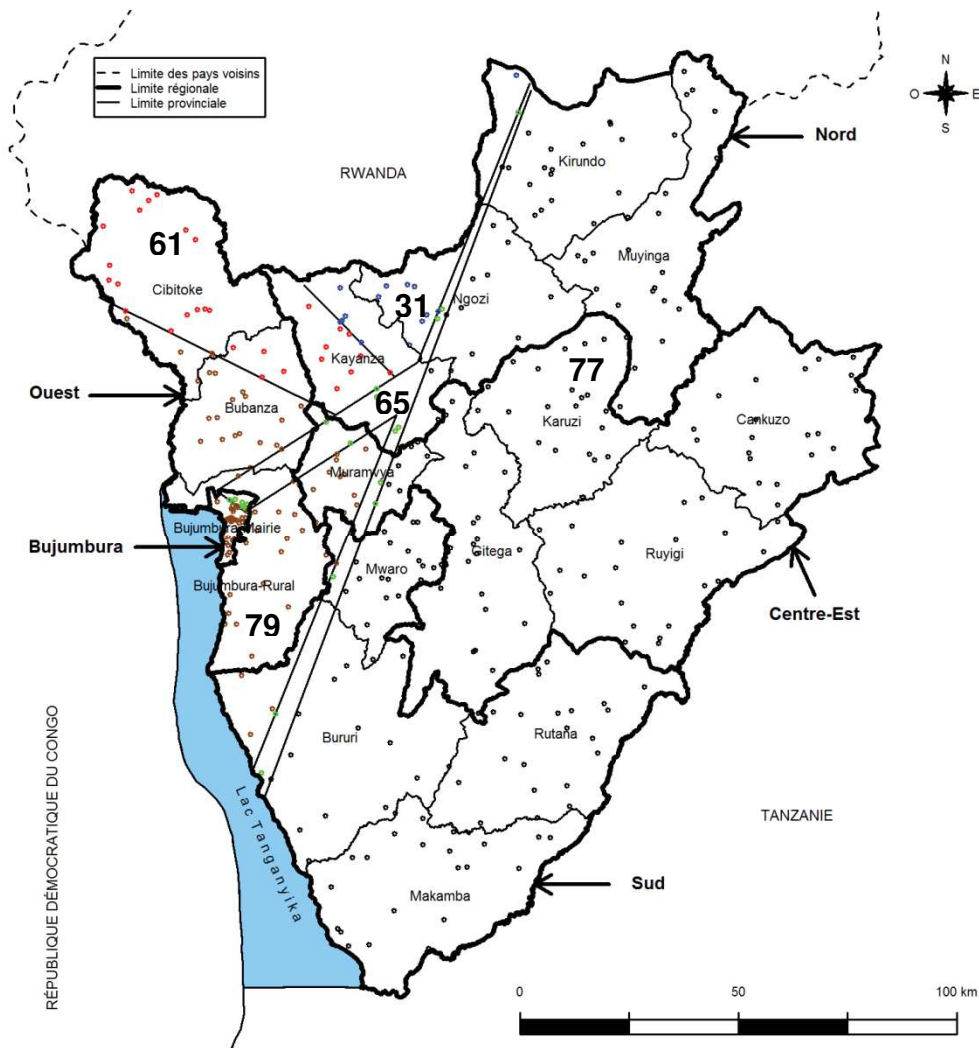
* Les effectifs pondérés ont été arrondis.

Nous avons ensuite détecté les agrégats de basse et de haute prévalence du VIH à l'aide de la méthode de Kulldorff basée sur le scan spatial [157].

En plus du scan spatial de Kulldorff, nous avons appliqué l'arbre de décision oblique (spODT : spatial oblique decision tree) pour partitionner le Burundi en zones de différents risques en les délimitant par des segments de droites [59]. Nous avons, au départ, choisi de le partitionner en 5 zones (**Figure 47**), le nombre maximal de zones à créer étant borné supérieurement par le nombre de distances non nulles entre les grappes. Ce nombre correspond au nombre de régions sanitaires. Nous avons, en

outre, choisi de mettre en référence la zone dans laquelle le cluster de haute prévalence du VIH a été détecté.

Figure 47 : Partitionnement oblique du Burundi en zones de différents risques



Nous avons fait dix essais en changeant les paramètres et nous avons gardé le modèle avec le R^2 le plus important. Nous avons également calculé le nombre de cas (VIH positif) et de témoins (VIH négatif) à l'intérieur, puis à l'extérieur de chaque zone. Le rapport de cotes (RC) était alors estimé par le rapport des produits croisés du nombre de cas dans une zone et du nombre de témoins à l'extérieur de la zone (numérateur) et du nombre de cas à l'extérieur de la zone et du nombre de témoins dans la zone, en considérant la zone englobant le cluster de haute prévalence du VIH comme zone de référence. Quant à la variance de l'estimateur du RC, il était

calculé comme étant la racine carrée de la somme des inverses des nombres étant intervenu dans le calcul du RC. Les bornes des intervalles de confiance à 95 % des RC ont été calculées par exponentiation de la différence (la somme) entre le logarithme népérien du RC et le produit de la variance par le quantile de Student ($t=1,96$). Les résultats de la comparaison des risques dans les zones obtenues lors du découpage du Burundi en zones de risques différents sont présentés dans le **tableau 8**.

Tableau 8 : Comparaison des risques dans les zones

Zone	Nombre de grappes	OR	IC 95 %
79	89	1,00	
61	29	0,06	0,01-0,45
77	220	0,31	0,20-0,48
31	14	1,55	0,84-2,87
65	24	2,20	1,31-3,70

La zone 79 recouvre la quasi-totalité de Bujumbura-Mairie (le cluster de haute prévalence englobait cette province) et les parties des provinces de Bujumbura-Rural, Mwaro, Muramvya, Bubanza et Cibitoke. En mettant cette zone en référence, la prévalence du VIH était significativement plus faible dans les zones 61 et 77 et significativement plus élevée dans la zone 65.

Enfin, nous avons effectué une modélisation spatiale à l'aide de la régression logistique spatiale à l'échelle de l'individu. Nous avons regroupé les modalités des variables pour lesquelles le nombre de cas de VIH était faible par rapport à celui des autres modalités ou les modalités avec aucun nombre de cas de VIH (si le regroupement était cohérent). Nous avons utilisé des modèles additifs de régression structurés et estimé les paramètres des modèles à l'aide du maximum de vraisemblance restreinte (MVR) du modèle mixte et de l'estimation empirique bayésienne. L'estimation par la méthode du maximum de vraisemblance (MV) ne tient pas compte de la perte de degrés de liberté engendrée par la prise en compte des effets fixes. La méthode du MVR permet alors de corriger les biais des

estimateurs de covariance. Pour les modèles spatiaux univariés, nous avons retenu un seuil de décision de $p < 0,20$ pour décider quelles variables mettre dans le modèle multivarié spatial complet.

Nous avons utilisé le Critère de l'Information Bayésienne (BIC – Bayesian Information Criterion) pour sélectionner le meilleur modèle. Nous avons rapporté les rapports de cotes *a posteriori* et leurs intervalles de crédibilité à 95 % ainsi que des *p*-values pour leurs significativités. Nous avons également comparé les risques d'infection par le VIH dans les provinces en faisant une régression logistique non spatiale et en mettant la province de Ruyigi qui a la plus faible prévalence du VIH comme catégorie de référence.

L'association entre le statut VIH et les caractéristiques sociales et comportementales des personnes a été évaluée à l'aide d'un modèle spatial binomial en ajustant sur les effets aléatoires spatiaux et non spatiaux des provinces avec le logiciel BayesX, version 2.1 [106].

2.6.4. Explicitation du modèle spatial utilisé dans l'étude

La variable réponse Y_{is} qui désigne le statut VIH (positif ou négatif) d'un individu i de la province s suit une loi de Bernoulli :

$$Y_{is} \sim \text{Bernoulli}(1, p_{is}) \text{ avec } p_{is} = E[Y_{is}] = P(Y_{is} = 1) \text{ une probabilité inconnue} \quad (2.1)$$

Le modèle de régression logistique binaire géo-additive convient pour établir des associations entre le statut VIH et une série de variables explicatives qualitatives Q_{i1}, \dots, Q_{iq} ou quantitatives X_{i1}, \dots, X_{ic} [91,92].

Dans notre étude, toutes les variables explicatives sont catégorielles. L'équation (1.47) du premier chapitre peut alors s'écrire [91,158]:

$$\log it(p_{is}) = \beta_0 + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + f_{spat}(\text{Province}_s) \quad (2.2)$$

où β_i sont les paramètres du modèle, X_i les variables indépendantes et où l'effet aléatoire est mis sur la province. Dans ce modèle, les variables indépendantes sont des effets fixes.

Le modèle logistique contenant un effet aléatoire qui désigne la province et les variables catégorielles X_j^{cat} sans le terme d'erreur s'écrit :

$$\log it(p_{is}) = \beta_0 + \sum_{j=1}^{v_1} \beta_j X_{ij}^{cat} + f_{spat}(\text{Province}_s) \quad (2.3)$$

où $f_{spat}(\text{Province}_s)$ est une fonction spatiale non linéaire des provinces et des effets aléatoires, p_{is} la probabilité, pour un individu i donné résidant dans la province s d'être de statut VIH+, étant donné que les individus sont distribués indépendamment l'un de l'autre dans la province, $\beta = (\beta_0, \beta_1, \dots, \beta_{v_1})^t$ le vecteur des paramètres associés aux variables catégorielles. Un a priori non informatif est mis sur les effets fixes, ce qui veut dire que les effets fixes sont constants :

$$p(X_j^{cat}) \propto \text{constante} \quad (2.4)$$

et les paramètres associés aux effets fixes sont gaussiens.

Aux variances inconnues τ_j^2 , il leur est assigné des lois a priori Gamma-inverse, conjuguées à la loi des β_j : $\tau_j^2 \sim IG(a_j, b_j)$ de paramètres $a_j, b_j > 0$ avec des valeurs par défaut $a_j = b_j = 0,001$ pour un a priori non informatif de densité [159] :

$$p(\tau_j^2) \propto \frac{1}{(\tau_j^2)^{a_j+1}} \exp\left(-\frac{b_j}{\tau_j^2}\right) \quad (2.5)$$

Différentes valeurs de a_j et b_j ont été utilisées en analyse de sensibilité. Cette analyse n'a pas montré de différences avec les résultats obtenus avec les paramètres par défaut, sachant que les valeurs de ces paramètres ont une faible influence sur la distribution a posteriori des paramètres [159]. L'effet aléatoire se décompose en une somme de l'effet aléatoire spatial $f_{struc}(\cdot)$, effet dit « structuré » qui tient compte de l'effet spatiale «Province » et d'un effet aléatoire non spatial $f_{unstruc}(\cdot)$, dit « non structuré » qui tient compte du fait que dans le modèle il y a des effets fixes :

$$f_{spat}(\text{Province}_s) = f_{struc}(\text{Province}_s) + f_{unstruc}(\text{Province}_s) \quad (2.6)$$

Les effets aléatoires spatiaux $f_{struc}(\text{Province}_s)$ sont modélisés comme des champs aléatoires de Markov (une généralisation d'une marche aléatoire de premier ordre) défini par la matrice des voisinages/proximité entre les provinces et non par la matrice des distances entre les grappes. La valeur du paramètre est donc constante dans une province donnée. Deux provinces sont voisines si elles ont une limite commune. Les effets aléatoires non spatiaux $f_{unstruc}(\text{Province}_s)$ sont, quant à eux, modélisés comme des effets aléatoires gaussiens.

Une loi a priori est définie sur les paramètres du modèle, appelés aussi effets spatiaux des provinces, de façon à ce que ces paramètres soient conditionnellement indépendants et normalement distribués avec une espérance mathématique

conditionnelle égale à la moyenne des effets des provinces voisines et une variance inversement proportionnelle au nombre N_s de provinces s voisines :

$$f_{spat}(s | s') = \beta_s | \beta_{s'}, s \neq s', \tau^2 \sim N\left(\frac{1}{N_s} \sum_s \beta_{s'}, \frac{\tau^2}{N_s}\right) \quad (2.7)$$

où τ^2 est la variance qui est un paramètre de lissage spatial.

Les effets fixes sont le sexe, l'âge (en 6 classes telles que définies dans l'EDS), le niveau d'éducation codé en deux catégories (aucune, primaire ou plus), le statut marital (célibataire, marié, concubin, veuf, divorcé ou séparé), la religion (catholique, protestant, adventiste, musulman, témoin de Jéhovah, autre, aucune), le quintile du bien-être économique codé en deux catégories (riche, autre), l'activité sexuelle récente durant les 4 dernières semaines codée en 3 catégories (aucune, active durant les 4 dernières semaines, non active durant les 4 dernières semaines), le nombre de partenaires sexuels autre que le conjoint durant les 12 derniers mois (0, 1, ≥ 2) et les infections sexuellement transmissibles (IST) durant les 12 derniers mois codées en deux catégories (non, oui). La variable qui indique la province n'a pas été utilisée dans un modèle spatial mais plutôt dans un modèle de régression logistique simple afin de comparer les prévalences du VIH observées dans les provinces par rapport à la province de Ruyigi, la province avec la plus faible prévalence du VIH.

Nous avons d'abord comparé un modèle avec uniquement un effet spatial (Provinces) et un modèle avec un effet aléatoire (Provinces) décomposé en une somme d'un effet structuré $f_{struc}(Province_s)$ et d'un effet non structuré $f_{unstruc}(Province_s)$.

Ces modèles peuvent s'écrire :

$$\eta_1 = \beta_0 + f_{struc}(Province_s) \quad (2.8)$$

$$\eta_2 = \beta_0 + f_{struc}(Province_s) + f_{unstruc}(Province_s) \quad (2.9)$$

Le premier modèle permet d'évaluer les variations géographiques de l'infection par le VIH dues uniquement à l'effet spatial « Provinces », c'est-à-dire uniquement au fait que les individus appartiennent à une province donnée, sans tenir compte de leurs autres caractéristiques décrites plus haut. Le deuxième modèle ajuste sur l'effet aléatoire non spatial « Provinces » et permet donc d'évaluer les variations géographiques de l'infection par le VIH dues à l'effet conjoint d'un effet spatial et d'un effet aléatoire liés à la province d'appartenance des individus.

Nous avons comparé ces deux modèles à l'aide d'un BIC conditionnel (cBIC). Le premier modèle a un cBIC égal à 2252.29 et le deuxième modèle un cBIC égal à 2240.95. Au vu de ces résultats, nous avons retenu le modèle avec un effet structuré et un effet non structuré, et cela pour tous les autres modèles.

Nous avons ensuite effectué des analyses spatiales univariées et les variables significatives au seuil de $p < 0.20$ ont été mises dans le modèle complet. Les individus présentant des observations manquantes ont été exclus de l'analyse.

Les 9 modèles spatiaux univariés sont donc :

$$\eta_3 = \beta_0 + \beta_1 \text{Sexe} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.10)$$

$$\eta_4 = \beta_0 + \beta_1 \text{Age} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.11)$$

$$\eta_5 = \beta_0 + \beta_1 \text{Education} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.12)$$

$$\eta_6 = \beta_0 + \beta_1 \text{Marital} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.13)$$

$$\eta_7 = \beta_0 + \beta_1 \text{Religion} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.14)$$

$$\eta_8 = \beta_0 + \beta_1 \text{Quintile} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.15)$$

$$\eta_9 = \beta_0 + \beta_1 \text{Activité} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.16)$$

$$\eta_{10} = \beta_0 + \beta_1 \text{Partenaire} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.17)$$

$$\eta_{11} = \beta_0 + \beta_1 \text{IST} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \quad (2.18)$$

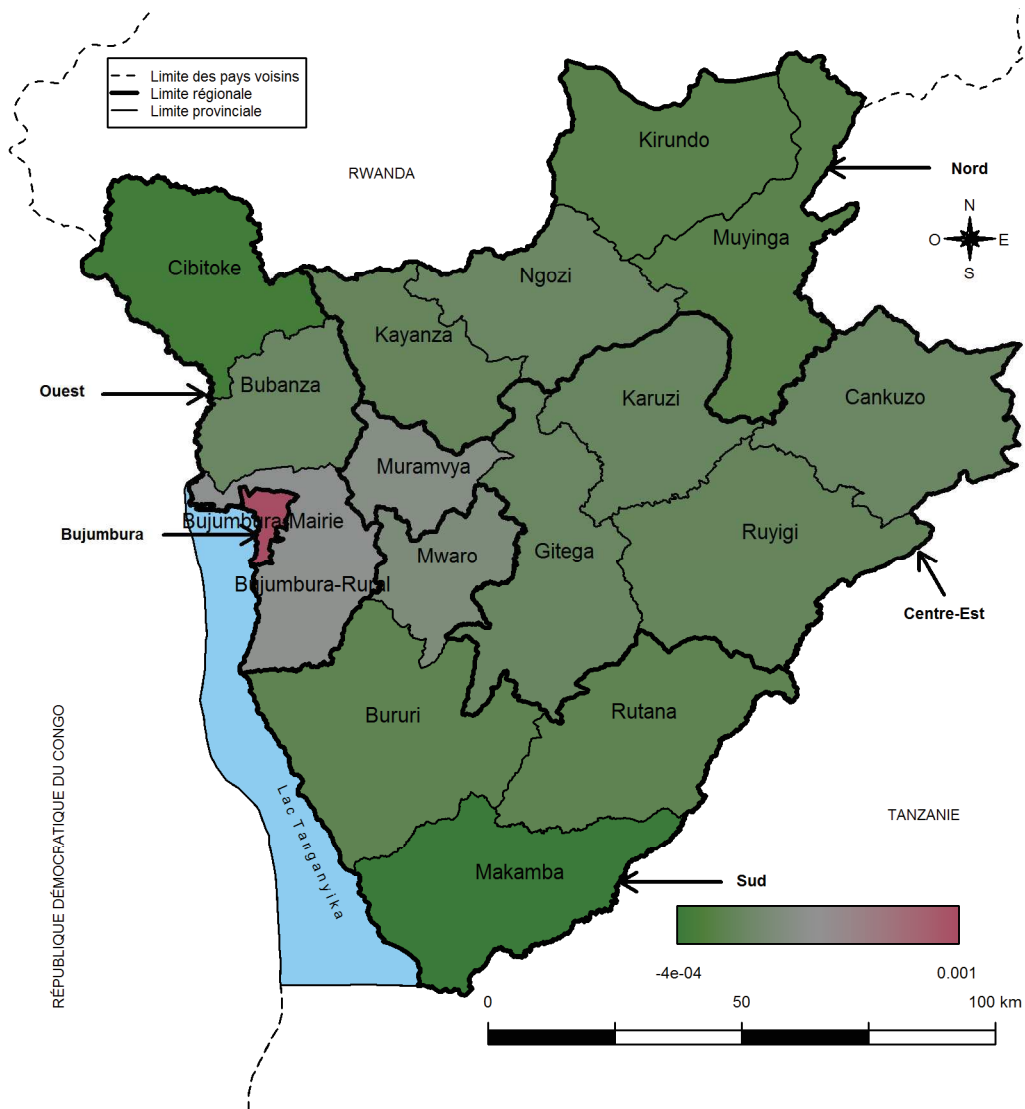
Les variables retenues dans le modèle complet sont le sexe, l'âge, le niveau d'éducation, le statut marital, le quintile du bien-être économique, l'activité sexuelle durant les 4 dernières semaines, et les IST.

Le modèle final s'écrit alors :

$$\begin{aligned} \eta_3 = & \beta_0 + \beta_1 \text{Sexe} + \beta_2 \text{Age} + \beta_3 \text{Education} + \beta_4 \text{Marital} + \beta_5 \text{Quintile} \\ & + \beta_6 \text{Activité} + \beta_7 \text{IST} + f_{\text{struc}}(\text{Province}_s) + f_{\text{unstruc}}(\text{Province}_s) \end{aligned} \quad (2.19)$$

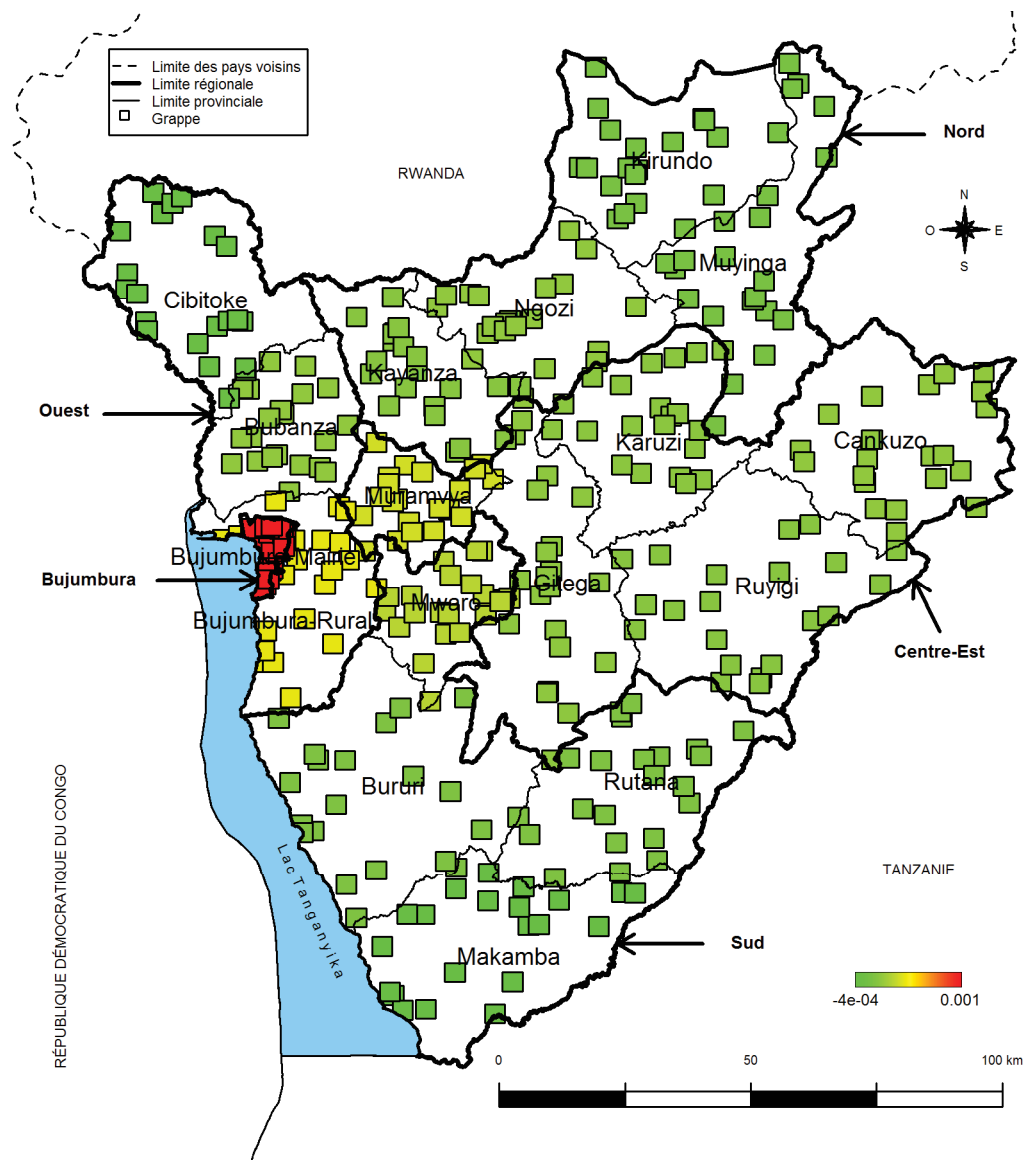
Pour tous ces modèles, chaque paramètre est un vecteur de paramètres à l'exception de la constante. Le cBIC de ce modèle vaut 2103.29. L'estimation des paramètres était basée sur les simulations de MCMC et le maximum de vraisemblance restreinte. L'inférence sur les paramètres était empirique bayésienne. Les résultats sont fournis sous forme d'un tableau contenant, pour chaque paramètre y compris l'intercept, le mode a posteriori, la déviation standard, la p-value et l'intervalle de crédibilité des modes des distributions a posteriori des paramètres. La **Figure 48** montre la distribution des modes a posteriori des effets spatiaux ajustés des provinces.

Figure 48 : Modes a posteriori des effets spatiaux



Cette figure montre que, en ajustant sur les déterminants sociodémographiques et comportementaux, le risque d'infection par le VIH est plus élevé dans les provinces de Bujumbura-Mairie, Bujumbura-Rural, Muramvya et Mwaro et plus faible dans les provinces de Makamba, Cibitoke, Musinga et Kirundo. Les provinces où le risque est élevé sont frontalières de Bujumbura (où la prévalence du VIH est élevée) ou bien lui sont reliées par des routes nationales caractérisées par une circulation intense des populations, des biens et des services. La **figure 49** montre la distribution des modes a posteriori des paramètres.

Figure 49 : Modes a posteriori des paramètres



2.7. Article « Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention »

Un article portant sur cette étude a été publié en 2016 dans BMC Public Health.

RESEARCH ARTICLE

Open Access



Spatial analysis of HIV infection and associated individual characteristics in Burundi: indications for effective prevention

Emmanuel Barankanira^{1,2,5*}, Nicolas Molinari³, Théodore Niyongabo⁴ and Christian Laurent²

Abstract

Background: Adequate resource allocation is critical in the battle against HIV/AIDS, especially in Africa. The determination of the location and nature of HIV services to implement must comply with the geographic, social and behavioral characteristics of patients. We therefore investigated the spatial heterogeneity of HIV prevalence in Burundi and then assessed the association of social and behavioral characteristics with HIV infection accounting for the spatial heterogeneity.

Methods: We used data from the 2010 Demographic and Health Survey. We analyzed these data with a geostatistical approach (which takes into account spatial autocorrelation) by i) interpolating HIV data using the kernel density estimation, ii) identifying the spatial clusters with high and low HIV prevalence using the Kulldorff spatial scan statistics, and then iii) performing a multivariate spatial logistic regression.

Results: Overall HIV prevalence was 1.4 %. The interpolated data showed the great spatial heterogeneity of HIV prevalence (from 0 to 10 %), independently of administrative boundaries. A cluster with high HIV prevalence was found in the capital city and adjacent areas (3.9 %; relative risk 3.7, $p < 0.001$) whereas a cluster with low prevalence straddled two southern provinces (0 %; $p = 0.02$). By multivariate spatial analysis, HIV infection was significantly associated with the female sex (posterior odds ratio [POR] 1.36, 95 % credible interval [CrI] 1.13-1.64), an older age (POR 1.97, 95 % CrI 1.26-3.08), the level of education (POR 1.50, 95 % CrI 1.22-1.84), the marital status (POR 1.86, 95 % CrI 1.23-2.80), a higher wealth index (POR 2.11, 95 % CrI 1.77-2.51), the sexual activity (POR 1.76, 95 % CrI 1.04-2.96), and a history of sexually transmitted infection (POR 2.03, 95 % CrI 1.56-2.64).

Conclusions: Our study, which shows where and towards which populations HIV resources should be allocated, could help national health policy makers develop an effective HIV intervention in Burundi. Our findings support the strategy of the Joint United Nations Programme on HIV/AIDS (UNAIDS) for country-specific, in-depth analyses of HIV epidemics to tailor national prevention responses.

Keywords: HIV, Prevalence, Heterogeneity, Spatial, Factors

* Correspondence: baremma2002@yahoo.fr

¹Département des Sciences Naturelles, Ecole Normale Supérieure, Bujumbura, Burundi

²TransVIHMI, IRD UMI 233 / INSERM U 1175 / Université de Montpellier, Montpellier, France

Full list of author information is available at the end of the article



Background

Adequate resource allocation is critical in the battle against HIV/AIDS, especially in sub-Saharan African countries where 70 % of the people living with HIV worldwide currently reside [1]. These countries face major financial constraints, shortages of healthcare workers, and poorly developed healthcare systems [1, 2]. The determination of the location and nature of HIV services to implement must therefore be made according to the state of the epidemic and to the geographic, social and behavioral characteristics of patients. The *'Know your epidemic'* strategy of the Joint United Nations Programme on HIV/AIDS (UNAIDS) underlines the need for country-specific, in-depth analyses of HIV epidemic features to tailor national prevention responses to the people most at risk [3].

Because half HIV cases only are diagnosed in sub-Saharan Africa [4], most countries rely on Demographic and Health Surveys (DHS) to estimate HIV prevalence along with other data including social and behavioral characteristics [5]. DHS are population-based surveys and use a standardized methodology. They are performed in large, random, clustered samples of people. Since 2001, DHS include informed, anonymous, and voluntary HIV testing in adult women and men. However, a major limitation of crude DHS estimations is that the spatial heterogeneity of HIV prevalence and the hotspots of the disease are hardly examined, although aggregated regional HIV prevalence data can mask large intra-regional differences. In addition, the spatial heterogeneity of HIV prevalence has seldom been taken into account when examining the factors associated with HIV infection.

By contrast to standard statistical tools, spatial analysis methods allow investigating the spatial heterogeneity and identifying the hotspots of diseases independently of administrative boundaries. They also allow accounting for the spatial heterogeneity in the assessment of risk factors. These methods thus provide crucial additional data to national health policy makers for developing effective interventions and allocating financial and human resources based on the local situations. They have therefore been increasingly used in the last years, especially in the field of HIV/AIDS (a disease with a well-known spatial epidemiology) [6–15].

Burundi, in Eastern Africa, is among the world's poorest countries and was severely affected by a civil war from 1993 to 2003. It is bounded by Rwanda to the north, the Democratic Republic of Congo to the west, and Tanzania to the east and south-east (Fig. 1). The population is approximately 11 million inhabitants. The national AIDS program, launched in 1988, includes prevention, testing, care and treatment activities. HIV services have been progressively decentralized to primary

health centers throughout the country [16]. The 2010 Burundi DHS reported an overall HIV prevalence of 1.4 % among adults and suggested regional differences with HIV prevalence of 0.9 % in the South, 1.0 % in the Centre-East, 1.3 % in the North, 1.6 % in the West, and 3.7 % in Bujumbura-Mairie (the capital city) [17]. In September 2014, UNAIDS reported estimations of HIV prevalence at the smaller provincial level ranging from 0.4 % in two rural provinces to 3.6 % in Bujumbura-Mairie [18]. Based on individual-level data collected in the 2010 Burundi DHS, we further investigated the spatial heterogeneity of HIV prevalence and then assessed the association of social and behavioral characteristics with HIV infection accounting for the spatial heterogeneity.

Methods

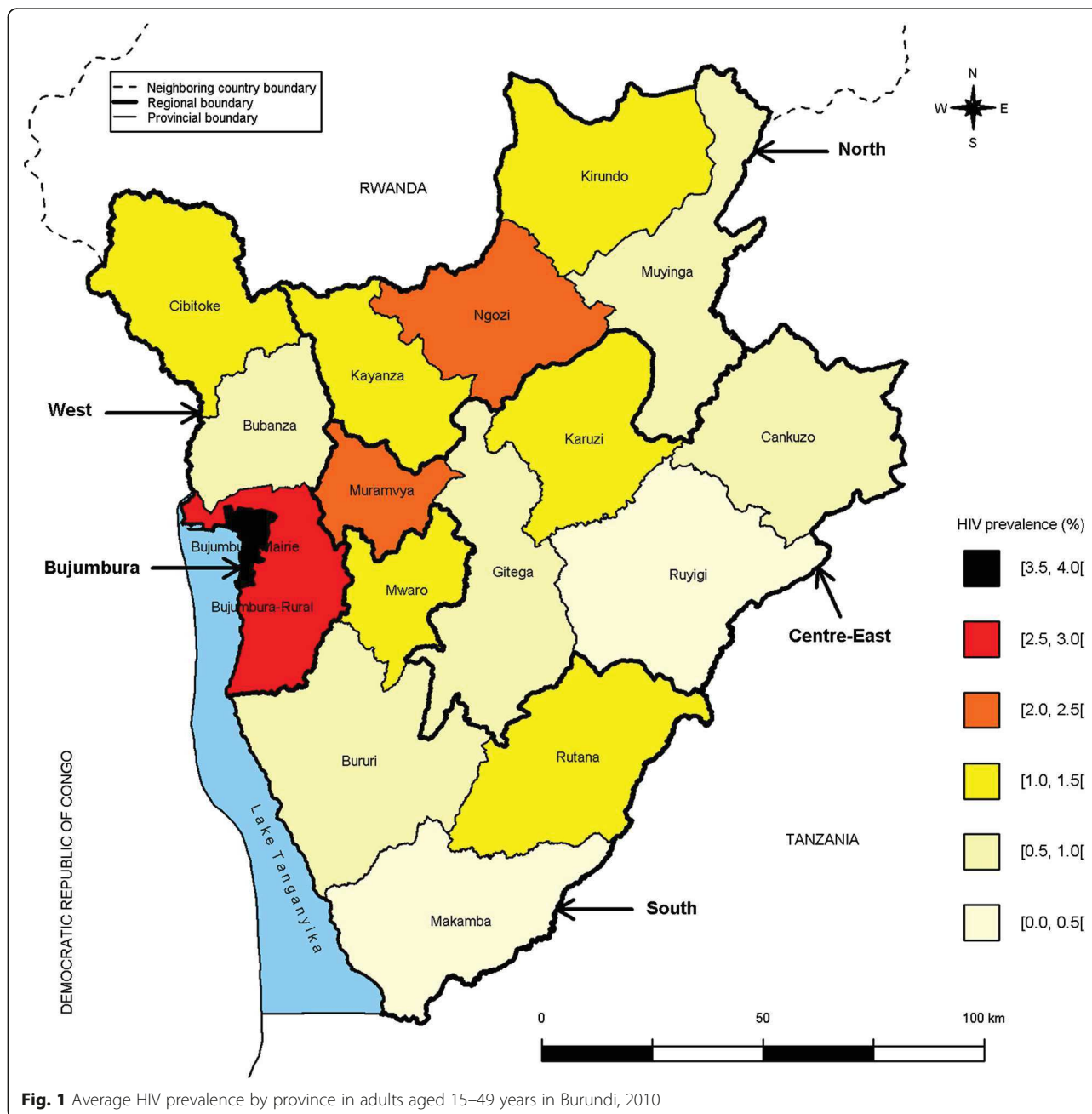
Study design

We performed a cross-sectional study based on a secondary analysis of the 2010 Burundi DHS data.

Design and procedures of the 2010 Burundi DHS

The 2010 Burundi DHS was conducted by national authorities and ICF International between August 29, 2010 and January 30, 2011 following the standardized DHS methodology [5]. This methodology and crude results are described extensively elsewhere [17]. Briefly, the 2010 Burundi DHS used a stratified two-stage random cluster sampling design. Stratification was made at two levels: the provincial level (17 provinces) and the urban or rural area. At the first stage, 376 of the 8104 enumeration areas (i.e. groupings of households) identified in the 2008 national population and housing census [19] were selected with a probability proportional to their size (i.e. the number of households in each enumeration area). At the second stage, 24 households were selected in each enumeration area with an equal probability, leading to a total of 9024 eligible households. HIV tests were proposed to 50 % of the 9024 households. All women aged 15–49 years and men aged 15–59 years living in or having spent the previous night in one of these households were eligible for HIV testing. The Institutional Review Board of ICF International and the National Ethics Committee of Burundi approved the study protocol. After being provided with information, respondents aged 18 years or older and parents or guardians of minors aged 15–18 years gave their written consent to participate in the survey. Minors gave their oral assent.

Blood spot samples were collected from consenting household residents' fingers and put on filter papers. Serologic screening for HIV infection was then performed on the dried blood spots at the Public Health National Institute in Bujumbura-Mairie using an enzyme-linked immunosorbent assay (ELISA; Vironostika HIV Uni-Form



Ag/Ab, Biomérieux, Marcy l’Etoile, France). All positive samples and 10 % of the negative samples (for quality control) were further tested with a second ELISA (Enzignost HIV Integral II, Siemens, Erlangen, Germany). Samples with discordant results between both tests were reanalyzed using a line immunoassay (INNO-LIA HIV I/II Score, Innogenetics, Gent, Belgium).

Social and behavioral characteristics of residents were collected using a standardized questionnaire and included gender (woman or man), age (continuous variable), level of education (no formal education, primary, secondary, or superior), marital status (single, married,

cohabiting, divorced, separated, or widowed), religion (Catholic, Protestant, Muslim, Adventist, Jehovah’s witness, other, or none), wealth quintile (poorest, poorer, middle, richer, or richest), sexual activity (never had sex, active in the last four weeks, or not active in the last four weeks), the number of extramarital sex partners in the last 12 months, and a history of sexually transmitted infection (STI) in the last 12 months (yes or no).

The geographic coordinates (latitude and longitude) of the 376 enumeration areas were collected using the geographic information system (GIS) and global positioning system (GPS) technologies. They were recorded at the

center of the enumeration areas. In order to preserve the confidentiality of the respondents, the GPS latitude/longitude positions were randomly displaced (up to 2 km for the urban enumeration areas and up to 5 km for the rural enumeration areas, with 1 % of the rural enumeration areas being displaced up to 10 km).

Statistical analysis

Study population

For the present study, the analyses were restricted to the 15–49 year age group because HIV testing had not been performed in women aged 50–59 years (by contrast to men) in the 2010 Burundi DHS. It is worth noting that UNAIDS also reports adult data in the 15–49 year age group [1, 18].

Data

All analyses used the individual-level data and their specific weights provided in the DHS databases (available from the DHS program website [5]). As usual, the weights took into account the survey design of the 2010 Burundi DHS and the proportion of respondents in each enumeration area. They were equal to the inverse of the probability for a given resident of being included in the survey.

Spatial heterogeneity of HIV prevalence

First, we computed and mapped the crude estimations of HIV prevalence in the 17 provinces. We then used a non-spatial logistic regression model to assess the relationship between HIV infection and the provinces, taking the province with the lowest HIV prevalence as the reference category. These analyses were performed using the Stata software version 11 [20].

Second, we analyzed the spatial autocorrelation of HIV prevalence data by performing a global Moran test. Moran's *I* statistic tests the null hypothesis that observed data at one location are independent of data at other locations. Its value ranges from -1 (data perfectly dispersed), 0 (data randomly dispersed) to 1 (data perfectly correlated). Because the Moran's *I* statistic showed the existence of a significant spatial autocorrelation of our HIV prevalence data, we subsequently analyzed the data using a geostatistical approach which takes into account this spatial autocorrelation.

Third, we mapped HIV prevalence throughout the country independently of provincial boundaries using a Gaussian kernel density estimation with adaptive bandwidths and the specific *prevR* package of R software (R Core Development Team, April 10, 2014) [21, 22]. This approach is promoted by UNAIDS and has been used to estimate HIV prevalence at a sub-national level in various countries including Burundi [18, 23]. It allows generating a smoothed surface of HIV prevalence based on

observed data. In our study, we interpolated HIV prevalence data in 449,065 points using the observed data in the 376 enumeration areas. As recommended by Larmarange and Bendaud [23], we set the number of observations at 500 so that the bandwidths adapt to capture this minimum number.

Finally, we identified the spatial clusters with high and low HIV prevalence using the Kulldorff spatial scan statistics (SaTScan software version 9.3) [24]. This method has been widely used in the last years, especially in the field of HIV/AIDS [7, 10, 12, 15, 25–27]. It allows finding the location of areas with higher or lower numbers of HIV cases than expected under the hypothesis of uniform spatial distribution of cases by gradually scanning circular windows of various sizes across the study area. We assumed that the number of HIV cases in each circular window was an independent Bernoulli random variable. For the circular windows, we used a maximum radius of 15 Km for the detection of clusters with high HIV prevalence and of 50 Km for the detection of clusters with low prevalence. We chose these radii because high HIV prevalence was more likely in small, densely populated areas such as in Bujumbura-Mairie where the maximum distance between the centroid and the city's boundaries is of 13 Km while low HIV prevalence was more likely in large, sparsely populated areas. We also used the default value of 50 % of the total study population for the maximal size of the clusters. The statistical significance of clusters was ascertained using the likelihood ratio test and its associated *p*-value obtained through 999 Monte Carlo simulations. The null hypothesis of uniform spatial distribution of HIV cases (no cluster) was rejected if the *p*-value was <0.05 . When a cluster was identified, the strength of the clustering was estimated using the relative risk of excess HIV cases.

Factors associated with HIV infection

The association of social and behavioral characteristics with HIV infection (infected or not) was investigated using a spatial logistic regression model performed with the BayesX software version 2.1 [28]. This model allowed adjusting for the spatial and non-spatial random effects of provinces. The parameters were estimated using 400 Markov chain Monte Carlo simulations in restricted maximum likelihood regression models. Independent covariates associated with HIV infection with a conservative *p*-value of <0.2 in univariate analysis were subsequently tested in multivariate analysis [29]. A backward elimination procedure was used to determine the final model containing only the covariates significantly associated with HIV infection. The strength of associations was estimated using the posterior odds ratios (PORs) and their 95 % credible intervals (CrIs). Finally,

Table 1 HIV prevalence by province and resident's characteristics in adults aged 15–49 years in Burundi, 2010

	Number*	HIV +*	Percent
Province			
Bubanza	379 (4.7 %)	3	0.9
Bujumbura-Mairie	731 (9.0 %)	27	3.7
Bujumbura-Rural	493 (6.1 %)	13	2.7
Bururi	654 (8.1 %)	6	0.9
Cankuzo	209 (2.6 %)	2	1.0
Cibitoke	489 (6.0 %)	5	1.1
Gitega	590 (7.3 %)	3	0.6
Karuzi	439 (5.4 %)	5	1.1
Kayanza	560 (6.9 %)	6	1.1
Kirundo	547 (6.8 %)	8	1.4
Makamba	474 (5.9 %)	2	0.3
Muramvya	313 (3.9 %)	8	2.4
Muyinga	610 (7.5 %)	5	0.8
Mwaro	288 (3.6 %)	4	1.3
Ngozi	613 (7.6 %)	13	2.2
Rutana	334 (4.1 %)	4	1.3
Ruyigi	362 (4.5 %)	1	0.1
Gender			
Women	4532 (56.0 %)	78	1.7
Men	3554 (44.0 %)	36	1.0
Age group (years)			
15–19	2014 (24.9 %)	5	0.3
20–24	1544 (19.1 %)	14	0.9
25–29	1341 (16.6 %)	13	0.9
30–34	962 (11.9 %)	20	2.1
35–39	864 (10.7 %)	24	2.8
40–44	702 (8.7 %)	23	3.3
45–49	659 (8.1 %)	15	2.3
Level of education			
No formal education	3067 (37.9 %)	34	1.1
Primary school	3706 (45.8 %)	61	1.6
Secondary school	1197 (14.8 %)	17	1.4
Superior	116 (1.5 %)	2	1.7
Marital status			
Single	3040 (37.6 %)	11	0.4
Married	3272 (40.5 %)	39	1.2
Cohabiting	1329 (16.4 %)	40	3.0
Widowed	201 (2.5 %)	16	8.1
Divorced or separated	243 (3.0 %)	6	2.6
Religion			
Catholic	5066 (62.7 %)	59	1.2
Protestant	2359 (29.2 %)	36	1.5

Table 1 HIV prevalence by province and resident's characteristics in adults aged 15–49 years in Burundi, 2010 (Continued)

Muslim	204 (2.5 %)	10	4.7
Adventist	185 (2.3 %)	3	1.5
Jehovah's Witnesses	20 (0.3 %)	1	5.3
Other	63 (0.8 %)	1	2.2
None	180 (2.2 %)	4	2.1
Wealth index			
Poorest	1469 (18.2 %)	17	1.2
Poorer	1558 (19.3 %)	16	1.0
Middle	1583 (19.6 %)	20	1.2
Richer	1645 (20.3 %)	12	0.7
Richest	1831 (22.6 %)	49	2.7
Sexual activity			
Never had sex	2457 (30.4 %)	5	0.2
Active in the last 4 weeks	4177 (51.7 %)	68	1.6
Not active in the last 4 weeks	1446 (17.9 %)	41	2.8
Number of extramarital sex partners in the last 12 months			
0	7815 (96.7 %)	106	1.4
1	238 (2.9 %)	7	2.8
≥ 2	29 (0.4 %)	1	4.6
Sexually transmitted infections in the last 12 months			
No	7915 (98.1 %)	102	1.3
Yes	153 (1.9 %)	12	7.8
Total	8086 (100.0 %)	114	1.4

*Weighted numbers were rounded

the goodness-of-fit of models was assessed using the conditional Bayesian Information Criterion (BIC).

All statistical tests were interpreted at the 0.05 significance level.

Results

Characteristics of the study population

Of 9503 residents eligible for HIV testing, 90.4 % were interviewed and tested (91.8 % of 4911 eligible women and 88.8 % of 4592 eligible men). Five hundred and one men, aged 50–59 years, were excluded from the present analysis. After weighting, 8086 residents aged 15–49 years enrolled in 3816 households from the 376 enumeration areas were included in the analysis. Median number of residents by enumeration area was 21 (interquartile range [IQR] 18–25). There were 4532 women and 3554 men (Table 1). Median age was 26 years (IQR 20–35). Most residents had attended at least primary school (62.1 %), were not single (62.4 %), were Catholic (62.7 %), had sex in the last 4 weeks (51.7 %), had no extramarital sex partner in the last 12 months (96.7 %), and had no STI in the last 12 months (98.1 %). There were 114 HIV cases (78

women and 36 men), leading to an overall HIV prevalence of 1.4 %.

Spatial heterogeneity of HIV prevalence

The crude estimations of HIV prevalence ranged from 0.1 % in the Ruyigi province to 3.7 % in Bujumbura-Mairie (Fig. 1 & Table 1). HIV prevalence was also high in the Bujumbura-Rural (2.7 %), Muramvya (2.4 %) and Ngozi (2.2 %) provinces. The logistic regression analysis further showed that HIV infection was significantly more frequent in six provinces (Bujumbura-Mairie, Bujumbura-Rural, Kirundo, Muramvya, Ngozi and Rutana) than in the Ruyigi province (Table 2).

Data of HIV prevalence were spatially auto correlated (Moran's $I = 0.03$, $p = 0.021$). The interpolated data showed the great spatial heterogeneity of HIV prevalence (from 0 to 10 %), independently of provincial boundaries (Fig. 2). The higher HIV prevalence was observed near Bujumbura-Mairie. Prevalence above 2.0 % was also observed in different locations throughout the country. By contrast, HIV prevalence was especially low in the Centre-East and South regions.

The spatial scan statistics analysis confirmed these findings (Fig. 3). It identified a cluster with high HIV prevalence with a 13.5 Km radius in Bujumbura-Mairie and adjacent areas. There were 37 HIV cases (32.5 % of all) giving a relative risk of 3.7 ($p < 0.001$). HIV prevalence was thus 3.9 % among the 943 residents from 50

enumeration areas. The spatial scan statistics analysis also identified a cluster with low HIV prevalence with a 29.1 Km radius which straddled the Makamba and Bururi provinces ($p = 0.02$). No HIV case was detected in this cluster although the study included 651 residents from 23 enumeration areas. By contrast, no cluster was identified in the Centre-East region including the Ruyigi province.

Spatial analysis of factors associated with HIV infection

After controlling for the spatial heterogeneity (Table 3), HIV infection was strongly associated in the univariate analysis with the female sex (POR 1.43, 95 % CrI 1.20–1.70, $p < 0.001$), an older age (e.g. POR 2.96, 95 % CrI 2.04–4.28, $p < 0.001$ for 35–39 years versus 15–19 years), the marital status (e.g. POR 4.31, 95 % CrI 3.18–5.86, $p < 0.001$ for widowed versus single), a higher wealth index (POR 1.98, 95 % CrI 1.67–2.35, $p < 0.001$ for richest versus others), the sexual activity (e.g. POR 2.68, 95 % CrI 1.86–3.85, $p < 0.001$ for active in the last 4 weeks versus never had sex), and a history of STI in the last 12 months (POR 2.55, 95 % CrI 1.98–3.30, $p < 0.001$). HIV infection also tended to be associated with the religion (POR 1.85, 95 % CrI 0.99–3.44, $p = 0.052$ for Muslim versus Catholic), and the number of extramarital sex partners in the last 12 months (POR 1.76, 95 % CrI 0.98–3.15, $p = 0.057$ for ≥ 2 extramarital sex partners versus none) but the statistical significance was not reached. By contrast, HIV infection was not associated with the level of education (POR 1.15, 95 % CrI 0.95–1.38, $p = 0.141$).

By multivariate spatial analysis (Table 3), HIV infection remained significantly associated with the female sex (POR 1.36, 95 % CrI 1.13–1.64, $p = 0.001$), an older age (e.g. POR 1.97, 95 % CrI 1.26–3.08, $p = 0.004$ for 35–39 years versus 15–19 years), the marital status (e.g. POR 1.86, 95 % CrI 1.23–2.80, $p = 0.004$ for widowed versus single), a higher wealth index (POR 2.11, 95 % CrI 1.77–2.51, $p < 0.001$ for richest versus others), the sexual activity (e.g. POR 1.76, 95 % CrI 1.04–2.96, $p = 0.034$ for active in the last four weeks versus never had sex), and a history of STI in the last 12 months (POR 2.03, 95 % CrI 1.56–2.64, $p < 0.001$). HIV infection was also significantly associated with the level of education, being higher in residents who had attended school than in those who did not (POR 1.50, 95 % CrI 1.22–1.84, $p < 0.001$).

Discussion

This spatial study allowed identifying populations at higher risk of HIV infection because of geographic, social or behavioral characteristics in Burundi.

Thus, our study added important information on the spatial heterogeneity of HIV infection in this country. We first found a significant association between HIV

Table 2 Relationship between HIV infection and provinces in adults aged 15–49 years in Burundi, 2010

Province	OR	95 % CI	p^*
Bubanza	6.32	0.60–66.50	0.124
Bujumbura-Mairie	28.07	3.63–217.04	0.001
Bujumbura-Rural	20.41	2.42–172.16	0.006
Bururi	6.22	0.63–61.17	0.117
Cankuzo	6.97	0.74–65.74	0.090
Cibitoke	7.85	0.79–77.69	0.078
Gitega	4.18	0.34–51.37	0.263
Karuzi	8.03	0.89–72.14	0.063
Kayanza	7.90	0.89–70.33	0.064
Kirundo	10.16	1.13–91.22	0.039
Makamba	2.44	0.23–25.38	0.456
Muramvya	18.12	2.21–148.54	0.007
Muyinga	5.52	0.48–63.60	0.170
Mwaro	9.23	0.89–95.74	0.063
Ngozi	16.17	1.86–140.79	0.012
Rutana	9.72	1.08–87.44	0.043
Ruyigi	1.00		

OR odds ratio, CI confidence interval. * P -value for the corresponding province versus the Ruyigi province

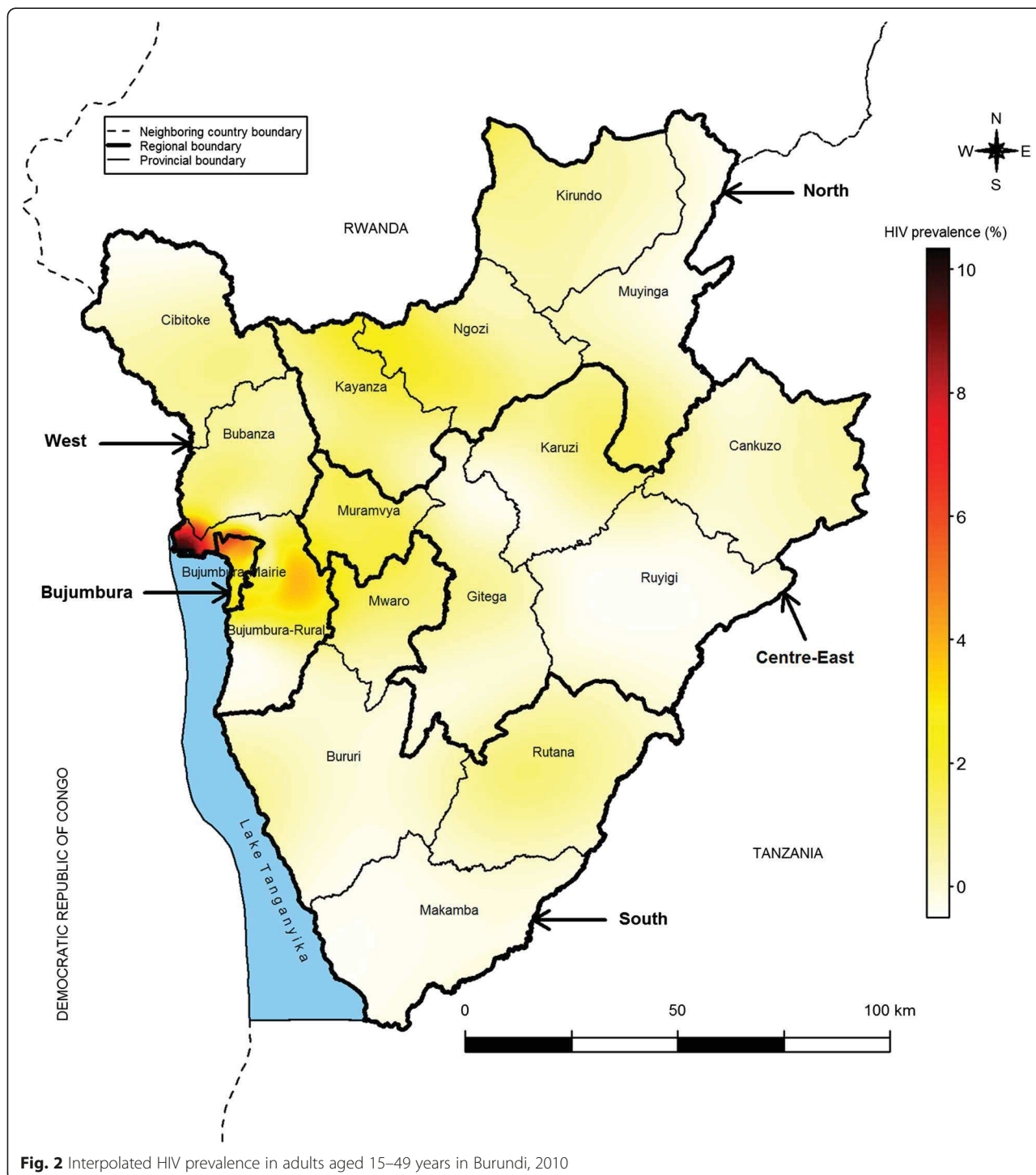
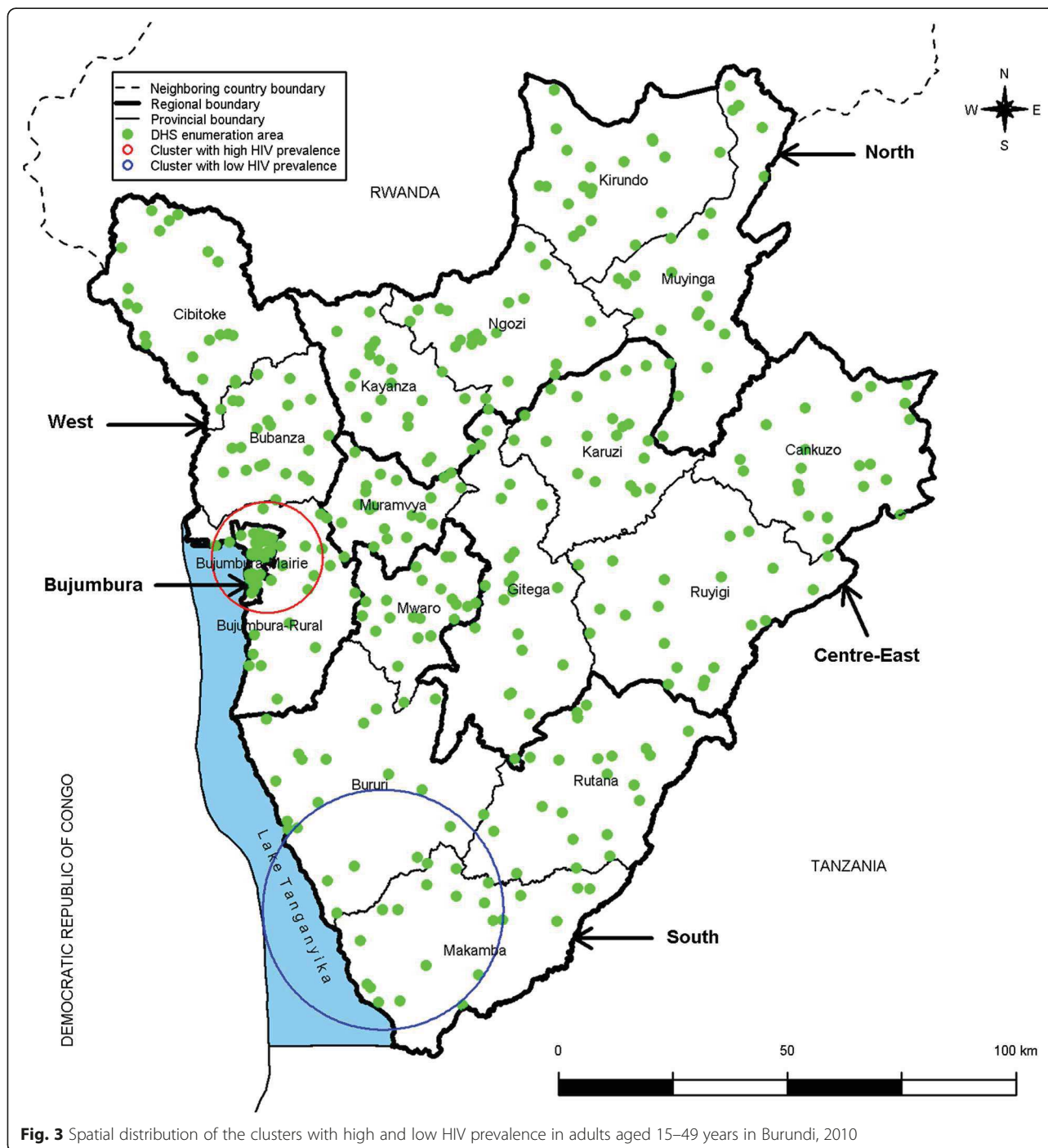


Fig. 2 Interpolated HIV prevalence in adults aged 15–49 years in Burundi, 2010

infection and the provinces. By contrast, the 2010 Burundi DHS report did not analyze HIV data by province [17]. UNAIDS recently reported HIV prevalence by province but did not assess the statistical significance of differences [18]. Our interpolated data further showed different locations with relatively high HIV prevalence, independently of provincial boundaries. Finally, we

identified a cluster with high HIV prevalence centered in Bujumbura-Mairie and a cluster with low prevalence in the southern part of the country. Cuadros et al. also found a significant cluster with high HIV prevalence in Bujumbura-Mairie but they did not detect any significant cluster with low prevalence ($p = 0.069$) probably because they performed their study in a slightly different



population including men aged 50–59 years and used unweighted data [7]. Overall, our findings in Burundi are consistent with the localized spatial clustering of HIV infection found in other countries [6–10, 15]. They highlighted that the crude regional estimations of the 2010 Burundi DHS report masked intra-regional heterogeneities in HIV prevalence [17].

In our study controlling for this spatial heterogeneity, HIV infection was significantly more frequent in the

residents who were women, older than 35 years, educated, widowed, divorced or separated, richest, sexually active, and in those who had had STIs in the last 12 months. These factors are well-known risk factors for HIV infection [30–34]. However, the novelty of our spatial analysis is that it provided better estimators of the strength of associations than a non-spatial analysis because the former takes into account the spatial autocorrelation of data.

Table 3 Univariate and multivariate spatial logistic regression analyses of factors associated with HIV infection in adults aged 15–49 years in Burundi, 2010

	Univariate			Multivariate		
	POR	95 % CrI	<i>p</i>	aPOR	95 % CrI	<i>p</i>
Gender						
Men	1.00			1.00		
Women	1.43	1.20–1.70	<0.001	1.36	1.13–1.64	0.001
Age group (years)						
15–19	1.00			1.00		
20–24	1.59	1.06–2.37	0.024	1.19	0.77–1.85	0.425
25–29	1.85	1.25–2.74	0.003	1.21	0.76–1.91	0.420
30–34	2.44	1.67–3.57	<0.001	1.56	0.99–2.45	0.056
35–39	2.96	2.04–4.28	<0.001	1.97	1.26–3.08	0.004
40–44	2.99	2.04–4.38	<0.001	2.14	1.34–3.42	0.002
45–49	3.06	2.09–4.50	<0.001	2.13	1.32–3.44	0.002
Level of education						
No formal education	1.00			1.00		
Primary school or higher	1.15	0.95–1.38	0.141	1.50	1.22–1.84	<0.001
Marital status						
Single	1.00			1.00		
Married	1.71	1.32–2.21	<0.001	0.91	0.61–1.34	0.622
Cohabiting	2.56	1.97–3.32	<0.001	1.46	0.99–2.15	0.057
Widowed	4.31	3.18–5.86	<0.001	1.86	1.23–2.80	0.004
Divorced or separated	3.08	2.17–4.36	<0.001	1.56	1.04–2.36	0.033
Religion						
Catholic	1.00					
Protestant	1.00	0.55–1.80	0.999			
Muslim	1.85	0.99–3.44	0.052			
Adventist	1.18	0.55–2.53	0.661			
Jehovah's Witnesses	1.94	0.79–4.80	0.150			
Other	1.30	0.53–3.18	0.570			
None	0.91	0.51–1.63	0.757			
Wealth index						
Richest	1.98	1.67–2.35	<0.001	2.11	1.77–2.51	<0.001
Others	1.00			1.00		
Sexual activity						
Never had sex	1.00			1.00		
Active in the 4 last weeks	2.68	1.86–3.85	<0.001	1.76	1.04–2.96	0.034
Not active in the 4 last weeks	3.50	2.42–5.06	<0.001	2.10	1.32–3.35	0.002
Number of extramarital sex partners in the last 12 months						
0	1.00					
1	1.13	0.78–1.62	0.515			
≥ 2	1.76	0.98–3.15	0.057			

Table 3 Univariate and multivariate spatial logistic regression analyses of factors associated with HIV infection in adults aged 15–49 years in Burundi, 2010 (*Continued*)

Sexually transmitted infections in the last 12 months						
No	1.00			1.00		
Yes	2.55	1.98–3.30	<0.001	2.03	1.56–2.64	<0.001

POR posterior odds ratio, CrI credible interval

Our findings confirmed that in-depth analyses of local HIV epidemics are crucial for national AIDS programs when designing the most effective prevention responses [3]. Indeed, the reduction of the number of new HIV infections implies the need for a greater understanding of 'where' and 'towards which populations' efforts should be concentrated, in terms of primary and secondary prevention activities such as counseling, availability and accessibility of condoms, HIV testing, linkage to care, early antiretroviral treatment, and support. For instance in Burundi, our findings suggest that HIV activities should be especially reinforced in and around Bujumbura-Mairie. With regard to social or behavioral characteristics, greater efforts should be focused on higher risk groups such as women, people who are older than 35 years, educated, widowed, divorced or separated, richest, sexually active, and those with STIs. In addition, the reasons for the spatial heterogeneity of HIV prevalence should be investigated.

One of the strengths of our study was the use of data collected in a DHS based on a standardized methodology. A second strength was that this DHS was large, involving 8086 residents enrolled in 3816 households from 376 enumeration areas. Thus, although the 2010 Burundi DHS was designed to provide estimates of HIV prevalence at the national and regional level (as the other DHS), estimates at the provincial level have been found to be good or moderately good [18, 23]. Finally, we analyzed data using standardized geostatistical methods which take into account the spatial autocorrelation of data.

Our findings should be interpreted taking into account several study limitations. First, as HIV prevalence is quite low in Burundi, the spatial variability was relatively modest. In addition, there was zero HIV case in certain enumeration areas. This may have limited our ability to find significant associations with HIV infection at the provincial level (for instance, between the Cibitoke, Karuzi, Kayanza and Mwaro provinces and the Ruyigi province). Second, the spatial logistic regression method used here to assess the association of social and behavioral characteristics with HIV infection provides one single posterior odds ratio by characteristic assuming that the strength of the association is uniform over the study area. However, this hypothesis might be incorrect, especially as our study area was large (the whole country).

An additional analysis using a geographically weighted regression method which would provide the posterior odds ratios specific to the cluster with high HIV prevalence might be useful for identifying the populations most in need of interventions [35].

Conclusion

In conclusion, this study could help health policy makers develop an effective intervention in Burundi by showing where and towards which populations HIV resources should be allocated. Our findings support the need for in-depth analyses of HIV epidemics in every countries to tailor national prevention responses, as promoted by UNAIDS. This should encourage program managers in other countries to perform such studies in their own settings. This kind of study is not costly and is relatively rapid thanks to the availability of recurrent DHS data.

Abbreviations

AIDS: acquired immune deficiency syndrome; CrI: credible interval; DHS: demographic and health survey; GIS: geographic information system; GPS: global positioning system; HIV: human immunodeficiency virus; IQR: interquartile range; POR: posterior odds ratio; STI: sexually transmitted infection; UNAIDS: Joint United Nations Programme on HIV/AIDS.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

E.B., N.M., and C.L. elaborated the study design. E.B. obtained the permission to use the 2010 Burundi DHS data, performed the statistical analysis and, with C.L., wrote the first draft of the paper. E.B., N.M., T.N. and C.L. contributed to the interpretation of the results, reviewed the draft, and approved the final manuscript.

Acknowledgments

We thank the Institut de Statistiques et d'Études Économiques du Burundi (ISTEEBU), the Ministère de la Santé Publique et de la Lutte contre le Sida [Burundi] (MSPLS), and ICF International which performed the 2010 Burundi DHS, Measure DHS which provided the databases, and Jude Sweeney (Rome, Italy) for the English revision and editing of the manuscript. Emmanuel Barankanira was the recipient of a doctoral fellowship from the Ministry of Higher Education and Scientific Research in Burundi.

Author details

¹Département des Sciences Naturelles, Ecole Normale Supérieure, Bujumbura, Burundi. ²TransVIHMI, IRD UMI 233 / INSERM U 1175 / Université de Montpellier, Montpellier, France. ³IMAG, UMR 519 / Centre Hospitalier Régional Universitaire de Montpellier / Université de Montpellier, Montpellier, France. ⁴Centre Hospitalier Universitaire de Kamenge, Bujumbura, Burundi. ⁵Institut de Recherche pour le Développement (UMI 233), 911 avenue Agropolis, BP 64501, Montpellier 34394 cedex 5, France.

Received: 20 July 2015 Accepted: 20 January 2016

Published online: 04 February 2016

References

- UNAIDS. Global report on the AIDS epidemic. 2013 [July 15, 2015]. [http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf].
- Laurent C. Scaling up HIV treatment in resource-limited countries: The challenge of staff shortages. *J Public Health Policy*. 2011;32(2):211–8.
- UNAIDS. Practical guidelines for intensifying HIV prevention: towards universal access. 2007 [July 15, 2015]. [http://data.unaids.org/pub/Manual/2007/20070306_Prevention_Guidelines_Towards_Universal_Access_en.pdf].
- UNAIDS. 90–90–90 - An ambitious treatment target to help end the AIDS epidemic. 2014 [November 10, 2015]. [http://www.unaids.org/en/resources/documents/2014/90-90-90].
- Measure DHS. Demographic and Health Surveys. [July 15, 2015]. [http://www.measuredhs.com].
- Tanser F, LeSueur D, Solarsh G, Wilkinson D. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Trop Med Int Health*. 2000;5(1):40–6.
- Cuadros DF, Awad SF, Abu-Raddad LJ. Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *Int J Health Geogr*. 2013;12(1):28.
- Wand H, Ramjee G. Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *J Int AIDS Soc*. 2010;13(1):41.
- Messina JP, Emch M, Muwonga J, Mwandagalirwa K, Edidi SB, Mama N, et al. Spatial and socio-behavioral patterns of HIV prevalence in the Democratic Republic of Congo. *Soc Sci Med*. 2010;71(8):1428–35.
- Tanser F, Barnighausen T, Cooke GS, Newell M-L. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int J Epidemiol*. 2009;38(4):1008–16.
- Kandala N-B, Ji C, Cappuccio PF, Stones RW. The epidemiology of HIV infection in Zambia. *AIDS Care*. 2008;20(7):812–9.
- Chimoyi LA, Musenge E. Spatial analysis of factors associated with HIV infection among young people in Uganda, 2011. *BMC Public Health*. 2014; 14(1):555.
- Niragire F, Achia TNO, Lyambabaje A, Ntaganira J. Bayesian mapping of HIV infection among women of reproductive age in Rwanda. *PLoS ONE*. 2015; 10(3):e0119944. Schillaci G, éditeur.
- Zulu LC, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994–2010. *BMC Infect Dis*. 2014;14(1):285.
- González R, Augusto OJ, Munguambe K, Pierrat C, Pedro EN, Sacoor C, et al. HIV incidence and spatial clustering in a rural area of Southern Mozambique. *PLoS ONE*. 2015;10(7):e0132053. Carr JK, éditeur.
- Ministère de la Santé Publique et de la Lutte contre le Sida. Plan Stratégique National de Lutte contre le sida 2012–2016. 2012 [February 3, 2016]. [http://www.ilo.org/wcmsp5/groups/public/—ed_protect/—protrav/—ilo_aids/documents/legaldocument/wcms_202048.pdf].
- Institut de Statistiques et d'Études Économiques du Burundi (ISTEEBU), Ministère de la Santé Publique et de la Lutte contre le Sida [Burundi] (MSPLS), ICF International. Enquête Démographique et de Santé Burundi 2010. 2012 [February 3, 2016]. [https://dhsprogram.com/pubs/pdf/FR253/FR253.pdf].
- UNAIDS. Burundi - Developing subnational estimates of HIV prevalence and the number of people living with HIV. 2014 [July 16, 2015]. [http://www.unaids.org/sites/default/files/media_asset/2014_subnationalestimatessurvey_Burundi_en.pdf].
- Institut de Statistiques et d'Études Économiques du Burundi (ISTEEBU). Recensement Général de la Population et de l'Habitat du Burundi 2008: Synthèse des résultats définitifs. Bujumbura, Burundi, 2011 [July 17, 2015]. [https://www.usaid.gov/sites/default/files/documents/1866/2008%20Burundi%20Population%20Census_Summary%20of%20Final%20Results.pdf].
- StataCorp, StataCorp. Stata statistical software: Release 11. College Station, TX: StataCorp LP; 2011.
- Larmarange J, Vallo R, Yaro S, Msellati P, Méda N. Methods for mapping regional trends of HIV prevalence from Demographic and Health Surveys (DHS). 2011 [November 6, 2015]. [http://cybergeo.revues.org/24606].
- Larmarange J. prevR: estimating regional trends of a prevalence from a DHS. Paris: IRD; 2013. [November 10, 2015]. [http://cran.r-project.org/web/packages/prevR/index.html].
- Larmarange J, Bendaub V. HIV estimates at second subnational level from national population-based surveys. *AIDS*. 2014;28:S469–76.
- Kulldorff M. A spatial scan statistic. *Commun Stat - Theory Methods*. 1997; 26(6):1481–96.
- Westercamp N, Moses S, Agot K, Ndinya-Achola JO, Parker C, Amolloh KO, et al. Spatial distribution and cluster analysis of sexual risk behaviors reported by young men in Kisumu, Kenya. *Int J Health Geogr*. 2010;9(1):24.
- Namosha E, Sartorius B, Tanser F. Spatial clustering of all-cause and HIV-related mortality in a rural South African population (2000–2006). *PLoS ONE*. 2013;8(7):e69279. Vermund SH, éditeur.
- Cuadros DF, Abu-Raddad LJ. Spatial variability in HIV prevalence declines in several countries in sub-Saharan Africa. *Health Place*. 2014;28:45–9.
- Belitz C, Brezger A, Kneib T, Lang S, Umlauf N. BayesX - Software for Bayesian Inference in Structured Additive Regression Models. Version 2.1. 21st edition. Munich, Germany; 2012 [May 21, 2015]. [http://www.statistik.lmu.de/~bayesx/bayesxdownload.html].
- Maldonado G, Greenland S. Simulation study of confounder-selection strategies. *Am J Epidemiol*. 1993;138(11):923–36.
- Barongo LR, Borgdorff MW, Moshia FF, Nicoll A, Grosskurth H, Senkoro KP, et al. The epidemiology of HIV-1 infection in urban areas, roadside settlements and rural villages in Mwanza Region, Tanzania. *Aids*. 1992;6:1521–8.
- Chao A, Bulterys M, Musanganire F, Abimana P, Nawrocki P, Taylor E, et al. Risk factors associated with prevalent HIV-1 infection among pregnant women in Rwanda. *Int J Epidemiol*. 1994;23(2):371–80.
- Bloom SS, Urassa M, Isingo R, Ng'weshemi J, Boerma JT. Community effects on the risk of HIV infection in rural Tanzania. *Sex Transm Infect*. 2002;78:261–6.
- Parkhurst JO. Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in African countries. *Bull World Health Organ*. 2010;88(7):519–26.
- Msisha WM, Kapiga SH, Earls F, Subramanian S. Socioeconomic status and HIV seroprevalence in Tanzania: a counterintuitive relationship. *Int J Epidemiol*. 2008;37(6):1297–303.
- Kaulh B, Heil J, Hoebe CJ, Schweikart J, Krafft T, Dukers-Muijers NH. The spatial distribution of hepatitis C virus infections and associated determinants—An application of a geographically weighted poisson regression for evidence-based screening interventions in hotspots. *PLoS ONE*. 2015;10(9):e0135656.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit



Chapitre 3 : Analyse spatiale et temporelle du retard de croissance en Côte d'Ivoire

3.1. Introduction

La malnutrition est un état pathologique qui résulte d'un manque ou d'un surplus de nutriments et micronutriments, ce qui traduit une alimentation déséquilibrée ou inadaptée pour assurer le développement de l'organisme et satisfaire les besoins nutritionnels. Dans les pays à ressources limitées, la dénutrition (malnutrition par carence) chez l'enfant de moins de 5 ans est non seulement un problème de société car elle affectera les générations futures, mais également un problème individuel de l'enfant car l'enfant dénutri aura un mauvais développement physique et intellectuel, une vulnérabilité importante aux infections, et un risque élevé de mortalité infanto-juvénile. Dans ce chapitre, la malnutrition sera utilisée dans le sens de « dénutrition ».

Deux mesures anthropométriques servent à évaluer la malnutrition d'un enfant d'âge et de sexe donnés : le poids et la taille (la taille d'un enfant âgé de moins de deux ans est mesurée en position couchée, tandis que celle d'un enfant âgé de deux ans et plus est mesurée en position debout).

Il existe trois types de malnutrition : la malnutrition chronique ou retard de croissance (une petite taille par rapport à l'âge), la malnutrition aigue ou maigreur ou encore émaciation (un faible poids par rapport à la taille) et l'insuffisance pondérale (un faible poids par rapport à l'âge). Ces trois types de malnutrition sont définis respectivement par les indices : taille-pour-âge, poids-pour-taille et poids-pour-âge. Sous la supposition que tous les enfants du monde ont le même potentiel de croissance, chaque indice est calculé par rapport aux z-scores (écarts-types) de la médiane de la norme internationale de croissance, constituée d'enfants bien nourris et en bonne santé, de même âge et de même sexe. Chaque z-score est calculé comme étant la différence entre la valeur observée (la taille par exemple) et la valeur médiane (taille moyenne par exemple) de la population de référence, divisée par l'écart-type de la variable dans la population de référence :

$$Z_i = \frac{X_i - Me}{\sigma} \quad (3.1)$$

où X_i est la mesure anthropométrique (taille par exemple) de l'enfant, Me et σ respectivement la médiane et l'écart-type de la mesure anthropométrique dans la population de référence. Pour chaque indice, il existe la forme modérée et la forme sévère. Ainsi, un enfant souffre de retard de croissance sous sa forme modérée si l'indice taille-pour-âge est compris entre -3 et -2 z-scores et sous sa forme sévère s'il est inférieur à -3 z-scores. Dans les enquêtes nutritionnelles en population, c'est généralement la forme modérée ou sévère, c'est-à-dire lorsque la taille est inférieure à -2 z-scores, qui est rapportée.

L'insuffisance pondérale est une forme mixte car elle peut résulter soit d'un retard de croissance, soit d'une maigreur, ou enfin de la combinaison des deux.

La lutte contre le retard de croissance chez les enfants commence par la lutte contre le retard de croissance intra-utérin en fournissant aux femmes enceintes une alimentation suffisante, variée et équilibrée afin d'éviter de mettre au monde des enfants de faible poids. La malnutrition maternelle et la malnutrition infantile sont alors deux catégories de malnutrition auxquelles les programmes de lutte contre la malnutrition en général et le retard de croissance en particulier s'attèlent à contrôler puisque le retard de croissance et l'émaciation chez l'enfant ont pour conséquence l'insuffisance pondérale.

Chez les personnes de plus de 18 ans, une faible valeur de l'indice de masse corporelle (IMC) traduit un état d'insuffisance pondérale. L'IMC est défini comme le poids divisé par la taille élevée au carré. Une personne est qualifiée de maigre si son IMC est inférieur à 18,5 kg/m² ; normale si son IMC est compris entre 18,5 et 25 ; en surpoids si son IMC est compris entre 25,0 et 30,0 et obèse si son IMC est supérieur à 30.

Plusieurs causes peuvent être à l'origine de la malnutrition de l'enfant. Les causes dites « primaires » sont les maladies comme les gastroentérites, les infections pulmonaires ou le paludisme, et l'alimentation au sens large (qualité insuffisante des aliments consommés et des pratiques d'allaitement, dont le sevrage précoce) Le retard de croissance a la particularité d'être le résultat d'un apport prolongé en

aliments insuffisants en quantité et/ou en qualité. Quand l'enfant est soumis à des infections graves et répétées, l'organisme subit une perte de nutriments et d'énergie (diarrhées, vomissements, fièvre) et un déficit d'apports (anorexie). Ainsi, les besoins nutritionnels augmentent pendant la convalescence et, s'ils ne peuvent être remplis, l'enfant devient plus vulnérable à de nouvelles infections et entre ainsi dans la spirale mortifère malnutrition – infections, qui peut conduire au décès.

Les causes distales qui déterminent le niveau sanitaire et l'alimentation sont soit spécifiques aux conditions de vie du ménage (le niveau de vie des membres du ménage, leur niveau d'éducation, leur type de travail), soit encore plus distales comme la situation géopolitique et écologique (insécurité alimentaire, guerre civile, manque ou surplus de précipitations, inflation des prix des produits alimentaires sur les marchés, pénurie des produits alimentaires, catastrophes naturels) dans laquelle évoluent les individus.

Le retard de croissance a la caractéristique d'être peu affecté par des fluctuations temporelles. Autrement dit, une fois constitué, il est stable au cours du temps. Un enfant qui souffre du retard de croissance durant ses premières années de la vie le restera le plus souvent durant toute sa vie. C'est pourquoi nous nous sommes intéressés au retard de croissance dans cette thèse. De plus, la prévalence du retard de croissance est plus élevée que celle de l'émaciation. En outre, l'utilisation du retard de croissance permet d'évaluer les progrès réalisés par les pays en développement dans l'amélioration de l'état nutritionnel chez l'enfant [160].

Le retard de croissance reflète le niveau de vie des ménages et des familles, et affecte le développement cognitif, physique et intellectuel de l'enfant. De plus, une fille ayant un retard de croissance a plus de risques de donner naissance à des enfants malnutris, ce qui va avoir des conséquences néfastes sur les générations futures, sur sa productivité et sur le développement social et économique de son pays. De ce fait, le retard de croissance constitue un problème majeur auquel les gouvernements doivent faire face afin de garantir le développement socio-économique de leurs pays. Le retard de croissance et l'émaciation peuvent coexister

chez l'enfant malnutri, notamment en Asie ou en Afrique. Par contre, le retard de croissance est le plus souvent isolé en Amérique Latine [161].

Une série d'études a montré l'absence de diversité inter-ethnique de la croissance ou, autrement dit, un potentiel de croissance similaire chez tous les enfants d'âge préscolaire du monde [162]. Cela nécessite de sélectionner des enfants vivant dans d'excellentes conditions socioéconomiques, bien nourris et en bonne santé, soit au sein d'élites des pays en développement, soit ayant migré dans des pays riches [163–166]. Par contre, les enfants provenant des milieux pauvres et ceux provenant des milieux riches des pays en développement n'ont pas la même croissance, parce que les enfants pauvres n'arrivent pas à exprimer leur potentiel. Cela justifie le fait d'utiliser une norme universelle pour évaluer l'état nutritionnel des enfants dans le monde.

3.2. Anciennes références et nouvelles normes de croissance de l'OMS

Les anciennes références de croissance de l'OMS pour les enfants de moins de 5 ans étaient celles du National Center for Health Statistics (NCHS) de 1977-1978 qui constituaient, jusqu'en 2006, une référence internationale de croissance chez les enfants [167]. Elles étaient basées uniquement sur la population américaine comme population de référence. Pire, de 0 à 24 mois, elle était basée sur une petite cohorte d'enfants de la classe moyenne suivis dans un seul état américain dans les années 1920 et essentiellement nourris au lait maternisé [168]. Cette population n'étant pas représentative de la population mondiale d'enfants bien nourris et en bonne santé, et pas en conformité avec les recommandations d'alimentation du nourrisson par l'OMS (allaitement maternel exclusif jusqu'à 6 mois et poursuite de l'allaitement partiel jusqu'à 1-2 ans), elle ne pouvait pas continuer à servir de base pour l'évaluation de l'état nutritionnel de l'enfant. C'est dans ce cadre qu'une étude multicentrique coordonnée par l'OMS a été menée sur 8500 enfants provenant de 6 pays (Brésil, Etats-Unis, Ghana, Inde, Norvège et Oman) de 1997 à 2003 en vue d'établir en 2006 les nouvelles normes de croissance permettant d'évaluer la croissance de tous les nourrissons et les jeunes enfants dans le monde [169]. Ces enfants étaient

sélectionnés dans des populations très favorisées socialement pour bénéficier d'une bonne santé général et devaient être nourris selon les recommandations de l'OMS. Cette diversité ethnique et spatiale garantissait une bonne représentativité de la norme et était aussi un argument politique en faveur de son acceptation par tous les états.

3.3. Situation nutritionnelle dans le monde

3.3.1. Période des 1000 jours, et la tranche d'âge 0-5 ans

La période des 1000 jours de la vie de l'enfant correspond à la somme de la durée de la grossesse (depuis la date de conception jusqu'à l'accouchement), c'est-à-dire 9 mois, et de la durée qui sépare la date de naissance de l'enfant et l'âge auquel l'enfant célèbre son deuxième anniversaire (24 mois, soit au total 33 mois, soit environ 1000 jours). Cette période est cruciale pour la croissance de l'enfant, car la vitesse de croissance est alors très élevée (+75 cm pendant les 9 mois de la vie intra-utérine et +25 cm pendant la première année de vie) [170]. Pendant cette période, le jeune enfant doit passer d'une alimentation exclusivement lactée de 0 à 6 mois (allaitement au sein) à la combinaison d'une nourriture liquide, semi-liquide, semi-solide et solide. De plus, l'alimentation de la mère durant la grossesse et dans la période post-partum doit être bien contrôlée afin que l'enfant puisse avoir tout ce dont il a besoin pour son développement physique et intellectuel, pour atteindre son potentiel de croissance, et pour le développement de son système immunitaire puisque, dans cette période, il a une grande susceptibilité aux infections. Une mauvaise alimentation de la mère avant et durant la grossesse peut avoir des conséquences sur le développement du fœtus, y compris le retard de croissance intra-utérin, et peut augmenter les risques de mettre au monde un enfant qui souffrira d'un retard de croissance et d'autres formes de sous-nutrition [171].

3.3.2. Sous-alimentation et retard de croissance dans le monde

En 1990, le groupe UNICEF-OMS-Banque Mondiale estimait à 255 millions le nombre d'enfants de moins de 5 ans souffrant de retard de croissance dans le monde, soit une prévalence de 39,6 % [172]. En 2013, cette estimation portait à 161 millions le nombre d'enfants souffrant de retard de croissance, soit une prévalence de 24,5 % [162]. La prévalence du retard de croissance en 2013 a donc diminué d'un tiers par rapport à l'année 1990.

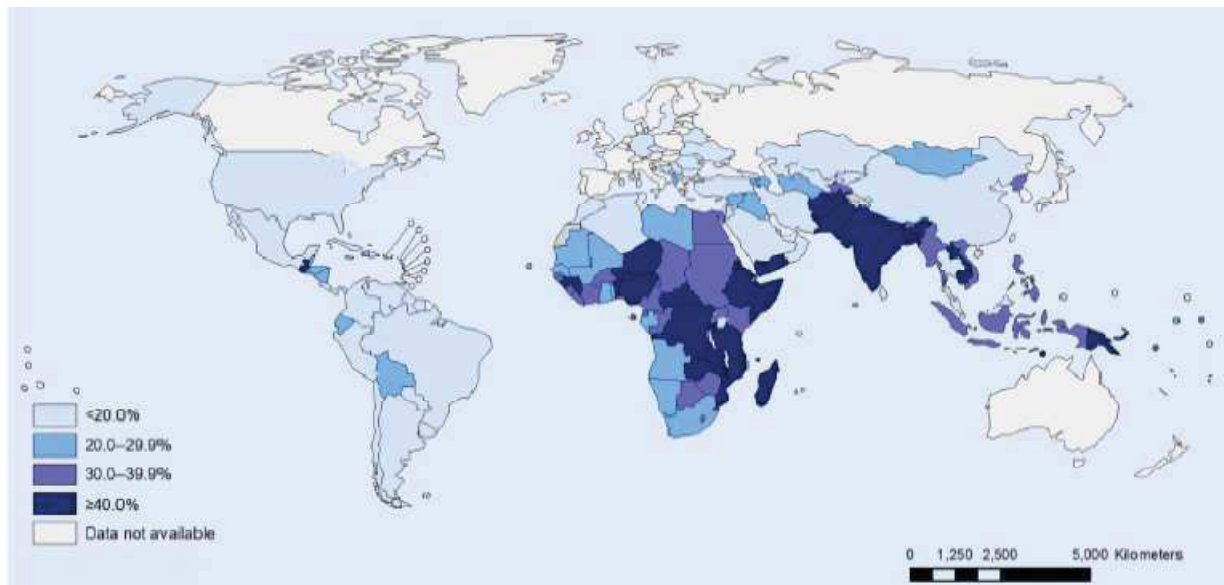
L'assemblée mondiale de la santé (AMS), qui est l'organe décisionnel de l'OMS, a fixé en 2012 un plan d'application de six objectifs visant à améliorer la nutrition chez la mère, le nourrisson et le jeune enfant avec notamment comme cibles le retard de croissance et l'émaciation chez l'enfant. Cette assemblée a fixé pour objectif de réduire de 40 % le nombre d'enfants de moins de 5 ans souffrant de retard de croissance en 2025 [173].

Le retard de croissance chez les enfants de moins de 5 ans dans le monde en général et en Afrique subsaharienne en particulier fait partie des Objectifs de Développement Durable (ODD). L'objectif lié à la malnutrition est le deuxième ODD qui consiste à « *éliminer la faim, assurer la sécurité alimentaire, améliorer la nutrition et promouvoir l'agriculture durable* » en 2030 et sa deuxième cible vise à « *atteindre d'ici 2025 les cibles adoptées au niveau international sur le retard de croissance et la malnutrition aigüe dont souffrent les enfants de moins de 5 ans* ». Le défi d'éliminer la faim dans le monde en 2030 vise également celui d'avoir aucun enfant de moins de 2 ans souffrant de retard de croissance comme le souligne le rapport final de consultation sur la sécurité alimentaire formulé par une Task Force en 2014 [173–175].

En 2011, l'examen de la distribution spatiale de la prévalence du retard de croissance chez les enfants de moins de 5 ans montrait que les prévalences étaient très élevées notamment en Papouasie Nouvelle-Guinée (en Océanie), en Asie du Sud et dans certains pays d'Afrique subsaharienne tels que la Guinée-Conakry, le Bénin, le Niger, le Nigéria, la République Centrafricaine (RCA), la RDC, le Burundi, le

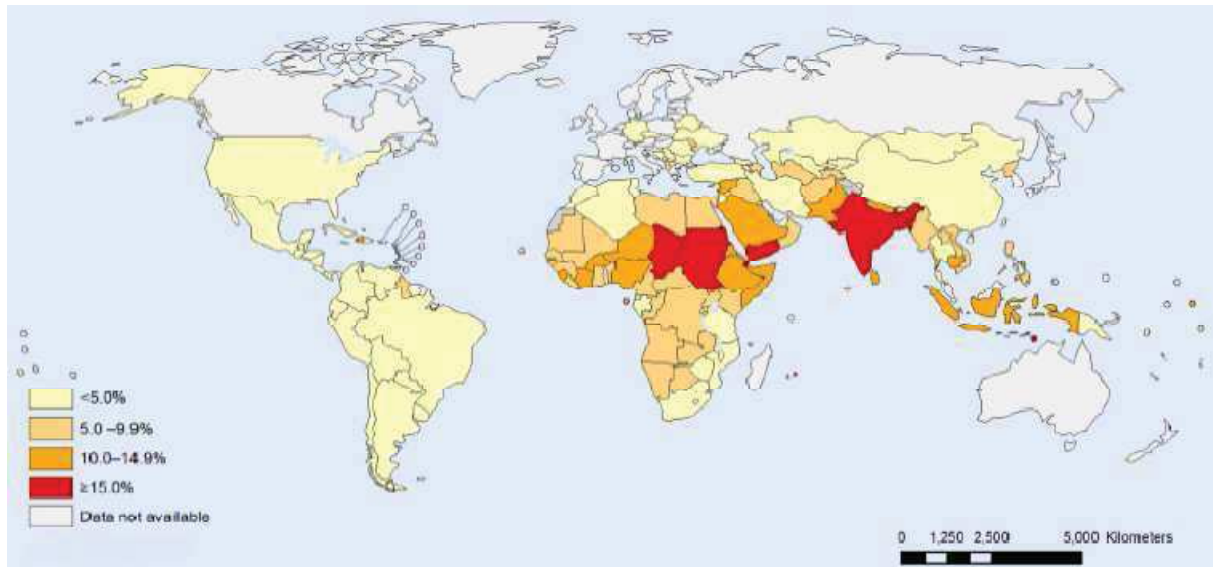
Rwanda, la Tanzanie, la Somalie, l’Ethiopie, l’Érythrée, la Zambie, le Mozambique, le Malawi, Madagascar et les Comores (**Figure 50**) [172]. Cette figure montre que le retard de croissance est plus fréquent en Afrique subsaharienne que dans la plupart des régions du monde.

Figure 50 : Prévalence du retard de croissance dans le monde, 2011



Source : UNICEF-OMS-Banque Mondiale, 2012

Au cours de la même année, l’examen de la distribution spatiale de la prévalence de l’émaciation montre que la prévalence de l’émaciation était très élevée notamment au Tchad, au Soudan (du Sud et du Nord), à Djibouti et en Asie du Sud (**Figure 51**).

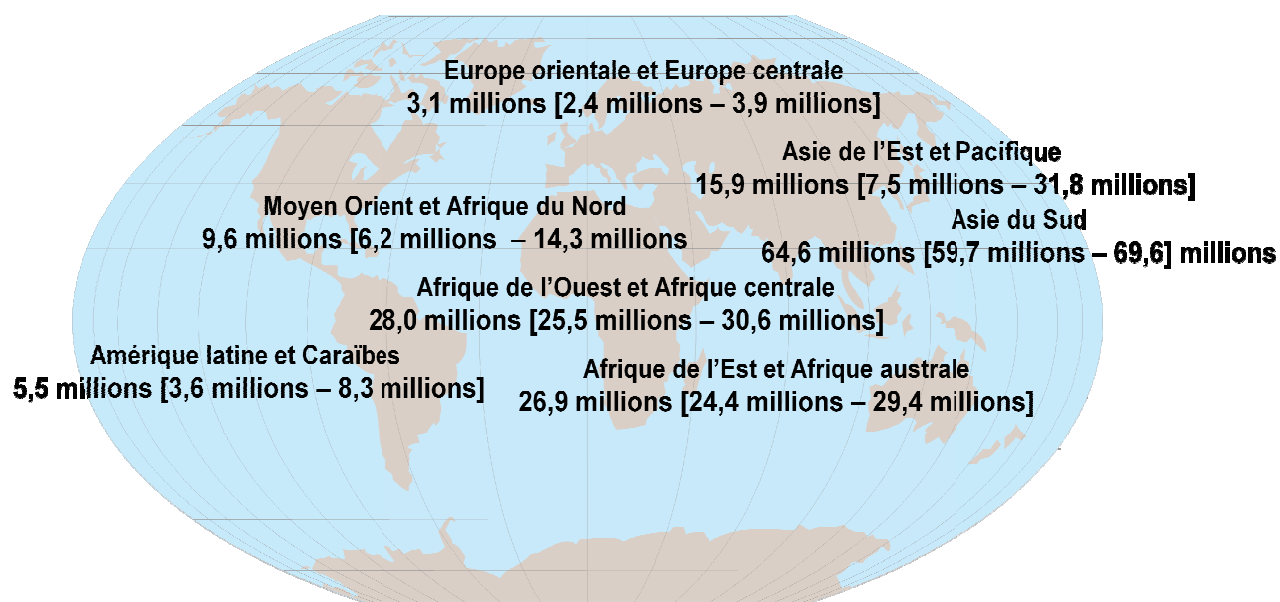
Figure 51 : Prévalence de l'émaciation dans le monde, 2011

Source : UNICEF-OMS-Banque Mondiale, 2012

Selon les estimations de l'OMS, 5 enfants de moins de 5 ans mourraient chaque minute en Afrique en 2013. Au cours de cette année, la malnutrition incluant le retard de croissance était responsable de la mort de 3 millions d'enfants de moins de 5 ans chaque année en Afrique subsaharienne, ce qui représentait 45,4 % des décès [176].

En 2014, l'UNICEF, l'OMS et la Banque Mondiale estimaient conjointement à environ 159 millions [146,1 millions-171,1 millions] le nombre d'enfants âgés de moins de 5 ans dans le monde qui souffraient de retard de croissance (**Figure 52**). La répartition du nombre d'enfants de moins de 5 ans souffrant de retard de croissance dans le monde était inégale. En effet, plus des trois quarts (77,2 %) se trouvaient en Asie du Sud et en Afrique subsaharienne, ces deux parties du monde représentant respectivement 40,7 % et 36,5 % des enfants souffrant de retard de croissance.

Figure 52 : Nombre estimé d'enfants de moins de 5 ans souffrant du retard de croissance dans le monde selon les régions des Nations Unies, 2014

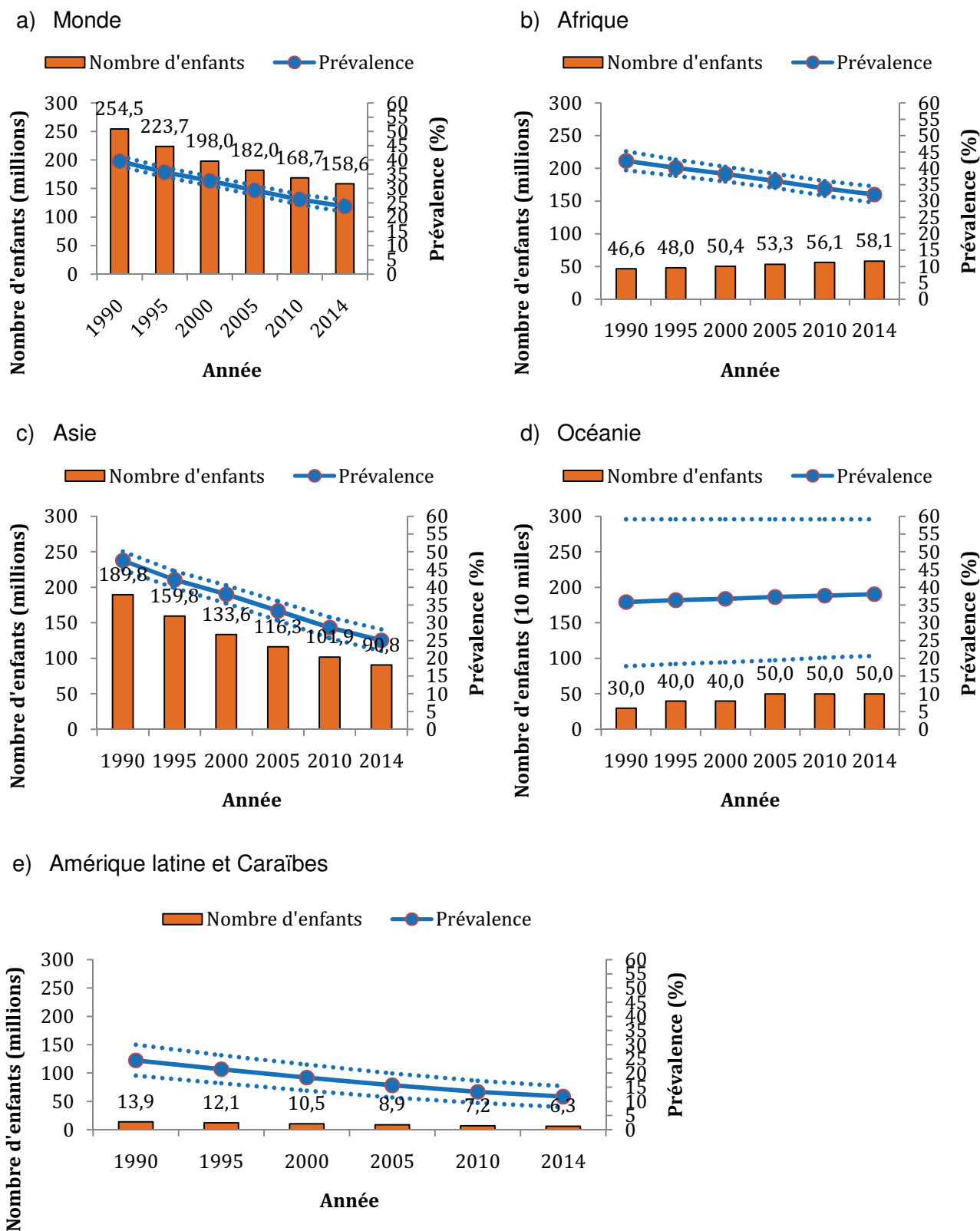


Total: 158,6 millions [146,1 millions – 171,1 millions]

Source : À partir des données de l'UNICEF-OMS-Banque Mondiale, 2015

Au cours de la même année, le nombre d'enfants de moins de 5 ans qui souffraient d'émaciation était estimé à 50 millions parmi lesquels 16 millions souffraient d'émaciation sévère. La prévalence de l'émaciation était estimée à 7,5 % [172]. Le nombre d'enfants de moins de 5 ans vivant dans le monde étant estimé à 667 millions, la prévalence du retard de croissance était estimée à 23,8 % [21,9 %-25,7 %] (**Figure 53**).

Figure 53 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant de retard de croissance dans le monde et par région, 1990-2014



Source : À partir des données UNICEF-OMS-Banque mondiale, 2015

Au cours de l'année 2014, les prévalences du retard de croissance chez les enfants de moins de 5 ans au-dessus de la moyenne selon les régions des Nations Unies étaient observées en Océanie excluant l'Australie et la Nouvelle Zélande (38,1 % ; 20,7 %-59,2 %), en Afrique (32,0 % ; 29,5 %-34,5 %) et en Asie (25,1 % ; 22,0 %-28,2 %). En 1990, le nombre d'enfants âgés de moins de 5 ans dans le monde qui souffraient du retard de croissance était estimé à 255 millions (**Figure 54**), ce qui représentait une prévalence de 39,6 % [37,8 %-41,4 %]. Le nombre d'enfants de moins de 5 ans souffrant du retard de croissance a donc diminué de 96 millions (38,0 %) et la prévalence du retard de croissance de 40,0 % de 1990 à 2014.

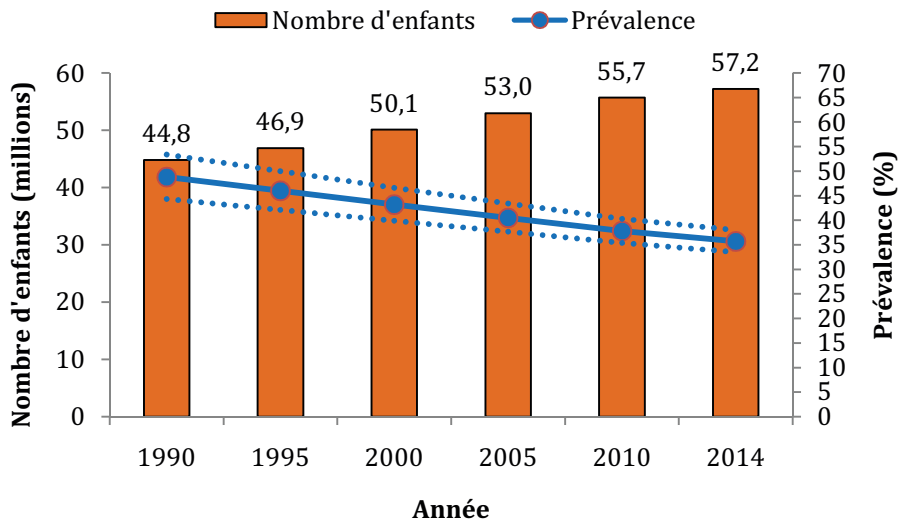
De plus, la malnutrition infanto-juvénile et les conditions maternelles sont parmi les 20 principales causes de mortalité et de morbidité en Afrique : la malnutrition était la 9^{ème} cause (2,8 %) et la 7^{ème} cause (2,9 %) d'années de vie perdues en bonne santé respectivement en 2000 et en 2012. Le fardeau de la malnutrition est passé de 19,83 millions de DALYs en 2000 à 18,83 millions de DALYs en 2012 en Afrique.

3.4. Situation nutritionnelle en Afrique subsaharienne

La prévalence du retard de croissance chez les enfants de moins de 5 ans en Afrique est passée de 42,3 % [39,4 %-45,3 %] en 1990 à 32,0 % [29,5 %-34,5 %] en 2014, soit une diminution de 24,3 %. L'Afrique subsaharienne reste la région du monde la plus touchée par le retard de croissance.

En Afrique subsaharienne, selon la classification de l'OMS, le nombre d'enfants de moins de 5 ans souffrant de retard de croissance est passé de 44,9 millions [40,7 millions-49,1 millions] en 1990 et à 57,3 millions [53,8 millions-60,9 millions] en 2014, soit une augmentation de 12,4 millions. La prévalence, quant elle, est passée de 48,8 % [44,3 %-53,4 %] en 1990 à 35,7 % [33,5 %-38,0 %] en 2014, soit une diminution de 27,0 % (**Figure 55**).

Figure 54 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant de retard de croissance en Afrique subsaharienne, 1990-2014



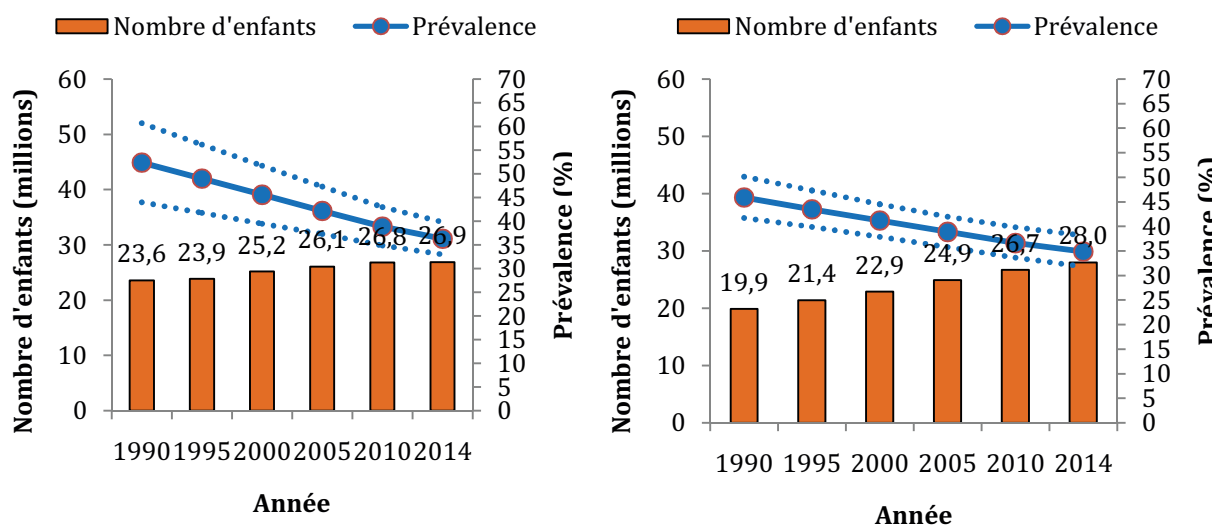
Source : À partir des données UNICEF-OMS-Banque mondiale, 2015

En Afrique subsaharienne, la prévalence du retard de croissance chez les enfants de moins de 5 ans est plus élevée dans la partie englobant l'Afrique de l'Est et l'Afrique australe que dans celle englobant l'Afrique de l'Ouest et l'Afrique centrale. En effet, selon la classification de l'UNICEF, le nombre d'enfants souffrant de retard de croissance en Afrique de l'Est et en Afrique australe est passé de 23,6 millions [19,8 millions-27,3 millions] en 1990 à 26,9 millions [24,4 millions-29,4 millions] en 2014, soit une augmentation de 3,3 millions (14,0 %) (**Figure 55**). Quant à la prévalence, elle est passée de 52,4 % [44,0 %-60,7 %] en 1990 à 36,3 % [33,3 %-39,8 %] en 2014, soit une diminution de 30,7 %. De même, le nombre d'enfants souffrant de retard de croissance en Afrique de l'Ouest et en Afrique centrale est passé de 19,9 millions [18,1 millions-21,7 millions] en 1990 en 1990 à 28,0 millions [25,5 millions-30,6 millions] en 2014, soit une augmentation de 8,1 millions (40,7 %). Quant à la prévalence, elle est passée de 45,9 % [41,7 %-50,1 %] en 1990 à 34,9 % [31,8 %-38,1 %] en 2014, soit une diminution de 24,0 %.

Figure 55 : Évolution de la prévalence et du nombre d'enfants de moins de 5 ans souffrant du retard de croissance en Afrique subsaharienne par région, 1990-2014

a) Afrique de l'Est et Afrique australe

b) Afrique de l'Ouest et Afrique centrale



Source : À partir des données UNICEF-OMS-Banque mondiale, 2015

3.5. Situation nutritionnelle en Côte d'Ivoire et dans les pays limitrophes

Le retard de croissance constitue un problème de santé publique en Côte d'Ivoire. En effet, les résultats de l'EDS conduite en 2010 ont montré des prévalences du retard de croissance et de l'émaciation chez les enfants de moins de 5 ans respectivement égales à 29,8 % et 7,5 % [177] (**Tableau 9**).

Tableau 9 : Prévalences du retard de croissance et de l'émaciation chez les enfants de moins de 5 ans en Côte d'Ivoire et dans les pays limitrophes

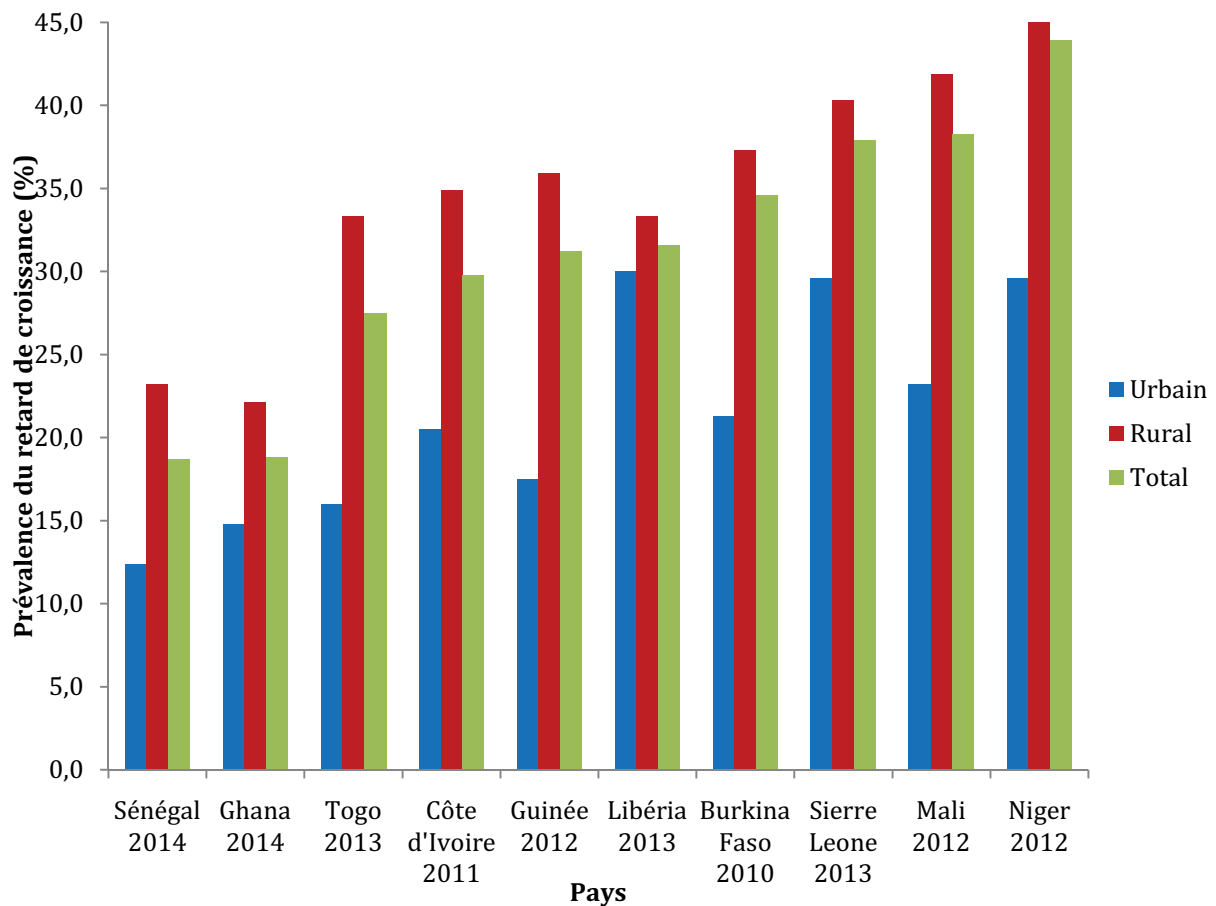
Pays	Année	Retard de croissance	Émaciation
Côte d'Ivoire	2011	29,8 %	7,5 %
Ghana	2014	18,8 %	4,7 %
Libéria	2013	31,6 %	6,0 %
Guinée	2012	31,2 %	9,6 %
Mali	2012	38,3 %	12,7 %
Burkina Faso	2010	34,6 %	15,5 %

Source : À partir des rapports des EDS

En 2014, la prévalence du retard de croissance était estimée à 29,6 % selon le Mouvement international SUN (Scaling Up Nutrition) [178]. Le nombre d'enfants souffrant du retard de croissance était estimé à 1,162 millions en 2013.

Une EDS conduite au Ghana en 2014 a montré une prévalence nationale du retard de croissance de 18,8 % avec des disparités selon le milieu de résidence (14,8 % en milieu urbain contre 22,1 % en milieu rural) (**Figure 56**) [179].

Figure 56 : Prévalence du retard de croissance chez les enfants de moins de 5 ans en Afrique de l'Ouest selon le milieu de résidence

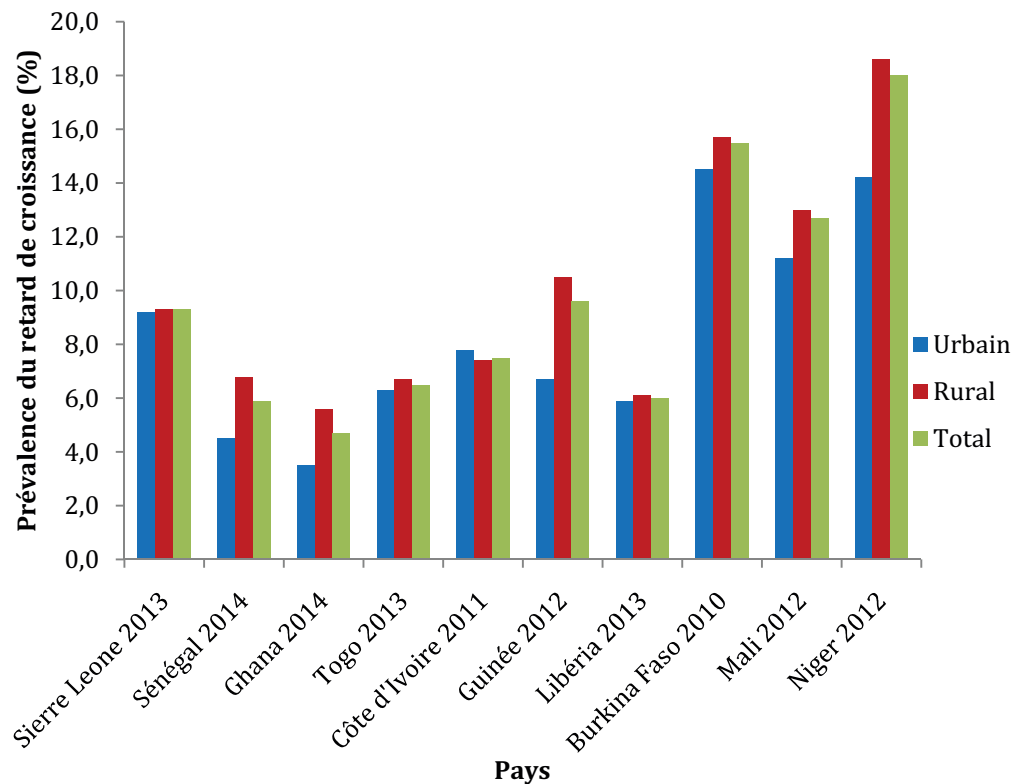


Source : À partir des rapports des EDS

La prévalence de l'émaciation était, quant à elle, estimée à 4,7 % au niveau national, 3,5 % en milieu urbain et 5,6 % en milieu rural (**Figure 57**). De même, au Libéria, une EDS réalisée en 2013 a montré une prévalence globale du retard de croissance de 31,6 % avec des inégalités selon le milieu de résidence (30,0 % en milieu urbain et 33,3 % en milieu rural) [180]. La prévalence de l'émaciation était, quant à elle,

estimée à 6,0 % globalement, 5,9 % en milieu urbain et 6,1 en milieu rural. Au Burkina Faso, l'EDS réalisée en 2010 a montré une prévalence du retard de croissance de 34,6 % au niveau national, 21,3 % en milieu urbain et 37,3 % en milieu rural. La prévalence de l'émaciation était de 15,5 % en population générale, 14,5 % en milieu urbain et 15,7 % en milieu rural [181]. Une EDS a été conduite au Mali en 2012 et a montré une prévalence du retard de croissance de 38,3 % au niveau national avec des inégalités selon le milieu de résidence (23,2 % en milieu urbain contre 41,9 % en milieu rural) [182]. Le mouvement SUN a estimé la prévalence du retard de croissance à 32,9 % au Burkina Faso, 18,8 % au Ghana, 31,3 % en Guinée, 32,1 % au Libéria et 38,5 % au Mali en 2014 [178].

Figure 57 : Prévalence de l'émaciation chez les enfants de moins de 5 ans en Afrique de l'Ouest selon le milieu de résidence

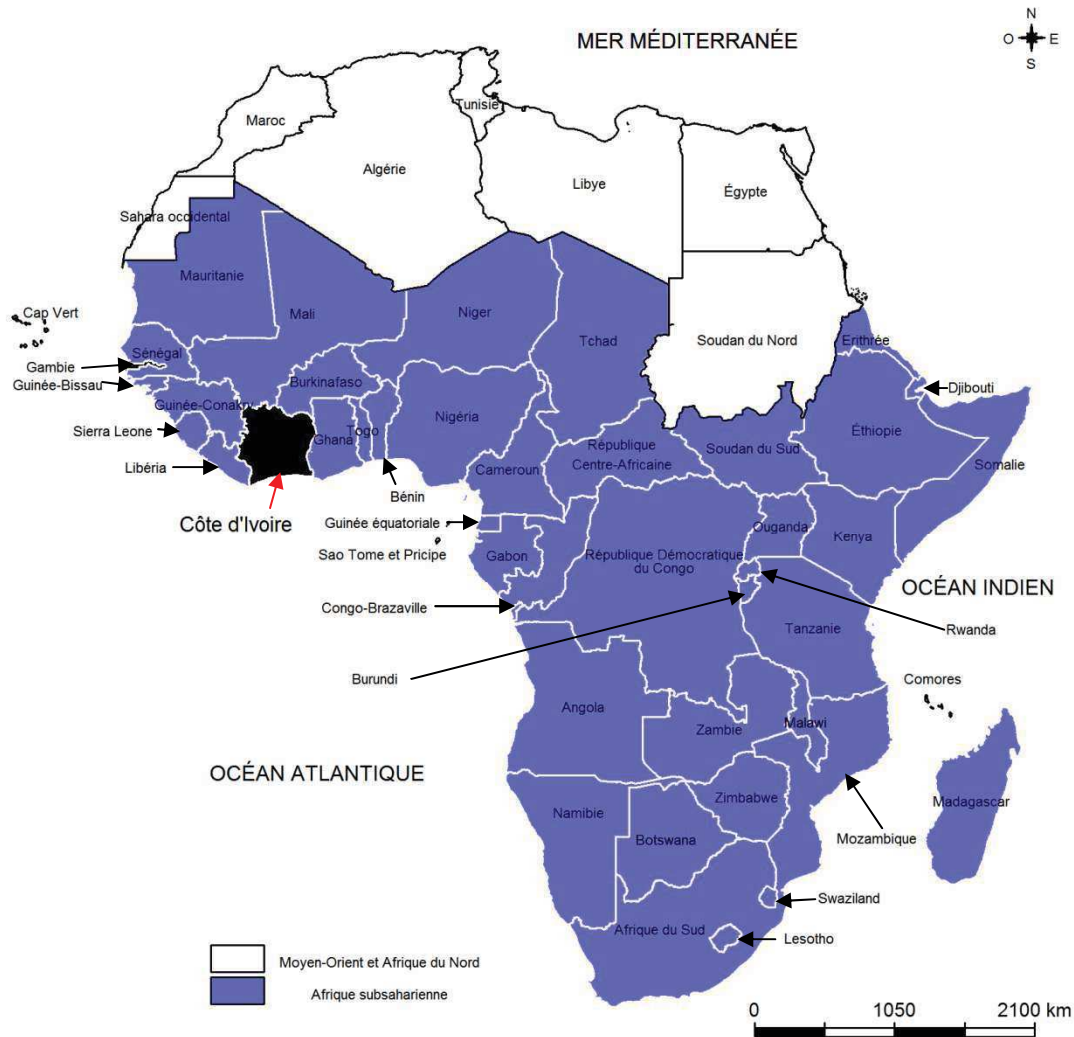


Source : À partir des rapports des EDS

3.6. Contexte de la Côte d'Ivoire

3.6.1. Situation géographique, administrative et démographique

La Côte d'Ivoire est un pays d'Afrique de l'Ouest situé entre 4°30' et 10°30' de latitude nord et entre 2°30' et 8°36' de longitude ouest. Abidjan et Yamoussoukro sont respectivement les capitales économique et administrative. Limitée entièrement au sud par l'Océan Atlantique à l'endroit du Golfe de Guinée, la Côte d'Ivoire est l'un des pays les plus vastes d'Afrique avec une superficie de l'ordre de 322462 km² dont plus de la moitié (167100 km²) servent aux productions agricoles. Elle a une frontière commune à l'ouest avec le Libéria et le Guinée Conakry, et au nord avec le Mali et le Burkina Faso. Un seul pays à l'est, le Ghana, lui est frontalier (**Figure 58**).

Figure 58 : Situation de la Côte d'Ivoire en Afrique subsaharienne

Sa latitude varie entre moins de 350 m (bas plateaux) et 1753 m (Mont Nimba). Les principaux fleuves de la Côte d'Ivoire sont le fleuve Bandama (1050 km de longueur), le fleuve Comoé (1160 km), le fleuve Sassandra (600 km) et le fleuve Cavally (700 km). En 2008, la Côte d'Ivoire était divisée en 12 districts (hormis les districts autonomes d'Abidjan et de Yamoussoukro) et 30 régions administratives (hormis Abidjan). Ces régions administratives étaient réparties en 10 régions statistiques (hormis Abidjan) qui constituaient les pôles de développement. Les régions administratives étaient divisées en 108 départements ou préfectures, les départements en 510 sous-préfectures et les sous-préfectures en 197 communes [177]. Le pays est divisé en deux principales zones : la zone forestière au Sud et la savane au Nord (**Figure 59**). Du Nord au Sud, le pays est divisé en 4 zones

climatiques et écologiques : la zone soudanaise méridionale (climat chaud et sec), la zone de savane herbeuse, la région forestière et la zone côtière de pêche et des cocoteraies au Sud du pays.

Figure 59 : Répartition spatiale des zones de savane et des zones de forêt en Côte d'Ivoire, 1994



Source : Adapté de l'EDSCI, 1994

3.6.2. Recensements de population en Côte d'Ivoire

Le Recensement Général de la Population et de l'Habitat (RGPH) de 1988 a montré une population de 10 815 694 habitants, soit une densité de population d'environ 34 habitants/km². La proportion de population urbaine était estimée à 39,0 %, celle des

moins de 15 ans à 46,8 %. Le RGPH suivant, en 1998, a dénombré 15 366 672 habitants, soit une densité de population d'environ 48 habitants/km². La proportion de population urbaine était estimée à 42,5 %, celle des moins de 15 ans à 42,9 %. Enfin, le dernier RGPH, réalisé au mois de mai 2014, a montré que la population ivoirienne s'élevait à 22 671 331 habitants, soit une densité d'environ 71 habitants/km² [183]. La proportion de la population urbaine était estimée à 50,3 % et la proportion des moins de 15 ans à 41,8 %.

3.6.3. Crise ivoirienne et ses conséquences

Depuis août 1960, date de l'indépendance du pays, la Côte d'Ivoire est dirigée comme une République. Le cacao (et la noix de cajou) d'une part et le café d'autre part dont la Côte d'Ivoire est respectivement le premier et le cinquième producteur mondial en 1998 sont des cultures qui en font le pays le plus influent économiquement en Afrique de l'Ouest. Ainsi, la Côte d'Ivoire a longtemps été une terre d'immigration des habitants des pays voisins soit à la recherche de terres ou d'emplois, soit fuyant les guerres dans leur pays d'origine. Les années 1990 furent caractérisées par l'émergence du multipartisme dans les pays africains. Après la mort du premier président ivoirien (Félix Houphouët-Boigny) en 1993, le pays se heurte à des problèmes de succession provoquant des troubles sur tout le territoire national. Le deuxième président (Henri Konan Bédié), élu en 1995, dirige le pays jusqu'en décembre 1999, date à laquelle un premier coup d'État est opéré. Ce coup d'État conduit à l'élection d'un troisième président (Laurent Bagbo) en octobre 2000 lors d'une élection largement controversée et boycottée par l'opposition. La guerre civile en Côte d'Ivoire, débutée en septembre 2002 suite à l'éclatement des conflits cumulés liés aux problèmes de succession au poste de président de la République, n'a pas manqué d'effets néfastes sur les familles et les ménages. Ces conséquences sont entre autres les déplacements des populations fuyant la guerre, la non-accessibilité aux champs, la destruction des maisons et des réserves de nourriture, le vol des biens et du bétail par les forces rebelles, la dégradation de l'état sanitaire des enfants et de leurs mères, et le manque de nourriture. Durant la période de guerre civile (2002-2007), le pays était divisé en deux zones : le nord contrôlé par les forces rebelles (les « forces nouvelles ») et le sud contrôlé par les forces

gouvernementales (loyalistes), les deux zones étant séparées par une zone de confiance contrôlée par les forces neutres de la communauté internationale (**Figure 60**).

Figure 60 : Position de la zone de confiance, flux et zones de concentration de populations déplacées de guerre



La signature de l'Accord Politique de Ouagadougou (APO) en 2007 avait donné un espoir sur la cessation des hostilités [184]. De 2000 à 2011, le pays est dirigé par Laurent Bagbo. Après la signature de l'APO du 4 mars 2007, les partis politiques intensifient les campagnes électorales et Laurent Bagbo fait campagne pour sa propre succession à l'élection présidentielle qui, initialement prévue en octobre 2008,

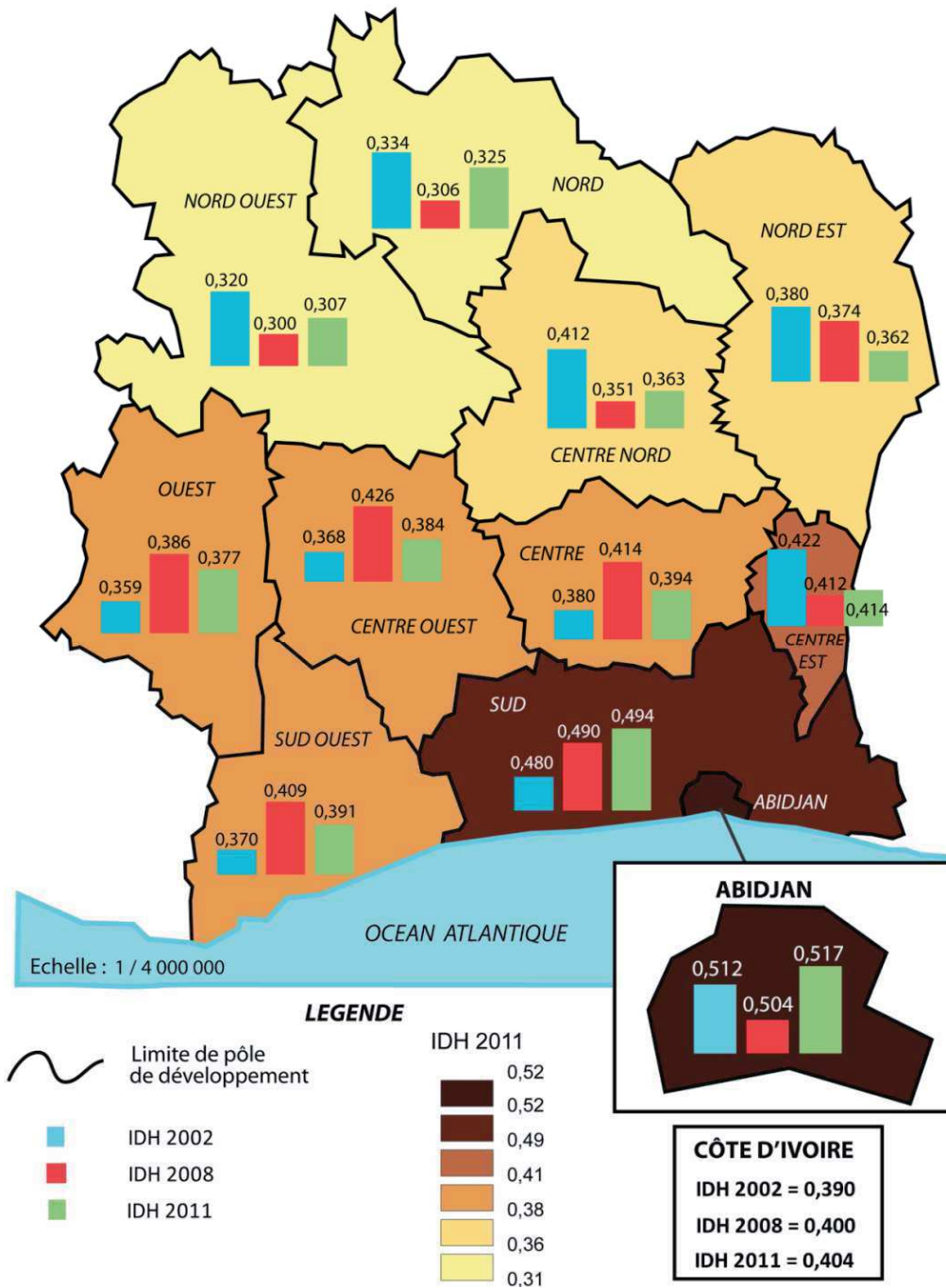
n'a pu se réaliser qu'en 2010 à cause du climat défavorable aux élections (tensions entre les communautés, grèves, massacres, manifestations violentes contre la vie chère et l'exploitation des ressources naturelles, vols, viols) [185]. Pendant la période de 2007 à 2010, la liste électorale a été élaborée, les forces rebelles ont été désarmées, les unes démobilisées et les autres réintégrées au sein de l'armée loyale, les milices ont été démantelées, la population a été rééduquée en matière de protection des droits de l'homme et de maintien de la paix, et le pays est réunifié [186–188]. Les élections présidentielles ont été organisées en octobre et novembre 2010. Leurs résultats ont été largement contestés, ce qui a provoqué de nouveau un climat d'instabilité politique et économique, des violences, des déplacements internes et externes des populations. Les effets de cette crise sociopolitique se sont fait sentir davantage dans la région de l'Ouest où les infrastructures tant publiques que privées ont été détruites durant les périodes de guerre.

Selon le Programme de Nations Unies pour le Développement (PNUD), l'IDH (Indice de Développement Humain), qui était de 0,39 en 1990, est passé de 0,40 en 2000 à 0,45 en 2011 et à 0,46 en 2014 [189]. Le Rapport National sur le Développement Humain (RNDH), quant à lui, montre que l'IDH s'est globalement amélioré en Côte d'Ivoire de 2002 à 2008, puis de 2008 à 2011 mais qu'il s'est détérioré dans les régions du Nord entre 2002 (avant la guerre civile) à 2011 (après la guerre civile) (**Figure 61**) [190].

La crise de 2002 a pu affecter la sécurité alimentaire des personnes (c'est-à-dire le nombre de repas par jour et leur diversification), étant donné que les déplacés de guerre notamment ceux qui ne sont pas regroupés, ont moins de chances d'avoir accès à l'aide alimentaire, ce qui affecte principalement l'état sanitaire et nutritionnel des enfants au bas âge. Les zones les plus au nord de la Côte d'Ivoire en général et la région du Nord en particulier, moins peuplées que les régions du Sud et du Sud-Ouest, ont été sujettes à la désertion massive des services publics et des personnels de l'éducation et des services de santé, ce qui les isole économiquement du reste du pays. Cette crise a également forcé les gens du Centre et de l'Ouest à fuir le pays ou à se déplacer vers les zones contrôlées par les forces gouvernementales [191]. Les

migrants internes étaient concentrés dans les villes (81,0 %) et principalement à Abidjan (67,5 %) [191].

Figure 61 : Évolution de l'IDH selon les régions, 2002-2011



Source : RNDH, 2013

3.6.4. Lutte contre la malnutrition en Côte d'Ivoire

La lutte contre la malnutrition a donné naissance à un Programme National de Nutrition (PNN). Le PNN organise des activités diverses de lutte contre la malnutrition telles que la sensibilisation à de bonnes pratiques alimentaires, les démonstrations culinaires et les publicités et encourage l'allaitement maternel exclusif des enfants de 0 à 6 mois et de l'alimentation de complément de 6 à 24 mois.

Le Comité National de Nutrition (CNN) et le Groupe Scientifique d'Appui (GSA) appuient la Direction de Coordination du PNN (DC-PNN) dans la lutte contre la malnutrition au sein du Ministère de la Santé et de l'Hygiène Publique (MSHP). Cette DC-PNN contribue à réduire la mortalité et la morbidité liées au manque de nutriments et de micronutriments par la population en général et par des groupes les plus vulnérables tels que les femmes en âge de procréer, les femmes enceintes et les enfants de moins de 5 ans en particulier [192]. Le CNN est un organe de coordination et de suivi des interventions nutritionnelles dans le pays. La DC-PNN est donc dotée d'un Plan National Stratégique (PNS), d'un plan d'action et des programmes. Le CNN est décentralisé en des Comités Régionaux de Nutrition (CRN) dont les secrétariats sont tenus par les Directions Régionales de la Santé et de l'Hygiène Publique (DRSHP) [192].

Dans la lutte contre la malnutrition, le CNN est appuyé par divers programmes tels que :

- le Programme Élargi de Vaccination (PEV) qui participe à la mise à jour du statut vaccinal des enfants d'âge scolaire et préscolaire et à la vaccination proprement dite des enfants des écoles secondaires ;
- la DC du Programme National de Santé Infantile et de Survie de l'Enfant (DC-PNSI/SE) qui sensibilise les mères et les enfants à consommer des aliments riches en micronutriments et à adopter des bonnes pratiques alimentaires ;
- la Direction de la Santé Communautaire et de la Médecine de Proximité (DSCMP) ;

- la DC du Programme National de la Santé Scolaire et Universitaire (DC-PNSSU) qui coordonne l'activité de supplémentation des élèves en vitamine A et en micronutriments afin de prévenir ou guérir les maladies;
- le centre SAVALGI de Nutrition de Korhogo qui dépiste et prend en charge la malnutrition, prodigue des conseils nutritionnels, fait la sensibilisation à la lutte contre la malnutrition [193];

Il est également appuyé par les organisations non gouvernementales ou projets tels que :

- l'Animation Rurale de Korhogo (ARK) qui intervient dans la prévention des effets de la malnutrition principalement dans les régions du Nord ;
- le Fonds des Nations Unies pour l'enfance (UNICEF), Action contre la Faim (ACF), le Programme alimentaire mondial (PAM) et l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) qui ont mis en place en 2008 un plan permettant d'intégrer la prise en charge des enfants souffrant de la malnutrition aigüe dans les services étatiques ;
- le projet « Approche communautaire pour la réhabilitation nutritionnelle et le changement des comportements alimentaires en Côte d'Ivoire » à travers lequel l'association « ACF/PNN » a utilisé en 2008 l'approche « Foyer d'Apprentissage et de Réhabilitation Nutritionnelle (FARN) » pour récupérer les enfants atteints de la malnutrition aigüe modérée [194];
- le programme intégré d'éducation nutritionnelle (PIEN), un projet qui fait bénéficier les populations d'une éducation nutritionnelle pour leur permettre d'améliorer et de diversifier leur alimentation afin de lutter contre l'insécurité alimentaire et la malnutrition. Cette éducation est appuyée par la mise en place des pépinières et des intrants pour le potager ainsi que des démonstrations culinaires [195]. Ce programme opère dans les Centres Nutritionnels thérapeutiques (CNT), les Centres Nutritionnels Supplémentaires (CNS) et les Centres Nutritionnels communautaires (CNC). Des CNT ont également été implantés dans les régions du Centre, Nord et Ouest (CNO) par ACF après le début de la guerre civile de 2002 afin de venir en aide aux populations de ces localités ;

- le centre de Nutrition Supplémentaire Focolari (CNS), né dans la ville de Man de la région de Tonkpi, a contribué à la sensibilisation aux causes et aux conséquences de la malnutrition et aux moyens de lutte contre la malnutrition, tout cela dans le but de contribuer à la réduction de la mortalité. Des programmes d'allaitement exclusif au sein et de prévention de la malnutrition sont respectivement dirigés vers les enfants de 0 à 6 mois et ceux de 6 à 24 mois. Les femmes enceintes et les enfants déjà atteints de malnutrition sont également suivis dans ce centre.

3.7. Contexte de notre recherche en Côte d'Ivoire

Une étude réalisée à partir des enquêtes conduites dans 141 pays en développement de 1985 à 2011 a permis d'estimer, pour chaque région, l'évolution des prévalences du retard de croissance et de l'émaciation et, pour chaque pays, les probabilités d'atteindre le premier OMD. Elle a montré une forte probabilité d'augmentation de la prévalence du retard de croissance chez les enfants de moins de cinq ans en Côte d'Ivoire entre 1986 et 2010 contrastant avec l'amélioration générale en Afrique [196]. A l'inverse, une étude centrée sur la Côte d'Ivoire suggérait une diminution de 4,2 % de la prévalence du retard de croissance chez les enfants âgés de moins de 2 ans entre 1994 à 2011 [197].

Des différences de prévalence du retard de croissance selon la place de résidence (savane versus milieu urbain) ont été rapportées lors d'une étude réalisée en 1990 en Côte d'Ivoire [198]. En 2003, une étude transversale réalisée dans le nord de la Côte d'Ivoire chez 560 enfants âgés de moins de cinq ans a montré une prévalence du retard de croissance plus élevée en milieu rural qu'en milieu urbain (39,9 % versus 16,7 %, $P < 0,001$) [199]. En milieu urbain, la prévalence du retard de croissance était significativement associée à l'absence d'allaitement jusqu'à l'âge de deux ans (OR=3,58, IC à 95 % : 0,83-13,35, $P=0,045$). En milieu rural, par contre, elle était significativement associée aux maladies diarrhéiques (OR=2,10, IC à 95 % : 1,01-4,35, $P=0,043$), à l'émaciation de la mère (OR=2,81, $P=0,028$) et à la consommation d'eau non potable (OR=1,69, IC à 95 % : 1,02-2,79, $P=0,042$). Une

étude sur l'état sanitaire de l'enfant en Côte d'Ivoire montre une détérioration entre 2002 et 2007 dans les endroits affectés par des conflits sociopolitiques tels que Guiglo, Bouaké et Korhogo [200].

À notre connaissance, il n'existe pas d'études qui se sont intéressées à l'analyse spatiale et temporelle de la prévalence du retard de croissance en Côte d'Ivoire.

3.8. Objectifs

L'objectif de notre étude était d'évaluer, d'une part, l'hétérogénéité spatiale de la prévalence du retard de croissance chez les enfants de moins de 3 ans en 1994, 1998 (avant la guerre civile) et 2011 (après la guerre civile). D'autre part, il s'agissait d'examiner les changements dans la prévalence du retard de croissance de 1994 à 2011, globalement et puis dans chacune des trois zones définies par nos soins (Abidjan, le Sud du pays hors Abidjan et le Nord du pays). Notre hypothèse de recherche était que la prévalence du retard de croissance avait augmenté de 1994 à 2011 dans certains endroits de la Côte d'Ivoire, et plus particulièrement dans la zone du Nord qui était économiquement isolée du reste du pays durant la guerre civile [201].

3.9. Méthodologie

3.9.1. Sélection de la population d'étude

Nous avons fait une analyse secondaire des données des EDS conduites en Côte d'Ivoire en 1994, 1998 et 2011.

3.9.1.1 Méthodologie de l'EDS de 1994

La première EDS en Côte d'Ivoire (EDSCI) a été réalisée entre les mois de juin et de novembre 1994 par l'Institut National de la Statistique (INS) de la Côte d'Ivoire et Macro International [202]. Elle était basée sur quatre questionnaires dont un

questionnaire ménage et un questionnaire individuel femme pour les femmes âgées de 15 à 49 ans. Cette EDS était stratifiée à deux ou à trois degrés selon la place de résidence. Il y avait cinq strates : (1) Abidjan, (2) forêt urbaine, (3) forêt rurale, (4) savane urbaine et (5) savane rurale. La zone de forêt (forêt dense, ombrophile et mésophile) est située au sud du pays tandis que la zone de savane (forêts clairsemées, savanes soudanaises, savanes pré-forestières) est située au nord [202]. La zone de forêt est, en grande partie, constituée des régions statistiques d'Abidjan, du Sud, du Sud-Ouest, du Centre-Ouest, de l'Ouest et du Centre-Est. Quant à la zone de savane, elle est constituée essentiellement des régions statistiques du Nord, du Nord-Ouest, du Nord-Est et du Centre-Nord.

Basée sur les 10333 zones de dénombrement (ZD) ou grappes du RGPH de 1988, cette EDS avait prévu de tirer un échantillon de 7000 femmes. À Abidjan, en forêt rurale et en savane rurale, un échantillon de grappes a été tiré au premier degré avec une probabilité proportionnelle à leur taille (i.e. le nombre de personnes) et un échantillon de ménages au deuxième degré avec probabilité égale. Quant aux zones de forêt urbaine et de savane urbaine, une autre stratification en grandes et petites villes a été faite en fonction du nombre de femmes âgées de 15 à 49 ans vivant dans chaque ville ou du nombre d'habitants de chaque ville, ce qui a conduit à 7 strates. Ainsi, une ville de la forêt urbaine était qualifiée de « grande ville » si elle contenait 9631 femmes ou 43000 habitants. De même, une ville de la savane urbaine était qualifiée de « grande ville » si elle contenait 8238 femmes ou 36 800 habitants. Ces effectifs étaient obtenus sur la base du nombre de femmes en 1993 en forêt urbaine (481564) et en savane urbaine (362468) respectivement, la proportion de femmes en milieu urbain autre qu'Abidjan selon des données du RGPH de 1988 (25,4 % à Abidjan, 22,4 % dans les autres villes et 21,5 % en milieu rural) et le nombre de grappes à tirer en forêt urbaine et en savane urbaine (50 et 44 respectivement). En forêt urbaine et en savane urbaine, 14 et 10 villes ont été respectivement identifiées. Un échantillon de villes a été tiré au premier degré, un échantillon de grappes au deuxième degré proportionnellement à leur taille et un échantillon de ménages au troisième degré avec une probabilité égale. Dans chaque grande ville tirée, le nombre de grappes à tirer était proportionnel à la taille de cette ville en 1993. Pour les petites villes, cependant, un échantillon de petites villes a été tiré en fonction du

nombre de grappes prévues dans la strate des petites villes et dans chaque petite ville tirée, une grappe a été tirée. Au total, 246 grappes (146 urbaines et 100 rurales) ont été tirées.

Sur 6348 ménages éligibles, 5935 ont été enquêtés avec succès, soit un taux de réponse de 97,2 %. Pour les enquêtes individuelles, 8271 femmes étaient éligibles. Parmi elles, 8099 femmes (97,9 %) ont été interviewées avec succès. Les mesures anthropométriques (taille, poids) ont été prises chez 3645 enfants de moins de trois ans (0-35 mois) nés des femmes de 15 à 59 ans qui ont été interviewées.

3.9.1.2 Méthodologie de l'EDS de 1998

La deuxième EDS s'est déroulée entre septembre 1998 et mars 1999 et était basée, comme la première, sur les grappes du RGPH de 1988 [203]. Il s'agissait d'une enquête stratifiée à deux degrés basée sur trois questionnaires : un questionnaire ménage, un questionnaire individuel femme pour les femmes âgées de 15 à 49 ans, et un questionnaire individuel homme pour les hommes âgés de 15 à 59 ans. Les strates étaient Abidjan, les autres villes et le milieu rural. Abidjan et les autres villes ont ensuite été stratifiés en zones d'intervention (I) et de non-intervention (NI) déjà définies par le programme Santé Familiale et Prévention du SIDA (SFPS), ce qui a donné naissance à 5 strates. Ces strates étaient Abidjan-I, Abidjan-NI, Autres villes-I, Autres villes-NI, milieu rural où les extensions I et NI font référence aux zones d'intervention et de non-intervention respectivement réparties dans les 19 régions de la Côte d'Ivoire.

Au premier degré, 138 grappes ont été sélectionnées parmi les 246 grappes de l'EDS 1994 avec des probabilités égales et 2 grappes (une dans la zone de non intervention d'Abidjan et une autre dans la ville de Sakassou qui ne tombait pas dans l'échantillon de l'EDS de 1994), soit 140 grappes (100 urbaines et 40 rurales). Au second degré, 2302 ménages ont été sélectionnés avec des probabilités égales. Parmi les ménages, 2122 ont pu être enquêtés avec succès, soit un taux de réponse de 97,9 %. L'échantillon prévu était composé de 3000 femmes. Dans les ménages enquêtés avec succès, 3155 femmes étaient éligibles et parmi elles, 3040 ont été

interviewées avec succès, soit un taux de réponse de 96,4 %. Au cours de cette enquête, l'état nutritionnel des enfants devrait être évalué chez 1909 enfants de moins de cinq ans. Cependant, seulement 1630 enfants de moins de 5 ans (85,4 %) ont pu être pesés et mesurés.

3.9.1.3 Méthodologie de l'EDS de 2011

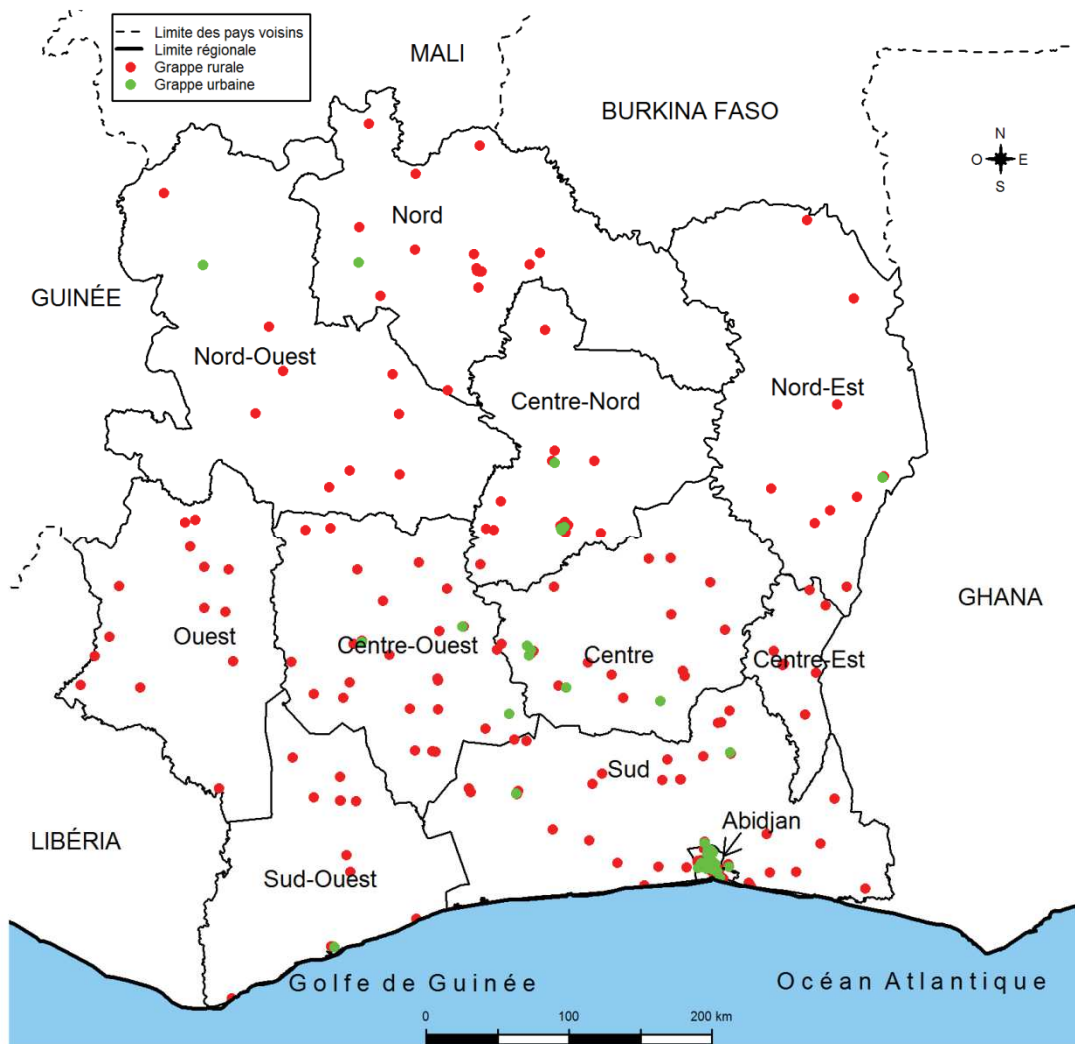
La troisième EDS a été conduite par le Ministère de la Santé et de la Lutte contre le SIDA (MSLS), l'INS et ICF International de décembre 2011 à mai 2012 et était basée sur le RGPH de 1998 [204]. Comme pour l'EDS 1998, il s'agissait d'une enquête stratifiée à deux degrés basée sur trois questionnaires : un questionnaire ménage, un questionnaire individuel femme pour les femmes âgées de 15 à 49 ans, et un questionnaire individuel homme pour les hommes âgés de 15 à 59 ans. Les strates étaient les 11 régions statistiques. Ces régions statistiques proviennent de la fusion des régions administratives (au nombre de 19) et contiennent 58 départements. Dans chacune des 11 régions statistiques à l'exception d'Abidjan, une stratification selon la place de résidence (urbain/rural) a été faite, conduisant ainsi à un total de 21 strates. Au premier degré, 352 grappes (161 urbaines et 191 rurales) ont été sélectionnées avec des probabilités proportionnelles à leur taille (le nombre de ménages) parmi lesquelles 351 ont été enquêtées. Au second degré, 27 ménages et 32 ménages ont été tirés dans chaque grappe urbaine et rurale respectivement avec des probabilités égales, conduisant ainsi à un total de 10413 ménages éligibles parmi lesquels 9686 ont été enquêtés avec succès, soit un taux de réponse de 98,1 %. Sur 10848 femmes éligibles, 10060 ont été interviewés avec succès, soit un taux de réponse de 92,7 %.

3.9.1.4 Notre méthodologie

Les données utilisées pour cette étude proviennent des trois EDS conduites respectivement en 1994, en 1998 et en 2011 [202,203,205]. Comme pour l'étude de la prévalence du VIH au Burundi (chapitre II), ces données nous ont été fournies par MEASURE DHS [152]. Puisque l'EDS 1994 comportait uniquement des enfants de moins de 3 ans alors que celles de 1998 et de 2011 comportaient des enfants de

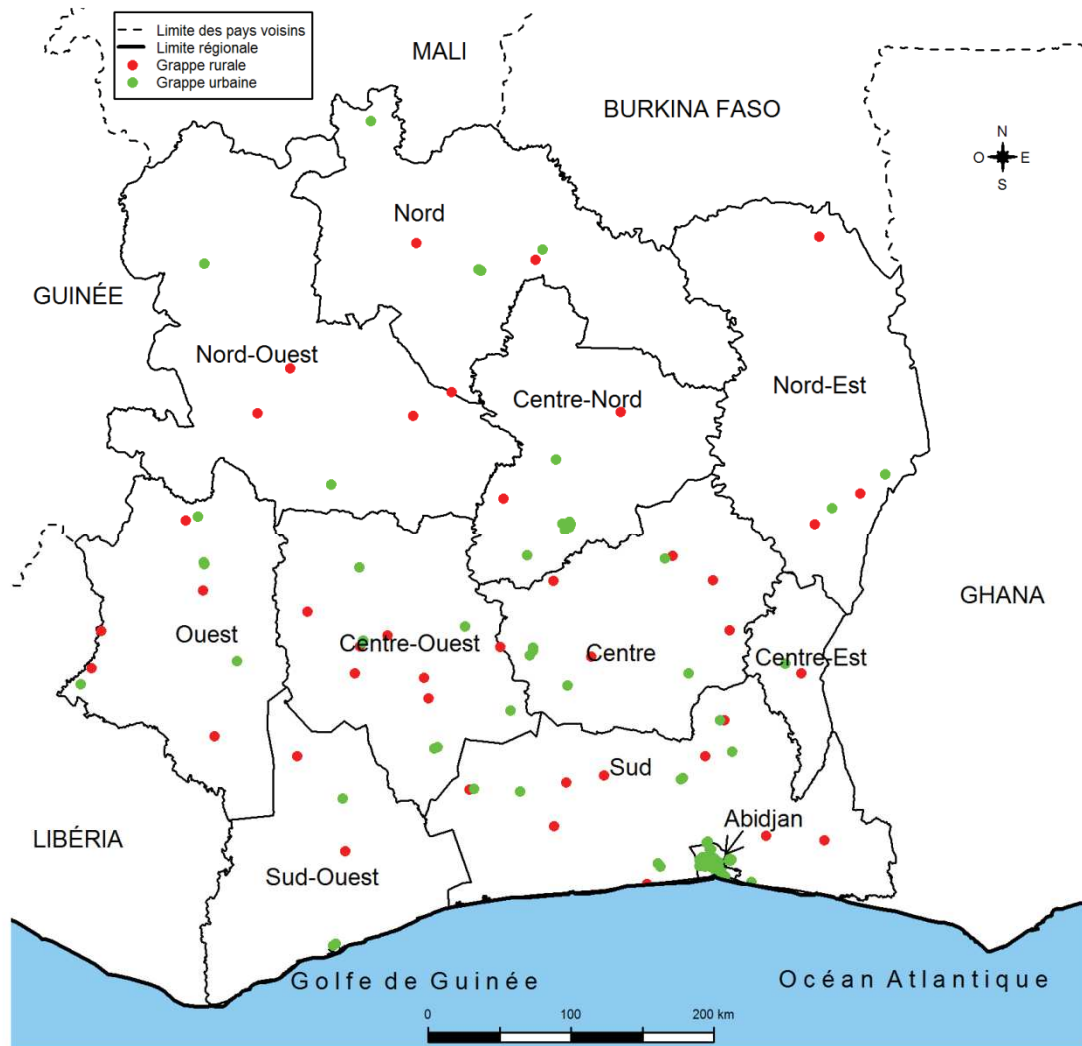
moins de 5 ans, nous nous sommes limités aux enfants de moins de 3 ans. Notre échantillon comportait alors 3356 enfants de moins de 3 ans répartis dans 246 grappes pour l'année 1994, 1113 enfants répartis dans 139 grappes pour l'année 1998 et 2240 enfants répartis dans 349 grappes pour l'année 2011. La localisation géographique des grappes des EDS de 1994, 1998 et 2011 dans lesquelles les enfants de moins de 3 ans appartenaient est représentée respectivement par les **Figures 62, 63 et 64**.

Figure 62 : Répartition géographique des grappes de l'EDS de 1994 selon le lieu de résidence



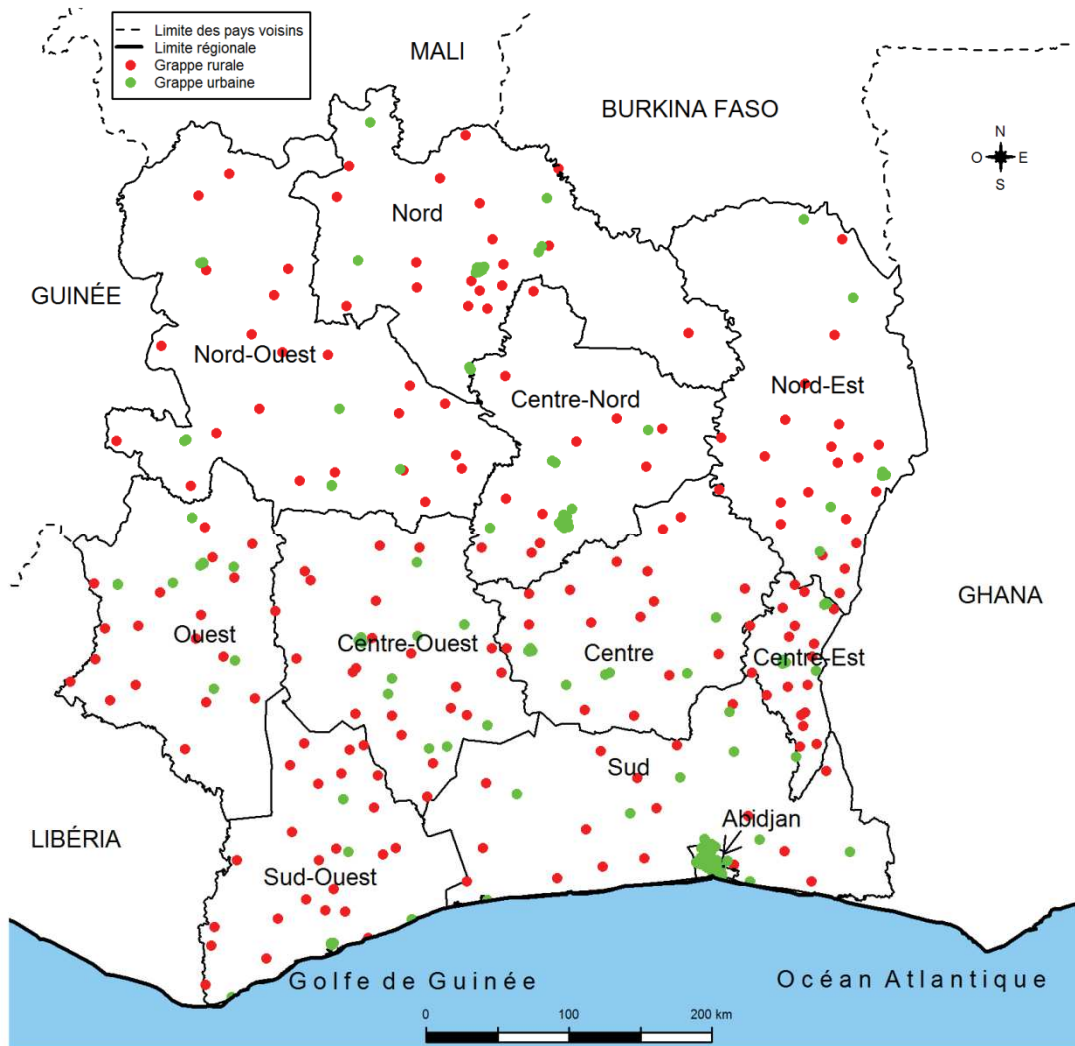
Source : À partir des données de l'EDS 1994

Figure 63 : Répartition géographique des grappes de l'EDS 1998 selon le lieu de résidence



Source : À partir des données de l'EDS 1998

Figure 64 : Répartition géographique des grappes de l'EDS 2011 selon le lieu de résidence



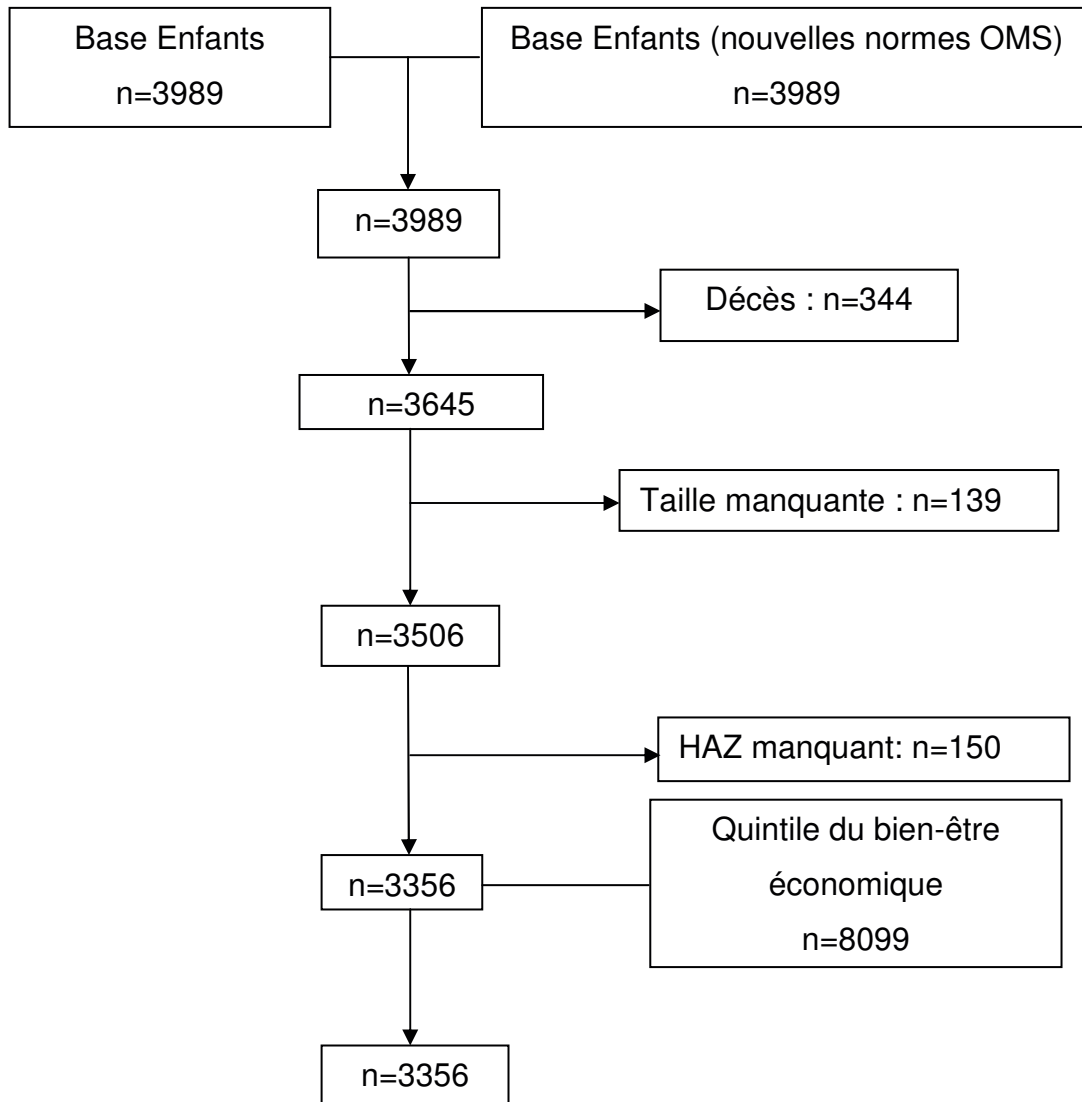
Source : À partir des données de l'EDS 2011

3.9.2. Diagrammes des flux des enfants de 0-35 mois

Pour l'EDS 1994, tous les enfants étaient âgés de moins de 3 ans. Cependant, les prévalences du retard de croissance avaient été calculées sur la base des références de croissance de 1977. Nous avons sélectionné les enfants sur base des z-scores de l'indice taille-pour-âge calculées à l'aide des nouvelles normes de

croissance de 2006 (**Figure 65**). En effet, le nombre d'enfants avec des valeurs manquantes pour le HAZ dépend de la référence utilisée.

Figure 65 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 1994



Pour l'EDS 1998, sur 1909 enfants de moins de cinq ans qui étaient éligibles aux mesures anthropométriques, 1113 enfants étaient âgés de moins de 3 ans (**Figure 66**). Ces enfants étaient observés dans 139 des 140 grappes de l'EDS. Pour l'EDS 2011, parmi 3581 enfants de moins de 5 ans qui ont été mesurés et pesés, 2240 étaient âgés de moins de trois ans (**Figure 67**). Rappelons que tous les enfants de l'EDS de 1994 étaient âgés de moins de 3 ans.

Figure 66 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 1998

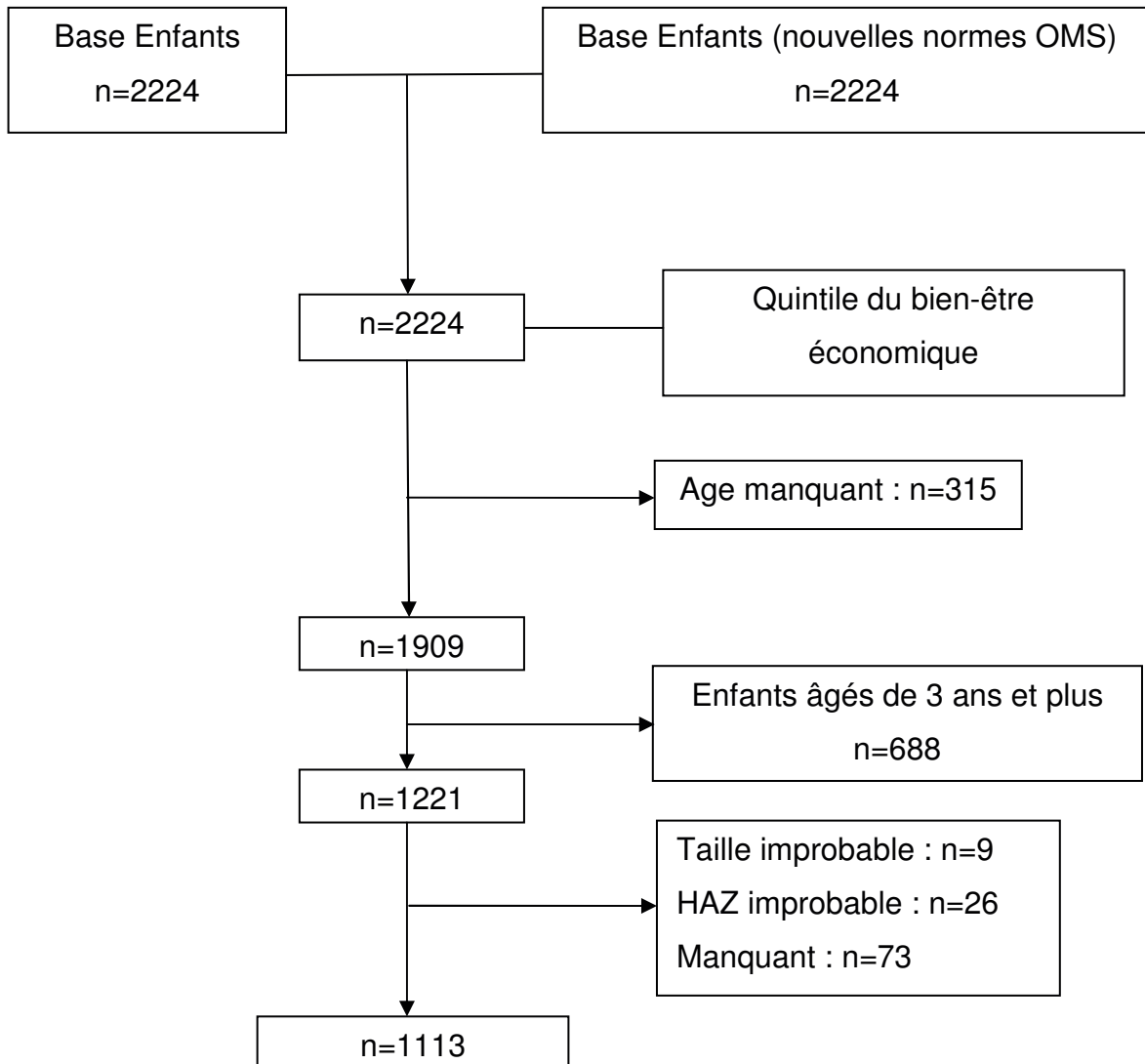
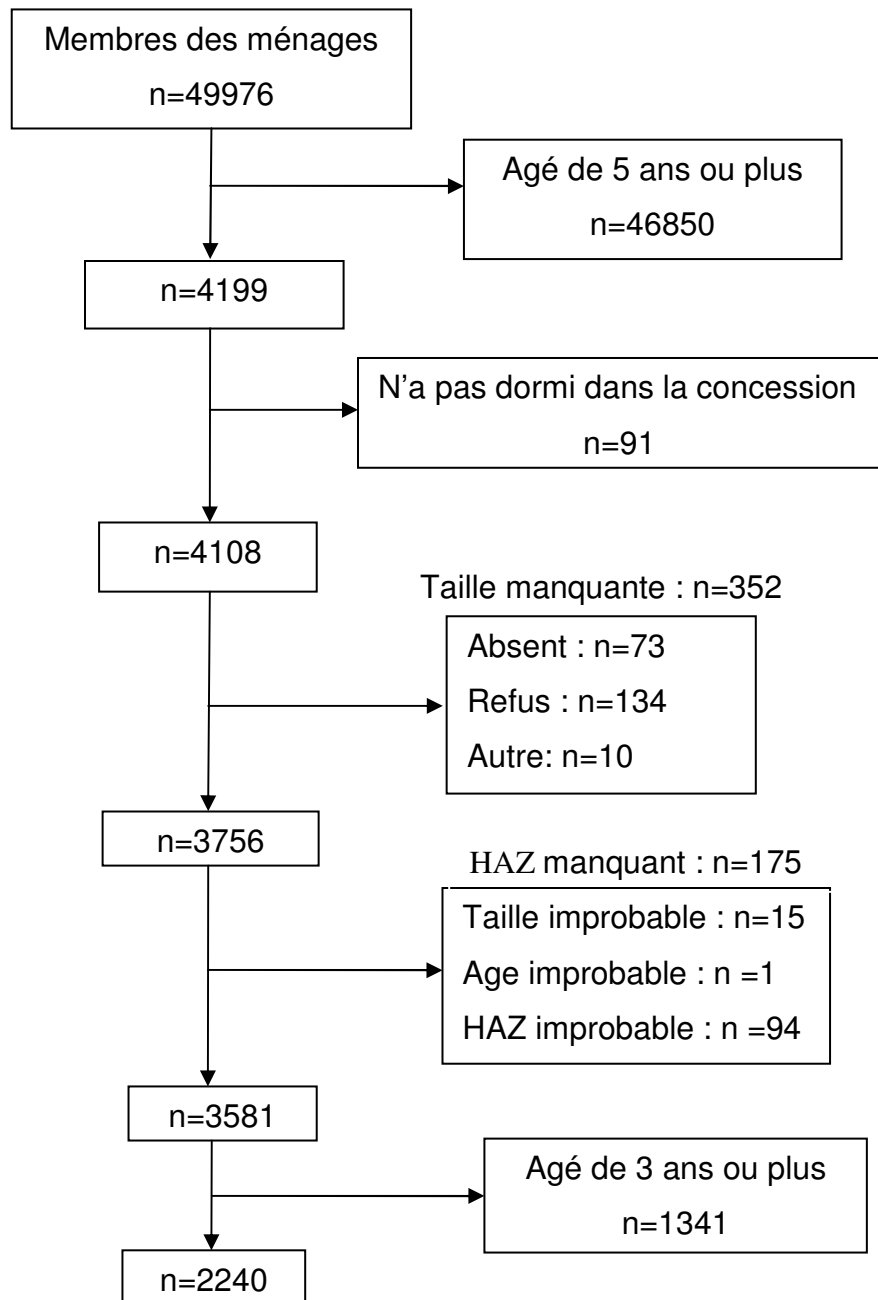


Figure 67 : Diagramme des flux des enfants de moins de 3 ans inclus dans l'EDS 2011



3.9.3. Définition des variables

La variable réponse était le retard de croissance définie par l'indice taille-pour-âge inférieur à -2 z-scores (oui/non). Rappelons que nous avons utilisé les normes de croissance de l'OMS de 2006 pour le calcul des indices taille-pour-âge et poids-pour-taille pour les trois EDS, en vue de leur comparaison. Les deux principales variables explicatives étaient l'année (1994, 1998 et 2011) et la zone (Abidjan, Sud et Nord). La zone Sud désigne la partie du pays (hors Abidjan) située en dessous de la ligne de démarcation et la zone Nord désigne la partie du pays située au nord de cette ligne. Les autres variables explicatives étaient le sexe, l'âge et le rang de naissance de l'enfant, le niveau d'instruction et l'IMC de la mère, et finalement le quintile du bien-être économique des ménages. D'autres variables explicatives potentiellement associées au retard de croissance n'ont pas été utilisées en raison d'un nombre important de valeurs manquantes, dans une ou plusieurs des enquêtes, telles que la présence de la fièvre chez l'enfant, le lieu d'accouchement, l'assistance à l'accouchement, les soins prénatals, le statut marital, le nombre d'injections antitétaniques avant l'accouchement, le nombre d'injections antitétaniques avant la grossesse, l'intervalle entre naissances, la religion, le fait que l'enfant a reçu ou non le vaccin contre la tuberculose, la première dose du vaccin contre la diphtérie, le tétanos et la poliomyélite, la première dose du vaccin contre la poliomyélite et le fait que l'enfant a ou n'a pas eu la diarrhée (oui/non).

L'IMC de la mère a été défini comme le rapport poids/taille² et exprimé en kg/m². Pour toutes les EDS, nous avons attribué des valeurs manquantes à l'IMC (en plus des valeurs manquantes pour cette variable) pour les femmes ayant été enceintes durant les 4 dernières semaines. Ainsi, pour l'EDS de 1994, il y avait 19 valeurs manquantes pour l'IMC et 292 femmes enceintes, ce qui a conduit à 3045 valeurs utilisables et 311 valeurs non utilisables pour l'IMC de la mère. Pour l'EDS 1998, deux femmes présentaient des valeurs manquantes pour l'IMC et il y avait 106 femmes qui avaient été enceintes durant les quatre dernières semaines. Cela a conduit à 1005 valeurs utilisables et 108 valeurs non utilisables pour l'IMC. Pour l'EDS 2011, 438 femmes présentaient des valeurs manquantes pour l'IMC. Parmi

elles, 181 femmes avaient été enceintes durant les 4 dernières semaines. Cela a alors conduit à 1802 valeurs utilisables et 438 valeurs non utilisables pour l'IMC.

L'indice du bien-être économique est une mesure composite qui est calculé à l'aide des possessions telles que la télévision, les sanitaires et le réfrigérateur ; de la source d'eau potable (robinet, puits, étang) ; de la nature du sol de la maison (cimenté, en terre, tapissé) et de la toiture (plafond en bois, en métal, en béton), et de la qualité des murs (briques adobes, briques en terre cuite, mur en béton). La construction de l'indice du bien-être économique se fait par l'intermédiaire des coefficients obtenus à l'aide d'une analyse en composante principale (ACP) des variables retenues. Une fois obtenu, cet indice est découpé en 5 intervalles d'effectifs égaux pour créer les quintiles du bien-être économique (le plus pauvre, pauvre, moyen, riche, le plus riche). Cependant, ces variables qui ont servi à construire l'indice du bien-être économique varient selon les années et ne permettent donc pas de recalculer un indice du bien-être économique comparable entre les enquêtes. Pour le rendre comparable entre les enquêtes, nous avons utilisé les paramètres α (constante) et β (coefficient) fourni par Measure DHS. Ces paramètres étaient issus d'une régression logistique entre la proportion des ménages qui avaient un bien donné et l'indice du bien-être économique et ont été obtenus par une analyse des données de 172 EDS conduites entre 1990 et 2012 dans 69 pays, parmi lesquelles figuraient les 3 EDS de la Côte d'Ivoire. L'indice du bien-être économique comparable entre les trois EDS de Côte d'Ivoire était alors obtenu à l'aide de la relation :

$$IBEC_i = \alpha + \beta \times IBE_i \quad (3.2)$$

où $IBEC_i$ est l'indice du bien-être comparable correspondant à l'année i et IBE_i l'indice du bien-être original à l'EDS réalisée au cours de l'année i . Nous avons ensuite découpé le nouvel indice du bien-être économique en 5 classes d'effectifs égaux que nous avons utilisé dans l'analyse « poolée ».

3.9.4. Analyse statistique

3.9.4.1 Dé-normalisation des poids d'échantillonnage

Les poids d'échantillonnage fournis dans les bases de données des EDS servent à corriger la représentativité de l'échantillon. Ils sont spécifiques à une EDS donnée et normalisés de telle sorte que leur somme soit égale à la taille de l'échantillon. Leur utilisation pour une analyse commune de plusieurs EDS conduit à des résultats erronés. Ils doivent donc être dé-normalisés pour que les EDS soient rendus comparables.

Nous disposons des poids d'échantillonnage des EDS réalisées respectivement en 1994, 1998 et 2011 en Côte d'Ivoire. Soit $V005_i$ la variable qui désigne le poids d'échantillonnage pour une EDS conduite au cours de l'année i notée EDS_i . Soient également P_{15-49}^i le nombre total de femmes âgées de 15 à 49 ans vivant dans le pays à la date de l'EDS et $P_{15-49}^{EDS_i}$ le nombre total de femmes âgées de 15 à 49 ans interviewées lors de l'enquête. Pour chaque EDS (année i) et selon le milieu de résidence (urbain/rural), les nouveaux poids d'échantillonnage (dé-normalisés) $V005_i^*$ pour l'année i sont liés aux poids d'échantillonnage dé-normalisés par la relation :

$$V005_i^* = V005_i \times \frac{P_{15-49}^i}{P_{15-49}^{EDS_i}} \quad (3.3)$$

Comme il n'y avait pas eu de RGPH en 1994 et en 2011, nous avons utilisé le nombre de femmes âgées de 15 à 49 ans obtenu dans les bases de données de l'UNDP concernant les projections de populations et le nombre de femmes de 15-49 ans se trouvant dans les bases de données des EDS. Cependant, les bases de données de l'UNDP ne sont pas disponibles par milieu de résidence. Pour l'année 1998, étant donné qu'un RGPH a été réalisée au cours de cette année là, nous avons utilisé le nombre de femmes (par milieu de résidence) âgées de 15 à 49 ans

se trouvant dans le RGPH ainsi que le nombre de femmes âgées de 15 à 49 ans se trouvant dans la base de données de l'EDS.

3.9.4.2 Analyse de l'autocorrélation spatiale et interpolation spatiale

Nous avons d'abord attribué les grappes des EDS de 1994 et 1998 aux régions statistiques utilisées pour l'EDS 2011, puis dans les trois zones créées (Abidjan, Sud et Nord). Nous avons ensuite calculé les prévalences du retard de croissance pour chaque année de l'EDS globalement, par grappe, par région et par zone. Pour chaque année, nous avons fait une analyse de l'autocorrélation spatiale à l'aide de l'indice global de Moran. Nous avons construit les corrélogrammes, c'est-à-dire les graphiques de l'indice global de Moran en fonction des distances seuil, afin de nous rendre compte de la distance à partir de laquelle l'autocorrélation spatiale n'est plus significative et de nous assurer de la valeur de la distance à utiliser dans l'identification des clusters de haute et de basse prévalence du retard de croissance.

Comme les indices globaux de Moran étaient significatifs, nous avons interpolé les données de prévalence du retard de croissance à des points non échantillonnés sur une grille régulière de 737 km × 776 km à l'aide du krigeage ordinaire [206–208]. Le nombre de points de prédiction était égal à 1896246.

L'incertitude sur les prévalences du retard de croissance interpolées, exprimée par les écart-types, était plus forte dans les régions du Nord-Ouest, du Sud-Ouest, du Nord, du Centre-nord et du Nord-Est que dans les autres régions en 1994 (**Figure 68**), tout près des frontières et dans certaines régions comme le Sud-Ouest, le Nord, le Nord-Ouest et le Nord-Est en 1998 (**Figure 69**) et plus faible à Abidjan que dans les autres régions en 2011 (**Figure 70**). Ces endroits de forte incertitude sur les prédictions correspondaient à des endroits où aucune grappe n'avait été tirée.

Figure 68 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 1994



Figure 69 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 1998

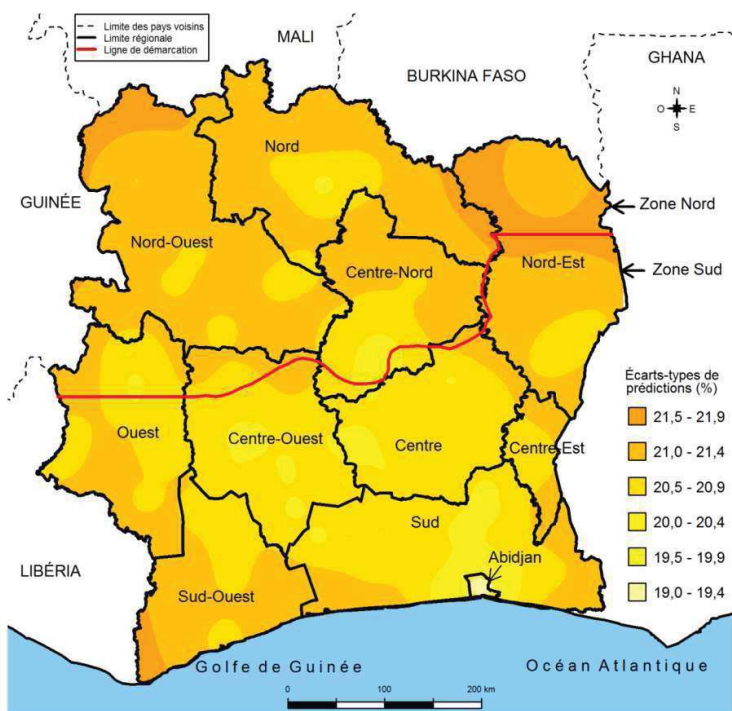


Figure 70 : Incertitude sur les valeurs prédites de la prévalence du retard de croissance en Côte d'Ivoire, 2011



3.9.4.3 Scan spatial de Kulldorff

Nous avons ensuite détecté les clusters de haute et de basse prévalence du retard de croissance à l'aide de la statistique du scan spatial de Kulldorff [209].

3.9.4.4 Modèles de régression logistique non spatiaux

Enfin, nous avons construit des modèles de régression logistique mixtes avec la variable binaire « retard de croissance » comme variable réponse, les variables année et zone et leur interaction comme variables explicatives principales et la grappe comme effet aléatoire. Les variables connues dans la littérature internationale comme étant fréquemment associées au risque de retard de croissance ont été incluses dans l'analyse comme variables de confusion potentielles. Il s'agit des variables sexe de l'enfant, rang de naissance de l'enfant, âge de l'enfant, niveau

d'éducation de la mère, IMC de la mère et le quintile du bien-être économique comme facteurs d'ajustement. L'interaction entre l'année et la zone permet de mesurer l'effet du temps sur la prévalence du retard de croissance dans chaque zone. Nous avons choisi d'intégrer ce terme d'interaction compte tenu de notre hypothèse d'une évolution temporelle différente selon les zones.

Comme l'interaction entre l'année et la zone était significative, nous avons construit des modèles univariés et multivariés dans chaque zone (Abidjan, Sud, Nord). Pour chaque modèle, la qualité de l'ajustement a été vérifiée en comparant un modèle avec uniquement l'intercept au modèle final à l'aide de la statistique F de Fisher [210]. Cet ajustement est vérifié pour les modèles évaluant l'effet ajusté du temps sur la prévalence du retard de croissance dans chacune des 3 zones (Abidjan, Sud et Nord) en mettant l'année 1994 comme référence (**Tableau 10**). L'hypothèse nulle selon laquelle l'ajustement est bon n'est donc pas rejetée au seuil de 0,05 uniquement dans la zone du Sud.

Tableau 10 : Qualité de l'ajustement des modèles logistiques dans les zones de la Côte d'Ivoire

Zone	F	P-value
Abidjan	1,02	<0,001
Sud	1,13	0,168
Nord	0,76	0,001

En fait, une analyse utilisant des méthodes statistiques prenant en compte le caractère spatio-temporel des données avait été tentée. Malheureusement, une estimation des paramètres de la fonction de vraisemblance pénalisée par la méthode bayésienne (méthode REML sur un modèle binomial) n'a pu être obtenue par faute de convergence. Ni les grappes, ni les régions n'ont pu être utilisées comme variables aléatoires : les grappes étaient trop nombreuses, et les observations trop éparses au sein des régions.

Par conséquent, la méthode statistique utilisée dans ce chapitre (régression logistique prenant en compte la dépendance des observations se trouvant dans une

même grappe) est une méthode d'analyse classique. Des méthodes spatio-temporelles utilisant une inférence bayésienne (basée sur les simulations de Monte-Carlo par chaînes de Markov) auraient également pu être utilisées pour obtenir des intervalles de crédibilité plus étroits que les intervalles de confiance, ce qui aurait amélioré la précision de nos estimations.

3.10. Article « Stunting among children under three years in Côte d'Ivoire: spatial and temporal variations between 1994 and 2011 »

La version révisée d'un article a été soumise à la revue Public Health Nutrition.

Stunting among children under three in Côte d'Ivoire: spatial and temporal variations between 1994 and 2011

Emmanuel Barankanira^{1,2,*}, Nicolas Molinari³, Philippe Msellati^{2,4}, Christian Laurent², and Kirsten A Bork²

¹ École Normale Supérieure, Département des Sciences Naturelles, Bujumbura, Burundi.

² IRD UMI 233, INSERM U 1175, Université de Montpellier, Unité TransVIHMI, Montpellier, France.

³ IMAG, UMR 519 / Centre Hospitalier Régional Universitaire de Montpellier / Université de Montpellier, Montpellier, France.

⁴ Institut National de Santé Publique, Abidjan, Côte d'Ivoire.

⁵ PACCI, Abidjan, Côte d'Ivoire.

*Corresponding author: Emmanuel Barankanira, Institut de Recherche pour le Développement (UMI 233), 911 avenue Agropolis, BP 64501, 34394 Montpellier cedex 5, France. E-mail: baremma2002@yahoo.fr

Running head: Stunting trends in Côte d'Ivoire

Abstract

Objective: We investigated the spatial heterogeneity of stunting prevalence among children in Côte d'Ivoire (CI) and examined changes in stunting between 1994 and 2011, so as to assess the impact of the 2002-2011 civil war which led to temporary partitioning of the country.

Design: Data from the 1994, 1998 and 2011 CI Demographic and Health Surveys (DHS) were analysed using a geostatistical approach taking into account spatial autocorrelation. Stunting data were interpolated using ordinary kriging; spatial clusters with high and low stunting prevalences were identified using Kulldorff spatial scan statistics. Multilevel multivariable logistic regression was then carried out, with year of survey as the main independent variable and an interaction term for time by geographic zone (Abidjan, South, North).

Setting: CI, West Africa.

Subjects: 0-35-month-old children included in three CI DHS ($n=6709$).

Results: Overall prevalence of stunting was 30.7%, 28.7% and 27.8% in 1994, 1998 and 2011, respectively ($P=0.32$). Clusters with high prevalence were found in 1994 (in the West region, $P<0.001$) and 1998 (in the West and North-West regions, $P<0.01$ and $P=0.01$, respectively), but not in 2011. Abidjan was included in a cluster with low prevalence in all surveys ($P<0.05$). Risk of stunting did not change between 1994 and 2011 at a national level (adjusted OR=1.39; 95% CI: 0.72, 2.64), but decreased in the South (0.74; 0.58, 0.94) and increased from 1998 to 2011 in Abidjan (1.96; 1.06, 3.64).

Conclusions: In CI, significant changes in stunting prevalence were observed at the sub-national level between 1994 and 2011.

Introduction

Health, nutrition, food security and care practices in developing countries are keys to development. The nutritional status of children is the most valid indicator of their well-being⁽¹⁾. Under-nourished children have slow physical growth and are exposed to severe illness and premature death⁽²⁾. Below normal height-for-age, also called stunting, is generally considered the best indicator of undernutrition, as it measures long-term deprivation of children's basic needs. Stunting in early life is a sign of cumulative effects of individual factors: the socio-eco-political environment in which the child evolves, a low household wealth index, inadequate food intake for mother and child, maternal stunting, repeated illness and infectious diseases⁽³⁾. Reducing stunting prevalence contributes to achieve the first of Millennium Development Goals (MDGs), target 1.C⁽⁴⁾, a challenging objective for sub-Saharan African countries in which 40% of the world's under-five stunted children currently reside⁽⁵⁾. Reducing it by 40% in 2025 and ending it by 2030 are part of the second Sustainable Development Goal (SDG2)⁽⁶⁾.

Political instability may affect a child's well-being⁽⁷⁾. Civil wars and armed conflicts in several West African countries (Côte d'Ivoire, Guinea-Bissau, Sierra Leone and Liberia) have caused population displacement⁽⁸⁾. Côte d'Ivoire was severely affected by a civil war from 2002 to 2011, creating a demarcation line between North and South. A recent analysis of stunting from 1994 to 2011 found no change in socio-economic inequality of stunting in 31 out of 53 countries, including Côte d'Ivoire⁽⁹⁾. However, worldwide systematic analysis of nutritional changes from 1985-2011 reported high probability of deterioration of the preschool height-for-age z-score (HAZ) in Côte d'Ivoire and Niger, in contrast to overall improvement in sub-Saharan Africa from the late 1990s to 2011⁽¹⁰⁾. Conversely, a recent overview found a 4.2% decrease in the prevalence of stunting among children below the age of two in Côte d'Ivoire between 1994 and 2011⁽¹¹⁾. The knowledge of stunting prevalence distribution at different spatial levels (country, region, zone) and its change over time may help decision policy makers to identify areas where urgent nutritional interventions are needed and to adapt national programs and strategies to reduce the burden of stunting in Côte d'Ivoire.

The main objective of this study was to examine stunting changes over time among children under three years of age across three defined zones (i.e., Abidjan, the South and North) in Côte d'Ivoire, based on Demographic and Health Survey (DHS) and Multiple Indicator Cluster Survey (MICS) data collected in 1994, 1998 and 2011. It also sought to describe spatial distribution of stunting across regions/zones, and to detect spatial clusters of low and high stunting prevalence.

Our study hypothesis was that stunting prevalence among children under three would have increased from 1994 to 2011 in some parts of the country, and especially the North, which was economically isolated during the civil war.

Methods

Study design

This secondary study of Côte d'Ivoire DHS data analysed changes in stunting prevalence and distribution among 0-35-month-old children in Côte-d'Ivoire over a period of 17 years, using nationally representative data from Côte d'Ivoire DHS in 1994, 1998-1999 and 2011-2012.

Côte d'Ivoire, the civil war 2002-2011

During the civil war which began in September 2002, the country was divided into two parts separated by a buffer zone: a northern part, occupied by rebel forces, and a southern part, including Abidjan, controlled by the government^(12,13). The civil war led to wide-scale displacement of populations, including internal and external migrations and displacement of civil servants, teachers and health staff from the North to the South. Abidjan was the most frequent destination of these internal migrations. Conversely, some inhabitants living in the South, and especially Burkinabe citizens, fled the country.

Inclusion criteria

Our analytical sample included all children under three years of age who had reliable HAZ values (i.e. in the range from -6 to 6).

Design and procedures of the 1994, 1998 and 2011 DHS in Côte d'Ivoire

The 1994, 1998-1999 and 2011-2012 Côte d'Ivoire DHS were conducted by the National Statistical Institute and ORC Macro between June and November, 1994, between September, 1998 and March, 1999, and by the National Statistical Institute and ICF Macro between December, 2011 and May, 2012, respectively⁽¹⁴⁻¹⁶⁾. All followed standardised DHS methodology, and all investigators were female⁽¹⁷⁾.

The 1994 Côte d'Ivoire DHS used a stratified two-stage cluster random sample design in Abidjan and rural areas. Enumeration areas (EA, i.e. groupings of households) were selected at the first stage and households at the second stage. For other areas, a three-stage cluster design was used: cities were selected at the first stage, EA at the second stage and households at the third stage. Overall, 246 EA (146 urban and 100 rural) were selected, leading to a total of 6348 eligible households, among which 5935 (2745 urban and 3190 rural) were included (Supplementary online material, Figure 1). All eligible women and their children under three were eligible for anthropometric measurements; for financial reasons, 3- and 4-year-old children were not eligible (Côte d'Ivoire DHS staff, personal communication, August 16, 2016). Weight and height information was collected for 8099 women interviewed; 3645 children under three years of age were eligible.

The 1998-1999 Côte d'Ivoire DHS used a stratified two-stage cluster random sample design. Among 2302 eligible households, 2122 (1421 urban and 701 rural) were included. Weight and height were collected for 3040 interviewed women; 1221 children under three were eligible. For the 1994 and 1998 Côte d'Ivoire DHS, the sampling frame of household groupings (n=246 and n=140, respectively) had been elaborated using the second general population and housing census (GPHC) performed in 1987-1988⁽¹⁸⁾.

The 2011-2012 Côte d'Ivoire DHS-MICS used a stratified two-stage cluster design. Stratification was based on 11 statistical regions and place of residence (urban or rural). The sampling frame of household groupings had been elaborated based on the third Côte d'Ivoire GPHC in 1998-1999⁽¹⁸⁾. Among 10 413 eligible households, 9686 were included. Half of the eligible households were selected for anthropometric measurements. Among 10 848 eligible women, 10 060 were interviewed; 2570 children under three were eligible.

Geographic data

Côte d'Ivoire DHS collected EA geographic coordinates (latitude and longitude) using a geographic information system (GIS) and global positioning system (GPS) technologies. In average, urban EA grouped 20 and rural EA 35 households in 1994. In 1998, urban EA grouped 14 and rural EA 18 households, as compared to 27 and 32 households, respectively, in 2011. The median household size was five inhabitants in all three surveys. Geographic coordinates were collected at the centre of EA for all Côte d'Ivoire DHS. Geographic, household and child datasets for all three surveys are available from the DHS program website⁽¹⁹⁾. Geospatial data manipulations were performed using Quantum Geographic Information System (QGIS) software, version 1.8.0⁽²⁰⁾.

Definition of variables

The dependent variable was stunting, defined as a height-for-age (HA) index below - 2 z-scores of the WHO international growth standards⁽²¹⁾. The main independent variables used were the year of survey and the zone in which the child resided. EA for the 1994 and 1998 Côte d'Ivoire DHS were allocated to the 11 statistical regions defined in the 2011-2012 DHS (Fig. 1). Next, we divided the country into three zones based on partitioning of the country during the 2002-2011 civil war⁽²²⁾. These were Abidjan, the South (below the demarcation line, excluding Abidjan) and the North (above the demarcation line, Fig. 1). Other variables were sex and birth order (1, 2-3, 4-5, or ≥ 6), child age group (<6, 6-11, 12-17, 18-23, 24-29, and 30-36 months), education level of the mother (no formal education, primary or secondary level), body mass index (BMI) of the mother (underweight, normal and overweight/obese for <18.5, 18.5-25.0, and >25.0 kg/m², respectively), and economic status of the household (as defined by each Côte d'Ivoire DHS) divided into fifths.

Assets used for construction of the wealth index varied across surveys, making survey-specific indexes inadequate for pooled analyses. Instead, we constructed a comparative wealth index (CWI) for each of the 1994, 1998 and 2011 Côte d'Ivoire DHS based on alpha and beta coefficients representing, respectively, the amount of adjustment and dispersion of the level of the survey-specific wealth index relative to

the baseline wealth index, for 172 DHS conducted worldwide from 1990 to 2012⁽²³⁾. The CWI was divided into fifths for analysis.

Sampling weights

Sampling weights, which are country- and sex-specific, are not valid for analysing pooled data⁽²⁴⁾. For 1994 and 2011 Côte d'Ivoire DHS, we de-normalised weights using the number of females aged 15-49 years in Côte d'Ivoire in 1994 and 2011, respectively (reference: United Nations Population Division (UNDP) website⁽²⁵⁾). For 1998, a national census was available. Thus, we used the total number of women aged 15-49 in the 1998 Côte d'Ivoire GPHC, stratified by place of residence, and the total number of women 15-49 years old in the 1998 Côte d'Ivoire DHS database.

Statistical analysis

Stunting prevalence mapping and cluster detection

We computed the EA-based stunting prevalence for each Côte d'Ivoire DHS, globally and across the three zones (Abidjan, North and South). The overall spatial pattern of stunting prevalence was analysed using the global Moran's test. Moran's I statistics test the hypothesis that observed data at one location are independent of observed data at other locations. When this hypothesis is rejected, the use of a geostatistical approach is justified. We mapped stunting prevalence throughout the country independently of regional boundaries using the geoR library of R software, version 3.1.0 (R Core Development Team, April 10, 2014)^(26,27). We fitted the sample variograms with Matern, spherical and exponential models, respectively for 1994, 1998 and 2011 Côte d'Ivoire DHS using the VARIOFIT function. We interpolated stunting prevalence in the 1 896 246 points using a regular grid of 737 km × 776 km and the ordinary kriging technique⁽²⁸⁾. The kriging technique consists of predicting nearest sampled data values (EA-based stunting prevalence in our case) at unsampled spatial locations. Predicted values are weighted means of the observed values, and weights are based on a variogram that measures the spatial dependence or autocorrelation between observed values. A full description of this technique can be found elsewhere⁽²⁶⁻²⁸⁾. The ordinary kriging technique, which assumes that the mean is an unknown constant over space, was chosen because it is the best linear unbiased estimator with known minimum variance⁽²⁹⁾.

Moreover, we used Kulldorff's spatial scan statistics (package SaTScan, version 9.3) to identify spatial clusters with high or low stunting prevalence for each Côte d'Ivoire DHS^(30,31). This method has been widely used in recent years, including in the field of child nutrition^(32,33). It enables locating areas with higher or lower numbers of stunted children than expected under the hypothesis of uniform spatial distribution of stunting. A scanning circular window of varying size gradually moves across the study area to detect areas with an excess number of stunted children. At each step, the number of observed cases in the moving window is compared to the expected number of cases assuming random spatial distribution. We used a default value of 50% of the total study population for detection of clusters and a maximum radius of 100 km. This distance was chosen based on the results of correlogram analyses: we created a set of radius from 50 to 750 km (i.e. 50, 100, 150, etc.), covering the range of the observed distances between EA for each DHS. Each radius was used as the threshold distance for neighborhood for all DHS in the correlogram analyses. The maximum radius beyond which the spatial autocorrelation was no longer significant was used as inputs in the spatial scan procedure. Based on Monte Carlo hypothesis testing⁽³⁴⁾, the likelihood ratio tests and their associated p-values obtained through 9,999 Monte Carlo replications in the Bernoulli probability model were used to ascertain the statistical significance of detected spatial clusters. The null hypothesis is that of a uniform distribution of observed number of stunted children in Côte d'Ivoire, showing equal child stunting risks across the country. When a cluster was identified, strength of clustering was ascertained using the relative risk (RR) of excess of stunted children. The estimated risk within a cluster was defined as the ratio between the observed and expected number of stunted children, and the RR was defined as the ratio between estimated risks inside the cluster vs. outside (i.e. in the remaining part of the country). A cluster was declared significant for $P < 0.05$.

Temporal variations in stunting

Assessment of trends in stunting and background variables was based on observed individual-level data recorded in the 1994, 1998 and 2011 Côte d'Ivoire DHS for the country as a whole and for each of the three zones, using a chi-square test for trend for the combined cross-sectional survey data. Maternal BMI was assessed for normality using the Shapiro-Wilk test and compared across surveys using the

Kruskal-Wallis test. The association between stunting and background variables was assessed using chi-square tests.

Multilevel logistic regression models were then used⁽³⁵⁾. The objective was to assess whether changes in stunting over time could be explained by changes in some of its major risk factors. Levels were individuals and EAs. EAs were used as random effects. The random effects variance-covariance structure was defined as independent. The dependent variable was stunting and the main explanatory variables were year of survey, zone and interaction term between year and zone. In case of a significant interaction, the study protocol included separate logistic models for each zone. Models were adjusted for maternal education and BMI, family wealth, child's birth order, age and sex. These variables were chosen based on their association with stunting in the literature and their availability in all surveys. Cases with missing data (i.e. for birth order, maternal education or maternal BMI) were excluded. For each zone-specific model, we computed intra-cluster correlation (ICC) and median odds ratio (MOR). ICC informs us on the proportion of total variance in the outcome that is attributable to the area level^(35,36). The MOR is defined as the median value of the odds ratio of the area at higher risk compared to that at lower risk, when randomly picking out two areas. The MOR can be conceptualised as the increased risk that an individual would have (in median) if moving to another area with higher risks⁽³⁷⁾. In our study, the MOR shows the extent to which the probability of being stunted is attributable to individual-level factors. The estimation was done using the maximum pseudo-likelihood for survey data with the svyset function⁽³⁸⁾. The association of independent variables with risk of stunting was assessed using odds ratios (OR) and their 95% confidence intervals (CI). Statistical software used was Stata[®] (Release 14, StataCorp LP, College Station, TX, USA).

All tests were considered statistically significant for $P < 0.05$.

Results

In the 1994 Côte d'Ivoire DHS, 3645 children under three were included in the survey. We excluded 289 children (7.9%) who had unreliable or missing height or age data, leading to 3356 children (Supplementary online Figure 1). The 1998 Côte

d'Ivoire DHS included 1221 children under three, among whom 108 (8.9%) had unreliable or missing height or age data. Thus, 1113 children were kept in the analysis. In the 2011 Côte d'Ivoire DHS, there were 2570 children under three. We excluded 330 children (12.8%) who had unreliable or missing height or age data. Thus, 2240 children were kept in the analysis.

The study sample thus comprised 6709 children of whom 5852 had complete data for all background variables. The median number of children per EA was 10 (interquartile range (IQR): 6-19), 4 (IQR: 2-11) and 5 (IQR: 2-8), for Côte d'Ivoire DHS 1994, 1998 and 2011, respectively.

Characteristics of the study population

Table 1 shows the distribution of under-three-year-old children by zone and other characteristics for the 1994, 1998 and 2011 Côte d'Ivoire DHS. The chi-square test for trend results showed that the proportion of children by zone did not change over time ($P=0.80$), but that the age distribution differed across surveys ($P=0.014$). The distribution of birth order also changed over time ($P<0.01$); in particular, fewer children had a birth order ≥ 6 in 2011.

Two-thirds of mothers had no formal education, and the level of schooling did not change between 1994 and 2011. Their median BMI was 21.6 kg/m² (IQR: 20.1-23.5), 21.9 kg/m² (IQR: 20.1-23.9) and 22.2 kg/m² (IQR: 20.5-24.5) in 1994, 1998 and 2011, respectively ($P<0.001$). The prevalence of maternal underweight decreased and that of overweight-obesity increased over time (Table 1). The CWI deteriorated in 2011; indeed, fewer children lived in the richest households compared to earlier years ($P<0.01$ for differences across surveys).

The prevalence of stunting was 30.7% (95% CI: 28.5, 32.9), 27.8% (24.0, 31.6) and 28.7% (26.2, 31.2) in 1994, 1998 and 2011, respectively ($P=0.32$). About 10% of children were wasted, and this prevalence did not change between 1994 and 2011 ($P=0.14$).

Spatial inequalities in stunting prevalence

The results of Moran's tests showed that stunting prevalence data were spatially autocorrelated (Moran's $I=0.12$ in 1994, Moran's $I=0.19$ in 1998 and Moran's $I=0.08$ in 2011, $P<0.05$ for all). Thus, the hypothesis of spatial homogeneity was rejected.

In 1994, stunting prevalence ranged between 15.5% (11.8, 19.3) in Abidjan and 43.8% (37.0, 50.5) in the West region (Fig. 1(a)). It was close to or above 40% in the North-West, the North-East and the North regions, and was 34.1% overall in the North zone. Interpolated data showed the spatial heterogeneity of stunting independently of regional boundaries (Fig. 2(a)). The highest prevalence was observed in the North-West, West and North regions, with internal variations in prevalence in all regions and zones. Spatial scan statistics detected one significant cluster of high stunting prevalence in a radius of 79.0 km which ran throughout the West region. This region contained 86 stunted children (i.e., 49.1% of children in the cluster), giving a RR of 1.64 ($P < 0.001$). Spatial scan also detected one significant cluster of low prevalence with a radius of 16.2 km, located around Abidjan. This cluster enclosed 78 stunted children (i.e., 14.5%), giving a RR of 0.43 ($P < 0.001$) (Fig. 3(a)).

In 1998, the prevalence was high in the West, North-West and Centre-North regions, and, globally, in the North zone (Fig. 1(b)). Interpolated data showed substantial spatial heterogeneity stunting, ranging from 10.0% to 44.9% across regions (Fig. 2(b)). There were areas with high risk of stunting in the Centre-West and Centre-North regions. Spatial scan statistics (Fig. 3(b)) revealed two significant clusters with high stunting prevalence, one with a radius of 94.4 km located in the West region ($P = 0.003$, RR=1.93, 39 (51.3%) cases) and another with a radius of 84.8 km, centred in the North-West region ($P = 0.011$, RR=1.90, 36 (50.7%) cases). There existed one significant cluster with low stunting prevalence with a radius of 17.2 km located around Abidjan ($P = 0.003$, RR=0.48, 25 cases, i.e. 14.8%).

In 2011, stunting had decreased in certain regions that had shown a high prevalence in 1994 and 1998, but not in the North, where it remained high (37.1%, Fig. 1(c)). Interpolated data exhibited areas of high risk of stunting in the South-West, South, North and Centre-North regions (Fig. 2(c)). Spatial scan statistics showed no clusters with high prevalence, but one cluster with low prevalence and a radius of 93.7 km ($P = 0.010$, RR=0.67, 95 cases, i.e. 20.6%) located under the secession line and covering Abidjan and parts of the South (Fig. 3(c)).

For all years under study, the prevalence was lowest in Abidjan (15.5%, 13.1% and 21.7%, respectively).

Temporal variations in stunting

Unadjusted logistic regression analyses based on all 6709 children showed no significant change in stunting prevalence from 1994 to 1998 or from 1994 to 2011 in Côte d'Ivoire as a whole ($P > 0.20$, Table 2). The chi-square test for trend results showed that there was no linear trend in stunting for Abidjan (15.5, 13.1 and 21.7% for the three surveys, respectively, $P = 0.11$), but the odds were significantly higher in 2011 than in 1998 (OR=1.85, 95% CI: 1.08, 3.15). In the South, there was a significant negative trend in stunting over time (33.2, 27.2 and 28.3%, respectively, for the three surveys, P for trend=0.024); compared to 1994, ORs of stunting in 1998 and 2011 were both significantly below 1, i.e. 0.75 (0.56, 0.995) and 0.79 (0.66, 0.96), respectively. There was no difference across years of survey for the North.

All co-variables were significantly associated with stunting, either in the whole country (sex, age, family wealth, maternal education and BMI) or in the South zone (birth order, results not shown).

Multivariable logistic regression analysis was conducted in 5852 complete cases. The proportion of excluded children was 9.2%, 9.6% and 19.6% in 1994, 1998 and 2011, respectively ($P < 0.001$); it did not differ by zone ($P = 0.29$). Excluded children were more likely to be stunted than complete cases in the South zone in 1994 (45.7% vs. 31.8%, $P < 0.001$) and 1998 (42.1% vs. 25.4%, $P = 0.017$), and in Abidjan in 1998 (33.3% vs. 10.5%, $P < 0.01$), but not in the remaining zones and years ($P > 0.10$ for all).

In multivariable regression, a significant interaction existed between year and zone ($P = 0.028$), with children living in the South and surveyed in 2011 having a lower risk (Table 3). Therefore, logistic regression analyses for complete cases were performed separately for the three zones, as per protocol (sample sizes were 845, 3630 and 1377 children in Abidjan, South zone and North zone, respectively). In the South, stunting decreased significantly over time. The adjusted ORs were 0.71 (0.53, 0.95) and 0.74 (0.58, 0.94) for 1998 and 2011, respectively (Table 3). There was no change in the prevalence of stunting between 1998 and 2011 (Table 4). In the North, there was no significant effect of year on stunting between 1994 and 1998, nor between 1994 and 2011, in adjusted analyses ($P > 0.20$ for both, Table 3). Thus, adjustment did not change the association between stunting and year of survey.

In Abidjan, the prevalence of stunting did not vary neither between 1994 and 1998 nor between 1994 and 2011 (Table 3), but the significant increase between 1998 and 2011 which existed in the unadjusted analyses remained very similar in the fully adjusted model (adjusted OR=1.96, 95% CI: 1.06, 3.64, Table 4). Between 1994 and 2011, birth order decreased in Abidjan whereas household wealth decreased from 1994 to 1998 and then increased again ($P<0.001$ for both, Supplementary online material, Table 1).

Final zone-based logistic models showed that ICC and MOR were respectively 2.9% (0.8%, 9.4%) and 1.44 (1.21, 1.98) in the South zone, 1.2% (0.0%, 24.5%) and 1.27 (1.05, 3.33) in the North zone, 9.0% (2.9%, 25.0%) and 1.94 (1.44, 3.38) in Abidjan. MOR of the empty model (i.e. without explanatory variables) was 1.72 (1.52, 2.03) and decreased when explanatory variables were taken into account in the South and North zones, showing a significant effect of clustering in these zones, in contrast to Abidjan.

Discussion

We examined the effect of time on stunting prevalence among children aged 0-35 months in Côte d'Ivoire, globally and within three large zones (South, North and Abidjan), using 1994, 1998 and 2011 DHS data. We identified spatial clusters with high and low stunting prevalence.

Our study hypothesis was that stunting prevalence would have increased during that period because of the 2002-2011 civil war, especially in the North, which was economically isolated. The analysis showed that there was no overall significant change in stunting. The North showed a fairly constant prevalence, and we even found a significant reduction in stunting between 1994 and 2011 in the large southern zone of the country (adjusted OR: 0.74; 0.58, 0.94). Only Abidjan had an increased risk from 1998 to 2011.

The analysis also confirmed our hypothesis of the existence of substantial spatial heterogeneity in stunting across regions and zones in all three surveys. The spatial dimension of stunting is often analysed using exploratory spatial data analysis techniques at the national and regional level⁽³⁹⁾. In addition, the analysis presented

here included spatial patterns of stunting prevalence on a local scale⁽⁴⁰⁾. This was achieved using the Kulldorff spatial scan test as a local clustering test⁽⁴¹⁾.

In 1994, stunting was most prevalent in the North (34.1%), in particular in the North and North-West regions and parts of the West and North-East regions. In spatial analyses, we identified a cluster with high prevalence in the West region (43.8%). These results are consistent with those of a study conducted in Côte d'Ivoire in 1990⁽⁴²⁾. In 1998, the highest stunting prevalence was found in the Centre-North, West and North-West regions, which contained two high-risk areas.

For all surveys, a cluster with low stunting prevalence was found in the South, centered on Abidjan. This was expected given the higher socio-economic level in the capital city. However, there was an increase in stunting between 1998 and 2011. This might be due to a massive arrival of refugees from the North during the civil war⁽⁴³⁾, and/or to a global negative impact of the war on the nutritional situation of children below the age of three.

At the time of the last survey in 2011, the civil war had just ended. The aftermath of the civil war, mainly in the North, involved theft of livestock, destruction of public infrastructures and crops, shortage of human resources, economic contraction, and displacement of health staff and populations towards the southern parts⁽⁴⁴⁻⁴⁶⁾. A recent study in Côte d'Ivoire provided evidence that children from conflict-affected regions (South-West, West, North-West, North, Abidjan, North-East regions) or households having experienced economic stress during the period 2002-2007 (measured by loss of productive assets such as jobs, farms and livestock, or a decrease in income) had lower mean HAZ than those from less conflict-affected regions⁽⁴⁷⁾. Our findings are also consistent with those of a study carried out in 2014 which showed that childhood health was affected in regions around Guiglo (West region), Bouaké (Centre-North region) and Korhogo (North region), where numerous conflicts occurred⁽⁴⁴⁾.

The civil war was not the only socio-economic event during the years under study. In addition, the western part, where stunting was highly prevalent in 1994 and 1998, suffered from land-related conflicts between indigenous ethnic groups and non-Ivoirians, among them Liberians, who fled the 1999-2003 Liberian civil war⁽⁴⁴⁾.

In the 1990s, the purchasing power of the population decreased due to a decline in international cocoa prices⁽⁴⁸⁾. Another major macro-economic change was the 50%

devaluation of the national currency, the Communauté Financière Africaine francs (CFAF), in February 1994⁽⁴⁹⁾. This change, which affected most French-speaking countries in West and Central Africa, was shown to have a negative impact on the nutritional status of preschool children in Brazzaville, Congo, between 1993 and 1996⁽⁵⁰⁾. However, it is doubtful that a macro-economic event in February 1994 would already have affected stunting of children below the age of three years at the time of the Côte d'Ivoire DHS, carried out from June-November 1994. Indeed, stunting is a cumulative process, and environmental insults have the greatest impact on linear growth from conception to age 12-18 months⁽⁵¹⁾. Thus, the high stunting prevalence in the West and North-West regions in 1994 was probably largely independent of the currency devaluation.

Spatial variations in preschool stunting have been reported previously in sub-Saharan countries, e.g. in the DRC, which suffered severe regional conflicts⁽⁵²⁾. In 2007, the prevalence was very high in the eastern provinces affected by conflicts (46.1%), but also in those relying on mining. In Nigeria, there was a strong South-North gradient in stunting prevalence among preschool children in 2008, ranging from 11.6% in one southern state to 59.7% in a northern state⁽³⁹⁾.

To our knowledge, this is the first study making use of spatial techniques to analyse the effect of time on the prevalence of stunting in Côte d'Ivoire. However, the study had limitations. One was the rather low sample size in the 1998 Côte d'Ivoire DHS, limiting the statistical power of tests. Second, multivariate logistic regression was conducted on complete cases after exclusion of 12.8% of children with missing data (for birth order, maternal education and BMI). It was not possible to estimate missing data using multiple imputations, due to the lack of other relevant background variables in the DHS databases. In some surveys and zones (three out of nine), stunting prevalence was higher for excluded children than for complete cases. Thus, although the results of multivariate analyses were very similar to those of unadjusted analyses, we cannot exclude that the observed change in stunting over time in the South zone or Abidjan was explained by changes in background variables such as birth order, maternal BMI or economic level.

Finally, we acknowledge that this secondary analysis of Côte d'Ivoire DHS data could not formally link sub-national trends in stunting to the civil war. Since data are cross-sectional, we were not able to distinguish 'true' changes in young child nutritional

status in a given population (due to changes in feeding and/or caring practices) from changes in population composition due to massive arrival of residents to certain areas such as Abidjan.

Our findings show the progress made by Côte d'Ivoire in reducing stunting among young children at a local scale before and after the civil war. The exploration of spatial and spatiotemporal variations of stunting prevalence should provide national public health decisions-makers with useful information for nutritional interventions. The recent increase in stunting in Abidjan is a matter of concern, but its prevalence remained much lower than in the South and North zones. Although the North zone contributed less to the total amount of stunted children in the country than the South zone, due to its lower population density, it requires attention. Indeed, the reduction of inequalities in nutrition within the country is an important development goal.

In conclusion, the analysis presented here shows that the overall prevalence of stunting remained stable in Côte d'Ivoire between 1994 and 2011 despite the civil war, but that this average hides significant and contrasting changes at the sub-national level. Further analyses are needed to assess how and where the prevalence of stunting has changed in Côte d'Ivoire since the civil war ended in 2011. In the meantime, public health efforts should be reinforced throughout the country because stunting prevalence remained high in all three zones.

References

1. de Onis M, Frongillo EA, Blössner M (2000) Is malnutrition declining? An analysis of changes in levels of child malnutrition since 1980. *Bull World Health Organ* **78**, 1222–1233.
2. Prendergast AJ, Humphrey JH (2014) The stunting syndrome in developing countries. *Paediatr Int Child Health* **34**, 250–265.
3. Black RE, Allen LH, Bhutta ZA et al. (2008) Maternal and child undernutrition: global and regional exposures and consequences. *Lancet* **371**, 243–260.
4. United Nations (2015) The Millennium Development Goals Report. Beacon Press.
http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20%28July%201%29.pdf (accessed 8 October 2015).
5. UNICEF (2013) Improving child nutrition: the achievable imperative for global progress. New York: United Nations Children's Fund.
http://www.unicef.org/gambia/Improving_Child_Nutrition_-_the_achievable_imperative_for_global_progress.pdf (accessed 12 October 2015).
6. UNDP. Transforming our world: the 2030 Agenda for Sustainable Development.
<https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf> (accessed 1 March 2016).
7. Milman A, Frongillo EA, de Onis M et al. (2005) Differential Improvement among Countries in Child Stunting Is Associated with Long-Term Development and Specific Interventions. *J Nutr* **135**, 1415–1422.
8. Marc A, Verjee N, Mogaka S (2015) The Challenge of Stability and Security in West Africa. The World Bank.
<http://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-0464-9> (accessed 8 October 2015).
9. Bredenkamp C, Buisman LR, Van de Poel E (2014) Persistent inequalities in child undernutrition: evidence from 80 countries, from 1990 to today. *Int J Epidemiol* **43**, 1328–1335.
10. Stevens GA, Finucane MM, Paciorek CJ et al. (2012) Trends in mild, moderate, and severe stunting and underweight, and progress towards MDG 1 in 141 developing countries: a systematic analysis of population representative data. *Lancet* **380**, 824–834.
11. Tzioumis E, Kay MC, Bentley ME et al. (2016) Prevalence and trends in the childhood dual burden of malnutrition in low- and middle-income countries, 1990–2012. *Public Health Nutr*. Published online: 24 February 2016. doi: 10.1017/S1368980016000276

12. Aurégan X (2014) Géopolitique de la Chine en Côte d'Ivoire. <https://tel.archives-ouvertes.fr/tel-01190744/> (accessed 20 October 2015).
13. Vidjannangni A (2011) La complexité de la question identitaire en Côte d'Ivoire. <http://www.archipel.uqam.ca/id/eprint/4180> (accessed 20 October 2015).
14. N'Cho S, Kouassi L, Koffi AK et al. (1995) Enquête Démographique et de Santé, Côte d'Ivoire 1994. Calverton, Maryland, U.S.A.:Institut National de la Statistique et Macro International Inc. <http://dhsprogram.com/publications/publication-FR66-DHS-Final-Reports.cfm> (accessed 15 November 2014).
15. Institut National de la Statistique [Côte d'Ivoire] et ORC Macro (2001). Enquête Démographique et de Santé, Côte d'Ivoire 1998-1999. Calverton, Maryland USA: Institut National de la Statistique et ORC Macro. <http://dhsprogram.com/publications/publication-FR129-DHS-Final-Reports.cfm> (accessed 15 November 2014).
16. Institut National de la Statistique (INS) et ICF International. 2012. Enquête Démographique et de Santé et à Indicateurs Multiples de Côte d'Ivoire 2011-2012. Calverton, Maryland, USA: INS et ICF International. <http://dhsprogram.com/publications/publication-FR272-DHS-Final-Reports.cfm> (accessed 15 November 2014).
17. Measure DHS. Demographic and Health Surveys. <http://dhsprogram.com/data/available-datasets.cfm> (accessed 15 July 2015).
18. Institut National de la Statistique; Comité National du Recensement (1998). Recensement Général de la Population et de l'Habitat - Côte d'Ivoire 1988. Volume 3: analyse des résultats définitifs. Tome 3: caractéristiques socio-économiques de la population, 108 p. https://www.odsef.fss.ulaval.ca/sites/odsef.fss.ulaval.ca/files/fonds_gp/c-doc_323_odsef.pdf (accessed 17 August 2015).
19. Measure DHS. Demographic and Health Surveys. <http://www.measuredhs.com> (accessed 15 July 2015).
20. QGIS Development Team: QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org/en/site/> (accessed 1 February 2015).
21. WHO (2006) WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age:Methods and development. Geneva, World Health Organization. http://www.who.int/childgrowth/standards/Technical_report.pdf?ua=1 (accessed 18 December 2014).
22. Hagos S, Lunde T, Mariam DH et al. (2014) Climate change, crop production and child under nutrition in Ethiopia; a longitudinal panel study. *BMC Public Health*, **14**:884. <http://www.biomedcentral.com/1471-2458/14/884> (accessed 12 November 2015).

23. Rutstein SO, Staveteig S (2014) Making the Demographic and Health Surveys wealth index comparable. DHS Methodological Reports no. 9. Rockville, MD: ICF International. <http://dhsprogram.com/pubs/pdf/MR9/MR9.pdf> (accessed 8 February 2016).
24. Measure DHS. Sampling and Household Listing Manual - Demographic and Health Surveys Methodology. http://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf (accessed 18 August 2015).
25. United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision, Data. <http://esa.un.org/unpd/wpp/Download/Standard/Population/> (accessed 25 January 2016).
26. Diggle P & Ribeiro PJ (2007) Model-based geostatistics. New York: Springer.
27. Ribeiro PJ, Diggle PJ (2001) geoR: a package for geostatistical analysis. *R-NEWS* **1**, 15–18.
28. Cressie N (1988) Spatial prediction and ordinary kriging. *Math Geol* **20**, 405–421.
29. Cressie NAC (1993) Statistics for Spatial Data, Revised Edition. New York: Wiley and Sons.
30. Kulldorff M (1997) A spatial scan statistic. *Commun Stat - Theory Methods* **26**, 1481–1496.
31. Kulldorff M, Nagarwalla N (1995). Spatial disease clusters: detection and inference. *Stat Med* **14**, 799–810. <http://www.satscan.org/papers/k-sm1995.pdf> (accessed 22 February 2015).
32. Dahly D, Gordon-Larsen P, Emch M *et al.* (2013) The spatial distribution of overweight and obesity among a birth cohort of young adult Filipinos (Cebu Philippines, 2005): an application of the Kulldorff spatial scan statistic. *Nutr Diabetes*. Published online: 1 July 2013. doi: 10.1038/nutd.2013.21.
33. Gebreyesus SH, Mariam DH, Woldehanna T *et al.* (2015) Local spatial clustering of stunting and wasting among children under the age of 5 years: implications for intervention strategies. *Public Health Nutr*. Published online: 23 Dec 2015. doi: 10.1017/S1368980015003377.
34. Dwass M (1957) Modified randomization tests for non-parametric hypothesis. *Ann Math Statist* **28**, 181–187.
35. Haile D, Azage M, Mola T *et al.* (2016) Exploring spatial variations and factors associated with childhood stunting in Ethiopia: spatial and multilevel analysis. *BMC Pediatr* **16**:49, Published online: 15 April 2016. doi: 10.1186/s12887-016-0587-9. <http://bmcpediatr.biomedcentral.com/articles/10.1186/s12887-016-0587-9> (accessed 23 August 2016).

36. Merlo J (2006) A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health* **60**, 290–297.
37. Larsen K, Merlo J (2005) Appropriate Assessment of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression. *Am J Epidemiol* **161**, 81–88.
38. StataCorp LP (2014). Stata Survey Data Reference Manual, Release 14, College Station, Texas: Stata Press Publication. <http://www.stata.com/manuals14/svy.pdf> (accessed 29 October 2016).
39. Adekanmbi VT, Uthman OA, Mudasiru OM (2013) Exploring variations in childhood stunting in Nigeria using league table, control chart and spatial analysis. *BMC Public Health*. Published online: 18 April 2013. doi: 10.1186/1471-2458-13-361.
40. Getis A, Ord JK (1992). The analysis of spatial association by use of distance statistics. *Geogr Anal* **24**, 189–206.
41. Anselin L (2004). Review of Cluster Analysis Software, in: *Report in Fulfilment of Consultant Agreement #2003-04-01 with the North Association of Central Cancer Registries, Inc.*
42. Sahn DE (1990) Malnutrition in Côte d'Ivoire: prevalence and determinants. Washington, D.C: World Bank.
43. Courtin F (2007) Les dynamiques de peuplement induites par la crise ivoirienne dans l'espace ivoiro-burkinabè, au regard de la maladie du sommeil (Population movements caused by the Côte d'Ivoire political crisis in the Côte d'Ivoire-Burkina area, in relation to sleeping sickness). PhD Thesis, University of Montpellier III.
44. Minoiu C, Shemyakina ON (2014) Armed conflict, household victimization, and child health in Côte d'Ivoire. *J Dev Econ* **108**, 237–55.
45. Bundervoet T, Verwimp P, Akresh R (2009) Health and civil war in rural Burundi. *J Hum Resour* **44**, 536–563.
46. Akresh R, Verwimp P, Bundervoet T (2011) Civil War, Crop Failure, and Child Stunting in Rwanda. *Econ Dev Cult Change* **59**, 777–810.
47. Minoiu C, Shemyakina ON (2012) Child Health and Conflict in Côte d'Ivoire. *Am Econ Rev* **102**, 294–299.
48. Cogneau D, Jedwab R (2012) Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Côte d'Ivoire. *Econ Dev Cult Change* **60**, 507–534.
49. Grimm M, Guenard C, Mesple-Soms S (2002) What has happened to the urban population in Côte d'Ivoire since the 1980s? An analysis of monetary

poverty and deprivation over 15 years of household data. *World Dev* **30**, 1073–1095.

50. Martin-Prével Y, Traissac P, Massamba J-P *et al.* (2000) Deterioration in the nutritional status of young children and their mothers in Brazzaville, Congo, following the 1994 devaluation of the CFA franc. *Bull World Health Organ* **78**, 108–118.
51. Black M, Perez-Escamilla R, Fernandez Rao S (2015) Integrating Nutrition and Child Development Interventions: Scientific Basis, Evidence of Impact, and Implementation Considerations. *Adv Nutr Int Rev J* **6**, 852–859.
52. Kandala N-B, Madungu TP, Emina J-BO *et al.* (2011) Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? *BMC Public Health*. Published online: 25 April 2011. doi: 10.1186/1471-2458-11-261.

Fig. 1 Average stunting prevalence in children aged 0-35 months in Côte d'Ivoire in: (a) 1994; (b) 1998; (c) 2011.

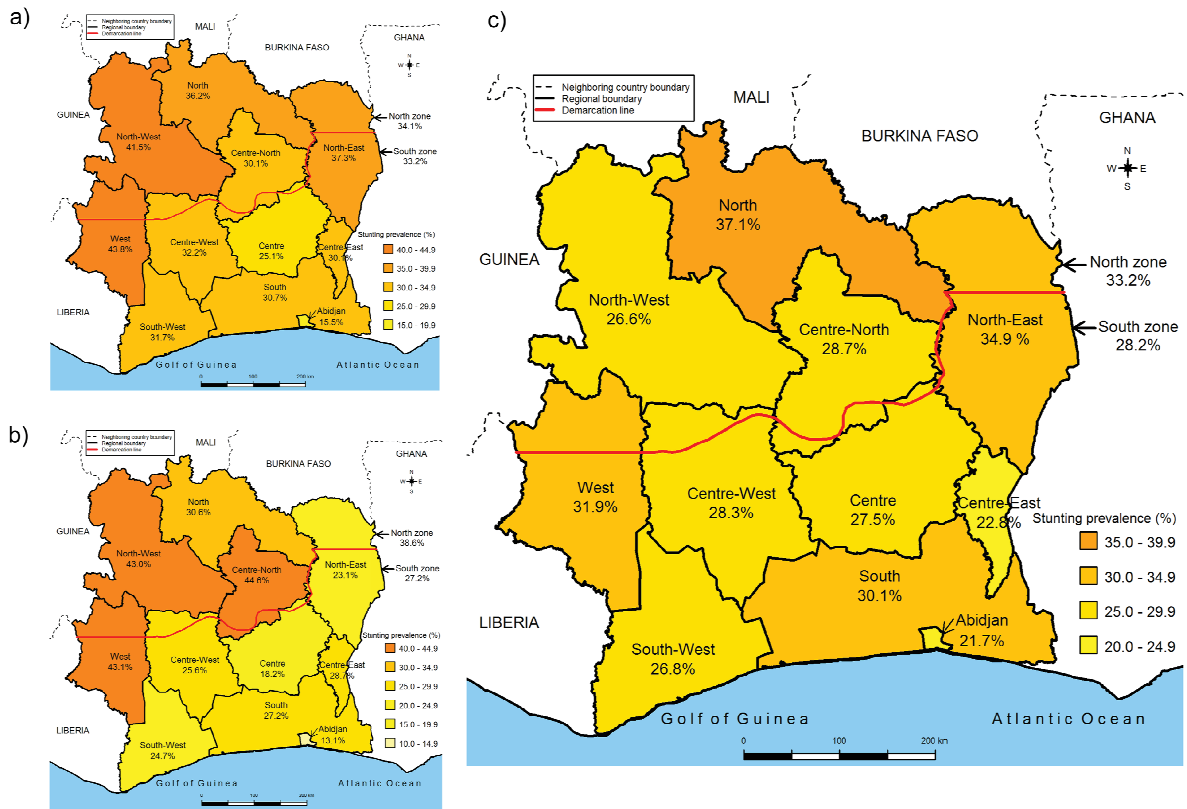


Fig. 2 Interpolated stunting prevalence in children aged 0-35 months in Côte d'Ivoire in: (a) 1994; (b) 1998; (c) 2011.

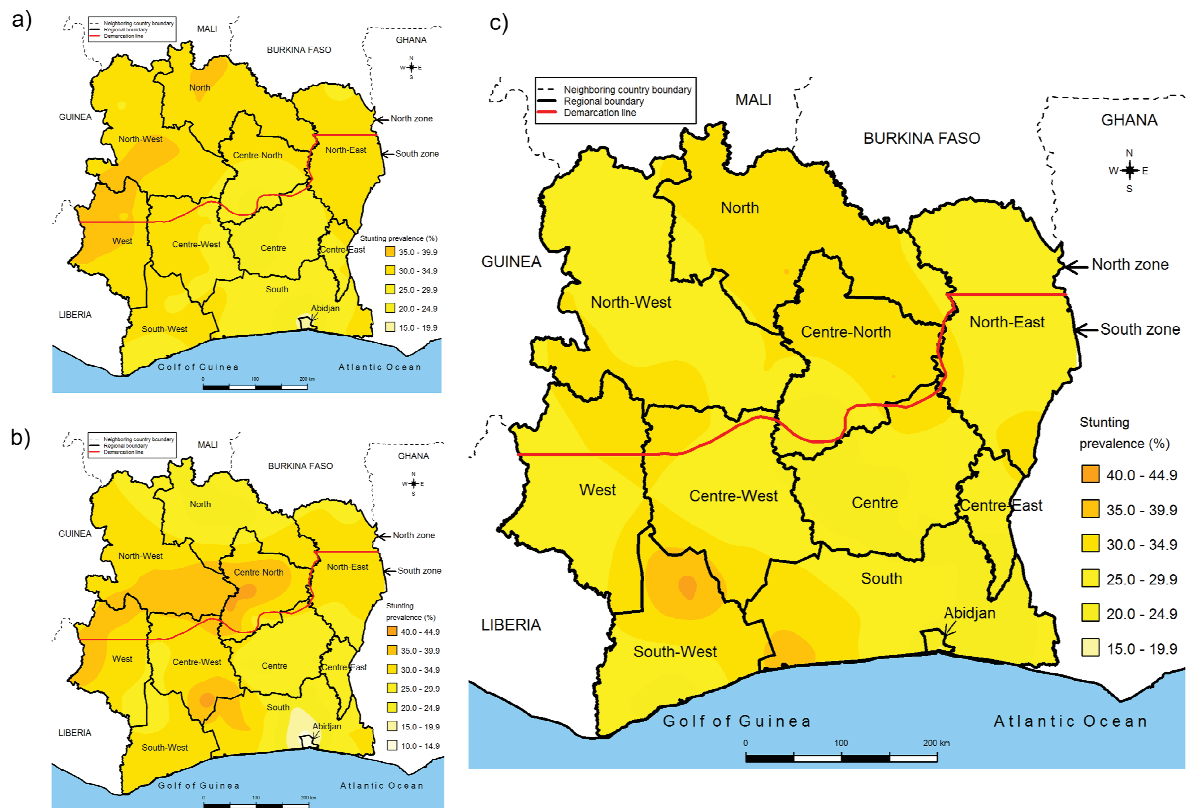


Fig. 3 Spatial distribution of the sampling enumeration areas and clusters with high and low stunting prevalence in children aged 0-35 months in Côte d'Ivoire in: (a) 1994; (b) 1998; (c) 2011.

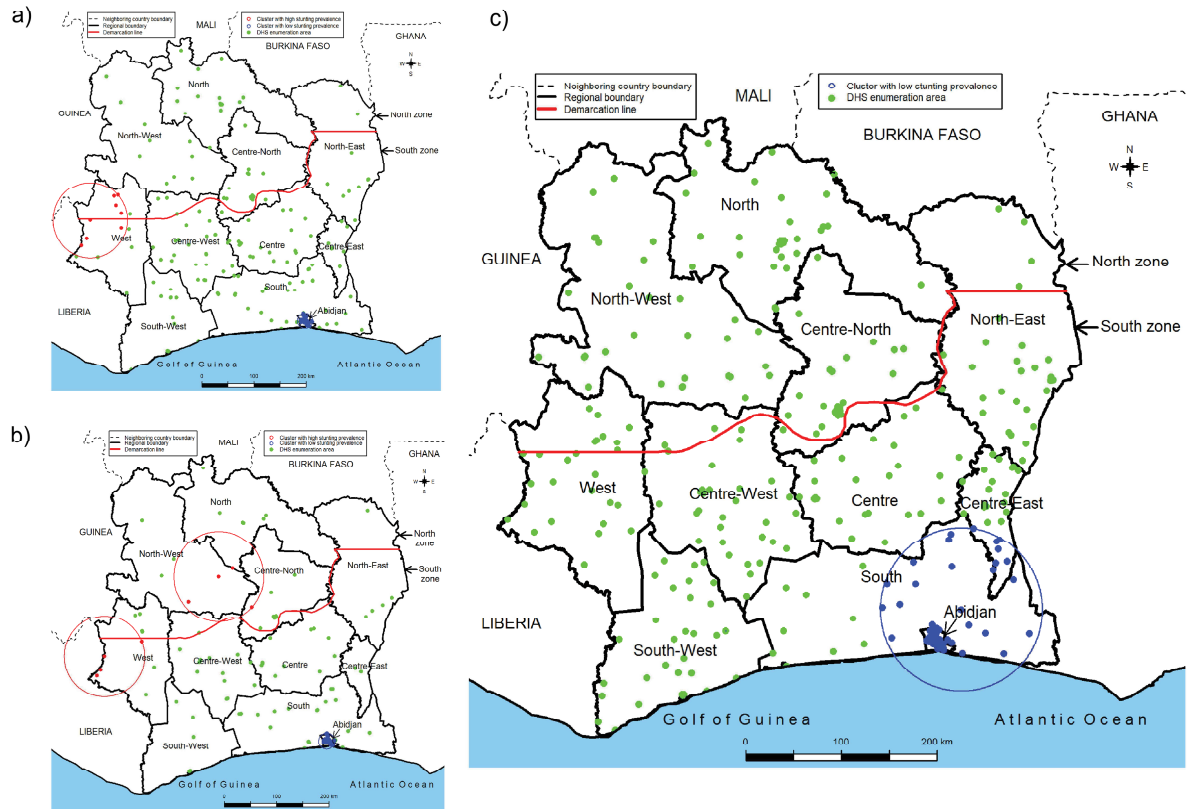


Table 1. Background characteristics of 0-35-month-old children in Côte d'Ivoire from 1994 to 2011.

Characteristics	Year			<i>P</i> [†]
	1994 (<i>n</i> =3356)	1998 (<i>n</i> =1113)	(<i>n</i> 2011 (<i>n</i> =2240)	
Zone				0.803
Abidjan	513 (15.3)	159 (14.3)	291 (13.0)	
South	2117 (63.1)	699 (62.8)	1352 (60.4)	
North	726 (21.6)	255 (22.9)	597 (26.6)	
Stunting				0.318
Yes	1030 (30.7)	309 (27.8)	644 (28.7)	
No	2326 (69.3)	804 (72.2)	1596 (71.3)	
Wasting				0.135
Yes	366 (10.9)	94 (8.4)	218 (9.7)	
No	2990 (89.1)	1019 (91.6)	2022 (90.3)	
Sex				0.436
Male	1704 (50.8)	551 (49.5)	1092 (48.8)	
Female	1652 (49.2)	562 (50.5)	1148 (51.2)	
Age group (months)				0.014
<6	641 (19.1)	161 (14.5)	362 (16.2)	
6-11	593 (17.7)	226 (20.3)	404 (18.0)	
12-17	570 (17.0)	217 (19.5)	391 (17.4)	
18-23	509 (15.2)	200 (18.0)	350 (15.6)	
24-29	551 (16.4)	166 (14.9)	364 (16.3)	
30-35	492 (14.6)	143 (12.8)	369 (16.5)	
Birth order [‡]				0.001
1	637 (19.0)	252 (22.6)	440 (21.8)	
2-3	1076 (32.1)	363 (32.6)	748 (37.0)	
4-5	830 (24.7)	240 (21.6)	457 (22.6)	
≥6	813 (24.2)	258 (23.2)	375 (18.6)	
Level of education [§]				0.652
No formal education	2239 (66.7)	734 (65.9)	1376 (64.4)	
Primary	826 (24.6)	298 (26.8)	562 (26.3)	
Secondary/higher	291 (8.7)	81 (7.3)	200 (9.3)	
Mother's BMI				<0.001
Normal	2371 (77.8)	766 (76.2)	1354 (75.1)	
Underweight	249 (8.2)	68 (6.8)	84 (4.7)	
Overweight/obese	425 (14.0)	171 (17.0)	364 (20.2)	

CWI [¶]				0.008
Poorest	744 (22.2)	318 (28.6)	490 (21.9)	
Poorer	685 (20.4)	270 (24.3)	522 (23.3)	
Middle	632 (18.8)	177 (15.9)	467 (20.8)	
Richer	575 (17.1)	143 (12.8)	491 (21.9)	
Richest	720 (21.5)	205 (18.4)	270 (12.1)	

Data are presented as *n* (%). DHS sampling weights were used.

[†] *P*-values of chi-square trend tests accounting for the complex sampling design.

[‡] For birth order, 220 values were missing, all in 2011.

[§] For level of education, 102 values were missing, all in 2011.

^{||} For maternal BMI, 311, 108 and 438 children had missing values, respectively, in 1994, 1998 and 2011.

[¶] The comparative wealth index (CWI) was computed based on 172 DHS conducted worldwide from 1990-2012⁽²³⁾

Table 2. Non-adjusted effect of time on stunting prevalence in 0-35-month-old children in Côte d'Ivoire and its three zones, from 1994 to 2011.

	Prevalence (%)			1998 versus 1994			2011 versus 1994			2011 versus 1998			Trends analysis
	1994	1998	2011	OR	95% CI	<i>P</i> [†]	OR	95% CI	<i>P</i> [†]	OR	95% CI	<i>P</i> [†]	<i>P</i> [‡]
Zone													
Abidjan	15.5	13.1	21.7	0.82	0.49, 1.38	0.455	1.51	0.88, 2.61	0.136	1.85	1.08, 3.15	0.024	0.106
South	33.2	27.2	28.3	0.75	0.56, 1.00	0.046	0.79	0.66, 0.96	0.017	1.06	0.78, 1.45	0.701	0.024
North	34.1	38.6	33.2	1.28	0.90, 1.81	0.173	0.96	0.72, 1.28	0.782	0.75	0.53, 1.07	0.117	0.429
Côte d'Ivoire	30.7	27.8	28.7	0.93	0.73, 1.17	0.525	0.91	0.78, 1.07	0.250	0.98	0.77, 1.25	0.889	0.318

Data are based on de-normalised sample weights; [†]*P*-values for logistic regression, [‡]*P*-values of chi-square tests for trend.

Table 3. Multivariable mixed-effects logistic regression analysis of factors associated with stunting in 0-35-month-old children in Côte d'Ivoire and in its three zones, from 1994 to 2011.

Characteristics	Côte d'Ivoire (n=5852)			Abidjan (n=845)			South zone (n=3630)			North zone (n=1377)		
	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>
Year			0.090			0.102			0.007			0.632
1994	1.00			1.00			1.00			1.00		
1998	0.64	0.34, 1.20	0.165	0.82	0.48, 1.41	0.478	0.71	0.53, 0.95	0.020	1.20	0.81, 1.79	0.361
2011	1.39	0.72, 2.64	0.323	1.62	0.87, 3.01	0.129	0.74	0.58, 0.94	0.012	1.00	0.72, 1.39	0.992
Zone			0.034	NA			NA			NA		
Abidjan	1.00											
South	1.81	1.14, 2.87	0.012									
North	1.87	1.14, 3.06	0.012									
Zone × year			0.028	NA			NA			NA		
Abidjan+1994	1.00											
South+1998	1.09	0.55, 2.16	0.811									
South +2011	0.53	0.27, 1.05	0.069									
North+1998	1.96	0.94, 4.12	0.074									
North+2011	0.69	0.33, 1.41	0.305									
Sex			0.019			0.060			0.002			
Male	1.00			1.00			1.00					
Female	0.77	0.62, 0.96	0.019	0.66	0.42, 1.02	0.060	0.70	0.56, 0.88	0.002			

Wealth index			0.004				0.049			0.119
Poorest	1.00			1.00			1.00			
Poorer	0.84	0.62, 1.13	0.246	0.90	0.64, 1.26	0.537	0.78	0.44, 1.37	0.388	
Middle	0.80	0.60, 1.06	0.113	0.89	0.62, 1.27	0.513	0.70	0.45, 1.09	0.114	
Richer	0.80	0.58, 1.12	0.196	0.96	0.62, 1.48	0.834	0.74	0.44, 1.24	0.249	
Richest	0.47	0.31, 0.70	<0.001	0.45	0.25, 0.79	0.006	0.41	0.21, 0.79	0.008	

aOR, adjusted odds ratio; BMI, body mass index; NA, not applicable.

Each model includes only variables that were significant at $P < 0.20$ in univariable and adjusted models.

Data are based on de-normalised sample weights.

Table 4. Multivariable mixed-effects logistic regression analysis of factors associated with stunting in 0-35-month-old children in Côte d'Ivoire and in its three zones from 1994 to 2011, with 1998 as the year of reference.

Characteristics	Côte d'Ivoire			Abidjan			South zone			North zone		
	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>	aOR	95% CI	<i>P</i>
Year			0.136			0.102			0.007			0.632
1998	1.00			1.00			1.00			1.00		
1994	1.18	0.94, 1.49	0.146	1.21	0.71, 2.08	0.478	1.41	1.06, 1.88	0.020	0.83	0.56, 1.24	0.361
2011	1.01	0.78, 1.31	0.936	1.96	1.06, 3.64	0.013	1.04	0.74, 1.46	0.842	0.83	0.55, 1.28	0.400
Sex			0.023			0.060			0.002			
Male	1.00			1.00			1.00					
Female	0.78	0.63, 0.97	0.023	0.66	0.42, 1.02	0.060	0.70	0.56, 0.88	0.002			
Age group (months)			<0.001			0.002			<0.001			0.001
<6	1.00			1.00			1.00			1.00		
6-11	1.05	0.72, 1.53	0.809	2.22	0.78, 6.31	0.134	0.94	0.57, 1.54	0.807	1.03	0.53, 2.02	0.927
12-17	1.79	1.22, 2.64	0.003	4.42	1.70, 11.46	0.002	1.67	1.08, 2.60	0.023	1.71	0.72, 4.04	0.222
18-23	2.80	1.85, 4.25	<0.001	4.51	1.69, 12.01	0.003	2.89	1.62, 5.17	<0.001	2.31	1.29, 4.13	0.075
24-29	2.65	1.79, 3.93	<0.001	4.80	2.01, 11.46	0.001	2.58	1.62, 4.12	<0.001	2.29	0.92, 5.71	0.144
30-35	3.74	2.52, 5.56	<0.001	7.05	2.77, 17.94	<0.001	3.70	2.17, 6.30	<0.001	3.18	1.56, 6.52	0.002
Birth order			0.559			0.003						
1	1.00			1.00								
2-3	0.87	0.68, 1.10	0.236	1.38	0.80, 2.37	0.242						
4-5	0.84	0.64, 1.09	0.192	0.44	0.21, 0.92	0.029						
≥ 6	0.84	0.61, 1.16	0.286	1.48	0.76, 2.91	0.249						

Level of education			0.007			0.200			0.114			0.177
No formal education	1.00			1.00			1.00			1.00		
Primary	0.74	0.59, 0.94	0.012	0.78	0.43, 1.39	0.391	0.76	0.58, 1.00	0.047	0.97	0.65, 1.46	0.893
Secondary/higher	0.65	0.46, 0.90	0.011	0.56	0.29, 1.06	0.072	0.81	0.51, 1.28	0.359	0.57	0.31, 1.04	0.066
Mother's BMI			0.001						0.031			0.001
Normal	1.00						1.00			1.00		
Underweight	1.49	1.05, 2.13	0.027				1.11	0.71, 1.74	0.658	2.20	1.35, 3.57	0.002
Overweight/obese	0.63	0.47, 0.85	0.002				0.58	0.38, 0.88	0.010	0.61	0.37, 0.99	0.045
Wealth index			<0.001						0.049			0.119
Poorest	1.00						1.00			1.00		
Poorer	0.77	0.57, 1.04	0.089				0.90	0.64, 1.26	0.537	0.78	0.44, 1.37	0.388
Middle	0.72	0.55, 0.95	0.021				0.89	0.62, 1.27	0.513	0.70	0.45, 1.09	0.114
Richer	0.71	0.51, 0.97	0.033				0.95	0.62, 1.48	0.834	0.74	0.44, 1.24	0.249
Richest	0.38	0.26, 0.55	<0.001				0.45	0.25, 0.79	0.006	0.41	0.21, 0.79	0.008

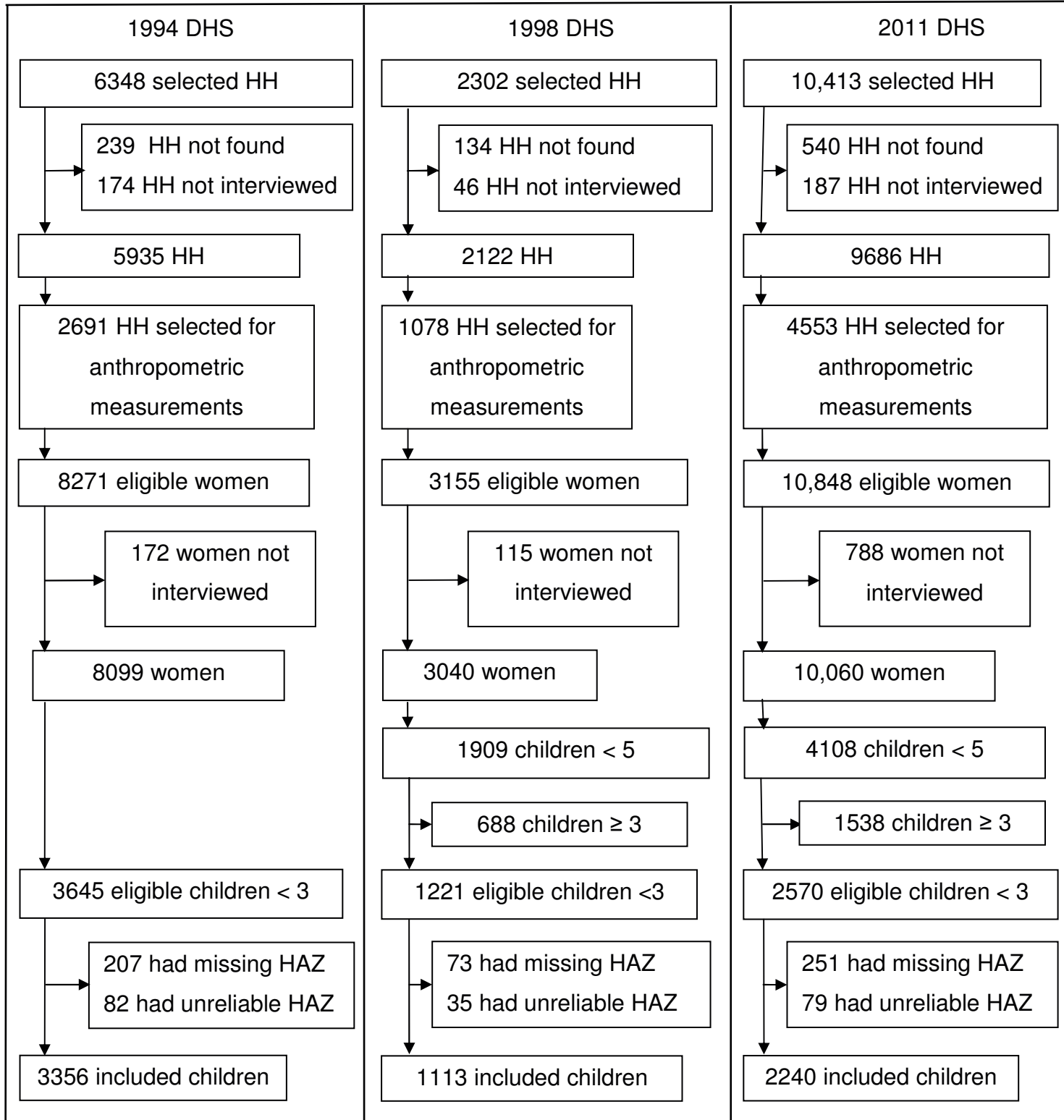
aOR, adjusted odds ratio. BMI, body mass index. NA, not applicable.

Each model includes only variables that were significant at $P < 0.20$ in univariable and adjusted models;

Data are based on de-normalised sample weights.

Supplementary online material

Flow chart for households, women and children under three, 1994, 1998 and 2011
Côte d'Ivoire DHS



DHS, demographic and health survey; HAZ, height-for-age z-score; HH, household.

Supplemental online material

Table 1. Background characteristics of children under the age of three in Abidjan, the South zone and the North zone, in DHS 1994, 1998 and 2011.

Characteristics	Abidjan				South zone				North zone			
	1994	1998	2011	P [†]	1994	1998	2011	P [†]	1994	1998	2011	P [†]
	N=513	N=159	N=291		N=2117	N=699	N=1352		N=726	N=255	N=597	
Stunting				0.11				0.02				0.43
Yes	80 (15.5)	21 (13.1)	63 (21.7)		702 (33.2)	190 (27.2)	382 (28.2)		248 (34.2)	98 (38.4)	198 (33.2)	
No	433 (84.5)	138 (86.9)	228 (78.3)		1415 (66.8)	509 (72.8)	970 (71.8)		478 (65.8)	157 (61.6)	399 (66.8)	
Wasting				0.46				0.36				0.16
Yes	54 (10.5)	11 (6.9)	28 (9.6)		218 (10.3)	56 (8.0)	137 (10.1)		93 (12.9)	27 (10.6)	53 (8.9)	
No	459 (89.5)	148 (93.1)	263 (90.4)		1899 (89.7)	643 (92.0)	1215 (89.9)		633 (87.2)	228 (89.4)	544 (91.1)	
Sex				0.12				0.33				0.36
Male	280 (54.6)	79 (49.7)	137 (47.1)		1060 (50.1)	357 (51.1)	644 (47.6)		364 (50.1)	115 (45.1)	311 (52.1)	
Female	233 (45.4)	80 (50.3)	154 (52.9)		1057 (49.9)	342 (48.9)	708 (52.4)		362 (49.9)	140 (54.9)	286 (48.9)	
Age group (mo)				0.30				0.11				0.25
<6	94 (18.3)	26 (16.3)	36 (12.4)		410 (19.4)	103 (14.7)	225 (16.6)		137 (18.9)	32 (12.6)	101 (16.9)	
6-11	82 (16.1)	28 (17.7)	46 (15.8)		381 (18.0)	138 (19.7)	255 (18.9)		129 (17.8)	60 (23.5)	102 (17.1)	
12-17	97 (18.8)	31 (19.5)	60 (20.6)		351 (16.6)	140 (20.1)	218 (16.1)		123 (16.9)	46 (18.0)	113 (18.9)	
18-23	98 (19.1)	25 (15.7)	55 (18.9)		291 (13.7)	123 (17.6)	211 (15.6)		120 (16.5)	51 (20.0)	85 (14.2)	
24-29	82 (16.1)	23 (14.5)	39 (13.4)		363 (17.1)	109 (15.6)	233 (17.3)		106 (14.6)	34 (13.3)	92 (15.4)	
30-35	60 (11.6)	26 (16.3)	55 (18.9)		321 (15.2)	86 (12.3)	210 (15.5)		111 (15.3)	32 (12.6)	104 (17.5)	

Birth order ¹				<0.0				0.02				0.36
1	123 (24.0)	45 (28.3)	57 (23.1)	01	396 (18.7)	159 (22.8)	286 (23.1)		118 (16.3)	48 (18.8)	97 (18.2)	
2-3	188 (36.6)	52 (32.7)	125 (50.6)		654 (30.9)	217 (31.0)	436 (35.2)		234 (32.3)	95 (37.2)	186 (34.9)	
4-5	108 (21.1)	35 (22.0)	53 (21.4)		532 (25.1)	159 (22.8)	268 (21.6)		189 (26.0)	45 (17.7)	135 (25.3)	
6+	94 (18.3)	27 (17.0)	12 (4.9)		535 (25.3)	164 (23.4)	249 (20.1)		185 (25.4)	67 (26.3)	115 (21.6)	
Maternal education ²				0.50				0.51				0.53
None	247 (48.2)	66 (41.5)	118 (42.9)		1428 (67.4)	462 (66.1)	831 (64.0)		563 (77.6)	206 (80.8)		
Primary	164 (31.9)	57 (35.9)	86 (31.3)		550 (26.0)	201 (28.8)	367 (28.2)		114 (15.7)	39 (15.3)		
Secondary	102 (19.9)	36 (22.6)	71 (25.8)		139 (6.6)	36 (5.1)	101 (7.8)		49 (6.7)	10 (3.9)		
Mother's BMI ³				0.15				<0.0				0.01
Normal	310 (64.3)	86 (61.0)	125 (56.3)		1547 (81.3)	486 (77.6)	861 (77.9)	01	514 (77.7)	195 (81.2)	368 (77.3)	
Underweight	24 (5.0)	10 (7.1)	6 (2.7)		149 (7.9)	39 (6.5)	53 (4.8)		75 (11.4)	20 (8.3)	25 (5.3)	
Overweight	148 (30.7)	45 (31.9)	91 (41.0)		206 (10.8)	100 (15.9)	191 (17.3)		72 (10.9)	25 (10.5)	83 (17.4)	
/obese												
Wealth index ⁴				<0.0				0.24				0.01
Poorest	0	0	0	01	501 (45.0)	222 (31.8)	354 (26.2)		233 (32.0)	138 (54.1)	123 (20.6)	
Poorer	11 (2.2)	6 (3.8)	0		607 (23.6)	240 (34.3)	362 (26.8)		108 (14.8)	41 (16.1)	101 (16.9)	
Middle	48 (9.4)	27 (17.0)	10 (3.4)		449 (8.7)	127 (18.2)	262 (19.4)		161 (22.2)	26 (10.2)	151 (25.3)	
Richer	154 (29.9)	67 (42.1)	76 (26.1)		342 (13.7)	58 (8.3)	219 (16.2)		113 (15.6)	23 (9.0)	125 (20.9)	
Richest	300 (58.5)	59 (37.1)	205 (70.5)		218 (9.0)	52 (7.4)	155 (11.4)		111 (15.4)	27 (10.6)	97 (16.3)	

Data are presented as n (%) or median (IQR). [†] P-values are survey design-based.

¹ For birth order, 43 values in Abidjan, 113 in South zone and 64 in North zone were missing, all in 2011

² For maternal education, 16 values in Abidjan, 53 in the South zone and 33 in the North zone were missing, all in 2011

³ For maternal BMI, (31, 215, 65), (18, 75, 15) and (69, 247, 122) children had missing values, respectively in 1994, 1998 and 2011 for (Abidjan, and the South and North zones, respectively)

⁴ Quintiles are based on the comparative wealth index of each DHS

Conclusion générale

Au cours de cette thèse, nous nous sommes intéressés aux techniques d'analyse spatiale utilisées dans le domaine de la santé. Nous avons appliqué ces techniques d'une part à la thématique du VIH/SIDA au Burundi et d'autre part à la thématique de la malnutrition en Côte d'Ivoire. Les techniques utilisées consistaient tout d'abord à prédire les niveaux de prévalence de ces deux problèmes sanitaires en des points non échantillonnés, soit par la méthode des noyaux, soit par la méthode du krigeage ordinaire. Nous avons ensuite procédé à la détection des agrégats spatiaux ou clusters par la méthode de Kulldorff puis à l'identification de zones de différents risques par la méthode du partitionnement oblique du territoire. Finalement, nous avons utilisé la modélisation spatiale pour identifier les facteurs associés.

Les techniques d'analyse spatiale disponibles actuellement présentent de nombreux avantages tels que la possibilité de produire des cartes descriptives indépendamment des limites administratives, d'identifier des zones de haute ou basse prévalence et de prendre en compte la localisation précise des données recueillies pour la modélisation du phénomène étudié. Elles présentent toutefois également de nombreuses limites.

La principale limite concerne la nécessité de fixer des hyper paramètres des modèles en faisant de nombreux choix sans critères objectifs. À titre d'exemple, l'interpolation spatiale suppose d'avoir choisi au préalable soit une des techniques d'interpolation spatiale utilisant le degré de similarité ou le degré de lissage, soit une des techniques d'interpolation géostatistiques tenant compte de la dépendance spatiale des données lors de l'interpolation. Elle suppose également d'avoir choisi la maille ou grille d'interpolation. Les techniques utilisant le degré de lissage font partie des techniques déterministes et exigent le choix de la forme mathématique de la fonction de lissage. Celles utilisant la notion de voisinage exigent la définition du degré de similarité des sites auxquels les données sont observées. Les techniques d'interpolation géostatistiques exigent, quant à elles, la connaissance des propriétés statistiques du phénomène étudié. Pour la méthode des noyaux utilisée dans notre étude sur le VIH/SIDA au Burundi, la surface estimée est influencée non seulement

par le choix de la fonction de lissage mais également par le choix de la fenêtre de lissage et ces choix sont basés sur des éléments subjectifs. Pour le krigeage ordinaire utilisé dans notre étude sur le retard de croissance en Côte d'Ivoire, l'analyste doit faire l'estimation de l'indice d'autocorrélation spatiale puis l'estimation du variogramme, ce qui demande un travail long, fastidieux et peu intuitif.

Une autre limite importante de ces techniques d'interpolation tient à l'absence de tests statistiques permettant de comparer les modèles de variogramme ou les techniques d'interpolation spatiale, ou de tester la significativité des différences de prévalence observées. Il n'existe pas non plus de test permettant de vérifier si la surface produite à l'aide des valeurs prédites s'ajuste bien aux données relatives au phénomène étudié.

En outre, les résultats de l'interpolation spatiale à proximité des limites de la zone d'étude (frontières du pays par exemple) sont biaisés par le fait que la valeur du phénomène étudié est inconnue et donc considérée comme nulle pour les points voisins d'un point d'interpolation spatiale situés à l'extérieur de la zone d'étude. Des méthodes de correction de ces effets de bords existent mais nécessitent également des choix a priori.

Par ailleurs, la représentation graphique des valeurs observées ou prédites peut différer fortement selon le type de symbole ou les gammes de couleurs utilisés, ce qui peut suggérer à tort des différences. L'utilisateur doit choisir, sans critères objectifs, le nombre et le type de couleurs qui permettent de représenter les configurations spatiales du phénomène étudié sur une carte lissée.

D'un autre côté, les résultats de la méthode du scan spatial de Kulldorff dépendent fortement du choix des paramètres tels que le rayon et la taille maximale de la fenêtre circulaire de balayage du territoire. Le choix du rayon de la fenêtre de balayage peut être guidé par les distances seuil fournies par le corrélogramme ou fait à partir d'un choix raisonné de l'utilisateur mais ce choix reste subjectif. La taille maximale de la fenêtre de balayage doit être choisie entre 0 % et 50 % de la population à risque.

Enfin, des logiciels spécialisés sont nécessaires pour réaliser les analyses spatiales. Ainsi, les techniques d'interpolation spatiale sont implémentées dans des logiciels d'analyse de l'information géographique tels que ArcGIS, QGIS, MapInfo, Sada, Crimestat, Clusterseer et GeoDA. De même, les fonctionnalités permettant de faire la modélisation spatiale pour rechercher les facteurs associés au phénomène d'intérêt sont implémentées dans le logiciel spécialisé BayesX. L'intégration de ces techniques dans les logiciels de statistique courants tels que Stata, SPSS et SPAD favoriserait leur utilisation plus fréquente.

A l'issue de cette thèse, il nous semble que les techniques d'analyse spatiale nécessitent encore des développements méthodologiques pour qu'elles deviennent plus performantes, attractives et utilisables par le plus grand nombre d'analystes.

Références bibliographiques

1. INSA (Institut National des Sciences Appliquées) de Strasbourg. Intégration de la photogrammétrie et de la télédétection dans les systèmes d'informations géographiques (SIG). Utilisation et qualité, Strasbourg 1990. [February 20, 2016]. Disponible sur : <http://retro.seals.ch/cntmng;jsessionid=B530C5A78287B81E39F3990DC046522B?pid=geo-006:1990:88::312>
2. De Joie C. Les Systèmes d'Information Géographique, des outils au service du raisonnement géographique. [February 22, 2016]. Disponible sur : http://www.clionautes.org/spip.php?page=article&id_article=553
3. Floch J-M. Détection des disparités socio-économiques, l'apport de la statistique spatiale. [February 23, 2016]. Institut National de la Statistique et des Études Économiques; 2012. Disponible sur : http://www.insee.fr/fr/publications-et-services/docs_doc_travail/H2012-04.pdf
4. Fradelos E, Papathanasiou I, Mitsi D, et al. Health Based Geographic Information Systems (GIS) and their Applications. *Acta Inform Medica*. 2014;22(6):402.
5. McLafferty SL. GIS and Health Care. *Annu Rev Public Health*. 2003;24(1):25-42.
6. Villain J. Conquête de l'espace - Succès et désillusions. *Encyclopædia Universalis*. [February 20, 2016]. Disponible sur : <http://www.universalis.fr/encyclopedie/espace-conquete-de-l-succes-et-desillusions/>
7. National imagery and mapping agency. Department of defense World Geodetic System 1984, Technical report, Third edition. 2000. [April 11, 2016]. Disponible sur : <http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf>
8. IGN. SEIG: Description des principales projections planes représentant la terre. [April 13, 2016]. Disponible sur : <http://seig.ensg.ign.fr/fichchap.php?NOCONT=&NOCHEM=CHEMS007&NOFICHE=FP52&NOLISTE=5&N=6&RPHP=&RCO=&RCH=&RF=&RPF=>
9. Lawson AB. *Statistical Methods in Spatial Epidemiology*. Second Edition. John Wiley & Sons, 2006. [December 30, 2015]. Disponible sur : <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470014849.html>
10. Morency C. Etude de méthodes d'analyse spatiale et illustration à l'aide de microdonnées urbaines de la Grande Région de Montréal, Département des génies civil, géologique et des mines, École Polytechnique de Montréal. *Cah Sci Transp*. 2006;49(7).
11. Gaetan C, Guyon X. *Modélisation et statistique spatiales*. Springer, 2008.

12. Arnaud M, Emery X. Estimation et interpolation spatiale: méthodes déterministes et méthodes géostatistiques. Paris: Hermes Science; 2000.
13. Cressie NAC. Statistics for spatial data, Revised Edition. New York: Wiley and Sons, 900p. 1993. 900 p.
14. Getis A. Spatial weights matrices. *Geogr Anal.* 2009;41(4):404–10.
15. Bailly A. Distances et espaces: vingt ans de géographie des représentations. *Espace Géographique.* 1985;14(3):197–205.
16. Tony ES. Spatial Weights Matrices. [December 24, 2015]. Disponible sur : http://www.seas.upenn.edu/~ese502/lab-content/extra_materials/SPATIAL%20WEIGHT%20MATRICES.pdf
17. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geogr Anal.* 1992;24(3):189–206.
18. Elliott P, Wekefield J. Disease clusters: should they be investigated, and if so, when and why? *Journal of the Royal Statistical Society 2002: Series A (Statistics in Society)*; 164(1).
19. Robinson TP. Spatial statistics and geographical information systems in epidemiology and public health. In: *Parasitology B-A* in, éditeur. Academic Press; 2000. (Remote Sensing and Geographical Information Systems in Epidemiology; vol. 47). [December 24, 2015]. Disponible sur : <http://www.sciencedirect.com/science/article/pii/S0065308X00470077>
20. Moran PAP. The interpretation of statistical maps. *Journal of the Royal Statistical Society 1948; Series B*, 10:243–51.
21. Besag J, Newell J. The Detection of Clusters in Rare Diseases. *J R Stat Soc Ser A Stat Soc.* 1991;154(1):143.
22. Zaninetti J-M. Statistique spatiale: méthodes et applications géomatiques. Paris: Hermès science publications; 2005. 320 p.
23. Sokal RR, Oden NL. Spatial autocorrelation in biology. 1. Methodology. *Biol J Linn Soc*, 1978;10: 199–228.
24. Cao R, Reis E. A new proposal to adjust Moran's I for population density. *Stat Med.* 1999;18:2147–62.
25. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med.* 1995;14(21-22):2323–34.
26. Tango T. Asymptotic Distribution of an Index for Disease Clustering. *Biometrics.* 1990;46(2):351.
27. Waller L, Gotway C. Applied spatial statistics for public health data. Hoboken, NJ: Wiley-Interscience; 2004. 494 p. (Wiley series in probability and statistics).

28. Oden NL. Assessing the significance of a spatial correlogram. *Geogr Anal.* 1984;16(1):1–16.
29. Geary RC. The Contiguity Ratio and Statistical Mapping. *Inc Stat.* nov 1954;5(3):115–45.
30. Lebart L. Analyse statistique de la contiguïté. Publication de l'Institut de Statistiques de l'Université de Paris 1969;28: 81–112.
31. Anselin L. Local indicators of spatial association. *Geographical Analysis* 1995; 27: 93–115.
32. Jayeth H. Économétrie et données spatiales: une introduction à la pratique. *Cahiers d'économie et sociologie rurales.* 2001; vol. 58-59, pp. 105–29.;
33. Cliff AD, Ord JK. *Spatial Autocorrelation.* Pion Ltd; 1973. 178 p.
34. Whittemore A, Friend N, Brown B, et al. A Test to Detect Clusters of Disease. *Biometrika.* 1987;74(3):631.
35. Tango T. The Detection of Disease Clustering in Time. *Biometrics.* 1984;40(1):15–26.
36. Whittemore A, Keller J. On Tango's index for disease clustering in time. *Biometrics.* 1986;42(1):218.
37. Whittemore AS, Keller JB. Properties of Tango's index for detecting clustering in time. *Stat Med,* 1993; 12(19-20): 1813–27.
38. Jackson M, Waller L. Exploring Goodness-of-Fit and Spatial Correlation Using Components of Tango's Index of Spatial Clustering. *Geogr Anal.* 2005;37(4):371–82.
39. Rogerson P. The Detection of Clusters Using a Spatial Version of the Chi-Square Goodness-of-Fit Statistic. *Geogr Anal.* 1999;31(2):130–47.
40. Rogerson P, Yamada I. *Statistical detection and surveillance of geographic clusters.* Boca Raton: CRC Press; 2009. 322 p. (Chapman & hall/ CRC Interdisciplinary statistics series).
41. Griffith D, Amrhein C, Huriot J-M. *Econometric Advances in Spatial Modelling and Methodology: Essays in Honour of Jean Paelinck.* Boston, MA: Springer US; 1998 [March 10, 2016]. Disponible sur : <http://dx.doi.org/10.1007/978-1-4757-2899-6>
42. Anselin L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. pp. 111–125 in M. M. Fischer, H. J. Scholten and D. Unwin (eds) *Spatial analytical perspectives on GIS,* London, Taylor and Francis. 1996.

43. Fischer M, Getis A. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer-Verlag Berlin and Heidelberg GmbH & Co. K; 2009. 811 p.
44. Kulldorff M, Nagarwalla N (1995). Spatial disease clusters: Detection and inference. *Stat Med* 14, 799-810. [February 22, 2015]. Disponible sur : <http://www.satscan.org/papers/k-sm1995.pdf>
45. Kulldorff M. Tests of Spatial Randomness Adjusted for an Inhomogeneity: A General Framework. *J Am Stat Assoc.* 2006;101(475):1289-305.
46. Elliott P, Wartenberg D. Spatial Epidemiology: Current Approaches and Future Challenges. *Environ Health Perspect.* 15 avr 2004;112(9):998–1006.
47. Wand H, Ramjee G. Targeting the hotspots: investigating spatial and demographic variations in HIV infection in small communities in South Africa. *J Int AIDS Soc.* 2010;13(1):41.
48. Sartorius B, Kahn K, Vounatsou P, Collinson MA, Tollman SM. Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992–2007. *Glob Health Action.* 2010; 3(0). [October 25, 2015]. Disponible sur : <http://www.globalhealthaction.net/index.php/gha/article/view/5225>
49. González R, Augusto OJ, Mungambe K, Pierrat C, Pedro EN, Sacoor C, et al. HIV Incidence and Spatial Clustering in a Rural Area of Southern Mozambique. Carr JK, éditeur. *PLOS ONE.* 6 juill 2015;10(7):e0132053.
50. Kulldorff M. SaTScan™ User Guide. 2006 [April 25, 2016]. Disponible sur : http://www.satscan.org/cgi-bin/satscan/register.pl/Current%20Version:%20SaTScan%20v9.1.1%20released%20March%209%202011.?todo=process_userguide_download
51. Abrams A, Kulldorff M, Kleinman K. Empirical/asymptotic p-values for Monte Carlo-based hypothesis testing: an application to cluster detection using the scan statistic. *Adv Dis Surveill.* 2006;1(1).
52. Chen J, Roth R, Naito A, et al. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int J Health Geogr.* 2008;7(1):57.
53. Tango T. A spatial scan statistic scanning only the regions with elevated risk. *Adv Dis Surveill.* 2007;4:117.
54. Jin C, MacEachren A, Lengerich E. Visual Analytics of Spatial Scan Statistic Results. [April 22, 2016]. Disponible sur : <http://geoanalytics.net/GeoVisualAnalytics08/a04.pdf>
55. Anselin L. Review of cluster analysis software. Report in fulfillment of consultant agreement #2003-04-01 with the North American Association of Central Cancer Registries, Inc 2004. 2007. [April 22, 2016]. Disponible sur :

<https://www.naacrr.org/LinkClick.aspx?fileticket=pScBROXkeTg%3D&tabid=239&mid=699>

56. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med.* 30 nov 2006;25(22):3929–43.
57. Dahly DL, Gordon-Larsen P, Emch M et al. (2013). The spatial distribution of overweight and obesity among a birth cohort of young adult Filipinos (Cebu Philippines, 2005): an application of the Kulldorff spatial scan statistic. *Nutr Diabetes.* 3(7):e80.
58. A scan statistic with a variable window. *Statistics in Medicine*, 1996; 15:845–50;
59. Gaudart J, Graffeo N, Coulibaly D, Barbet G, Rebaudet S, Dessay N, et al. SPODT: An R Package to Perform Spatial Partitioning. *J Stat Softw.* 2015;63(16):1–23.
60. Gaudart J, Poudiougou B, Ranque S, Doumbo O. Oblique decision trees for spatial pattern detection: optimal algorithm and application to malaria risk. *BMC Med Res Methodol.* 2005;5(1):22.
61. Breiman L, Friedman J, Stone C, Olshen R. *Classification and Regression Trees.* 1 edition. Chapman and Hall/CRC; 1984. 368 p.
62. Gaudart J, Giorgi R, Poudiougou B, Touré O, Ranque S, Doumbo O, et al. Détection de clusters spatiaux sans point source prédéfini: utilisation de cinq méthodes et comparaison de leurs résultats. *Rev DÉpidémiologie Santé Publique.* 2007;55(4):297–306.
63. Genton M. Highly robust variogram estimation. *Math Geol.* 1998;30(2):213–21.
64. Diggle P, Ribeiro PJ. *Model-based geostatistics.* New York, NY: Springer; 2007.
65. Cressie N, Hawkins D. Robust estimation of the variogram: I. *J Int Assoc Math Geol.* 1980;12(2):115–25.
66. Olea R, Pardo-Igúzquiza E, Dowd P. Robust and resistant semivariogram modelling using a generalized bootstrap. *J South Afr Inst Min Metall.* 2015;115(1):37–44.
67. Cressie N. Fitting variogram models by weighted least squares. *J Int Assoc Math Geol.* 1985;17(5):563–86.
68. Ripley BD. *Spatial statistics.* New York: Wiley; 1981. 252 p. (Wiley series in probability and mathematical statistics).
69. Myers, D. Spatial interpolation: an overview. *Geoderma.* 1994; 62:17–28.

70. Okabe A. Spatial tessellations: concepts and applications of Voronoi diagrams. 2nd ed. Chichester; New York: Wiley; 2000. 671 p. (Wiley series in probability and statistics).
71. Wahba G. Spline models for observational data. Philadelphia, Pa: Society for Industrial and Applied Mathematics; 1990. 169 p. (CBMS-NSF Regional Conference series in applied mathematics).
72. Bosser P. Interpolation spatiale. Département Positionnement Terrestre et Spatial, École Nationale des Sciences Géographiques, France. 2012.
73. Goovaerts P. Geostatistics for natural resources evaluation. New York: Oxford University Press; 1997. 483 p. (Applied geostatistics series).
74. Wackernagel H. Multivariate geostatistics: an introduction with applications. 3., completely revised ed. Berlin: Springer; 2003. 387 p.
75. Margai FM. Geographical targeting of risk zones for childhood stunting and related health outcomes in Burkina Faso. *World Health Popul.* 2007;9:64–82.
76. Monestiez P, Dubroca L, Bonnin E, Durbec J-P, Guinet C. Geostatistical modelling of spatial distribution of *Balaenoptera physalus* in the Northwestern Mediterranean Sea from sparse count data and heterogeneous observation efforts. *Ecol Model.* mars 2006;193(3-4):615–28.
77. Gary LG, Cort J.W. Spatial statistics and models. Spatial statistics and models, D. Reidel Publishing Company, Boston, 1984.
78. Diggle P. Statistical Analysis of Spatial Point Patterns. 2^e éd. Hodder Education; 2002. 288 p.
79. Gatrell A, Bailey T, Diggle P, et al. Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Trans Inst Br Geogr.* 1996;21(1):256.
80. Larmarange J, Vallo R, Yaro S, Msellati P, Méda N. Methods for mapping regional trends of HIV prevalence from Demographic and Health Surveys (DHS). *Cybergegeo.* 26 oct 2011 [November 6, 2015]. Disponible sur : <http://cybergegeo.revues.org/24606>
81. Wand MP, Jones MC. Kernel smoothing. 1st ed. London; New York: Chapman & Hall; 1995. 212 p. (Monographs on statistics and applied probability).
82. Silverman BW. Density estimation for statistics and data analysis. CRC press; 1986.
83. Richardon S, Montfort C. Ecological correlation studies, in: *Spatial Epidemiology: Methods and Applications.* Oxford University Press, New York, 2000. [December 30, 2015]. Disponible sur : <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198515326.001.0001/acprof-9780198515326>

84. Walter SD. Disease mapping: a historical perspective, in: *Spatial Epidemiology: Methods and Applications*. Oxford University Press, New-York, 2000; pp. 223-39. [December 27, 2015]. Disponible sur : <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198515326.001.0001/acprof-9780198515326>
85. Koch T. *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, 2005; CA: ESRI Press.
86. Cadoret I, Benjamin C, Martin F et al. *Econométrie appliquée : Méthodes, applications, corrigés*. 2e édition. Bruxelles; Paris: De Boeck; 2009. 462 p.
87. Magnus JR. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *J Econom*. 1978;7:281-312.
88. Heijmans RDH, Magnus JR. On the first-order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Stat Neerlandica*. 1986;40(3):169-88.
89. Banos A. A propos de l'analyse spatiale exploratoire des données. *Cybergeog*. 2001. [May 3, 2016]. Disponible sur : <http://cybergeog.revues.org/4056>
90. Box G, Tiao G. *Bayesian inference in statistical analysis*. Wiley classics library ed. New York: Wiley; 1992. 588 p.
91. Carlin B, Louis T. *Bayes and empirical Bayes methods for data analysis*. 2. ed. Boca Raton, Fla.: Chapman & Hall/CRC; 2000. 419 p.
92. Samaniego F. *A comparison of the Bayesian and frequentist approaches to estimation*. New York, NY: Springer; 2010. 225 p.
93. Kammann EE, Wand MP. Geoaddivitive models. *J R Stat Soc Ser C Appl Stat*. 2003;52(1):1-18.
94. Hastie T, Tibshirani R. Generalized additive models for medical research. *Stat Methods Med Res*. 1995;4(3):187-96.
95. Cox D. Regression Models and Life-Tables. *J R Stat Soc*. 1972;34(2):187-220.
96. Hastie T, Tibshirani R. Generalized Additive Models. *Stat Sci*. 1986;1(3):297-318.
97. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *J R Stat Soc B*. 1999;61(2):381-400.
98. Breslow NE, Clayton DG. Approximate Inference in Generalized Linear Mixed Models. *J Am Stat Assoc*. 1993;88(421):9.
99. Fahrmeir L, Kneib T, Lang S. Penalized structured additive regression for space-time data: a Bayesian perspective. *Stat Sin*. 2004;14(3):731-62.

100. Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines. *Comput Stat Data Anal.* 2006;50(4):967–91.
101. Belitz C, Lang S. Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Comput Stat Data Anal.* 2008;53(1):61–81.
102. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology.* New York: CRC Press, Chapman & Hall Taylor & Francis Group; 2009.
103. Fahrmeir L, Lang S. Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data. *Annals of the Institute of Statistical Mathematics*, 53, 10–30.;
104. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math.* mars 1991;43(1):1–20.
105. Fahrmeir L, Lang S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Appl Stat.* 2001;50(2):201–20.
106. Kandala N, Lang S, Klasen S, Fahrmeir L. Semiparametric analysis of the socio-demographic and spatial determinants of undernutrition in two african countries. *Res Offi Cial Stat EUROSTAT* 4 81–100. 2001 [cité 30 août 2016]; Disponible sur : <https://epub.ub.uni-muenchen.de/1626/>
107. Kandala N, Fahrmeir L, Klasen S, Priebe J. Geo-additive models of childhood undernutrition in three sub-Saharan African countries. *Popul Space Place.* 2009;15(5):461–73.
108. Belitz C, Brezger A, Kneib T et al. *BayesX - Software for Bayesian Inference in Structured Additive Regression Models.* Version 2.1. 21st edition. Munich, Germany. 2012. [May 21, 2015]. Disponible sur : <http://www.statistik.lmu.de/~bayesx/bayesxdownload.html>
109. Belitz C, Brezger A, Kneib T et al. *BayesX - Software for Bayesian Inference in Structured Additive Regression Models.* Version 3.0.2. Methodology. 2015. [May 21, 2015]. Disponible sur : http://www.statistik.lmu.de/~bayesx/manual/methodology_manual.pdf
110. Fahrmeir L, Kneib T. Property of posteriors in structured additive regression models: Theory and Empirical Evidence. 2009;139(3):843–59.
111. Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci.* 1996;89–102.
112. Klein N, Kneib T, Lang S. *Bayesian Structured Additive Distributional Regression.* Working paper 2013–23, working papers in economics and statistics, research platform empirical and experimental economics, University of Innsbruck; 2013.

113. Kneib T. Mixed model based inference in structured additive regression. Dissertation, LMU München: Faculty of Mathematics, Computer Science and Statistics; 2006. [December 29, 2015]. Disponible sur : <http://edoc.ub.uni-muenchen.de/5011>
114. Kneib T, Fahrmeir L. Structured Additive Regression for Multicategorical Space-Time Data: A Mixed Model Approach. *Biometrics*, 2006; 62: 109-18;
115. Knox E. The Detection of Space-Time Interactions. *J R Stat Soc Ser C Appl Stat*. 1964;13(1):25–30.
116. Baker R. Identifying space–time disease clusters. *Acta Trop*. 2004;91(3):291–9.
117. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967;27(2 Part 1):209–20.
118. Jacquez G. A k nearest neighbour test for space–time interaction. *Stat Med*. 1996;15(18):1935–49.
119. UNAIDS. Global statistics: fact sheet 2016. [August 8, 2016]. Disponible sur : http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf
120. UNAIDS. 90–90–90 - An ambitious treatment target to help end the AIDS epidemic. 2014 [November 10, 2015]. Disponible sur : <http://www.unaids.org/en/resources/documents/2014/90-90-90>
121. UNAIDS. The gap report 2014. [January 6, 2016]. Disponible sur : http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/unaidspublication/2014/UNAIDS_Gap_report_en.pdf
122. ONUSIDA. Rapport ONUSIDA sur l'épidémie mondiale de SIDA 2010 [March 21, 2016]. Genève: ONUSIDA; Disponible sur : http://www.unaids.org/globalreport/documents/20101123_GlobalReport_Full_Fr.pdf
123. UNDP. Transforming our world: the 2030 Agenda for Sustainable Development. [March 1, 2016]. Disponible sur : <https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>
124. UNAIDS. Getting to Zero: HIV in Eastern and Southern Africa. 2013. [February 2, 2016]. Disponible sur : <http://www.unicef.org/esaro/Getting-to-Zero-2013.pdf>
125. UNAIDS. Rwanda: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/rwanda>
126. National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], and ICF International. 2012. Rwanda Demographic and Health Survey 2010. Calverton, Maryland, USA: NISR, MOH, and ICF

- International. [February 2, 2016]. Disponible sur : <http://dhsprogram.com/pubs/pdf/FR259/FR259.pdf>
127. Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC), National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and ICF International. 2013. Tanzania HIV/AIDS and Malaria Indicator Survey 2011-12: Key Findings. Dar es Salaam, Tanzania: TACAIDS, ZAC, NBS, OCGS, and ICF International. [February 2, 2016]. Disponible sur : <http://dhsprogram.com/pubs/pdf/SR196/SR196.pdf>
128. UNAIDS. United Republic of Tanzania: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/unitedrepublicoftanzania>
129. Ministère du Plan et Suivi de la Mise en oeuvre de la Révolution de la Modernité (MPSMRM), Ministère de la Santé Publique (MSP) et ICF International, 2014. Enquête Démographique et de Santé en République Démocratique du Congo 2013-2014. Rockville, Maryland, USA: MPSMRM, MSP et ICF International. [January 21, 2016]. Disponible sur : <http://dhsprogram.com/pubs/pdf/SR218/SR218.pdf>
130. UNAIDS. Democratic Republic of the Congo: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/democraticrepublicofthecongo>
131. Kenya National Bureau of Statistics (KNBS) and ICF Macro. 2010. Kenya Demographic and Health Survey 2008-09. Calverton, Maryland: KNBS and ICF Macro. [August 9, 2016]. Disponible sur : <http://dhsprogram.com/pubs/pdf/FR229/FR229.pdf>
132. UNAIDS. Kenya: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/kenya>
133. UNAIDS. Uganda: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/uganda>
134. Institut de Statistiques et d'Études Économiques du Burundi (ISTEEBU), Ministère de la Santé Publique et de la Lutte contre le Sida [Burundi] (MSPLS), ICF International. Enquête Démographique et de Santé Burundi 2010. 2012 [July 15, 2015]. Disponible sur : <http://www.measuredhs.com/pubs/pdf/FR253/FR253.pdf>.
135. Segamba L. Enquête démographique et de santé au Burundi, 1987 [January 13, 2016]. République du Burundi, Ministère de l'intérieur, Dép. de la population; 1988.
136. WHO: Stratégie de coopération de l'OMS avec les pays République du Burundi 2008–2013. 2009. [January 20, 2016]. Disponible sur : http://www.who.int/entity/countryfocus/cooperation_strategy/ccs_cod_2009_2013_fr.pdf

137. Central Intelligence Agency (US). The World Factbook: Africa, Burundi 2015. [March 15, 2016]. Disponible sur : <https://www.cia.gov/library/publications/the-world-factbook/geos/by.html>
138. MSPLS. Plan National de Développement Sanitaire 2011-2015. [March 16, 2016]. Disponible sur : <https://www.minisante.bi/images/Documents/PNDS%202011-2015.pdf>
139. SEP/CNLS. Plan Stratégique National de Lutte contre le VIH 2007-2011. [March 16, 2016]. 2006. Disponible sur : http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/legaldocument/wcms_126635.pdf
140. MSPLS. Décret n° 100/ 254 du 04 octobre 2011 portant organisation et fonctionnement du Ministère de la Santé Publique et de la Lutte contre le SIDA. [March 17, 2016]. Disponible sur : <https://www.minisante.bi/index.php/organigramme>
141. MSPLS. Evaluation de la qualité des prestations dans les formations sanitaires du burundi, 2ème édition. 2013. [July 17, 2015]. Disponible sur : http://www.fbpsanteburundi.bi/cside/contents/docs/enquete_fosa.pdf
142. SEP/CNLS. Politique Nationale de Prévention de la Transmission du VIH de la Mère à l'Enfant. 2010. [March 17, 2016]. Disponible sur : <http://www.emtct-iatt.org/wp-content/uploads/2013/05/Burundi-National-PMTCT-Guidelines-2010.pdf>
143. ONUSIDA. Les trois principes, un concept clé. [March 15, 2016]. Disponible sur : http://data.unaids.org/una-docs/three-ones_keyprinciples-flyer_fr.pdf
144. SEP/CNLS. Plan Stratégique National de Lutte contre le VIH 2012-2016. 2012. [January 29, 2016]. Disponible sur : http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/legaldocument/wcms_202048.pdf
145. UNAIDS. Rapport d'activités sur la lutte contre le SIDA et rapport sur les progrès enregistrés vers un accès universel (Burundi). 2015. [March 19, 2016]. Disponible sur : http://www.unaids.org/sites/default/files/country/documents/BDI_narrative_report_2015.pdf
146. Ministère de la Fonction Publique, du Travail et de la Sécurité Sociale. Politique nationale de lutte contre le SIDA et le VIH sur le lieu de travail. 2011. [February 2, 2016]. Disponible sur : http://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---ilo_aids/documents/legaldocument/wcms_172468.pdf
147. UNAIDS. United Nations General Assembly Special Session on HIV/AIDS. Declaration of Commitment on HIV/AIDS. Resolution A/Res/S-26/2, 27 June 2001. [February 2, 2016]. Disponible sur : http://www.unaids.org/sites/default/files/sub_landing/files/aidsdeclaration_en.pdf

148. UNAIDS. Burundi: HIV and AIDS estimates, 2015. [August 9, 2016]. Disponible sur : <http://www.unaids.org/en/regionscountries/countries/burundi>
149. Tanser F, Barnighausen T, Cooke GS, Newell M-L. Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *Int J Epidemiol.* 2009;38(4):1008-16.
150. Cuadros DF, Awad SF, Abu-Raddad LJ. Mapping HIV clustering: a strategy for identifying populations at high risk of HIV infection in sub-Saharan Africa. *Int J Health Geogr.* 2013;12(1):28.
151. UNAIDS. Burundi - Developing subnational estimates of HIV prevalence and the number of people living with HIV. 2014. [July 16, 2015]. Disponible sur : http://www.unaids.org/sites/default/files/media_asset/2014_subnationalestimatessurvey_Burundi_en.pdf
152. Larmarange, J. prevR: estimating regional trends of a prevalence from a DHS. Paris: IRD; 2013. [November 10, 2015]. <http://cran.r-project.org/web/packages/prevR/index.html>.
153. UNAIDS. Methodology – understanding the HIV estimates. Geneva: Joint United Nations Programme on HIV/AIDS; 2013. [May 8, 2016]. Disponible sur : http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/20131118_Methodology.pdf
154. Measure DHS. Demographic and Health Surveys. [July 15, 2015]. Disponible sur : <http://www.measuredhs.com>
155. Burgert CR, Colston J, Roy T et al. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spatial Analysis Report No. 7. Calverton, Maryland, USA: ICF International.; 2013.
156. StataCorp. 2009. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP. 2011.
157. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [February 1, 2016]. Disponible sur : <http://ftp.igh.cnrs.fr/pub/CRAN/>
158. Larmarange J, Bendaud V. HIV estimates at second subnational level from national population-based surveys: AIDS. 2014;28:S469–76.
159. Kulldorff M (1997). A spatial scan statistic. *Commun Stat - Theory Methods.* 26(6):1481–96.
160. Ngesa O, Mwambi H, Achia T. Bayesian spatial semi-parametric modeling of HIV variation in Kenya. 2014. [January 28, 2016]. Disponible sur : <http://dx.plos.org/10.1371/journal.pone.0103299>

161. Fahrmeir L, Kneib T. Propriety of posteriors in structured additive regression models: Theory and empirical evidence. *J Stat Plan Inference*. 2009;139(3):843–59.
162. UNICEF. Improving child nutrition: the achievable imperative for global progress. New York: United Nations Children's Fund. 2013. [October 12, 2015]. Disponible sur : http://www.unicef.org/gambia/Improving_Child_Nutrition_-_the_achievable_imperative_for_global_progress.pdf
163. Black RE, Allen LH, Bhutta ZA, Caulfield LE, de Onis M, Ezzati M, et al. Maternal and child undernutrition: global and regional exposures and health consequences. *The lancet*. 2008;371(9608):243–60.
164. de Onis M, Branca F. Childhood stunting: a global perspective. *Matern Child Nutr*. 2016;12 (S1):12–26.
165. Habicht J-P, Martorell R, Yarbrough C et al. Height and weight standards for preschool children. How relevant are ethnic differences in growth potential? *Lancet*. 1974;1:611–5.
166. Bogin B, Loucky J. Plasticity, political economy, and physical growth status of Guatemala Maya children living in the United States. *Am J Phys Anthropol*. 1997;102(1):17–32.
167. Stewart C, Iannotti L, Dewey KG et al. Contextualising complementary feeding in a broader framework for stunting prevention: Complementary feeding in stunting prevention. *Matern Child Nutr*. 2013;9:27–45.
168. Bhandari N, Bahl R, Taneja S et al. Growth performance of affluent Indian children is similar to that in developed countries. *Bull World Health Organ*. 2002;80(3):189–195.
169. de Onis M, Onyango AW, Borghi E, Garza C, Yang H. Comparison of the World Health Organization (WHO) Child Growth Standards and the National Center for Health Statistics/WHO international growth reference: implications for child health programmes. *Public Health Nutr*. 2006; 9(07). [March 28, 2016]. Disponible sur : http://www.journals.cambridge.org/abstract_S1368980006001534
170. de Onis M, Yip R. The WHO growth chart: historical considerations and current scientific issues. *Bibliotheca Nutritio et Dieta* 1996; 53: 74–89. [March 28, 2016]. Disponible sur : http://www.who.int/childgrowth/publications/WHO_growth_charts.pdf?ua=1
171. WHO. Multicentre Growth Reference Study Group. WHO Child Growth Standards: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body mass index-for-age: Methods and Development. Geneva: World Health Organization, 2006. WHO. [March 28, 2016]. Disponible sur : http://www.who.int/childgrowth/standards/technical_report/en/

172. 1,000 Days. 1,000 Days. 2010. [April 21, 2016]. Disponible sur : <http://thousanddays.org/>
173. UNICEF. Améliorer la nutrition de l'enfant. Un objectif impératif et réalisable pour le progrès mondial. 2013. [June 12, 2016]. Disponible sur : https://www.unicef.fr/sites/default/files/userfiles/UNICEF_NutritionReport_FR.pdf
174. UNICEF, WHO, World Bank. Levels and Trends in Child Malnutrition. Joint Child Malnutrition Estimates. New York: United Nations International children's Fund; Geneva: World Health Organization; DC: World Bank, 2015. 2015. [June 12, 2016]. Disponible sur : http://www.unicef.org/media/files/JME_2015_edition_Sept_2015.pdf
175. FAO, UNICEF, WFP, WHO. Zero stunted children less than 2 years: Compendium - final report zero hunger challenge working groups. [June 12, 2016]. Disponible sur : <http://www.un.org/en/issues/food/taskforce/pdf/Zero%20stunted%20children%20less%20than%202.pdf>
176. FAO, UNICEF, WFP, WHO. High Level Task Force on Global Food and Nutrition Security (HLTF). [June 12, 2016]. Disponible sur : <http://www.un.org/en/issues/food/taskforce/pdf/Zero%20stunted%20children%20less%20than%202.pdf>
177. Food & Business Knowledge Platform. Final Report on Dutch Food Security Policy Consultation. [June 12, 2016]. Disponible sur : http://knowledge4food.net/wp-content/uploads/2014/09/140929_fbkp-consultation-report.pdf
178. Stuart G, Lawrence H, Venkatesh M et al. Maternal and Child Nutrition. The Lancet; 2013. [June 12, 2016]. Disponible sur : <http://www.thelancet.com/pb/assets/raw/Lancet/stories/series/nutrition-eng.pdf>
179. Institut National de la Statistique (INS) et ICF International. 2012. Enquête Démographique et de Santé et à Indicateurs Multiples de Côte d'Ivoire 2011-2012. Calverton, Maryland, USA: INS et ICF International. [November 15, 2014]. Disponible sur : <http://dhsprogram.com/publications/publication-FR272-DHS-Final-Reports.cfm>
180. Scaling Up Nutrition (SUN). Annual Report 2015. [June 12, 2016]. Disponible sur : http://scalingupnutrition.org/wp-content/uploads/2015/06/Orange_Internal_InOutline_FR_20140416_web.pdf
181. Ghana Statistical Service (GSS), Ghana Health Service (GHS), and ICF International. Ghana Demographic and Health Survey 2014. Rockville, Maryland, USA: GSS, GHS, and ICF International. 2015.
182. Liberia Institute of Statistics and Geo-Information Services (LISGIS), Ministry of Health and Social Welfare [Liberia], National AIDS Control Program [Liberia], and ICF International. Liberia Demographic and Health Survey 2013.

- Monrovia, Liberia: Liberia Institute of Statistics and Geo- Information Services (LISGIS) and ICF International. 2014.
183. Institut National de la Statistique et de la Démographie (INSD) et ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples du Burkina Faso 2010. Calverton, Maryland, USA : INSD et ICF International. 2012;
 184. Cellule de Planification et de Statistique (CPS/SSDSPF), Institut National de la Statistique (INSTAT/MPATP). INFO-STAT et ICF International. Enquête Démographique et de Santé au Mali 2012-2013. Rockville, Maryland, USA □: CPS, INSTAT, INFO-STAT et ICF International. 2014.
 185. Institut National de la Statistique (INS). Recensement Général de la Population et de l'Habitat 2014: Principaux résultats préliminaires (Côte d'Ivoire). 2014.
 186. Alain Handy SP, Charles T. L'Accord Politique de Ouagadougou: vers la sortie de crise pérenne en Côte d'Ivoire. [June 14, 2016]. Disponible sur : http://www.afri-ct.org/IMG/pdf/42_Alain_Handy_Charles_Ouaga.pdf
 187. Nations Unies. Dix-huitième rapport du Secrétaire général sur l'Opération des Nations Unies en Côte d'Ivoire. 2008. [June 22, 2016]. Disponible sur : http://www.un.org/fr/documents/view_doc.asp?symbol=S/2008/645
 188. Nations Unies. Vingt-deuxième rapport du Secrétaire général sur l'Opération des Nations Unies en Côte d'Ivoire. 2009. [June 22, 2016]. Disponible sur : http://www.un.org/fr/documents/view_doc.asp?symbol=S/2009/495
 189. Nations Unies. Vingt-troisième rapport du Secrétaire général sur l'Opération des Nations Unies en Côte d'Ivoire. 2011. [June 22, 2016]. Disponible sur : http://www.un.org/fr/documents/view_doc.asp?symbol=S/2010/15
 190. Nations Unies. Vingt-sixième rapport du Secrétaire général sur l'Opération des Nations Unies en Côte d'Ivoire. 2011. [June 22, 2016]. Disponible sur : http://www.un.org/fr/documents/view_doc.asp?symbol=S/2010/600
 191. UNDP, éditeur. Work for human development. New York, NY: United Nations Development Programme; 2015. 272 p. (Human development report). [June 12, 2016]. Disponible sur : http://hdr.undp.org/sites/default/files/2015_human_development_report.pdf
 192. UNDP. Tendances et cartographie du développement humain en Côte d'Ivoire. [June 12, 2016]. Disponible sur : http://www.ci.undp.org/content/dam/cote_divoire/docs/publications/Tendance%20et%20Carto%20Corrig%C3%A918.pdf
 193. Merabet O. Termes de Référence pour une étude sur le profil migratoire de la Côte d'Ivoire: Rapport final. 2006. [June 12, 2016]. Disponible sur : http://www.eeas.europa.eu/delegations/cote_ivoire/documents/more_info/7_doc_fr.pdf

194. Ministre de la Santé et de l'Hygiène Publique. Politique Nationale de Nutrition. [June 13, 2016]. Disponible sur : <http://faolex.fao.org/docs/pdf/lvc146623.pdf>
195. Centre SAVALGI de nutrition de Korhogo. Rapport semestriel d'activité sur la malnutrition et le paludisme. 2012. [June 23, 2016]. Disponible sur : <http://fondationsplarci.org/telechargement/RAPPORT%20SEMESTRIEL%20PLAR.pdf>
196. ACF. Une approche communautaire pour la réhabilitation nutritionnelle et le changement des comportements alimentaires en Côte d'Ivoire. 2013. [June 23, 2016]. Disponible sur : <http://www.alnap.org/pool/vacancies/cote-d'ivoire-evaluation-trd.pdf>
197. FAO. Programme intégré d'éducation nutritionnelle en Cote d'Ivoire. 2009. [June 12, 2016]. Disponible sur : http://www.fao.org/fileadmin/user_upload/wa_workshop/Factsheet/11._Fiches_FSNL__bonnes_pratiques_Education_Nutritionnelle_en_Cote_d_Ivoire.pdf
198. Stevens G, Finucane M, Paciorek C, Flaxman S, White R, Donner A, et al. Trends in mild, moderate, and severe stunting and underweight, and progress towards MDG 1 in 141 developing countries: a systematic analysis of population representative data. *The Lancet*. 2012;380(9844):824–34.
199. Tzioumis E, Kay M, Bentley M, Adair LS. Prevalence and trends in the childhood dual burden of malnutrition in low- and middle-income countries, 1990–2012. *Public Health Nutr*. 2016;19(08):1375–88.
200. Sahn D. Malnutrition in Côte d'Ivoire: prevalence and determinants. Washington, D.C: World Bank; 1990. 30 p. (Social dimensions of adjustment in Sub-Saharan Africa).
201. Aké-Tano O, Tiembré I, Konan YE, Donnen P, N'Cho Dagnan S, Dramaix M, et al. Malnutrition chronique chez les enfants de moins de 5 ans au nord de la Côte d'Ivoire. *Santé Publique*. 2010;22(2):213–20.
202. Minoiu C, Shemyakina O. Armed conflict, household victimization, and child health in Côte d'Ivoire. *J Dev Econ*. 2014;108:237–55.
203. Kandala N-B, Madungu T, Emina J-BO, Nzita K-PD, Cappuccio FP. Malnutrition among children under the age of five in the Democratic Republic of Congo (DRC): does geographic location matter? *BMC Public Health* 2011; 11:261.
204. N'Cho S, Kouassi L, Koffi AK, Shoemaker J, Barrère M, Barrère B, et al. Enquête Démographique et de Santé, Côte d'Ivoire 1994. Calverton, Maryland, U.S.A.:Institut National de la Statistique et Macro International Inc. 1995. [November 15, 2014]. Disponible sur : <http://www.dhsprogram.com/pubs/pdf/FR66/FR66.pdf>
205. Institut National de la Statistique [Côte d'Ivoire] et ORC Macro. Enquête Démographique et de Santé, Côte d'Ivoire 1998-1999. Calverton, Maryland

- USA: Institut National de la Statistique et ORC Macro. Institut National de la Statistique; 2001. [November 15, 2014]. Disponible sur : <http://www.dhsprogram.com/pubs/pdf/FR129/FR129.pdf>
206. Institut National de la Statistique; Direction Générale du Plan. Recensement Général de la Population et de l'Habitat - Côte d'Ivoire 1998.
207. Institut National de la Statistique et de la Démographie (INSD) et ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples de Côte d'Ivoire 2011-2012. Calverton, Maryland, USA: INS et ICF International. 2012. [November 15, 2014]. Disponible sur : <http://www.dhsprogram.com/pubs/pdf/FR272/FR272.pdf>
208. Cressie N. Spatial prediction and ordinary kriging. *Math Geol.* 1988;20(4):405–21.
209. Cressie N. *Statistics for Spatial Data, Revised Edition.* New York, NY: Wiley; 1993. 900 p. (Wiley series in probability and mathematical statistics).
210. Margai F. Geographic Targeting of Risk Zones for Childhood Stunting and Related Health Outcomes in Burkina Faso. *World Health Popul.* 2007;9(2):64–82.
211. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med.* 1995;14(8):799–810.
212. Kellie JA, Stanley L. Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal* 6;1, 97–105; 2006. [June 24, 2016]. Disponible sur : <http://www.stata-journal.com/sjpdf.html?articlenum=st0099>

Résumé

Cette thèse porte sur l'analyse spatiale de problèmes de santé publique rencontrés en Afrique subsaharienne. Dans une première étude, nous avons analysé les variations spatiales de la prévalence du VIH au Burundi et les déterminants socio-comportementaux chez les personnes âgées de 15 à 49 ans à partir des données de l'Enquête Démographique et de Santé (EDS) de 2010. Dans une deuxième étude, nous avons analysé l'évolution de la répartition spatiale de la prévalence du retard de croissance chez les enfants âgés de moins de 3 ans en Côte d'Ivoire à partir des données des EDS de 1994, 1998-1999 et 2011-2012. Pour ces deux études, nous avons tout d'abord prédit en des points non échantillonnés les niveaux de prévalence du problème étudié, soit par la méthode des noyaux, soit par la méthode du krigeage ordinaire. Nous avons ensuite procédé à la détection des agrégats spatiaux ou clusters par la méthode de Kulldorff puis à l'identification de zones de différents risques par la méthode du partitionnement oblique du territoire. Finalement, nous avons utilisé la modélisation spatiale pour identifier les facteurs associés. Ces techniques d'analyse spatiale présentent de nombreux avantages mais également de nombreuses limites que nous soulignons.

Mots clés : Analyse spatiale, VIH, retard de croissance, Burundi, Côte d'Ivoire.

Spatial analyses of public health issues in sub-Saharan Africa: examples of HIV/AIDS and malnutrition

Summary

This thesis is about the spatial analysis of public health issues in sub-Saharan Africa. In a first study, we analysed the spatial variations of HIV prevalence and the socio-behavioral determinants among people aged 15 to 49 in Burundi using 2010 Demographic and Health Survey (DHS) data. In a second study, we analysed the evolution of the spatial distribution of stunting prevalence among under-three children in Côte d'Ivoire using 1994, 1998-1999 and 2011-2012 DHS data. For both studies, we first predicted the prevalence levels of the public health issue under study in unsampled points by the kernel method or ordinary kriging. We then identified clusters using the Kulldorff's method and zones of different risks using the oblique partitioning method of the study area. Finally, we used spatial modeling to identify associated risk factors. These spatial analysis techniques have not only many advantages but also many limitations that we emphasize.

Key words: Spatial analysis, HIV, stunting, Burundi, Côte d'Ivoire.

Discipline : Biologie-Santé (Biostatistique)

Laboratoire : Unité Mixte Internationale 233 Transitions épidémiologiques, recherches translationnelles appliquées au VIH et aux maladies infectieuses (UMI 233 TransVIHMI), Institut de Recherche pour le Développement (IRD), Montpellier
